EMERGING 2011

The Third International Conference on Emerging Network Intelligence

ISBN: 978-1-61208-174-8

November 20-25, 2011

Lisbon, Portugal

IMMM 2011 Editors

Tulin Atmaca, IT/Telecom&Management SudParis, France

Michael D. Logothetis, University of Patras, Greece

Krishna Murthy, Global IT Solutions at Quintiles - Raleigh, USA

EMERGING 2011

Foreword

The Third International Conference on Emerging Network Intelligence [EMERGING 2011], held between November 20 and 25, 2011 in Lisbon, Portugal, constituted a stage to present and evaluate the advances in emerging solutions for next-generation architectures, devices, and communications protocols. Particular focus was aimed at optimization, quality, discovery, protection, and user profile requirements supported by special approaches such as network coding, configurable protocols, context-aware optimization, ambient systems, anomaly discovery, and adaptive mechanisms.

Next-generation large distributed networks and systems require substantial reconsideration of exiting 'de facto' approaches and mechanisms to sustain an increasing demand on speed, scale, bandwidth, topology and flow changes, user complex behavior, security threats, and service and user ubiquity. As a result, growing research and industrial forces are focusing on new approaches for advanced communications considering new devices and protocols, advanced discovery mechanisms, and programmability techniques to express, measure and control the service quality, security, environmental and user requirements.

We take here the opportunity to warmly thank all the members of the EMERGING 2011 Technical Program Committee, as well as the numerous reviewers. The creation of such a broad and high quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and efforts to contribute to EMERGING 2011. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

Also, this event could not have been a reality without the support of many individuals, organizations, and sponsors. We are grateful to the members of the EMERGING 2011 organizing committee for their help in handling the logistics and for their work to make this professional meeting a success.

We hope that EMERGING 2011 was a successful international forum for the exchange of ideas and results between academia and industry and for the promotion of progress in emerging network intelligence.

We are convinced that the participants found the event useful and communications very open. We also hope the attendees enjoyed the historic charm of Lisbon, Portugal.

EMERGING 2011 Chairs:

Tadashi Araragi
Tulin Atmaca
David Carrera
Robert Foster
Nuno M. Garcia
Raj Jain
Michael D. Logothetis
Krishna Murthy
Daniel Scheibli
Phuoc Tran-Gia

EMERGING 2011

Committee

EMERGING Advisory Chairs

Raj Jain, Washington University in St. Louis, USA
Michael D. Logothetis, University of Patras, Greece
Tulin Atmaca, IT/Telecom&Management SudParis, France
Phuoc Tran-Gia, University of Wuerzburg, Germany
Nuno M. Garcia, Universidade Lusófonas de Humanidades e Tecnologias, Lisboa, Portugal

EMERGING 2011 Industry Liaison Chairs

Krishna Murthy, Global IT Solutions at Quintiles - Raleigh, USA
Tadashi Araragi, Nippon Telegraph and Telephone Corporation – Kyoto, Japan
Robert Foster, Edgemount Solutions - Plano, USA

EMERGING 2011 Research Chairs

David Carrera, Barcelona Supercomputing Center (BSC) / Universitat Politecnica de Catalunya (UPC), Spain
Daniel Scheibli, SAP Research, Germany

EMERGING 2011 Technical Program Committee

Khalid Al-Begain, University of Glamorgan, UK
Artur Andrzejak, University of Heidelberg, Germany
Richard Anthony, University of Greenwich, UK
Tadashi Araragi, Nippon Telegraph and Telephone Corporation - Kyoto, Japan
Tulin Atmaca, IT/Telecom&Management SudParis, France
M. Ali Aydin, Istanbul University, Turkey
Magdy Bayoumi, University of Louisiana at Lafayette, USA
Andreas Berl, University of Passau, Germany
Robert Bestak, Czech Technical University in Prague, Czech Republic
Christian Blum, Universitat Politècnica de Catalunya, Spain
Chih-Yung Chang (張志勇), Tamkang University, Taiwan
Dong Ho Cho, Korea Advanced Institute of Science and Technology (KAIST), Republic of Korea
Alberto Dainotti, University of Napoli "Federico II", Italy
Carl James Debono, University of Malt, Malta
Rolf Drechsler, University of Bremen, Germany
Jean-Michel Dricot, Université Libre de Bruxelles, Belgium
Mohamed Eltoweissy, Virginia Tech, USA
Anna Förster, University of Lugano, Switzerland
Nuno M. Garcia, Universidade Lusófonas de Humanidades e Tecnologias, Lisboa, Portugal
Kamini Garg, University of Applied Sciences Southern Switzerland - Lugano, Switzerland
Christophe Guéret, Vrije Universiteit Amsterdam, The Netherlands
Go Hasegawa, Osaka University, Japan
Eva Hladka, Masaryk University & CESNET, Czech Republic
Raj Jain, Washington University in St. Louis, USA
Henrik Karstoft, Aarhus School of Engineering, Denmark

Copyright Information

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission or reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article is does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

Table of Contents

Model-Based Distributed On-line Safety Monitoring

Amer Dheedan, Yiannis Papadopoulos
Department of Computer Science
University of Hull
Hull, United Kingdom,
{A.A.Aloqaili@2007. , Y.I.Papadopoulos@}hull.ac.uk

***Abstract*— On-line safety monitoring, i.e. the tasks of fault detection and diagnosis, alarm annunciation, and fault controlling, is an essential task in the operational phase of critical systems. Although current safety monitors deliver this task to some extent, the problem of effective and timely safety monitoring is still largely unresolved. In this paper, we propose a Distributed On-line Safety Monitor (DOSM) that can achieve a range of real-time safety monitoring tasks: fault detection and diagnosis, alarm annunciation and control of hazardous failures. The monitor consists of a Multi-agent Monitoring System (MaMS) operating on a Distributed Monitoring Model (DMM) that contains reference knowledge derived from off-line safety assessments, and a number of Distributed Data Structures (DDSs) that provide up-to-date sensory measurements. Guided by the knowledge contained in the DMM and real-time observations of the system provided by the DSSs, agents are hierarchically deployed and work collaboratively to integrate and deliver safety monitoring tasks, both locally at the sub-system levels and globally overseeing the overall behaviour of the system.**

Keywords-Fault Detection and Diagnosis; Optimal Alarm Annunciation; Fault Controlling; Multi-agent Monitoring System;

I. INTRODUCTION

Over the last 30 years, considerable work on model-based safety monitoring, has resulted in approaches that exploit knowledge about the normal operational behaviour and failure of a system. In the context of this work, models such as state-machines, goal trees, goal hierarchies and fault trees have been exploited and demonstrated their benefits as reference knowledge for system monitoring (for a comprehensive see [1]). Typically, these models incorporate deep knowledge of the target system and enable qualitative and quantitative (often probabilistic) reasoning about behavioural transitions, symptoms, causes and possible effects of faults [2, 3].

Recently, a centralised safety monitor [4] that exploits knowledge derived from the application of a semi-automated off-line safety assessment method and tool called Hierarchically Performed Hazard Origin and Propagation Studies (HiP-HOPS) [5] has been proposed. That knowledge is composed of two elements: (a) a hierarchy of state-machines describing the behaviour of the system, effectively capturing the normal and abnormal mode and state transitions of the system and its sub-systems; (b) a set of fault trees, which effectively represent diagnostic models that relate the symptoms of failure to ultimate root causes.

The motivation for that work has been the observation that, in the current industrial practice, vast amounts of knowledge derived in off-line safety assessments cease to be useful following the certification and deployment of a system. A key contribution of this work is that it brings this knowledge forward to the operational phase of a system and usefully exploits it for the purposes of on-line safety monitoring. The concept is potentially very useful. However, the monitor described in [4] is limited in its potential because it is monolithic and centralised, and therefore, has limited applicability in systems that have a distributed nature and incorporate large numbers of components that interact collaboratively in dynamic cooperative structures.

Recent work on Multi-agent Systems (MaS) shows that the distributed reasoning paradigm could cope with the nature of such systems. In [6], for example, a MaS has been exploited to increase the capacity of a diagnostic scheme of a large-scale system. MaS have also demonstrated prompt responses in detecting faults and diagnosing the underlying causes of failures in complex distributed chemical processes [7]. Despite these encouraging developments, serious operational hazards are still recorded in safety critical systems and disastrous failures do not seem out of the question. Accordingly, the problem of developing a robust on-line monitor is still debated mainly in terms of two aspects. One aspect concerns the type of knowledge that is required to inform the on-line reasoning of the monitor: should it be, for example, a set of rules defined by experts or should it be knowledge based on engineering models, and in the latter case, what kind of knowledge should such models contain [1, 8]? The second aspect of the problem arises from the increasingly distributed nature of modern systems and the inevitably complicated collaboration among their components. This aspect is concerned with overcoming the limitations of centralised and rigidly distributed monitors, and is, looking into employing intelligent monitoring agents as means for delivering flexible, timely, consistent and effective monitoring [1, 9].

In order to address the issues discussed above, this paper proposes a DOSM which combines the benefits of using knowledge derived in off-line safety assessments with the benefits of a collaborative distribution of MaS. The DOSM consists of a DDM derived from the HiP-HOPS safety assessment model, a MaMS incorporating a number of Belief-Desire-Intention (BDI) agents, and a set of DDSs. According to the architectural model of the target system, agents are hierarchically deployed as monitoring agents (MAGs) and each is provided with its portion of the DMM

 ISBN: 978-1-61208-174-8

and appropriate DSSs. By exploiting their portions of the DMM, MAGs reason on the operational parameters held by DDSs, to detect and assess the effects of deviations, diagnose the underlying causes of the detected deviations and automatically apply corresponding fault controlling measures. Moreover, in order to avoid alarm avalanches and latent alarms that may mislead the system operators [10, 11], MAGs are also able to optimise alarm annunciation by (a) suppressing unimportant and false alarms; (b) filtering spurious sensory measurements; (c) incorporating helpful alarm information, such as assessment of the operational conditions after the occurrence of the fault, guidance on controlling the occurred fault, and diagnostics of the underlying causes of failures.

Benefit of the proposed DOSM ranges from increasing the flexibility, composability and extensibility of on-line safety monitoring to ultimately developing an effective and cost-effective monitor for safety critical systems.

The rest of this paper is organised in the following sections: section two briefly describes the nature of modern critical systems and the requirements for representation of such systems for the purpose of safety monitoring. Section three presents the approach, and the role and architecture of the DOSM. To demonstrate the effectiveness of the delivered monitoring tasks, in section four, the DOSM is applied to an aircraft fuel system and some failure scenarios are discussed. Finally, section five draws a conclusion and proposes further work.

II. MODELLING SYSTEMS FOR MONITORING

Large scale and dynamic behaviour are two common aspects of modern critical systems, for example, modern transportation systems, manufacturing systems, chemical and power plants. While the large scale of these systems calls into question the ability of a monitor to deliver consistent monitoring over an architecture that may integrate thousands of components, dynamic behaviour mainly calls into question the ability of a monitor to distinguish between normal and abnormal operational conditions. More specifically, what is considered as normal in one mode or phase of operation of the system may simply be abnormal in another mode. A typical example of a "phased mission" system is an aircraft system which delivers a trip mission through a number of phases, which include pre-flight, taxiing, take-off, climbing, cruising, approaching, and landing. Thorough knowledge about the architectural components and the dynamic behaviour in each phase is essential to achieve effective safety monitoring.

In order to model the mutual relations among sub-systems and components in a system model, a hierarchical organisation is commonly used to arrange them in a number of hierarchical levels. Across those levels components appear as parents, children and siblings. As shown in Fig. 1, we classify those levels into three different types as follows: the lowest level (level*0*) is classified as the basic components (BC) level. The upper levels, which extend from level*1* to level*n-1*, are classified as sub-system (Ss) levels. Finally, the top level (level*n*) is classified as the system (S) level.

In order to model dynamic behaviour, one needs to understand the behaviour itself and how it is initiated. Typically, dynamic behaviour is an outcome of, normal operational conditions in which the system engages its components in different operational functions and structures, so that it can deliver different functionalities in different phases of operation. Given that sub-systems are abstractions that represent aggregations of BCs, signals upon which that structure of the system is altered are always initiated by BCs (even operators will initiate changes through components in a graphical user interface or hardware panel). Typically, upon a signal from a BC, a system controller may instruct other BCs to be engaged in a certain structure and deliver certain functions in collaboration. For example, during the cruising of an aircraft, the navigation sensors may convey signals to the navigator sub-system (NS) which in turn calculates and passes those signals to the flight control computer sub-system (FCCS). Assuming that it is time for launching the approach, the FCCS accordingly instructs the powerplant sub-systems to achieve the required thrust and the surface hydraulic controller sub-system to achieve the required body motions. Accordingly, we define the case in

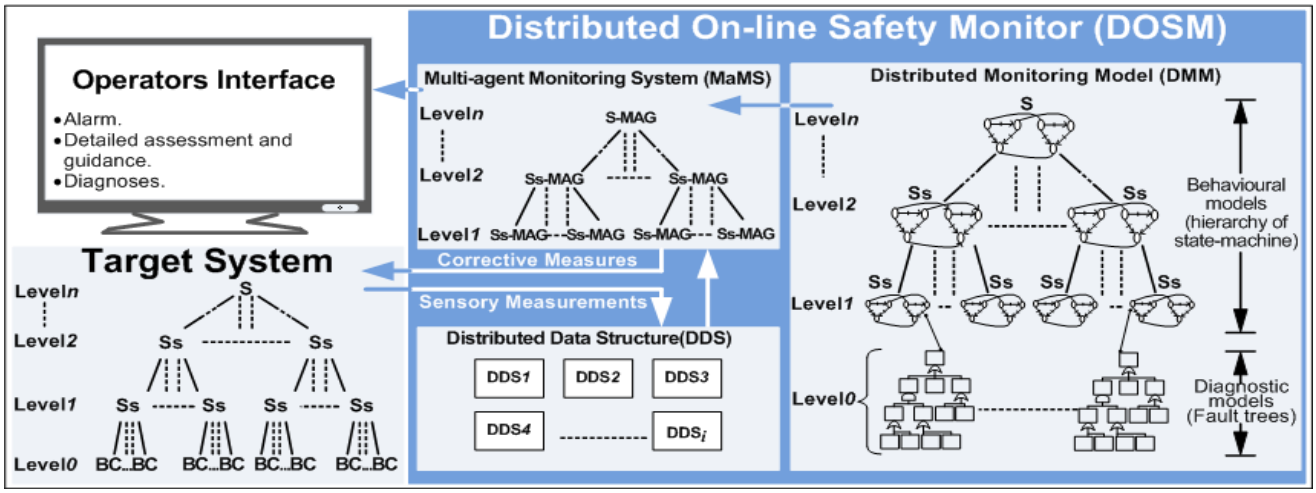


Figure 1. Target System and DOSM Position and Basic Constituents.

which the system uses a certain operational structure to deliver certain functionality as a *mode*.

Beyond being the result of normal changes in function and structure, dynamic behaviour also arises from the need to respond to and tolerate the faults of basic components. Fault tolerance is typically achieved using the following strategies: (a) recovery from a permanent fault usually achieved with functional or hardware redundancy, e.g. the fault of one engine of a two-engine aircraft can be compensated by the other engine; (b) the ability of tolerating the fault for a while until recovery of a healthy state can be achieved, e.g. temporary faults that are caused by ionisation, radiation, electromagnetic interference, or transient hardware failures, are often self-correcting, while certain types of tolerable software faults may be corrected with a controller restart.

It could, therefore, be said that during a mode, a system may appear in different health *states* which can be classified into two types. The first type is the *Error-Free State* (EFS) in which the system or a sub-system functions healthily. The second type is the *Error State* (ES), which in turn is classified into three different states: (a) Temporary Degraded or Failure State (TDFS) in which there is one or more functional failure, but corrective measures can be taken to resume a healthy state; (b) Degraded State (DS) in which a permanent fault has occurred, but part of the intended functionality is still delivered; (c) Failed State (FS) in which the component or system has lost its entire functionality.

In order to track dynamic behaviour, events that result in the normal and abnormal behavioural changes should be continuously monitored. The hierarchical level at which such events could be monitored most effectively is level*1*, i.e. one level above the level of BCs. This can easily be justified, because at that level low level events can be contextualised and could be identified as either normal or abnormal. For instance, the decreasing of velocity and altitude seem normal during the approaching mode of an aircraft, since the FCCS has already launched that mode. Excluding knowledge about the mode and focusing only on the measurements provided by the relevant sensors would certainly result in misinterpreting system behaviour. Specifically, decreasing velocity and altitude would appear as a malfunction and thus a misleading alarm would be released. This being the case, level*1* is preferable rather than any higher level, since it is the level at which a malfunction is detected while in its early stages. Finally, due to the number of the basic components, which is potentially huge, monitoring and reasoning about those events at level*0* is computationally expensive or even unworkable, whereas level*1* offers the required context and knowledge of local mode.

III. DISTRIBUTED ON-LINE SAFETY MONITOR (DOSM)

As shown in Fig. 1, the DOSM lies between the target system and the interface of the system operators. During normal operation, the role of the DOSM is confined to providing simple feedback about those conditions. The DOSM plays its role during abnormal operating conditions, which are triggered by and follow the occurrence of faults. In that role the DOSM achieves three real-time time safety tasks: fault detection and diagnosis, optimising alarm annunciation and automatic control of faults. In order to achieve those tasks, the DOSM employs three elements (see also Fig. 1): (a) a DMM which holds the reference monitoring knowledge, in other words, the DMM references the MAGs which in turn reason and achieve the three safety tasks; (b) Distributed Data Structures (DDS), which hold the necessary sensory measurements used by MAGs in order to monitor operational parameters and reason on the operational conditions of the monitored system; (c) A MaMS which is a set of BDI agents that are deployed over the components of the system to reason locally and collaborate globally towards achieving the three safety tasks.

A. Distributed Monitoring Model (DMM)

MAGs should be, in the first place, able to track the operational behaviour of the monitored components over different states, i.e. EFSs and ESs. Accordingly, both normal and abnormal behaviour should be modelled and recorded in the DMM. For that purpose, state-machines provide the means of recording behaviour at all levels of the architectural hierarchical decomposition of the system (see Fig. 1.). Accordingly, the spine of the DMM is a hierarchy of state-machines that describes dynamic behaviour. In those state-machines, every EFS or ES is represented as a state and every event whose occurrence results in a state transition is represented as a trigger event.

Practically, there are relationships among every sub-system and its parent and child components. For instance, the failure of a component within a sub-system may trigger a transition of the sub-system in a recovery state where another component changes function to compensate for the initial failure. Such relationships can be implemented in the state-machines in a similar way to the following example:

Let us assume that the flight control computer sub-system (FCCS) and power plant sub-system (PPS) are siblings and have the same parent, the aircraft control sub-system (ACS). During the cruising mode, an event may trigger a state transition to EFS of the approaching mode in the state-machine of the FCCS. That EFS appears as a trigger event whose occurrence triggers a state transition in the state-machine of the ACS, i.e. the parent, to the EFS of the approaching mode. The latter EFS appears, similarly, as a trigger event whose occurrence triggers a state transition in the state-machine of the PPS, i.e. a child, to the EFS of the approaching mode.

Similarly, ESs of the children could also trigger state transitions in the state-machines of the parents and vice versa. Consider, for example, when an engine of a two-engine aircraft fails; the FS of that engine triggers a state transition to the DS in the state-machine of the PPS. That DS, in turn, triggers a state transition to new EFS of the operative engine in which the lost functionality of the faulty engine is compensated.

In the state-machine of the sub-systems of level*1*, trigger events appear as (a) events that are originated by the BCs of level*0*, which might be failure, corrective or normal events; (b) events that are originated by the parent states, such as the EFS or ESs of the parent. In the state-machine of a sub-

 ISBN: 978-1-61208-174-8

system of the levels extending from level*2* to level*n-1*, trigger events appear as EFSs and ESs of the parent and the children. Finally, in the state-machine of the system, i.e. level*n*, trigger events appear as EFSs and ESs of the children.

Knowledge about the normal behaviour, i.e. EFS and normal events, of the system and its sub-systems can be obtained from design models, such as Data Flow Diagrams (DFD), Functional Flow Block Diagram (FFBD) and models in the Unified Model Language (UML) that model the system during the design life cycle. Knowledge about abnormal behaviour, i.e. ESs, abnormal events, assessment, guidance, and corrective measures, can be obtained by applying the Functional Failure Analysis (FFA) or HAZard and OPerability study (HAZOP) techniques on those models.

During the monitoring time, MAGs monitor only trigger events whose occurrence triggers transitions from the current state in the state-machine. As such, the computational load of the MAGs would be less and prompt responses to the occurrence of the events would be obtained.

In the state-machines of the sub-systems of level*1*, every failure event would be associated with (a) an alarm statement that would be quoted and provided to the operators upon the occurrence of the failure event; (b) corrective measures that can be applied to control the failure; (c) diagnosis, if the failure and the underlying cause are in a one-to-one relationship the cause would be associated, otherwise a diagnostic process should take place. Note, some corrective measures might be achievable only after diagnosing the underlying causes. In the state-machines of higher level sub-systems, normal and abnormal events are associated with a field of (a) assessment of the consequent operational conditions; (b) guidance on directing the hazards at that level. Knowledge of those fields can be obtained from the HAZOP.

A failure event and its underlying cause might not always be in a one-to-one relationship. Therefore, a diagnostic model that can relate failure events to their underlying cause is needed. The fault tree, a popular model used in safety assessments, can be used as a diagnostic model as it logically records the propagation paths and the associated symptoms of failure a long with underlying causes. In HiP-HOPS, fault trees are automatically constructed from the topology of a system and local failure logic specified at component level. This method can be applied to construct diagnostic fault trees for failure events that appear as trigger events in the state-machines of level*1* sub-systems. Corrective measures could also be incorporated in the failure mode nodes of the fault tree.

As shown above, knowledge encoded in the DMM is obtained from the design models and by applying classical manual safety analysis techniques (FFA, HAZOP, FMEA, Fault Tree analysis) [12] or more modern semi-automatic safety analysis techniques (HiP-HOPS) [5]. Hence, it could be said that a safety assessment model could be useful to derive a DMM after (a) associating the abnormal events with the alarm, controlling and diagnosis knowledge; (b) augmenting the states of the state-machines by assessment and guidance fields and the diagnostic model with the required corrective measures; (c) formalising the trigger events of the state-machine and the symptoms of the diagnostic model as monitoring expressions that could be evaluated computationally in real time. The deriving process would contribute essentially to providing the DOSM with thorough and consistent monitoring knowledge. Note that in this paper we adopt the HiP-HOPS as a safety assessment tool to produce the DMM.

B. Formal Monitoring Expressions and Distributed Data Structure (DDS)

Low level events that monitor the physical process and trigger state transition in the state-machines at Level*1*, should be formalised as monitoring expressions that reference parameters of the physical process. Through evaluating those expressions, the occurrence of the corresponding trigger events or symptoms could be verified. In the formalisation process, an event or a symptom is expressed as a constraint. In its simple form, a constraint consists of three main parts: (a) the status of operational conditions which is either a state of a child or the parent or a sensory measurement defined by the identifier of the relevant sensor; (b) a relational operator – equality or inequality; (c) a threshold whose violation results in evaluating that expression with a true truth value, i.e. the relevant event or symptom occurs. Thresholds might appear as a numerical or Boolean value.

Simple constraints may suffice for simple monitoring tasks. In general, though, events may require more complicated forms of constraints to be evaluated. In turn, such constraints might require (a) the status of a parameter to be calculated over a number of sensory measurements; (b) two operational operators, when the threshold is a range of values rather than a single value; (c) a threshold that represents a sensory measurement or a calculation of more than one measurement. Moreover, the status of parameters and the threshold might be calculated to find the average of the change of a quantity over an interval (Δt), i.e. differentiation, or the volumes from different sensory measurements at definite timings, i.e. integral calculus.

For the evaluation process of such monitoring expressions, we pre-declare a number of data structures that could hold satisfactory sensory measurements and the result of the calculation and the evaluation process. For every sub-system of level*1*, a DDS would be allocated to hold those structures; as shown in Fig. 1.For holding historical sensory measurements we use an updatable buffer of one-dimension array data structure that could hold two or more up-to-date sensory measurements. Such a structure is updated every Δt by (a) inserting the current measurement, which is collected at the current time (T) from the relevant sensor; (b) shifting out the earliest measurement, which is collected at T-2Δt in the past. As such, that structure holds two (or more) measurements collected at current time T and T- Δt in past.

Sensors may deliver spurious measurements because of (a) their own transient failures; (b) mode changes, which might be followed by an interval of unsteady behaviour in which the monitored parameter may temporarily fluctuate outside normal thresholds. One way of filtering out such spurious sensory measurements is by evaluating the

monitoring expressions successively over a filtering interval and based on a number of measurements. The final result of that evaluation is obtained by making accumulative conjunctions among the successive evaluations. If the final result is s true truth value, this means that the delivered measurements remain the same over the filtering interval, which is a confirmation that a parameter is persistently out of threshold and a sign of a persistent anomaly present - as opposed to a spurious measurement or a transient anomaly. The filtering interval of every expression is defined by examining both the conditions that may result in spurious measurements and the time intervals at which the involved sensors are requested by the monitor.

A three-value technique: 'True', 'False', and 'Unknown', is also employed to save evaluation time and produce earlier results in filtering spurious measurements and in the context of incomplete sensory data without violating the evaluation logic. Consider, for example, the following two expression forms:

Expression OR (Expression, ⊿t) (1)
Expression AND (Expression, ⊿t) (2)

Evaluating either of those expressions; (1) or (2), may require waiting time equal to Δt, i.e. until evaluating *(Expression, ⊿t)* part, regardless of the instant evaluation of the '*Expression'* part of either of the expressions. Knowing that the disjunction of 'True' with 'Unknown' is 'True' and the conjunction of 'False' with 'Unknown' is 'False', both expressions; (1) and (2), can be evaluated instantly. Therefore, in cases in which the '*Expression'* part of expression (1) is evaluated to 'True' and the '*Expression'* part of expression (2) is evaluated to False, both (1) and (2) could be evaluated instantly to 'True' and 'False', respectively.

C. Multi-agent Monitoring System (MaMS)

As shown in Fig. 1, MAGs are deployed over the sub-systems and the system, and appear as a number of subsystem MAGs (Ss-MAGs) and a system MAG (S-MAG), respectively. Fig. 2 shows a general illustration of the MAG. By perceiving the operational conditions and exchanging messages with other MAGs, a MAG obtains the up-to-date belief, deliberates among its desires to commit to an intention and achieves a means-ends process to select a course of action, i.e. plan. The selected plan is implemented by the MAG as actions towards achieving the monitoring tasks locally and as messages sent to other MAGs towards achieving those tasks globally. Upon having a new belief, MAG achieves a reasoning cycle; deliberation and means-ends processes.

Each Ss-MAG of level*1* would have its perception by perceiving (a) its own portion of the DMM which consists of a state-machine and a set of fault trees; (b) the corresponding DDS in which events and symptoms appear as expressions and are evaluated; (c) messages that are received from the parent to inform the Ss-MAG about the new states and the siblings, in which they either ask for or tell the given Ss-MAG about global sensory measurements, as they share their DDSs whenever needed. The main desires of a Ss-MAG of level*1* are to achieve local safety monitoring tasks and global collaboration and coordination. On the former desire, the intentions are to track the behaviour of the assigned sub-system and to provide the operators with alarms, assessment, guidance, and diagnostics and achieve automatic fault controlling. On the latter desire, the intention would be achieved by (a) informing the parent about the new states; (b) telling or asking the siblings about global sensory measurements.

Each Ss-MAG of the levels extending from level*2* to level*n-1*, would have its perception by (a) perceiving its own portion of the DMM which consists of a state-machine of the assigned sub-system, and (b) messages received from the parent and the children to tell about their new states. The main desires of the Ss-MAGs of those levels are to achieve local safety monitoring tasks and global collaboration. On the former desire, the intentions are to track the operational behaviour of the assigned sub-system and to provide the operators with assessment and guidance of their levels. On the latter desire, the intention would be achieved by telling, i.e. sending messages to, the parent and the children about the new states. The perceptions, desires and intentions of the S-MAG are similar to those of the Ss-MAGs of the levels extending from level*2* to level*n-1*. The only difference is that S-MAG has no parent to exchange messages with.

According to the Prometheus approach and notation for developing MaS [13], Fig. 3 shows the collaboration protocols among MAGs to track the operational behaviour of the monitored system. Fig. 4, similarly, shows the collaboration protocol among the Ss-MAGs of level*1* in which they share their sensory measurements globally.

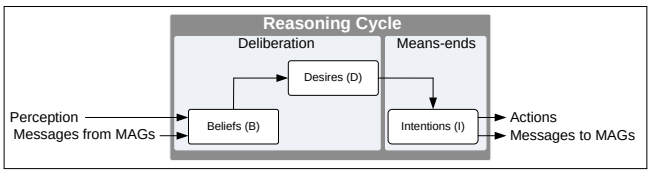


Figure 2. A general illustration of the MAG.

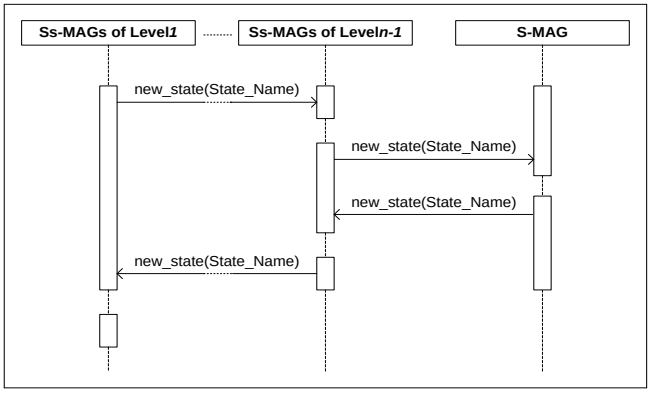


Figure 3. MAGs' collaboration protocol across the hierachical levels.

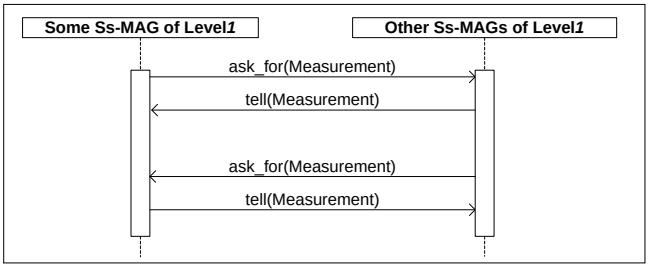


Figure 4. Collaboration protocol among Ss-MAGs of level*1*.

IV. CASE STUDY: AIRCRAFT FUEL SYSTEM (AFS)

Fig. 5 shows the physical illustration and the basic components of the AFS. The AFS functions to maintain safe storage and even distribution of fuel in two operational modes. The first is the consuming mode in which the AFS provides fuel to the port and starboard engines of a two-engine aircraft. The second is the refuel mode. During the consuming mode, and to maintain the central gravity and stability, a control scheme applies a feedback-control algorithm to ensure even fuel consumption across the tanks.

Another algorithm is applied similarly to control the even distribution of fuel injected from the refuelling point to the tanks during the refuel mode. The AFS is arranged in four sub-systems: a central deposit (CD), left and right wing (LW, RW) deposits and an engine feed (EF) deposit which connects fuel resources to the two engines.

In order to tolerate faults, an active fault-tolerant controller strategy is implemented. More specifically, in the presence of faults there are alternative flow paths, i.e. different configurations can potentially connect the two engines to the available fuel resources.

As shown in Fig. 6, five monitoring agents are deployed over the AFS as follows: four MAGs monitor the four sub-systems; EF-MAG, CD-MAG, LW-MAG, and RW-MAG. The fifth is AFS-MAG which monitors the entire FS. The DOSM is implemented by Jason interpreter; it is an extended version of AgentSpeak programming language [14].

In order to achieve fault detection, the four MAGs update their DDSs and evaluate the monitoring expressions and thus detect any parametric deviations at level*1*. Consider, for example, the deviation "no fuel flow to starboard engine"; this deviation could be detected locally by the EF-MAG, while the deviation of imbalance between the LW and RW deposits is detected through communicating sensory measurements globally between LW-MAG and RW-MAG.

Diagnosis of the underlying causes of a deviation is triggered when a deviation is detected. Through exploiting fault tree models and combining between depth-first and heuristic parse strategies, Ss-MAGs traverse and relate the top event in a tree to its bottom faulty components. Consider, for example, the deviation "no fuel flow to starboard engine". EF-MAG evaluates symptoms in the relevant fault tree to track the propagation path. The expected diagnosed causes could be one or more of the fault modes of EF basic components; likely, a pipe blockage, an inadvertent closure or a fault of a valve, a pump fault, or no fuel in the rear tank.

To achieve automatic fault controlling, corrective measures are provided across the DMM, some of which could be taken directly by a MAG and others may require global collaboration. Consider, for example, when the deviation "no fuel flow to starboard engine" is diagnosed with the cause of an inadvertent closure of valve VF5; the possible corrective measure that can be taken locally by EF-MAG is to instruct the controller to reopen that valve. However, if that fails to rectify the situation, then a global action should be taken through the following steps: (a) EF-MAG transits to a FS and tells the AFS-MAG about that state; (b) AFS-MAG, in turn, executes that state on its state-machine, achieves the corresponding transition and tells the four MAGs about the resulted state; (c) the four MAGs, in turn, execute the received state on their state-machines. According to their new states the four MAGs apply new flow rates for every sub-system. Thus, the FSA configuration will be changed and both engines will be fed from the front tank.

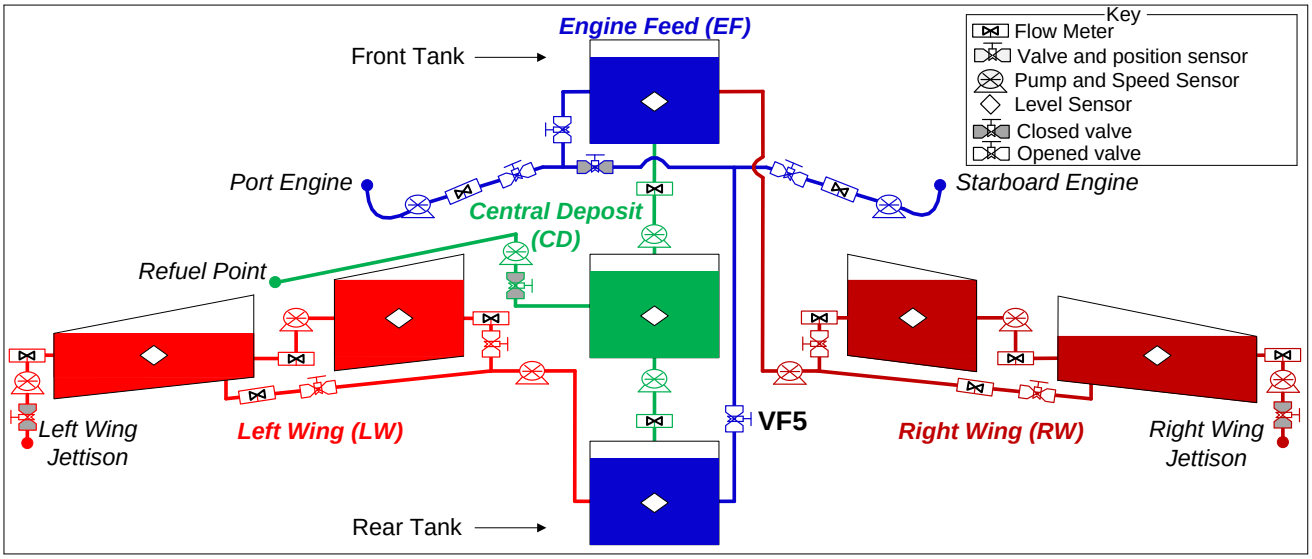


Figure 5. Physical illustration of aircraft fuel system.

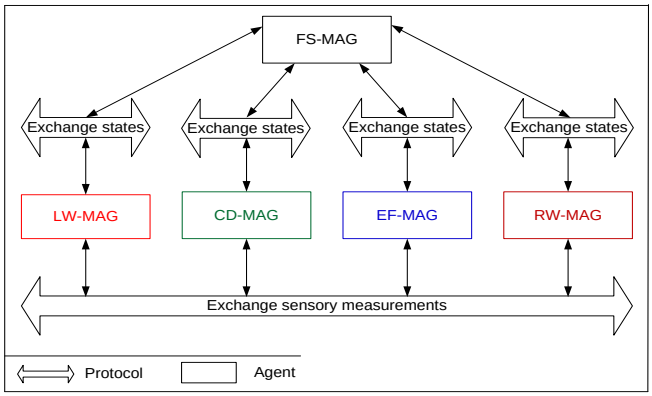


Figure 6. Architecture of the MaMS of the AFS.

When the occurrence of the failure event "no fuel flow to starboard engine" is verified, EF-MAG alarms and provides the pilots with the corresponding assessment and guidance. Moreover, as that failure triggers a state transition in the state-machine of the EF sub-system, which in turn triggers a transition in the state-machine of the AFS, assessment and guidance from the new state of the AFS would also be provided to the crew pilots, i.e. multi-level assessment and guidance.

V. CONCLUSION AND FUTURE WORK

This paper proposed a distributed on-line safety monitor (DOSM) based on a multi-agent system and knowledge derived from model-based safety assessment. Agents exploit that knowledge to deliver a range of real-time safety monitoring tasks which have been briefly discussed in the context of a study of an aircraft fuel system. The monitor can detect symptoms of failure on process parameters as violations of simple constraints, or deviations from more complex relationships among process parameters, and then diagnose the causes of such failures. With appropriate timed expressions, the monitor can filter normal transient behaviour and spurious measurements. By exploiting knowledge about dynamic behaviour, the monitor can also determine the functional effects of low-level failures and provide a simplified and easier to comprehend functional view of failure. Finally, by knowing the scope of a failure, the monitor can apply successive corrections at increasingly abstract levels in the hierarchy of a system.

Despite encouraging results certain research issues remain to be investigated. The first is that the quality of the monitoring tasks and the correctness of the inferences drawn by the monitor depend mainly on the integrity and consistency of the DMM. The validation of the DMM, therefore, is an area for further research. Secondly, more work is needed on uncertainty of the diagnostic model and the application of the three-value logic. For that purpose, the incorporation of Bayesian Networks will be investigated in the future.

VI. ACKNOWLEDGEMENT

This work was supported partly by the EU Project MAENAD (Grant 260057).

REFERENCES

[1] A. Dheedan and Y. Papadopoulos, "Multi-Agent Safety Monitoring System," Proc. of 10th IFAC Workshop on Intelligent Manufacturing Systems (IMS'10), Portugal, Lisbon, 1-2 July 2010, pp. 93-98.

[2] V. Venkatasubramanian, R. Rengaswamy, K. Yin and N. S. Kavuri, "A review of process fault detection and diagnosis: Part II: Qualitative Models and Search Strategies," Computers and Chemical Engineering 27(3), 2003, pp. 313-326.

[3] I. Monroy, R. Benitez, G. Escudero and M. Graells, "A semi-supervised approach to fault diagnosis for chemical processes," Computers and Chemical Engineering, vol. 34(5), 2010, pp. 631-642.

[4] Y. Papadopoulos, "Model-based system monitoring and diagnosis of failures using state-charts and fault trees," Reliability Engineering and System Safety, vol. 8(3), 2003, pp. 325-341.

[5] Y. Papadopoulos, J. McDermid, R. Sasse and G. Heiner, "Analysis and synthesis of the behaviour of complex programmable electronic systems in conditions of failure," Reliability Engineering & System Safety, 71 (3), 2001, pp. 229-247.

[6] M. Mendes, B. Santos, and J. Costa, "A matlab/Simulink multi-agent toolkit for distributed networked fault tolerant control systems," Proc. 7th IFAC symposium on Fault Detection, Supervision and Safety of Technical Processes, 30 June - 3 July 2009, in Barcelona, Spain, pp. 1073-1078.

[7] Y. S. Ng, and R. Srinivan, "Multi-agent based collaborative fault detection and identification in chemical processes," Engineering Applications of Artificial Intelligence, vol 23(6), 2010, pp 934-949.

[8] V. Venkatasubramanian, R. Rengaswamy, K. Yin and N. S. Kavuri, "A review of process fault detection and diagnosis: Part III: Process History based methods," Computers and Chemical Engineering, 27(3), 2003, pp. 327-346.

[9] C. J. Wallace, G. J. Jajn and S. D. J. Mcarthur, "Multi-Agent System for Nuclear Condition Monitoring' Proc. of 2nd International Workshop on Agent Technologies for Energy System (ATES'11), a workshop of the 10th International Conf. of Agent and Multi-agent System (AAMAS'11), 2nd of May 2011, in Taipei, Taiwan, pp. 17-23.

[10] E. J. Trimble, "Report on the Accident to Boeing 737-400 G-OBME near Kegworth, Leicestershire on 8 January 1989" Department of Transport Air Accidents Investigation Branch, Royal Aerospace Establishment. London, HMSO, 1990. Available from http://www.aaib.gov.uk/cms_resources.cfm?file=/4-1990%20G-OBME.pdf [Accessed 5th of March 2011].

[11] U.S.NRC, "Fact sheet: Backgrounder on the Three Mile Island Accident," United State nuclear regulatory commission (U.S.NRC), 2008. Available from: http://www.nrc.gov/reading-rm/doc-collections/fact-sheets/3mile-isle.pdf [Accessed 5th of March 2011].

[12] D. Pumfrey, The Principled Design of Computer System Safety Analyses, DPhil Thesis. University of York, 1999.

[13] L. Padgham, and M. Winikoff, Developing Intelligent Agent Systems: a Practical Guide. UK, Chichester:Wiley, 2004.

[14] R. Bordini, J. Hubner, and M. Woorldridge, Programming Multi-Agent Systems in AgentSpeak using Jason. UK, Chichester: Wiley, 2007.

 ISBN: 978-1-61208-174-8

Distributed Service Discovery Architecture: A Bottom-Up Approach with Application Oriented Networking

Haja M. Saleem*
Computer and Information Sciences
Universiti Teknologi PETRONAS
Perak, Malaysia
*Dual Affiliation with Faculty of Information and Communication Technology
Universiti Tunku Abdul Rahman
Perak, Malaysia
saleem@utar.edu.my

Mohd Fadzil Hassan[1], Vijanth S. Asirvadam[2]
Computer and Information Sciences[1], Fundamental and Applied Science[2]
Universiti Teknologi PETRONAS
Perak, Malaysia
{mfadzil_hassan[1], vijanth_sagayan[2]}@petronas.com.my

***Abstract*— Peer-to-Peer service discovery is the norms of today's Service Oriented Architecture. Efficiency and scalability of these systems are adversely affected by the type of distributed architecture, the query routing mechanism and the effective usage of underlying network topology. Traditional query routing mechanisms in distributed P2P systems function purely at the overlay layer by isolating itself from the underlying IP layer that degrades the performance due to large amount of inter-ISP traffic and unbalanced utilization of underlay network links. In this paper we address this problem by proposing novel distributed service discovery architecture, which enhances underlay awareness without the involvement of the overlay peers. Our design starts from the underlay layer, which is built on top of Application Oriented Networking (AON) backbone that exploits message level routing. This feature is further leveraged in the overlay layer with industry based classification of published services, which complements the process of message level routing. We present the conceptual design of our framework and analyze its effectiveness. We argue that both performance and scalability of the system are drastically improved by moving down the overlay query routing mechanism to the IP layer.**

Keywords-Web services; service discovery; AON; P2P; multicasting; clustering; SOA.

I. INTRODUCTION

In Service Oriented Architecture (SOA) the complexity of service composition increases proportionately with the increase in number of services. Efficient service discovery is one of the key aspects in automating the service composition process. Initially, SOA started with centralized service discovery. As more and more services are made available both from within and outside organizations, the centralized service discovery proved to be unsuccessful in terms of scalability and single point of failure [1], which paved the way for distributed approach.

Many approaches have been proposed earlier for the distributed service discovery which has its roots from Peer-to-Peer (P2P) file sharing systems. Amongst various P2P approaches, only few are suitable to be implemented in the service discovery domain, as their target is file sharing applications, where file download efficiency is one of the major concerns. However this is not the case for the service discovery where other constrains such as range queries, service cost and multiple matches are taken into account. The current P2P systems have been mainly classified into three categories; unstructured, structured and hybrid. The main shortcoming of the unstructured systems is their scalability, whereas the structured systems are prone to complex administration and poor performances in dynamic environment [2]. On the other hand hybrid systems are focused towards key mapping mechanisms which are inclined towards the tightly controlled structured approach. In this paper our focus is towards unstructured systems which are widely deployed due to their flexibility with dynamic entry and exit options for the peers.

Currently, most of the query routing mechanisms are implemented in the overlay layer which result in IP-oblivious query forwarding. This leads to three major problems. First, it heavily increases the inter-ISP network traffic [3], which is expensive for the service providers. Second, the network links are not loaded in a balanced manner which ends up with poor performance and thirdly it introduces interlayer communication overhead. To alleviate these problems several contributions have been made in making the peers underlay aware while choosing their neighbors. However, these solutions just provide the network knowledge to the peers in the overlay and let the peers decide on their own [4]. Letting the peers aware of network related parameters may lead to privacy and security issues both for the peers and the ISPs.

To this end, our contribution in this paper focuses on enhancing underlay awareness without involving the overlay peers; and to implement IP layer multicasting with message

 ISBN: 978-1-61208-174-8

level intelligence. We argue that the performance can be drastically improved if the search mechanism is moved down to the underlay layer and seamlessly integrated with the Internet protocol (IP) layer.

We also argue that our architecture enhances the following characteristics of distributed service discovery, which are lacking in the current systems,

1. Non-involvement of peers in the locality aware query forwarding that results in improved efficiency.
2. Increased peer privacy.
3. Increased response time with the elimination of interlayer communication overhead.

The rest of the paper is organized as follows. Section 2 discusses the related work, Section 3 demonstrates our design and analyzes the performance and Section 4 concludes the paper with future work.

II. RELATED WORK

Various approaches have been proposed and investigated towards improving the network layer awareness in query routing mechanisms.

TOPLUS [5] organizes peers into group of IP addresses and uses longest prefix match to determine the target node to forward the query. Here the peers use an API in the application layer to find the neighbor to forward the query. Their scope is not moving the routing decision functionality to the IP layer.

PIPPON [6] is closer to our effort in trying to match the overlay with the underlying network. The clustering mechanism in the overlay layer of PIPPON is based on network membership and proximity (latency). However, the similarity of the services provided is not taken into consideration in the cluster formation. Moreover, it ends up in a highly structured system with high administrative overhead.

The contribution made in [7] is a technique called *biased neighbor selection*. This technique works by selecting a certain number of neighboring peers within the same ISP and the remaining from different ISPs. This helps in reducing the inter-ISP traffic. This approach is well suited for file sharing systems like BitTorrent. However, the neighbors still functions at the overlay layer.

In [3], authors have discussed the problem space for the Application Layer Traffic Optimization (ALTO) working group, which is initiated by IETF. This approach allows the peer to choose the best target peer for file downloading by querying the ALTO service which has static topological information from the ISP. Here the ALTO service is provided as a complementary service to an existing DHT or a tracker service. The problem space here is the final downloading of the file and not the query search mechanism itself.

A framework that is used for conveying network layer information to the P2P applications has been proposed by P4P [4]. Peers make use of *iTrackers* and *appTrackers* to obtain the network information. The network information is used for the overlay query routing mechanism. However, the scope of the work is not in moving the query routing to the network layer which is the focus of our contribution.

Plethora [8] proposes a locality enhanced overlay network for semi-static peer to peer system. It is designed to have a two-level overlay that comprises a global overlay spanning all nodes in the network, and a local overlay that is specific to the autonomous system (AS). This is highly structured and is not suitable for highly dynamic environments.

There has been substantial contribution made in clustering as well. One such recent contribution is [9]. Our contribution contrasts with this and all others in making network provider based clustering, which aids in reduction of number of super registries that needs to be handled by Application Oriented Networking (AON) multicasting.

Deploying message level intelligence in network layer multicasting and dealing with QoS requirements in the service discovery domain are discussed in [10-12]. In [13], authors have initiated the discussion of employing AON in the service discovery but have not given a concrete implementation model, which is where our contribution fits in. The increasing trends in deployment of AON in other areas of SOA are provided in [14].

III. ARCHITECTURE AND DESIGN

A. Design goals

The following goals are considered in our design.

1. **Enhanced Security:** To enhance the security of the discovery system, network aware query routing is delegated to the underlying layer and is kept transparent to the overlay layer.
2. **Reduced inter-ISP traffic:** Another design focus is to let the query forwarding traffic enter the ISP domain only if it hosts the targeted service registry.
3. **Minimized overlay process overhead:** To eliminate the involvement of intermediate registries (peers) in query processing.
4. **Interoperability:** To integrate seamlessly with non-AON routers, if encountered, along the path of query forwarding.

To implement these goals a conceptual framework has been designed as shown in Figure 1. In layer 2 AON is employed for carrying query messages to the targeted peers with the help of message level intelligence. As redundant query forwarding in the underlay is minimized by AON with message level multicasting, the performance gain is very close to the IP level multicasting [15]. AS-based clustering and service classifications are implemented at layer 3, which leverages the message level multicasting by AON.

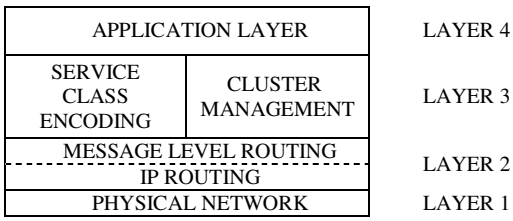


Figure 1. Layers of P2P service discovery.

B. Registry Clustering

Our approach in clustering is with respect to AS so as to reduce the inter-ISP traffic. A super registry (SR) is elected in each AS as shown in Figure 2 which demonstrates the

connectivity of SRs to the underlying physical network. The SRs are responsible for accepting queries for the whole AS.

The services published in the registries are classified in accordance with Global Industry Classification Standard (GICS) [16]. Our architecture uses these coding for the implementation of AON routing at the underlying layer. A sample of GICS industry classification is shown in Table 1.

This AS-based SR approach leverages the following characteristics of our system.

1. Enables the AON router in layer 2 to learn the query forwarding interface(s) that are specific to particular class of services.
2. Improves the scalability and dynamism of the system as new registries can enter and exit the cluster with minimal overhead.
3. AON routes learned by the routers are restricted to SRs which minimize the routing entries.

We also propose to use crawling technique [17] to update the entries in the super registries so that queries forwarded within the AS could be minimized as well.

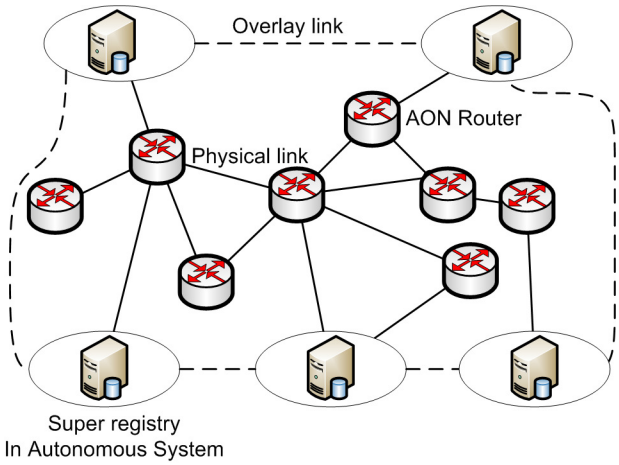


Figure 2. Autonomous system based clustering of service registries.

TABLE 1. SAMPLE INDUSTRY TYPES AND THEIR CODES

Industries	Class codes
Air Freight & Logistics	20301010
Airlines	20302010
Hotels, Resorts & Cruise	25301020
Leisure Facilities	25301030
Restaurants	25301040

C. AON Implementation

AON routers are capable of taking routing decisions based on application level messages [13]. We find this feature fits quite nicely into the distributed service discovery. Any query generated from an AS needs to be forwarded to the super registry in the AS which has an interface for classifying the query into one or more of its service classes. Then the message is constructed by encapsulating the query and its class and forwarded to the neighbors in the overlay routing table. Our packet structure in the IP layer is designed to record the interface(s) of the intermediate routers through which a particular router has received its query and reply, along with the intended source and destination IP addresses. The AON router uses this feature to inspect and update these fields and its AON specific routing table which is used for query multicast.

Possible scenarios that could be encountered during query forwarding are depicted in Table 2.

TABLE 2. SCENARIOS ENCOUNTERED IN QUERY FORWARDING

Router	Packet	Remark
AON	AON	Routing based on message level intelligence
AON	Non-AON	Classical routing based on IP header
Non-AON	AON	Classical routing based on IP header
Non-AON	Non-AON	Classical routing based on IP header

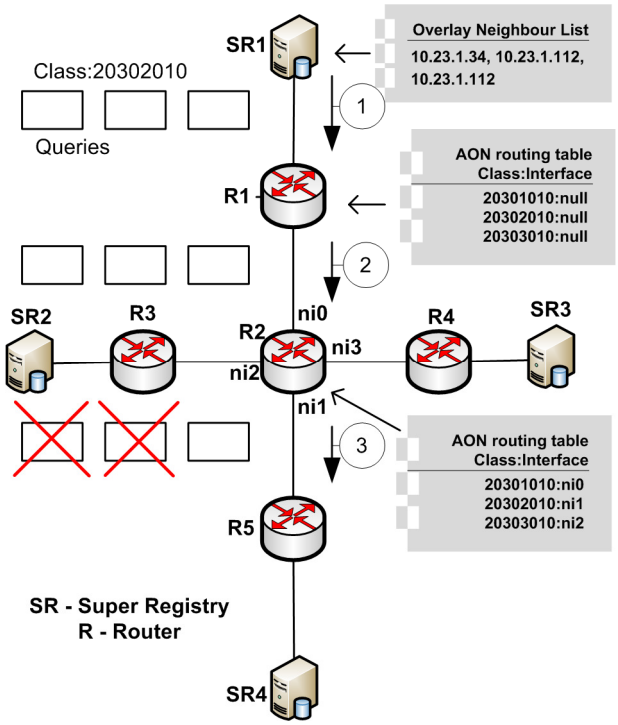


Figure 3. Sample scenario – Query forwarding.

D. Query forwarding

Figure 3 depicts an illustrative scenario of four ASs each with its own super registries connected via AON and non-AON routers. A query forwarded from an AS is received by the border routers of ASs, in this case router R1. Router R1 uses classical routing either if it is a non-AON router or an AON router in bootstrapping state. During the bootstrapping process the AON router would not have learned the service-classification specific routing and hence uses the IP header for the classical routing. Figure 3 demonstrates a sample query

 ISBN: 978-1-61208-174-8

forwarding from SR1 to SR2, SR3 and SR4 as per its overlay routing table (neighbor list). However, AON-router R2 inspects the query and finds that this query could be answered by SR4 alone as per its AON routing table. Hence it forwards the query via ni1 and drops the other queries intended for SR2 and SR3, which are connected via ni2 and ni3, respectively.

E. Analysis

In a pure overlay-based routing, for instance, Gnutella like systems, considering the worst case scenario (Flooding), query from SR1 to SR2, SR3 and SR4 would generate traffic along the paths SR1→SR2, SR1→SR3, SR1→SR4, SR2→SR4 and SR3→SR4 each having four physical links from the source to the destination. However, if AON routing is employed the traffic generated is just along the path SR1→SR4. This clearly illustrates that ineffective query propagation could be effectively limited by AON to improve the efficiency of search mechanism. The same can be visualized in terms of inter-ISP traffic as well. In our illustration the only inter-ISP traffic is from the source to the AS in which the target SR resides. Performance can also be improved in case of other query forwarding heuristics like random or probabilistic selection of neighbors which is summarized in Table 3.

TABLE 3. PERFORMANCE ANALYSIS FOR THE GIVEN SCENARIO

Query propagation method	No. of peers involved	No. of paths involved
Flooding (select all 3 neighbors)	4 (SR1,SR2, SR3,SR4)	5 (SR1→SR2, SR1→SR3, SR2→SR4, SR3→SR4, SR1→SR4)
Random/Probabilistic(selects 2/3 neighbors)	3 (SR1, SR2, SR4)	3 (SR1→SR2, SR1→SR4, SR2→SR4)
AON based	2 (SR1, SR4)	1 (SR1→SR4)

F. Current issues

1. **Security issues:** It is possible that routers could be compromised and mislead query forwarding which could result in performance degradation. In our design the system functions even if some routers along the path are non-AON-routers. In the event of an attack the ISP could detect it and switch the respective router(s) to classical routing until the attack is neutralized.
2. **Router performance:** There could be overhead in the router which processes the AON packets and in maintaining the second routing able. However, we argue that with tremendous increase in processing power and memory capacity of current routers, this issue can be resolved.
3. **Involvement of ISPs:** It needs to be studied that how ISPs could be encouraged to provide AON service. The reduction of cost due to reduced inter-ISP traffic could be the incentive.
4. **File sharing and downloading:** Our focus in this paper has been in resource discovery process. Its applicability in file sharing and downloading such as BitTorrent like systems need to be studied.

IV. CONCLUSION AND FUTURE WORK

We have introduced underlay-aware distributed service discovery architecture with message level intelligence and analyzed its effectiveness in terms of privacy and security of peers in the overlay, efficient query forwarding, scalability and performance. Currently a mathematical model for a full scale system is being developed. The system is also being modeled using J-Sim, a Java based network modeling and simulation tool. In future, we are planning to simulate the system with real case studies and compare its performance with relevant literatures as in [5-7].

REFERENCES

[1] P. T. Michael P. Papazoglou, Schahram Dustdar, Frank Leymann and Bernd J. Krämer, "Service-Oriented Computing Research Roadmap," Dagstuhl Seminar Proceedings 05462, Service Oriented Computing (SOC), 2006, http://drops.dagstuhl.de/opus/volltexte/2006/524, retrived: September, 2011.

[2] E. Meshkova, J. Riihijärvi, M. Petrova, and P. Mähönen, "A survey on resource discovery mechanisms, peer-to-peer and service discovery frameworks," Computer Networks, vol. 52, pp. 2097-2128, 2008.

[3] J. Seedorf, S. Kiesel, and M. Stiemerling, "Traffic localization for P2P-applications: The ALTO approach," presented at P2P '09. IEEE Ninth International Conference on Peer-to-Peer Computing, pp. 171-177, 2009.

[4] H. Xie, Y. R. Yang, A. Krishnamurthy, Y. G. Liu, and A. Silberschatz, "P4p: provider portal for applications," in Proceedings of the ACM SIGCOMM 2008 conference on Data communication. Seattle, WA, USA: ACM, 2008, pp. 351-362.

[5] L. GarcÃ©s-Erice, K. Ross, E. Biersack, P. Felber, and G. Urvoy-Keller, "Topology-Centric Look-Up Service," in Group Communications and Charges. Technology and Business Models, vol. 2816, Lecture Notes in Computer Science, B. Stiller, G. Carle, M. Karsten, and P. Reichl, Eds.: Springer Berlin / Heidelberg, 2003, pp. 58-69.

[6] D. B. Hoang, H. Le, and A. Simmonds, "PIPPON: A Physical Infrastructure-aware Peer-to-Peer Overlay Network," presented at TENCON 2005 2005 IEEE Region 10, pp. 1-6, 2005.

[7] R. Bindal, C. Pei, W. Chan, J. Medved, G. Suwala, T. Bates, and A. Zhang, "Improving Traffic Locality in BitTorrent via Biased Neighbor Selection," presented at 26th IEEE International Conference on Distributed Computing Systems (ICDCS), pp. 66, 2006.

[8] R. Ferreira, A. Grama, and S. Jagannathan, "Plethora: An Efficient Wide-Area Storage System," in High Performance Computing - HiPC 2004, vol. 3296, Lecture Notes in Computer Science, L. Bougé and V. Prasanna, Eds.: Springer Berlin / Heidelberg, 2005, pp. 305-310.

[9] S. Xin, L. Kan, L. Yushu, and T. Yong, "SLUP: A Semantic-Based and Location-Aware Unstructured P2P Network," presented at 10th IEEE International Conference on High Performance Computing and Communications (HPCC), pp. 288-295, 2008.

[10] D. A. Menasce'; and L. Kanchanapalli, "Probabilistic scalable P2P resource location services, http://doi.acm.org/10.1145/588160.588167," SIGMETRICS Perform. Eval. Rev., vol. 30, pp. 48-58, 2002.

[11] D. Tsoumakos and N. Roussopoulos, "Adaptive probabilistic search for peer-to-peer networks," presented at Third International Conference on Peer-to-Peer Computing, (P2P 2003). pp. 102-109, 2003.

[12] V. Kalogeraki, D. Gunopulos, and D. Zeinalipour-Yazti, "A local search mechanism for peer-to-peer networks, http://doi.acm.org/10.1145/584792.584842 " in Proceedings of the

 ISBN: 978-1-61208-174-8

eleventh international conference on Information and knowledge management McLean, Virginia, USA ACM, 2002 pp. 300-307

[13] I. I. o. T. Yu Cheng, U. o. T. Alberto Leon-Garcia, and U. o. C. Ian Foster, "Toward an Autonomic Service Management Framework: A Holistic Vision of SOA, AON, and Autonomic Computing," IEEE Communications Magazine, vol. 46, pp. 138-146, 2008.

[14] "http://soa.sys-con.com/node/155657", retrived: September, 2011.

[15] X. Tian, Y. Cheng, K. Ren, and B. Liu, "Multicast with an Application-Oriented Networking (AON) Approach," presented at IEEE International Conference on Communications, ICC '08, pp. 5646-5651, 2008.

[16] "http://www.mscibarra.com/products/indices/gics/", retrived: September, 2011.

[17] E. Al-Masri and Q. H. Mahmoud, "WSCE: A Crawler Engine for Large-Scale Discovery of Web Services," presented at IEEE International Conference on Web Services, ICWS 2007, pp. 1104-1111, 2007.

 ISBN: 978-1-61208-174-8

The Interoperability Challenge for Autonomic Computing

Richard John Anthony
The University of Greenwich
Park Row, Greenwich
London SE10 9LS, UK
+44 (0) 208 331 8482
R.J.Anthony@gre.ac.uk

Mariusz Pelc
The University of Greenwich
Park Row, Greenwich
London SE10 9LS, UK
+44 (0) 208 331 8588
M.Pelc@gre.ac.uk

Haffiz Shuaib
The University of Greenwich
Park Row, Greenwich
London SE10 9LS, UK
+44 (0) 208 331 8588
H.Shuaib@gre.ac.uk

Abstract - **Interoperability is an emerging need for autonomic computing systems, which stems from the very success of these systems. Autonomic computing is increasingly popular; soon autonomic control components will be commonplace, and present in almost every large or complex application. This inevitably leads to situations where multiple autonomic components coexist and interact either directly or indirectly within the same application or system. Problems can arise when numerous *independently* designed autonomic components interact. We advocate a service-based approach to interoperability and present a set of requirements for such an approach. We briefly present a universal interoperability service which automatically discovers and manages potential conflicts between manager components.**

Keywords - Autonomic systems, Interoperability, Services

I. INTRODUCTION

Autonomic Computing (AC) is increasingly popular, and has become a mainstream concept. Autonomic components will soon be commonplace and it is inevitable that there will be an increasing trend of co-existence amongst autonomic managers. As there are currently no universal standards for autonomic systems design, or for the provision of interoperability amongst managers, there can be no guarantees that separately-designed managers will operate harmoniously together. Almost all systems use multi-vendor software solutions and this implies that there will be a great variety of potential manager components existing, even for any one specific function of a system. For many systems, autonomic management will arrive incrementally; as new functionality is introduced, and through upgrades of non-managed components to new managed versions. In some cases the introduction of management capabilities will not be obvious – third party developers may deliver components with internal management that is not exposed at interfaces to other components.

Any multi-manager scenario leads to potential conflicts. Direct conflicts occur where Autonomic Managers (AMs) attempt to manage the same explicit resource. Indirect conflicts arise when AMs control different resources, but the management effects of one have an undesirable impact on the management function of the other. This latter type of conflict is expected to be the most frequent and problematic, as there are such a wide variety of unpredictable ways in which such conflicts can occur. The effects of indirect conflict will also be less obvious to detect and harder to diagnose than the direct conflicts. The effects of conflicts can vary widely, including e.g., a cancellation effect of opposing managers, and serious performance or stability problems. The problem is illustrated with an example: consider a system with two AMs: a Power Manager (PM1) which shuts down servers that have been idle for a short time; and a Performance Manager (PM2) which attempts to maintain a pool of idle servers to ensure high responsiveness to high priority applications. Each service was developed and evaluated in isolation and both performed perfectly, however the respective vendors did not envisage that they would co-exist. Bringing a shutdown server back on line has a latency of several seconds, thus PM1's 'locally correct' behaviour defeats PM2's contribution. As each manager is unaware of the presence and behaviour of the other, the problem can only be resolved if an external agent (such as a human system manager) can detect, diagnose, and identify a solution to the problem.

The contributions of this paper include: firstly we evaluate the nature and scope of the interoperability challenge for autonomic systems and identify a set of requirements for a universal solution (section III). We present a work-in-progress service-based interoperability service which enables exploration of these requirements (section IV). Section V outlines a management description language which is intended for use by developers to ensure consistent description of AMs' management capabilities. Automatic detection of management conflicts is discussed in section VI. The interoperability service is evaluated in section VII and finally we conclude (section VIII).

II. BACKGROUND

A clear demonstration of the need for interoperability mechanisms is provided in [1] where two independently-developed autonomic managers were implemented. The first dealt with application resource management, specifically CPU usage optimization. The second, the power manager, was responsible for modulating the operating frequency of the CPU to ensure that the power cap was not exceeded. It was shown that without a means to interact, both managers throttled and sped up the CPU without recourse to one another, thereby failing to achieve their intended

 ISBN: 978-1-61208-174-8

optimisations and potentially destabilising the system. We envisage widespread repetition of this problem until a universal approach to interoperability is implemented.

Early work has focussed on bespoke interoperability solutions for specific systems. [2] proposes a distributed management framework that seeks to achieve system-wide Quality of Service (QoS) goals. Autonomic controllers are added and removed from the system based on applications' QoS requirements. The controllers communicate indirectly with one another using the system variables repository. If a controller were to fail, other controllers reading this repository take over the responsibilities of the failed controller. Other works take a more direct approach to autonomic element interaction. For instance, in [3] the autonomic elements that enable the proposed data grid management system communicate directly with one another to ensure that management obligations are met. The relationship between each type of autonomic element is peer-to-peer – potentially leading to high interaction complexity. In contrast, [4] adopts a three-level hierarchical relationship to autonomic element interactions. Individual autonomic elements form the lowest level of the hierarchy. Multiple devices are grouped into servers and servers are further grouped into clusters. The autonomic element at each level interacts with the autonomic elements above and below it to achieve autonomic power and performance management.

Several works deal with interoperability from the viewpoint of homogenous competing managers. [5] implements a two-level autonomic data management system that optimizes the managed system so jobs are not starved of resources. A global manager is tasked with allocation of physical resources to a number of virtual servers in an optimal and equitable manner. Local managers oversee each virtual server, using fuzzy logic to infer the expected resource requirements of the applications that run on the virtual servers. [6] describes an experiment to separate out the Monitoring and Analysis stages of the MAPE loop into distinct autonomic elements, with designed-in interactions between them. Monitoring capabilities are implemented in a node called an agent, with the analysis aspect implemented in a node called a broker. Information received from the environment are processed by the agents and forwarded to the broker where it is further analyzed. One or more agents feed information to a specific broker. An example of bespoke designed-in interaction between autonomic elements is provided in [7]. Three types of autonomic elements work hierarchically to provide scalable management, differentiated in terms of their operating timescale and scope of responsibility. This example serves to differentiate interaction between components which is achieved here, from the concept of interoperability which has stricter requirements. The fact that the various elements are part of a single coherent service with designed-in support for interaction means that the full challenge of interoperability is not encountered in this situation. [8] illustrates the complexity of combining multiple management domains into a single controller. In this work a joint QoS and Energy manager is developed using a design-time oriented approach tuned for a specific environment and is thus highly sensitive to its operating conditions. This tight integration approach is not generalisable and the resulting combined manager would appear to be more costly to develop and test than two independent managers.

The majority of work to date has targeted planned interoperability between designed-for-collaboration AMs working towards a common goal. This is a valuable step towards AM interoperability, although these solutions generally lack a formal definition of the interfaces or where defined, these interfaces are specific to the system in question; preventing wide applicability and reusability. Custom solutions are expensive to develop and are sensitive to changes in target systems, and thus generally restrictive and not future-proof. A significant issue is that they do not tackle the problem of unintended or unexpected interactions that can occur when independently developed AMs co-exist in a system.

This challenge has been recognised for some time, for example [9] defines a number of interfaces to aid autonomic element interactions. Several 'vision' papers [10], [11], [12] identify interoperability as a key challenge for future autonomic systems. [10] argues that mechanisms that define interoperability between autonomic elements must be reusable to limit complexities i.e., it must be generic enough to capture all communications across the board but also prevent bloatedeness. A standard means must exist for exchanging contexts between communicating elements to allow one autonomic element to understand the basis for the action of another. [10] also identifies the need for a function to translate the output of one element to the format understood by another. [11] identifies some necessary components for autonomic element interaction, including: a name service registry for autonomic elements; a system interaction broker and a negotiator. An interface specification must also take cognizance of hierarchy amongst autonomic elements. [12] observes that a strict and specified communication behaviour should be enforced, to prevent interoperating autonomic elements from communicating through undocumented or backdoor interfaces.

III. INTEROPERABILITY ISSUES

We posit that interoperability support (or lack of it) will become a make-or-break issue for future autonomic systems which inevitably contain multiple AM components. Bespoke or application-specific approaches to interoperability only offer a temporary respite at best, as they suffer a number of significant limitations which include:

1. Lack of flexibility and ability to scale - it is unrealistic to keep adding signals and functionality to deal with each possible interaction between any combination of AM's.

 ISBN: 978-1-61208-174-8

2. Having many isolated pools of interoperability is too complex. AC became popular fundamentally as a means of controlling, or hiding, complexity. It is undesirable from maintainability and stability perspectives to actually add excessive complexity in the process of solving the complexity problem.

3. It is not technically feasible to achieve close-coupled interoperability (i.e., where specific actions in one AM react to, or complement those of another) unless the source code and detailed functional spec. is available for each AM.

4. It will not be cost effective or timely. The cost and complexity of a bespoke solution spirals exponentially as the number of interacting AM's increase (consider a near-future cloud computing facility with multi-vendor management software systems and with autonomic management embedded into platforms, operating software, application software and also infrastructure such as power management and cooling systems – this is a complexity and stability storm just waiting to happen).

5. Re-development of managers to facilitate specific interoperability, and especially to deal with conflicts that arise unexpectedly, is reactive and incremental (and thus always ongoing).

6. It is not possible to know the nature of AMs not yet built, or to predict exactly where conflict will materialise in advance of adding a particular AM into a running system.

The issues highlighted above strongly suggest that it is necessary to deal with interoperability proactively by developing managers that are interoperability-enabled from the outset. We propose a service-based approach to interoperability, in which an Interoperability Service (IS) is responsible for detecting possible conflicts of management interest, and granting or withholding management rights to specific AMs as appropriate. In this way the IS performs all of the active interoperability management, and AMs only participate passively by providing information and following control commands from the IS. The IS interacts with AMs via a special interface which they must support.

We identify a number of requirements for a universal IS solution:

- Be application-domain independent and system independent.
- Able to represent AMs' management interests in a standard way that facilitates accurate conflict detection. This includes recognising resources which are not directly managed, but are nevertheless impacted by the behaviour of the manager.
- Have variable conflict-detection sensitivity which is run-time configurable to suit specific system requirements.
- Have a hierarchical architecture so as to deal with both local and global conflicts, and conflicts that occur across different levels in a complex system.
- Be proactive and automated; these are mandatory qualities for sustainable systems containing dynamic combinations of AM's with potentially complex interaction patterns.
- Able to automatically suspend and resume AM management activity on the basis of conflict detection and resolution.
- Support independently developed and tested AMs which in the presence of other AMs are susceptible to conflicts that they cannot locally detect or handle.
- Sufficiently trustworthy that compliant AM's are *certifiable* for safe co-existence – regardless of platform, vendor etc.

IV. AN INTEROPERABILITY SERVICE

This section presents an initial IS for exploration of the requirements identified above. The IS maintains a database of all registered AMs along with a mapping of the resources they manage and their scope of operation and management. AMs register with the service via a standard interface and provide details of their management capabilities using a standardised description language. The IS detects potential conflicts and sends appropriate signals to one or more AMs to e.g., stop or suspend their management activity. The strengths of this approach are that it is scalable, generalisable, has low component-interaction complexity and because conflict management is handled within the IS, the AMs are not involved in negotiation with peers. The service has a hierarchical structure for scalability, enabling conflict detection at both global level (such as system-wide security management) and local level (such as platform-wide, or VM-wide, resource management) with respect to a particular AM. Additional levels can be added, with a communication infrastructure resembling that of a typical hierarchical service such as DNS. It is important that conflict-detection is performed at the correct level. For example, an autonomic VM scheduler only has a potential conflict with an autonomic memory manager if they are both operating on the same processor unit.

The architecture is formed around a number of regular interfaces and a communication protocol which define the interaction between the components of the system, as outlined in figure 1. A number of interfaces are specified, and form three groups:

IS-AM interaction is supported by two interfaces.

IAdvertise {*Advertise*, *Unregister*, *Heartbeat*} is used by AMs to signal joining (register), leaving and heartbeat messages to the IS. *Advertise* is accompanied by a list of resources that the AM either wishes to manage directly, or that the developer has identified might be impacted by the manager's behaviour. *Unregister* is used by an AM to signal an orderly shutdown, and *Heartbeat* (normally invoked periodically) enables (when absent) the IS to detect when a manager crashes or leaves abruptly. In either case, the AM's management interests are unregistered and the conflict detection analysis is triggered, so that any AMs which were suspended but are no longer in conflict with the system can be resumed.

IInteroperate {*Run*, *Stop*, *Suspend*, *Resume*, *Throttle*} is used to receive directives from the IS. The AM developer

 ISBN: 978-1-61208-174-8

uses the IS API to map these directives onto the AM-internal behaviour. *Run* is accompanied by a sub-list of the requested resources that the AM can manage, so partial conflicts can be handled without suspending the entire manager. *Stop* shuts down the AM. *Suspend* backgrounds the AM (part or all of its management activity). *Resume* reactivates a suspended AM. The IS uses *Throttle* to specify different rates of activity to potentially conflicting AMs to prevent certain oscillatory patterns developing.

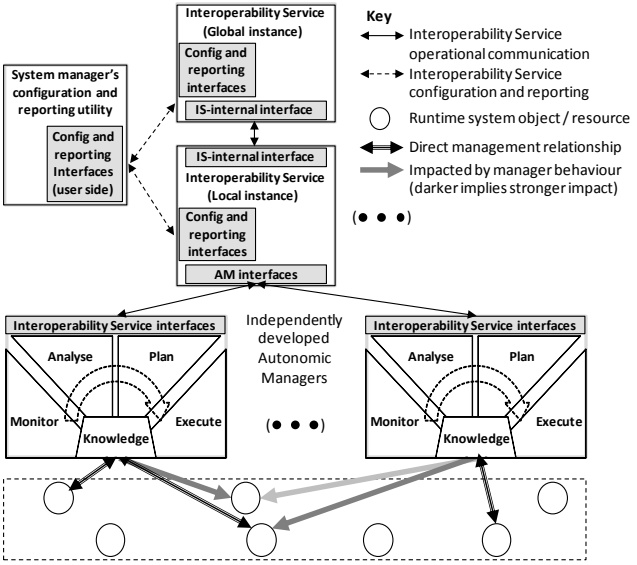


Figure 1. The Interoperability Service (IS) architecture.

IS-IS interaction is facilitated by a single interface.

ICommunicate {*Forward*, *Locate*, *Elect*, *SetISLevel*, *GetISLevel*} supports hierarchical operation. *Forward* is used to pass messages between local ISs which want to control global resources and the Global IS instance; this is the basis of system-wide and cross-level conflict detection. The remaining functions support the hierarchical IS structure itself including leader election for robustness. *Locate* returns the current service coordinator IS instance (which also performs the role of global conflict detection). *Elect* initiates an election if no coordinator instance is found. *SetISLevel* sets the IS level to be either Local or Coordinator. *GetISLevel* is used by each IS instance to determine its status during Locate and Elect events.

The IS provides an external management interface.

IConfigure {*SetMode*, *GetMode*, *SetSensitivity*, *GetSensitivity*, *StatusReport*} is a configuration and reporting interface which allows external system management utilities to perform system-specific configuration and generate status reports. *SetMode* and *GetMode* allow configuration of the service to allow different levels of safety; 'Safe' requires that all of a particular AM's management activity is suspended when it is found to be involved in a conflict, whilst 'Permissive' allows partial suspension. *SetSensitivity* and *GetSensitivity* are used to configure the conflict detection sensitivity level. *StatusReport* collects status information and statistics for report generation and IS performance monitoring.

The IS architecture specification precisely defines the interfaces, and with its accompanying communication protocol, defines the message formats and sequences that form the inter-component communication. It also specifies the semantics of this communication. Figure 2 shows how the IS functionality is integrated with the various components of the system.

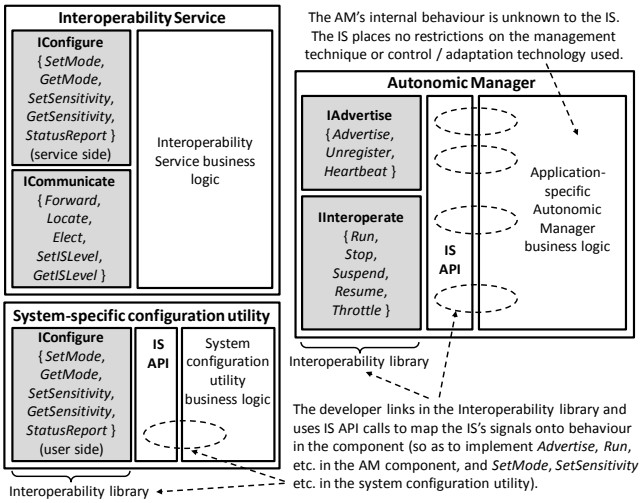


Figure 2. Internal architecture of the system components and the integration of the IS interfaces with these components.

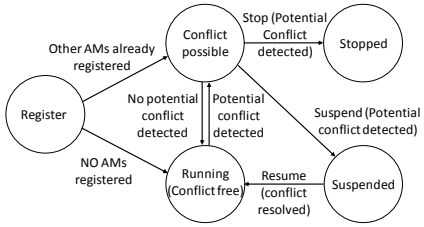


Figure 3. State diagram held by IS, for each registered AM.

The software developer retains flexibility with respect to the internal design and behaviour of the business logic of AM components and system configuration utilities. The architecture specification does not restrict the management approach, internal structure or control / adaptation techniques used within an AM component. The AM developer must integrate the API calls into the manager such that the control behaviour meets the IS specification. Where an AM manages multiple resources the developer can choose to implement Suspend such that it is effective at the level of the AM itself, or only on the management activity that has been notified as being in conflict. Similarly, the developer can decide the AM-internal semantics of Suspend so as to isolate the management output (effecter output) of the manager whilst still running the monitor, analyse and plan parts if desired. This approach facilitates the IS' regulatory control over the AM when conflicts occur, whilst enabling 'warm' start-ups of components when conflicts are resolved.

An instance of a state model is maintained for each

 ISBN: 978-1-61208-174-8

registered AM (see figure 3). The information held in these models drives the IS' conflict management behaviour and is the basis on which AMs' management rights are governed. During AM registration, if no other AMs are registered the new AM is granted management rights for the resources requested and signalled that it can Run. If other AMs are already registered, the IS evaluates whether or not there is a possible conflict of interest, and if so signals the AM to either Stop (in which case the AM must attempt re-registration at a later time driven by some external event) or Suspend (in which case the IS will signal the AM that it can Resume, i.e., manage, once the conflict has been resolved).

V. MANAGEMENT DESCRIPTION LANGUAGE

We discuss the need for a standard description of AMs' management interests, and briefly introduce our current language which is extensible to accommodate improvements in our understanding of ways actual and potential conflicts arise.

The IS facilitates interoperability amongst (unknown in advance) AMs which have been developed independently of each other, and thus do not directly support interoperability amongst themselves. The overall goal is to maximise the management freedom of AMs whilst at the same time ensuring that the system remains stable; requiring that the IS must also:

- Detect AMs and learn their characteristics (via registration);
- Identify potential conflict, determine the consequences and the level of risk, and achieve a system-specific balance when taking decisions to resolve conflicts by suspending or stopping AMs' management activities;
- Automatically resume suspended AMs when conflicts are resolved (e.g., when other AMs leave the system);
- Enable cooperation between AMs. For example to share learnt knowledge concerning system state, volatility etc.

To perform these functions, the IS needs certain information detailing each AMs' management domain and specific resources of interest. This information must use a standard language format, and a fixed vocabulary of key terms so that automated searching for overlaps of interest can be performed effectively. The information will be provided at run time by the AM via the IS API (the information is provided ultimately by the AM developer).

Conflicts can arise in several ways. Direct conflicts occur where multiple AMs attempt to manage the same resource or object. However conflicts can be indirect (and less obvious) because a manager's activity may impact resources other than those directly managed. Categories of this include cross-application conflicts, for example increasing a specific application's use of a particular resource such as network bandwidth reduces the availability of bandwidth available to other applications. Another category of indirect conflicts are cross-resource conflicts, for example increasing processor speed to maximise throughput increases direct power usage and may also increase power requirements for cooling systems (which may have their own autonomic management systems). Some system characteristics such as security policy, power usage, server provisioning strategy etc. may be managed at both the system-wide level, and locally at the level of individual computing node or cluster. This can lead to conflicts between global and local managers, resulting in parts of the system being out-of step with global policy, and/or inefficient behaviour. It will be difficult to identify every possible case of indirect conflict with certainty, and the extent of management impact in such cases is also highly variable. Therefore the description information provided by AMs must be sufficient to derive a similarity measure between their management interests and effects. The language needs to contain appropriate categories to express areas of management concern in a structured way, i.e., from high-level domain in which the manager operates down to specific resources that are managed, and also to express characteristics including the management scope (global or local) and specificity (e.g., organisation specific, application specific).

Given these requirements, the standard management description should include (see figures 4 and 5 for an example):

Category. Mandatory. The highest-level and most generic descriptor used to identify the AM's domain of interest. Terms include: {*Power general*, *Performance general*, *Security general*, ... }

Zone. Mandatory. A second level, more specific sub-category enabling developers to differentiate between specific management functions. Terms include: {*Power system*, *Power platform*, *Power cooling ... Performance system*, *Performance CPU*, *Performance disk*, *Scheduling*, *VM management*, ... }

Impact. Mandatory. A numerical indicator Impact Factor (IF), (where $0 < IF \leq 1$), is defined to express the strength of the management influence. A directly controlled resource is assigned the value 1. A value close to 0 indicates that the particular AM has a weak influence on the resource whilst values close to 1 indicate that the resource is closely impacted by changes to one that is directly managed by the AM; for example an AM directly controlling CPU speed (IF = 1) has a strong indirect influence on VM performance (IF ≈ 0.8). Term: { *ImpactFactor*(value) }

Scope. Mandatory. Whether the manager has local or global impact. Terms: { *Local*, *Global* }

Specificity. Optional. The extent of manager operation. Terms include: { *System-wide*, *Application-wide*, *Platform-wide*, *Process-wide*, *User-specific*, ... }

Trigger. Optional. Facilitates expression of temporal aspects such as periodicity or operating timescale, as well as specific events that invoke the management activity. Such characteristics can potentially be used to detect combinations of AMs at risk of causing of instability in the form of oscillation or control divergence. Terms include:

{*Period*(value), *Event*(name) , ... }
Parameter. Optional. Identifies specific context parameters that are of interest to the AM. Term: { *Name*(value) }
Envelope. Optional. Expresses range of, and/or the number of dimensions of, control freedom. This can potentially help to avoid false positive detections of conflict, when managers operate in the same domain but have non-overlapping envelopes of operation. Terms include: { *Name*(range, value) }

VI. CONFLICT DETECTION

For the initial exploration we use a conflict detection technique based on pair-wise fuzzy similarity measures of AMs' management interests. This uses a summation of weighted terms, derived from AMs' management descriptions (see sections V and VII). Conflict detection activity is triggered by events such as the registration of a newly-discovered AM, or the departure of an AM from the system. The items that comprise the management description form a vector. Weights are allocated to the items to signify relative importance.

A dynamically configurable conflict threshold ($0 < \text{ThreshC} \leq 1$) is used to tune the conflict detection sensitivity (via SetSensitivity, on IConfigure). A potential conflict is detected if the similarity measure of a pair of vectors exceeds ThreshC. It is intended that the sensitivity level is configured by the facility manager, via a control console application (or automated), and can be changed at run time as necessary. This enables safety critical systems to operate with very low tolerance to potential conflicts, whereas in domains where only e.g., efficiency is at stake, a higher tolerance can lead to benefits of having more AMs working simultaneously (bearing in mind that a 'potential conflict' may not be realised).

VII. EVALUATION

We demonstrate the operation and benefit of the IS in a data centre scenario in which two independently developed AMs coexist. A scheduling manager (AM1) has a main goal of maximising throughput by keeping all resources utilised where possible. A power manager (AM2) is designed to minimise power usage by slowing down processor speed or by shutting down entire processor units where possible. The co-existence of these AMs creates a high potential for conflict. For example AM2 will attempt to shutdown an underutilised resource as soon as load level starts to fall, whilst AM1 will attempt to bring unused resources into play as soon as load levels increase (or a backlog develops). Depending on the sequence of load level changes it is possible that oscillation will build up between the actions of these two managers.

Operation: During its initialisation each AM registers with the IS. The management capabilities of each AM are described using the standard language and categories described earlier. AM1 directly controls a parameter performance within the general management category performance general, and specific sub-zone CPU performance; and indirectly influences a parameter power within the general category performance general, and sub-zone system performance. AM2 directly controls a parameter power within the general category power general, and the specific zone of interest system power; and indirectly influences a parameter performance within the general category performance general, and the specific zone of interest CPU performance.

```
a) AddACItem   ("Performance", "Performance General",
                "CPU Performance", "1.0", "Local");
   AddACItem   ("Power", "Performance General",
                "System Performance", "0.5", "Local");
   RegisterAsAM  ();

b) AddACItem   ("Power", "Power General",
                "System Power","1.0","Local");
   AddACItem   ("Performance", "Performance General",
                "System Performance", "0.5", "Local");
   RegisterAsAM  ();

c) bool AddACItem(char *ParameterName, char *Category,
         char *Zone, char *Impactfactor, char *Scope);
```

Figure 4. API calls to register AMs' management interests.

The API calls for manager registration are shown in Figure 4a (for AM1), and 4b (for AM2), where AddACItem means 'Add autonomically controlled item'; its template is shown in figure 4c. Figure 5 shows the XML equivalent representation for AM1.

```
<!-- Autonomic Manager Configuration Specification
Language -->
<MetaData>
  <ConfigAuthor Name="Mariusz Pelc" Organisation="UoG" />
  <TimeStamp Time="12:00" Date="20/12/2010" />
    <AMDescription>
      <AM ID="AM1">
        <ACItems>
          <ACItem ID="Performance" Scope="Local">
            <Category>Performance General</Category>
            <Zone>CPU Performance</Zone>
            <ImpactFactor>1.0</ImpactFactor>
          </ACItem>
          <ACItem ID="Power" Scope="Local">
            <Category>Performance General</Category>
            <Zone>System Performance</Zone>
            <ImpactFactor>0.5</ImpactFactor>
          </ACItem>
        </ACItems>
      </AM>
    </AMDescription>
</MetaData>
```

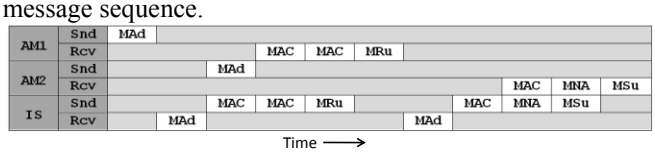
Figure 5. XML representation of the Management Description Language.

Scenario 1: Each manager registers separately in the system in the absence of the other. ThreshC = 0.6. AM1 requests management rights for CPU performance, and also notifies a potential impact on system power. As there are no other AMs present, the IS grants AM1 permission to manage unimpeded. Similarly, for AM2 (in the absence of AM1) the IS grants rights to manage system power level and also to have an indirect impact on system performance.

Scenario2: AM1 registers and is granted rights to manage the resources it requested. AM2 then registers whilst AM1 is still present. ThreshC=0.6. The IS performs

 ISBN: 978-1-61208-174-8

conflict detection analysis, based on the AMs' announced Impact Factors (IFs) for each requested managed item. This determines whether AM2 can be granted the requested management rights: Power directly managed (IF=1.0), and Performance potentially affected indirectly (IF=0.5). An indirect conflict is detected: AM1 already manages a system performance characteristic (specifically CPU performance), when AM2 registers, requesting to manage system power, but also announcing a potential impact on system performance. The IS does not detect a direct conflict with the power management, but the weighted conflict level for system performance (found to be 0.6875) exceeds the current ThreshC (0.6). The IS suspends the newly registering manager to prevent possible instability (this manager will be automatically resumed if AM1 leaves the system and there are no other conflicts with other AMs registered in the meantime). Figure 6 shows the resulting message sequence.



Key: Snd - Send message Rcv - Received message MNA - MNACK Mad – MAdvertise MAC - MACK MSu - MSuspend MRu – Mrun

Figure 6. Message sequence for scenario 2.

Scenario 3: As scenario 2, but with ThreshC = 0.8, i.e., the IS is less sensitive to potential conflicts (this configuration may be better suited to non-critical systems where some *potential* for conflict may be acceptable, i.e., the tradeoff between safety and management flexibility is shifted). The resulting message sequence is shown in Figure 7. In this case no conflicts are detected and the newly arriving AM2 is granted rights to manage system power level, and to have an impact on system performance, thus potentially interacting with AM1.

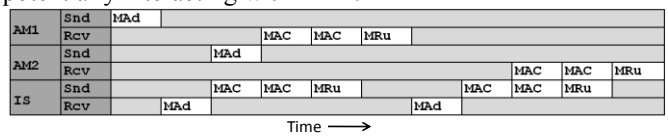


Figure 7. Message sequence for scenario 3.

VIII. CONCLUSION

We have outlined the case for greater research effort in the area of interoperability of autonomic managers. We have discussed why bespoke and custom solutions will not work in the long term and argued for a universal standard for interoperability. In line with this we have identified requirements for a service-based approach.

We presented initial work towards a service-based automatic and proactive interoperability service, being integrated into autonomic components and making them 'interoperability ready' in advance of their deployment. Our approach enables AMs to be developed independently, requiring that the developer uses a management description language to describe the component's management characteristics. This approach has the main advantage of not requiring an AM developer to have knowledge of future AM's that may exist in the target system, and thus supports agility i.e., configuration changes, expansion and upgrades.

IX. REFERENCES

[1] Kephart J. O., Chan H., Das R., Levine D. W., Tesauro G., Rawson F. and Lefurgy C. Coordinating multiple autonomic managers to achieve specified power-performance tradeoffs. *4th Intl. Conf. on Autonomic Computing* (Jacksonville, FL, USA, June 2007). ICAC'07. IEEE, 1–9.

[2] Wang M., Kandasamyt N., Guezl A. and Kam M. Adaptive performance control of computing systems via distributed cooperative control: Application to power management in computing clusters. *3rd Intl. Conf. on Autonomic Computing* (Dublin, Ireland, June 2006). ICAC'06. IEEE, 165–174.

[3] Zhao M., Xu J. and Figueiredo R. J. Towards autonomic grid data management with virtualized distributed file systems. *3rd Intl. Conf. on Autonomic Computing* (Dublin, Ireland, June 2006). ICAC'06. IEEE, 209–218.

[4] Khargharia B., Hariri S. and Yousif M. S. Autonomic power and performance management for computing systems. *3rd Intl. Conf. on Autonomic Computing* (Dublin, Ireland, June 2006). ICAC'06. IEEE, 145–154.

[5] Xu J., Zhao M., Fortes J., Carpenter R. and Yousif M. On the use of fuzzy modeling in virtualized data center management. *4th Intl. Conf. on Autonomic Computing* (Jacksonville, FL, USA, June 2007). ICAC '07. IEEE, 25-34.

[6] Kutare M., Eisenhauer G. and C. Wang. 2010. Monalytics: Online monitoring and analytics for managing large scale data centers. *7th Intl. Conf. on Autonomic Computing* (Washington DC, USA, June 2010). IEEE, 141–150.

[7] Zhu X., Young D., Watson B. J., Wang Z., Rolia J., Singhal S., McKee B., Hyser C., Gmach D., Gardner R., Christian T., and Cherkasova L. 1000 islands: Integrated capacity and workload management for the next generation data center. *5th Intl. Conf. on Autonomic Computing* (Chicago, IL, USA, 2008). ICAC '08. IEEE, 172–181.

[8] Poussot-Vassal C., Tanelli M. and Lovera M. 2010. A Control-Theoretic Approach for the Combined Management of Quality-of-Service and Energy in Service Centres. In *Run-time Models for self-managing Systems and Applications*. Ardagna D and Zhang L, Eds). Springer Basel AG. 73-96.

[9] White S. R., Hanson J. E., Whalley I., Chess D. M. and Kephart J. O. An architectural approach to autonomic computing. *1st Intl. Conf. on Autonomic Computing* (New York, NY, USA, May 2004). ICAC'04. IEEE. 2-9.

[10] Kennedy C. 2010. Decentralised metacognition in context-aware autonomic systems: some key challenges. In *Proc. 24th American Institute of Aeronautics and Astronautics* (AIAA) *Workshop on Metacognition for Robust Social Systems* (Atlanta, Georgia,) AAAI-10, AIAA. 34-41.

[11] Salehie M. and Tahvildari L. Autonomic computing: Emerging trends and open problems. *Workshop on the Design and Evolution of Autonomic Application Software* (New York, NY, USA, 2005). DEAS'05. ACM Special Interest Group on Software Engineering. 30. 1–7.

[12] Quitadamo R. and Zambonelli F. Autonomic communication services: a new challenge for software agents. *SpringerLink Journal of Autonomous Agents and Multi-Agent Systems*. 17, 3 (2008), 457–475.

An Alamouti Coding Scheme for Relay-Based Cooperative Communication Systems

Youngpo Lee[1], Youngje Kim[2], Sun Yong Kim[3], Gyu-In Jee[3], Jin-Mo Yang[4], and Seokho Yoon[1],[†]
[1]School of Information and Communication Engineering, Sungkyunkwan University, Suwon, Korea
[2]Samsung Thales Co. LTD., Seongnam, Korea
[3]Department of Electronics Engineering, Konkuk University, Seoul, Korea
[4]Agency for Defense Development (ADD), Daejeon, Korea
[†]Corresponding author
(E-mail: leeyp204@skku.edu, youngje99.kim@samsung.com, {kimsy, gijee}@konkuk.ac.kr,
jmy1965@dreamwiz.com, and [†]syoon@skku.edu)

***Abstract*—This paper addresses a novel Alamouti coding scheme for asynchronous cooperative communication systems over frequency selective fading channels. In the proposed scheme, the Alamouti coded form at the destination node is constructed through a simple combination of symbols at the source node, instead of the time-reversal operation at the relay nodes used in the conventional scheme. Numerical results show that the proposed scheme achieves a higher order cooperative diversity than that of the conventional scheme.**

***Index Terms*—Asynchronous cooperative communication systems, Alamouti coding, frequency selective fading channels**

I. INTRODUCTION

A spatial diversity is an efficient technique to improve the reliability of transmission in wireless environments [1]. The classical approach for achieving the spatial diversity is to use multiple-input multiple-output (MIMO) systems employing multiple antennas at transmitter and receiver. However, due to the limitations of size, cost, and power, it may be impractical to accommodate multiple antennas on mobile devices [2]. Cooperative communication systems can provide the spatial diversity referred to as the cooperative diversity by forming the virtual MIMO systems via distributed nodes with only a single antenna, thus overcoming the limitations of MIMO systems [3], [4]. In the cooperative communication systems, however, symbols from different nodes are received at different time instants, resulting in an asynchronous environment. Several transmission schemes have been presented to achieve the cooperative diversity for the asynchronous cooperative communications [5]-[8]; however, they are difficult to implement since a symbol decoding step is required at the relay nodes. In [9], a transmission scheme without any symbol decoding step at the relay nodes has been proposed for the asynchronous cooperative communications. The scheme in [9] can construct the well-known Alamouti coded form at the destination node using only the simple time-reversal and conjugate operations at the relay nodes, achieving the cooperative diversity. However, in the frequency selective fading channels, bit error rate (BER) performance of the scheme in [9] significantly degrades since the decoding process at the destination node is not able to construct the Alamouti coded form due to the multipath components. Recently, in [10], a modified version of the scheme in [9] has been introduced for the asynchronous cooperative communications, combating the influence of the frequency selective fading channels. With the modified operations at the relay nodes, each relay node performs either time-reversal or conjugate operation only, the scheme in [10] overcomes the influence of the frequency selective fading channels, and thus, the Alamouti coded form is constructed at the destination node.

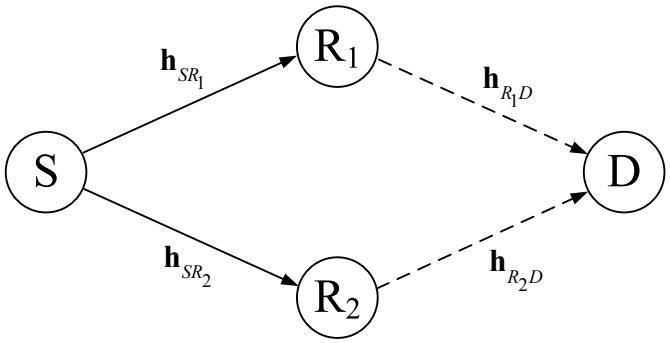


Fig. 1. A cooperative system model.

In this paper, we propose a novel Alamouti coding scheme for asynchronous cooperative communication systems over the frequency selective fading channels. By transmitting the combinations of the data symbols at the source node, the proposed scheme achieves a higher order cooperative diversity than that of the conventional scheme in [10], where each data symbol passes through the different channel. It is also demonstrated that the proposed scheme provides a higher order cooperative diversity, and thus, better BER performance than the conventional scheme in [10]. The rest of this paper is organized as follows. Section II introduces the system model of the cooperative systems. The conventional scheme is explained in Section III. In Section IV, a novel cooperative transmission scheme is proposed. Simulation results are presented in Section V, and conclusion is given in Section VI.

 ISBN: 978-1-61208-174-8

II. System Model

Fig. 1 illustrates a cooperative system model with one source node S, one destination node D, and two relay nodes R_1 and R_2, where each node has a single antenna.

It is assumed that the channels between the source node and each relay node and between each relay node and the destination node are frequency selective fading channels with L and Q independent propagation paths, respectively. Then, the nth channel impulse response coefficient $h_{SR_m}(n)$ from the source node to the mth relay node can be expressed as

$$h_{SR_m}(n) = \sum_{l=0}^{L-1} \alpha_{SR_m,l}\delta(n - \tau_{SR_m,l}), \tag{1}$$

where $\delta(n)$ denotes the delta function, and $\alpha_{SR_m,l}$ and $\tau_{SR_m,l}$ are the channel coefficient and path delay of the lth propagation path from the source node to the mth relay node, respectively. Similarly, the nth channel impulse response coefficient $h_{R_mD}(n)$ from the mth relay node to the destination node can be expressed as

$$h_{R_mD}(n) = \sum_{q=0}^{Q-1} \alpha_{R_mD,q}\delta(n - \tau_{R_mD,q}), \tag{2}$$

where $\alpha_{R_mD,q}$ and $\tau_{R_mD,q}$ are the channel coefficient and path delay of the qth propagation path from the mth relay node to the destination node, respectively. The channel coefficients $\alpha_{SR_m,l}$ and $\alpha_{R_mD,q}$ are modeled as complex Gaussian random variables with mean zero and variance $\sigma^2_{SR_m,l}$ and $\sigma^2_{R_mD,q}$, respectively, where $\sum_{l=0}^{L-1}\sigma^2_{SR_m,l} = \sum_{q=0}^{Q-1}\sigma^2_{R_mD,q} = 1$. It is assumed that the channel coefficients and path delays are constant during two orthogonal frequency division multiplexing (OFDM) symbol intervals, and a relative timing difference between the symbols arriving at the destination node from the relay nodes R_1 and R_2 is denoted by τ.

III. Conventional Scheme

In the conventional scheme [10], first, the source node generates the following two complex-valued data symbol blocks $\mathbf{X}_1$ and $\mathbf{X}_2$:

$$\mathbf{X}_i = [X_i(0), X_i(1), \cdots, X_i(N-1)]^T_{,} \text{ for } i = 1 \text{ and } 2, \tag{3}$$

where $X_i(k)$ and $(\cdot)^T$ denote the complex-valued phase shift keying (PSK) or quadrature amplitude modulation (QAM) data symbol on the kth sub-carrier of $\mathbf{X}_i$ and the transpose operation, respectively, and N is the number of the sub-carriers. For transmission from the source node to the relay nodes, complex-valued baseband samples are generated as

$$\begin{aligned} x_i(n) &= \mathrm{DFT}_N\{\mathbf{X}_i\} \\ &= \sqrt{\frac{1}{N}}\sum_{k=0}^{N-1} X_i(k)e^{-j2\pi kn/N}, \text{ for } n = 0, \cdots, N-1, \end{aligned} \tag{4}$$

where $\mathrm{DFT}_N(\cdot)$ denotes the N-point discrete Fourier transform (DFT). Then, the ith OFDM symbol $\mathbf{s}_i$ is obtained as

$$\begin{aligned} \mathbf{s}_i &= [s_i(0), s_i(1), \cdots, s_i(L_s-1)]^T \\ &= [x_i(N-N_G), x_i(N-N_G+1), \cdots, x_i(N-1), \\ &\quad x_i(0), \cdots, x_i(N-1)]^T \end{aligned} \tag{5}$$

by inserting the cyclic prefex (CP), where $L_s \triangleq N + N_G$ with N_G the length of CP. It is assumed that N_G is longer than the sum of total propagation path delay from the source node to the destination node and the maximum relative timing difference at the destination node. Next, the source node transmits two consecutive OFDM symbols to the relay nodes, and then, the ith received symbol $\mathbf{r}_{i,m}$ of the mth relay node can be written as

$$\mathbf{r}_{i,m} = \sqrt{P_1}\ \mathbf{s}_i * \mathbf{h}_{SR_m} + \mathbf{n}_{i,m}, \tag{6}$$

where P_1 is the transmission power at the source node, $\mathbf{h}_{SR_m} = [\alpha_{SR_m,0}, \alpha_{SR_m,1}, \cdots, \alpha_{SR_m,L-1}]^T$ is the $L \times 1$ impulse response vector of the channel between the source node and the mth relay node, $\mathbf{n}_{i,m}$ is the additive white Gaussian noise (AWGN) vector with zero mean and unit variance added to the ith received symbol of the mth relay node, and $*$ denotes the linear convolution.

Each relay node allows the Alamouti coded form to be constructed at the destination node by transmitting symbol $\widetilde{\mathbf{s}}_{i,m}$ obtained as in Table I to the destination node during two consecutive time slots, where $\widetilde{\mathbf{s}}_{i,m}$, P_2, $\zeta(\cdot)$, and $(\cdot)^*$ denote the ith transmit symbol of the mth relay node, the transmission power at the relay nodes, time-reversal operation, and conjugate operation, respectively. The time-reversal operation $\zeta(\cdot)$ is given by

$$\zeta\{r_{i,m}(n)\} = \begin{cases} r_{i,m}(0), & \text{for } n = 0 \\ r_{i,m}(L_s - n), & \text{for } n = 1, \cdots, L_s - 1, \end{cases} \tag{7}$$

where $r_{i,m}(n)$ denotes the nth sample of $\mathbf{r}_{i,m}$.

At the destination node, as in general OFDM systems, the CP is removed first for each received symbol. After the CP removal, the last $\tau_1{}' = N_G - (\max(\tau_{SR_1,l}) - 1)$ samples of the received symbols are shifted to the front of the received symbols. Then, after the CP removal and sample shifting process, the received symbols can be expressed as

$$\begin{aligned} \mathbf{z}_1 = \lambda\big[&\{\sqrt{P_1}\zeta(\mathrm{DFT}_N(\mathbf{X}_1)) \otimes \zeta(\mathbf{h}'_{SR_1}) + \mathbf{n}_{1,1}\} \otimes \mathbf{h}'_{R_1D} \\ &- \{\sqrt{P_1}(\mathrm{DFT}_N(\mathbf{X}_2))^* \otimes \mathbf{h}'^{*}_{SR_2} + \mathbf{n}_{2,2}\} \otimes \mathbf{h}'_{R_2D} \\ &\otimes \mathbf{\Gamma}_\tau \otimes \mathbf{\Gamma}'_1\big] + \mathbf{w}_1 \end{aligned} \tag{8}$$

TABLE I
Processing at the relay nodes of the conventional scheme.

	Relay node 1	Relay node 2
Time slot 1	$\widetilde{\mathbf{s}}_{1,1} = \sqrt{\frac{P_2}{P_1+1}}\zeta(\mathbf{r}_{1,1})$	$\widetilde{\mathbf{s}}_{1,2} = -\sqrt{\frac{P_2}{P_1+1}}\mathbf{r}^*_{2,2}$
Time slot 2	$\widetilde{\mathbf{s}}_{2,1} = \sqrt{\frac{P_2}{P_1+1}}\zeta(\mathbf{r}_{2,1})$	$\widetilde{\mathbf{s}}_{2,2} = \sqrt{\frac{P_2}{P_1+1}}\mathbf{r}^*_{1,2}$

 ISBN: 978-1-61208-174-8

for the first received symbol and

$$\begin{aligned}\mathbf{z}_2 = \lambda\big[\{\sqrt{P_1}\zeta(\mathrm{DFT}_N(\mathbf{X}_2)) \otimes \zeta(\mathbf{h}_{SR_1}^{'}) + \mathbf{n}_{2,1}\} \otimes \mathbf{h}_{R_1D}^{'} \\ + \{\sqrt{P_1}(\mathrm{DFT}_N(\mathbf{X}_1))^* \otimes \mathbf{h}_{SR_2}^{'*} + \mathbf{n}_{1,2}\} \otimes \mathbf{h}_{R_2D}^{'} \\ \otimes \mathbf{\Gamma}_\tau \otimes \mathbf{\Gamma}_1^{'}\big] + \mathbf{w}_2\end{aligned} \tag{9}$$

for the second received symbol, where $\lambda = \sqrt{\frac{P_2}{P_1+1}}$, $\otimes$ and $\mathbf{w}_i$ denote circular convolution and AWGN vector with zero mean and unit variance added to the ith received symbol of the destination node, respectively. $\mathbf{h}_{SR_m}^{'}$, $\mathbf{h}_{R_mD}^{'}$, $\mathbf{\Gamma}_\tau$, and $\mathbf{\Gamma}_1^{'}$ are $N \times 1$ vectors defined as $\mathbf{h}_{SR_m}^{'} = [\alpha_{SR_m,0}, \alpha_{SR_m,1}, \cdots, \alpha_{SR_m,L-1}, 0, \cdots, 0]^T$, $\mathbf{h}_{R_mD}^{'} = [\alpha_{R_mD,0}, \alpha_{R_mD,1}, \cdots, \alpha_{R_mD,Q-1}, 0, \cdots, 0]^T$, $\mathbf{\Gamma}_\tau = [\mathbf{0}_\tau, 1, 0, \cdots, 0]^T$, and $\mathbf{\Gamma}_1^{'} = [\mathbf{0}_{\tau_1'}, 1, 0, \cdots, 0]^T$, respectively, where $\mathbf{0}_\tau$ and $\mathbf{0}_{\tau_1'}$ denote all-zero vectors with dimension $1 \times \tau$ and $1 \times \tau_1^{'}$, respectively. Using the properties $(\mathrm{DFT}_N(\mathbf{X}))^* = \mathrm{IDFT}_N(\mathbf{X}^*)$ and $\mathrm{DFT}_N(\zeta(\mathrm{DFT}_N(\mathbf{X}))) = \mathbf{X}$, where $\mathrm{IDFT}_N(\cdot)$ denotes the inverse DFT, the DFT outputs of $\mathbf{z}_1$ and $\mathbf{z}_2$ are obtained as

$$\begin{aligned}Z_1(k) = \lambda\big[\sqrt{P_1}X_1(k)H_{SR_{1,c}}(k)H_{R_1D,_c}(k) \\ - \sqrt{P_1}X_2^*(k)H_{SR_{2,c}}(k)H_{R_2D,_c}(k)e^{-j2\pi k(\tau+\tau_1^{'})/N} \\ + N_{1,1}(k)H_{R_1D,_c}(k) \\ - N_{2,2}(k)H_{R_2D,_c}(k)e^{-j2\pi k(\tau+\tau_1^{'})/N}\big] + W_1(k)\end{aligned} \tag{10}$$

and

$$\begin{aligned}Z_2(k) = \lambda\big[\sqrt{P_1}X_2(k)H_{SR_{1,c}}(k)H_{R_1D,_c}(k) \\ + \sqrt{P_1}X_1^*(k)H_{SR_{2,c}}(k)H_{R_2D,_c}(k)e^{-j2\pi k(\tau+\tau_1^{'})/N} \\ + N_{2,1}(k)H_{R_1D,_c}(k) \\ + N_{1,2}(k)H_{R_2D,_c}(k)e^{-j2\pi k(\tau+\tau_1^{'})/N}\big] + W_2(k),\end{aligned} \tag{11}$$

respectively, where $N_{i,m}(k)$, $H_{SR_{1,c}}(k)$, $H_{SR_{2,c}}(k)$, $H_{R_mD,_c}(k)$, and $W_i(k)$ are the DFT outputs of $\mathbf{n}_{i,m}$, $\zeta(\mathbf{h}_{SR_1}^{'})$, $\mathbf{h}_{SR_2}^{'*}$ $\mathbf{h}_{R_mD}^{'}$, and $\mathbf{w}_i$, respectively. (10) and (11) can be expressed as the following matrix form:

$$\begin{bmatrix} Z_1(k) \\ Z_2^*(k) \end{bmatrix} = \lambda H_c(k) \begin{bmatrix} \sqrt{P_1}X_1(k) \\ \sqrt{P_1}X_2^*(k) \end{bmatrix} + \begin{bmatrix} G_{1,c}(k) \\ G_{2,c}(k) \end{bmatrix}, \tag{12}$$

where $G_{1,c}(k)$ and $G_{2,c}(k)$ denote the noise component of $Z_1(k)$ and $Z_2(k)$, respectively. $H_c(k)$ is the channel matrix defined as

$$H_c(k) = \begin{bmatrix} H_{1,c}(k) & H_{2,c}(k) \\ H_{2,c}^*(k) & -H_{1,c}^*(k) \end{bmatrix}, \tag{13}$$

where $H_{1,c}(k) = H_{SR_1,c}(k)H_{R_1D,c}(k)$ and $H_{2,c}(k) = H_{SR_2,c}(k)H_{R_2D,c}(k)e^{-j2\pi k(\tau+\tau_1^{'})/N}$. The channel matrix $H_c(k)$ is the Alamouti coded form, and thus, the estimates $\widehat{X}_1(k)$ and $\widehat{X}_2(k)$ can be obtained as

$$\begin{bmatrix} \widehat{X}_1(k) \\ \widehat{X}_2^*(k) \end{bmatrix} = H_c^H(k) \begin{bmatrix} Z_1(k) \\ Z_2^*(k) \end{bmatrix} \tag{14}$$

for $X_1(k)$ and $X_2(k)$, respectively, where $(\cdot)^H$ denotes the Hermitian transpose operation.

IV. Proposed Scheme

Combining the complex-valued data symbol blocks $\mathbf{X}_1$ and $\mathbf{X}_2$, the source node first generates the following four symbol blocks $\mathbf{C}_1$, $\mathbf{C}_2$, $\mathbf{C}_3$, and $\mathbf{C}_4$:

$$\begin{aligned}\mathbf{C}_d = [C_d(0), C_d(1), \cdots, C_d(N-1)]_{,}^T \\ \text{for } d = 1, 2, 3, \text{ and } 4\end{aligned} \tag{15}$$

with

$$C_d(k) = \begin{cases} \frac{1}{\sqrt{2}}\{X_1(k) + jX_2(k)\}, & \text{when } d = 1 \\ -\frac{1}{\sqrt{2}}\{X_2^*(k) + jX_1^*(k)\}, & \text{when } d = 2 \\ \frac{1}{\sqrt{2}}\{X_1(k) - jX_2(k)\}, & \text{when } d = 3 \\ \frac{1}{\sqrt{2}}\{X_2^*(k) - jX_1^*(k)\}, & \text{when } d = 4, \end{cases} \tag{16}$$

where $C_d(k)$ denotes the complex-valued data symbol on the kth sub-carrier of $\mathbf{C}_d$. The generated symbol blocks satisfy the following property used in the processes at the destination node:

$$j\mathbf{C}_d = \begin{cases} \mathbf{C}_{d+1}^*, & \text{when } d = 1 \text{ and } 3 \\ \mathbf{C}_{d-1}^*, & \text{when } d = 2 \text{ and } 4. \end{cases} \tag{17}$$

Then, complex-valued baseband samples corresponding to $\mathbf{C}_d$ are generated as

$$\begin{aligned}c_d(n) &= \mathrm{IDFT}_N\{\mathbf{C}_d\} \\ &= \sqrt{\frac{1}{N}}\sum_{k=0}^{N-1} C_d(k)e^{j2\pi kn/N}, \text{ for } n = 0, \cdots, N-1\end{aligned} \tag{18}$$

for transmission from the source node to the relay nodes, and the dth OFDM symbol $\mathbf{u}_d$ is obtained as

$$\begin{aligned}\mathbf{u}_d &= [u_d(0), u_d(1), \cdots, u_d(L_s-1)]^T \\ &= [c_d(N-N_G), c_d(N-N_G+1), \cdots, c_d(N-1), \\ &\quad c_d(0), \cdots, c_d(N-1)]^T\end{aligned} \tag{19}$$

by inserting the CP. Next, the source node transmits four consecutive OFDM symbols to the relay nodes, and thus, the dth received symbol $\mathbf{v}_{d,m}$ of the mth relay node can be written as

$$\mathbf{v}_{d,m} = \begin{cases} \sqrt{P_1/2}\ \mathbf{u}_d * \mathbf{h}_{SR_m} + \mathbf{n}_{d,m}, & \text{when } d = 1 \text{ and } 2 \\ \sqrt{P_1/2}\ \mathbf{u}_d * \mathbf{g}_{SR_m} + \mathbf{n}_{d,m}, & \text{when } d = 3 \text{ and } 4, \end{cases} \tag{20}$$

where $\mathbf{n}_{d,m}$ is the AWGN vector with zero mean and unit variance added to the dth received symbol of the mth relay node. Unlike the conventional scheme transmitting two OFDM symbols at the source node with the transmission power of P_1, the proposed scheme transmits four OFDM symbols at the source node with the transmission power of $P_1/2$, that is, the total transmission power of the proposed scheme is the same as that of the conventional scheme. $\mathbf{g}_{SR_m} = [\beta_{SR_m,0}, \beta_{SR_m,1}, \cdots, \beta_{SR_m,L-1}]^T$ is the $L \times 1$ impulse response vector of the channel between the source node and the mth relay node for the last two transmit symbols

 ISBN: 978-1-61208-174-8

TABLE II
PROCESSING AT THE RELAY NODES OF THE PROPOSED SCHEME.

	Relay node 1	Relay node 2
Time slot 1	$\widetilde{\mathbf{u}}_{1,1} = \sqrt{\frac{P_2}{P_1+2}}\mathbf{v}_{1,1}$	$\widetilde{\mathbf{u}}_{1,2} = \sqrt{\frac{P_2}{P_1+2}}\mathbf{v}_{2,2}$
Time slot 2	$\widetilde{\mathbf{u}}_{2,1} = -j\sqrt{\frac{P_2}{P_1+2}}\mathbf{v}_{1,1}$	$\widetilde{\mathbf{u}}_{2,2} = j\sqrt{\frac{P_2}{P_1+2}}\mathbf{v}_{2,2}$
Time slot 3	$\widetilde{\mathbf{u}}_{3,1} = \sqrt{\frac{P_2}{P_1+2}}\mathbf{v}_{3,1}$	$\widetilde{\mathbf{u}}_{3,2} = \sqrt{\frac{P_2}{P_1+2}}\mathbf{v}_{4,2}$
Time slot 4	$\widetilde{\mathbf{u}}_{4,1} = -j\sqrt{\frac{P_2}{P_1+2}}\mathbf{v}_{3,1}$	$\widetilde{\mathbf{u}}_{4,2} = j\sqrt{\frac{P_2}{P_1+2}}\mathbf{v}_{4,2}$

($d = 3$ and 4), where the channel coefficient $\beta_{SR_m,l}$ has the same distribution as the $\alpha_{SR_m,l}$ of $\mathbf{h}_{SR_m}$. Similarly, the impulse response vector $\mathbf{g}_{R_mD}$ of the channel between the mth relay nodes and the destination node for the last two transmit symbols is the $Q \times 1$ vector defined as $\mathbf{g}_{R_mD} = [\beta_{R_mD,0}, \beta_{R_mD,1}, \cdots, \beta_{R_mD,Q-1}]^T$, where the channel coefficient $\beta_{R_mD,q}$ has the same distribution as the $\alpha_{R_mD,q}$ of $\mathbf{h}_{R_mD}$. That means the channels are constant during the first two transmit symbol intervals and the last two transmit symbol intervals, respectively.

Finally, each relay node allows the Alamouti coded form to be constructed at the destination node by transmitting symbol $\widetilde{\mathbf{u}}_{d,m}$ obtained as in Table II to the destination node during four consecutive time slots, where $\widetilde{\mathbf{u}}_{d,m}$ is the dth transmit symbol of the mth relay node.

Now, without loss of generality, we describe the demodulation and decoding steps at the destination node with the received symbols during the first two time slots (that is, the received symbols during the last two time slots can be demodulated and decoded in the same manner). After the CP removal, the received symbols can be expressed as

$$\begin{aligned}\mathbf{y}_1 = \gamma\big[\{\sqrt{P_1/2}(\mathrm{IDFT}_N(\mathbf{C}_1)) \otimes \mathbf{h}'_{SR_1} + \mathbf{n}_{1,1}\} \otimes \mathbf{h}'_{R_1D} \\ + \{\sqrt{P_1/2}(\mathrm{IDFT}_N(\mathbf{C}_2)) \otimes \mathbf{h}'_{SR_2} + \mathbf{n}_{2,2}\} \otimes \mathbf{h}'_{R_2D} \\ \otimes \mathbf{\Gamma}_\tau\big] + \mathbf{w}_1\end{aligned} \tag{21}$$

for the first received symbol and

$$\begin{aligned}\mathbf{y}_2 = \gamma\big[\{\sqrt{P_1/2}(-j\mathrm{IDFT}_N(\mathbf{C}_1)) \otimes \mathbf{h}'_{SR_1} + \mathbf{n}_{1,1}\} \otimes \mathbf{h}'_{R_1D} \\ + \{\sqrt{P_1/2}(j\mathrm{IDFT}_N(\mathbf{C}_2)) \otimes \mathbf{h}'_{SR_2} + \mathbf{n}_{2,2}\} \otimes \mathbf{h}'_{R_2D} \\ \otimes \mathbf{\Gamma}_\tau\big] + \mathbf{w}_2,\end{aligned} \tag{22}$$

for the second received symbol, where $\gamma = \sqrt{\frac{P_2}{P_1+2}}$, $\mathbf{w}_d$ is the AWGN vector with zero mean and unit variance added to the dth received symbol at the destination node. Then, the DFT output is obtained as

$$\begin{aligned}Y_1(k) = \gamma\big[\sqrt{P_1/2}C_1(k)H_{SR_1,p}(k)H_{R_1D,p}(k) \\ + \sqrt{P_1/2}C_2(k)H_{SR_2,p}(k)H_{R_2D,p}(k)e^{-j2\pi k\tau/N} \\ + N_{1,1}(k)H_{R_1D,p}(k) \\ + N_{2,2}(k)H_{R_2D,p}(k)e^{-j2\pi k\tau/N}\big] + W_1(k)\end{aligned} \tag{23}$$

for $\mathbf{y}_1$ and

$$\begin{aligned}Y_2(k) = \gamma\big[\sqrt{P_1/2}\{-jC_1(k)\}H_{SR_1,p}(k)H_{R_1D,p}(k) \\ + \sqrt{P_1/2}\{jC_2(k)\}H_{SR_2,p}(k)H_{R_2D,p}(k)e^{-j2\pi k\tau/N} \\ + N_{1,1}(k)H_{R_1D,p}(k) \\ + N_{2,2}(k)H_{R_2D,p}(k)e^{-j2\pi k\tau/N}\big] + W_2(k)\end{aligned} \tag{24}$$

for $\mathbf{y}_2$, where $N_{d,m}(k)$, $H_{SR_m,p}(k)$, $H_{R_mD,p}(k)$, and $W_d(k)$ are the DFT outputs of $\mathbf{n}_{d,m}$, $\mathbf{h}'_{SR_m}$, $\mathbf{h}'_{R_mD}$, and $\mathbf{w}_d$, respectively. Using the property in (17), we can rewrite (23) and (24) in the following matrix form:

$$\begin{bmatrix} Y_1(k) \\ Y_2^*(k) \end{bmatrix} = \gamma H_p(k) \begin{bmatrix} \sqrt{P_1/2}C_1(k) \\ \sqrt{P_1/2}C_2(k) \end{bmatrix} + \begin{bmatrix} G_{1,p}(k) \\ G_{2,p}(k) \end{bmatrix}, \tag{25}$$

where $G_{1,p}$ and $G_{2,p}$ denote noise term of $Y_1(k)$ and $Y_2(k)$, respectively. $H_p(k)$ is the channel matrix defined as

$$H_p(k) = \begin{bmatrix} H_{1,p}(k) & H_{2,p}(k) \\ H_{2,p}^*(k) & -H_{1,p}^*(k) \end{bmatrix}, \tag{26}$$

where $H_{1,p}(k) = H_{SR_1,p}(k)H_{R_1D,p}(k)$ and $H_{2,p}(k) = H_{SR_2,p}(k)H_{R_2D,p}(k)e^{-j2\pi k\tau/N}$. Clearly, the channel matrix $H_p(k)$ is the Alamouti coded form, and thus, we can obtain the estimates $\widehat{C}_1(k)$ and $\widehat{C}_2(k)$ as

$$\begin{bmatrix} \widehat{C}_1(k) \\ \widehat{C}_2(k) \end{bmatrix} = H_p^H(k) \begin{bmatrix} Y_1(k) \\ Y_2^*(k) \end{bmatrix} \tag{27}$$

for $C_1(k)$ and $C_2(k)$, respectively. The estimates $\widehat{C}_3(k)$ and $\widehat{C}_4(k)$ corresponding to $C_3(k)$ and $C_4(k)$, respectively, can be obtained in the same manner.

Lastly, we can obtain the estimates $\widehat{X}_1(k)$ and $\widehat{X}_2(k)$ as

$$\begin{aligned}\widehat{X}_1(k) = \frac{1}{2}\big[\mathrm{Re}\{\widehat{C}_1(k) + \widehat{C}_3(k)\} - \mathrm{Im}\{\widehat{C}_2(k) + \widehat{C}_4(k)\}\big] \\ + \frac{j}{2}\big[\mathrm{Im}\{\widehat{C}_1(k) + \widehat{C}_3(k)\} - \mathrm{Re}\{\widehat{C}_2(k) + \widehat{C}_4(k)\}\big]\end{aligned} \tag{28}$$

and

$$\begin{aligned}\widehat{X}_2(k) = \frac{1}{2}\big[\mathrm{Im}\{\widehat{C}_1(k) - \widehat{C}_3(k)\} - \mathrm{Re}\{\widehat{C}_2(k) - \widehat{C}_4(k)\}\big] \\ - \frac{j}{2}\big[\mathrm{Re}\{\widehat{C}_1(k) - \widehat{C}_3(k)\} - \mathrm{Im}\{\widehat{C}_2(k) - \widehat{C}_4(k)\}\big]\end{aligned} \tag{29}$$

for the data symbols $X_1(k)$ and $X_2(k)$, respectively, where $\mathrm{Re}\{\cdot\}$ and $\mathrm{Im}\{\cdot\}$ denote real and imaginary parts, respectively.

From (27), we can see that $\widehat{C}_1(k)$ and $\widehat{C}_2(k)$ ($\widehat{C}_3(k)$ and $\widehat{C}_4(k)$) are obtained from $Y_1(k)$ and $Y_2^*(k)$ ($Y_3(k)$ and $Y_4^*(k)$), and thus, destination node requires four OFDM symbols to demodulate two data symbol blocks, resulting in a trade-off between the cooperative diversity order and transmission rate. Specifically, the proposed scheme has the half transmission rate compared with the conventional scheme, while achieving the higher diversity order by averaging more channels.

 ISBN: 978-1-61208-174-8

V. Simulation Results

In this section, the proposed scheme is compared with the conventional scheme in terms of the BER over the frequency selective fading channels. In evaluating the performance, we assume the following parameters as in [10], [11]: an FFT size of $N = 64$ samples, a CP length of $N_G = 16$ samples, binary PSK data modulation. The transmission power at the source node P_1 and at the relay nodes P_2 are assumed as $P_1 = 2P_2$. It is also assumed that the channel has a two path equal-power delay profile with a relative delay of 3 samples between the two paths, and the relative timing difference τ between the symbols arriving at the destination node from the relay nodes is distributed uniformly over $[0, 6]$ samples. From Fig. 2, it is clearly observed that, the conventional scheme demonstrates the same slope of the BER curve as that of the Alamouti scheme for 2×1 multiple-input single-output (MISO) systems when the value of P_1 is large, meanwhile, the proposed scheme shows the same slope of the BER curve as that of the Alamouti scheme for 2×2 MIMO systems, which demonstrates that the proposed scheme achieves a higher order cooperative diversity than that of the conventional scheme. This is due to the fact that, by transmitting the combinations of the data symbols at the source node, each data symbol undergoes more channels than the conventional scheme, and thus, the proposed scheme achieves a higher order cooperative diversity than that of the conventional scheme by averaging more channels than the conventional scheme. From the figure, it is also shown that the BER performance of the proposed (conventional) scheme is degraded compared to the 2×2 MIMO (2×1 MISO) systems while achieving the same order of cooperative diversity. This is due to the fact that the signal is contaminated by noise at both channels between source and relay nodes and relay and destination nodes in the cooperative communication systems, meanwhile, in the MIMO (MISO) systems, the noise is added at the channels between transmitter and receiver only.

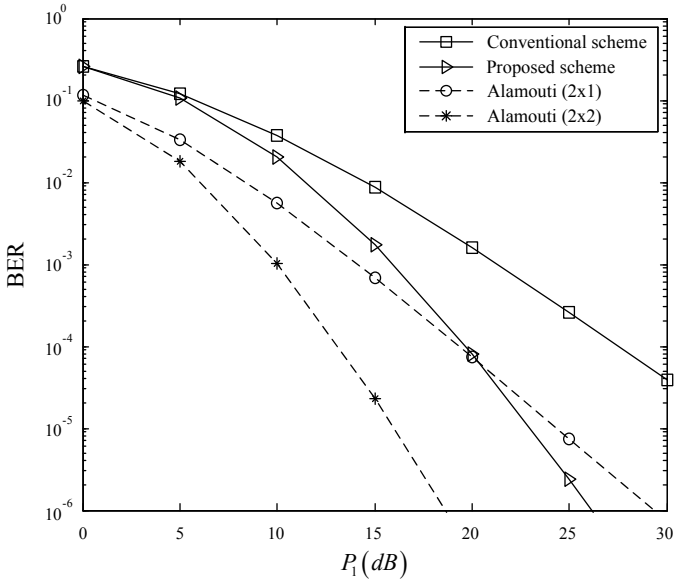


Fig. 2. BER performance of the proposed and conventional schemes as a function of P_1 in the frequency selective fading channels.

VI. Conclusion

In this paper, we have proposed a novel Alamouti coding scheme for asynchronous cooperative communication systems over frequency selective fading channels. The proposed scheme construct the Alamouti coded form at the destination node by using a simple combination of symbols at the source node, resulting in achieving a higher cooperative diversity than the conventional scheme. From the simulation results, it is confirmed that the proposed scheme provides a higher order cooperative diversity than that of the conventional scheme.

Acknowledgment

This work was supported by the National Research Foundation (NRF) of Korea under Grants 2011-0018046 and 2011-0002915 with funding from the Ministry of Education, Science and Technology (MEST), Korea; by a Grant-in-Aid of Samsung Thales; by the Information Technology Research Center (ITRC) program of the National IT Industry Promotion Agency (NIPA) under Grant NIPA-2011-C1090-1111-0005 with funding from the Ministry of Knowledge Economy (MKE), Korea; and by National GNSS Research Center program of Defense Acquisition Program Administration and Agency for Defense Development.

References

[1] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, no. 2, pp. 744-765, Mar. 1998.

[2] Y. Li, W. Zhang, and X.-G. Xia, "Distributive high-rate full-diversity space-frequency codes for asynchronous cooperative communications," *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, pp. 2612-2616, Seattle, WA, July 2006.

[3] A. Nosratinia, T. E. Hunter, and A. Hedayat, "Cooperative communication in wireless networks," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 74-80, Oct. 2004.

[4] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Trans. Inform. Theory*, vol. 50, no. 12, pp. 3062-3080, Dec. 2004.

[5] S. Wei, D. L. Goeckel, and M. C. Valenti, "Asynchronous cooperative diversity," *IEEE Trans. Wireless Commun.*, vol. 5, no. 6, pp. 1547-1557, June 2006.

[6] Y. Li and X.-G. Xia, "Full diversity distributed space-time codes for asynchronous cooperative communications," *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, pp. 911-915, Adelaide, Australia, Sep. 2005.

[7] Y. Shang and X.-G. Xia, "Shift-full-rank matrices and applications in space-time trellis codes for relay networks with asynchronous cooperative diversity," *IEEE Trans. Inform. Theory*, vol. 52, no. 7, pp. 3153-3167, July 2006.

[8] Y. Mei, Y. Hua, A. Swami, and B. Daneshrad, "Combating synchronization errors in cooperative relays," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process. (ICASSP)*, pp. 369-372, Philadelphia, PA, Mar. 2005.

[9] Z. Li and X.-G. Xia, "A simple Alamouti space-time transmission scheme for asynchronous cooperative systems," *IEEE Sig. Process. Lett.*, vol. 14, no. 11, pp. 804-807, Nov. 2007.

[10] Z. Li, X.-G. Xia, and M. H. Lee, "A simple orthogonal space-time-coding scheme for asynchronous cooperative systems for frequency selective fading channels," *IEEE Trans. Commun.*, vol. 58, no. 8, pp. 2219-2224, Aug. 2010.

[11] Y. Jing and B. Hassibi, "Distributed space-time coding in wireless relay networks," *IEEE Trans. Wireless Commun.*, vol. 5, no. 12, pp. 3524-3536, Dec. 2006.

 ISBN: 978-1-61208-174-8

Low Complexity Long PN Code Acquisition Scheme for Spread Spectrum Systems

Jeehyeon Baek[1], Jonghun Park[2], Youngpo Lee[1], Sun Yong Kim[3], Gyu-In Jee[3], Jin-Mo Yang[4], and Seokho Yoon[1],[†]
[1]School of Information and Communication Engineering, Sungkyunkwan University, Suwon, Korea
[2]Samsung Thales Co. LTD., Seongnam, Korea
[3]Department of Electronics Engineering, Konkuk University, Seoul, Korea
[4]Agency for Defense Development (ADD), Daejeon, Korea
[†]Corresponding author
(Email: bjh1987@skku.edu, johnjh.park@samsung.com, leeyp204@skku.edu,
{kimsy, gijee}@konkuk.ac.kr, jmy1965@dreamwiz.com, and [†]syoon@skku.edu)

***Abstract*—In this paper, a low complexity long pseudo noise (PN) code acquisition scheme is proposed for spread spectrum systems including global positioning system (GPS) and code division multiple access (CDMA) system. By using a phase-shift-network, the proposed scheme has less complexity than the conventional dual correlating sequential estimation scheme. From the analytic and numerical results, it is confirmed that the proposed scheme has the lower hardware complexity and the same mean acquisition time performance compared with the conventional scheme.**

***Index Terms*—acquisition, sequential estimation, phase-shift-network, PN code**

I. Introduction

In spread spectrum (SS) receivers, the pseudo noise (PN) code synchronization is one of the most important tasks, which is generally carried out in two stages: code acquisition and tracking [1]. In the code acquisition stage, the coarse alignment between the received and locally generated PN codes is performed, and subsequently, in the code tracking stage, the fine alignment between the two codes is performed.

In recent SS-based mobile communication systems including global positioning system (GPS) and code division multiple access (CDMA) system, a long PN code is essential to provide reliable positioning service or user distinction [2], [3]. However, acquisition of a long PN code leads to excessive acquisition time and hardware complexity for serial and parallel acquisition methods, respectively, which are typical acquisition schemes in SS-based systems [4].

To deal with acquisition of a long PN code, several schemes [5]-[7] have been proposed. Ward proposed an interesting code acquisition scheme based on sequential estimation of the received PN code [5], which is referred to as the traditional sequential estimation (TSE) scheme in this paper. The TSE scheme has a shorter mean acquisition time (MAT) and a lower hardware complexity than the serial and parallel acquisition schemes, respectively; however, the TSE scheme cannot achieve code acquisition when the PN code is inverted due to data modulation. To overcome this drawback, Chiu and Lee proposed an improved sequential estimation (ISE) scheme, which can achieve code acquisition regardless of whether the PN code is inverted or not by using an expanded primitive polynomial [6]. By incorporating the TSE and ISE schemes, Koller and Belkerdid proposed a dual correlating sequential estimation (DCSE) scheme, which can not only achieve acquisition for both inverted and non-inverted PN codes, but also demodulate the SS signal [7], whose hardware complexity is, however, inevitably high.

In this paper, thus, we propose a novel sequential estimation scheme for acquisition of a long PN code, which can achieve acquisition for both inverted and non-inverted PN codes and can also demodulate data as in the DCSE scheme, yet with only half the complexity of the DCSE scheme. We refer to the proposed scheme as phase-shift-network-based differential sequential estimation (PDSE) scheme since it employs a differential operator and a phase-shift-network. The numerical results demonstrate that the proposed scheme has a lower complexity regardless of the length of PN code while maintaining the same level of code acquisition performance compared to that of the conventional DCSE scheme.

This paper is organized as follows. Details of the system model and proposed PDSE scheme are described in Section II. The performance analysis and complexity comparison of conventional and proposed schemes are delineated in detail in Section III. Section IV concludes this paper with a brief summary.

II. Proposed Scheme

Fig. 1 shows the structure of the PDSE receiver proposed in this paper. At the receiver, as in the conventional DCSE scheme, PDSE scheme requires $n+1$ consecutive chips out of $L\ (=\ 2^n-1)$ chips of PN sequence period for PN code acquisition. The kth chip of the received signal can be expressed as

$$r_k = \sqrt{P}T_cD(2s_k-1)+w_k,\ \text{for } k=0,\ 1,\ \cdots,\ n, \quad (1)$$

where P, T_c, and D are the power of the PN signal, duration of a PN chip, and modulated data taking a value in $\{-1,1\}$ with equal probability, respectively, $s_k \in \{0,1\}$ is the kth chip

 ISBN: 978-1-61208-174-8

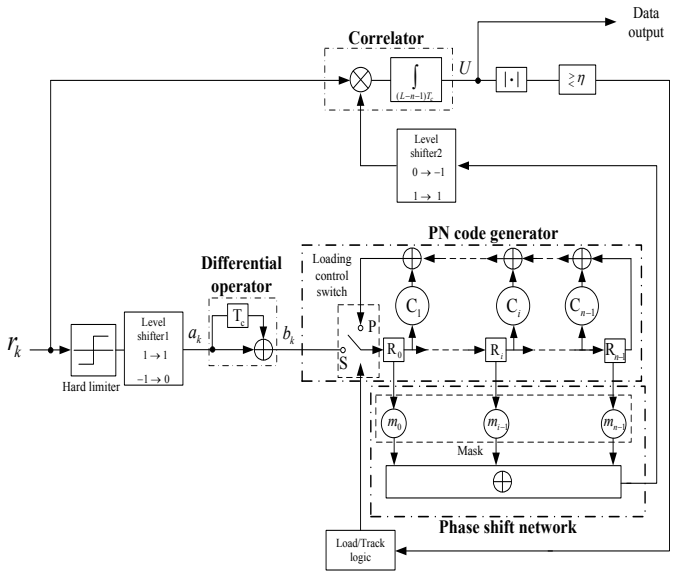


Fig. 1. A structure of the PDSE receiver.

of the PN code with period of L, and w_k is the kth additive white Gaussian noise (AWGN) sample with mean zero and variance of $\sigma_n^2 = N_0 T_c/2$.

First, r_k is mapped into 1 or -1 at the hard limiter with the threshold 0, and subsequently, the kth output a_k of level shifter1 is obtained by converting the value -1 (1) into 0 (1). To achieve code acquisition regardless of whether the PN code is inverted (i.e., $D = -1$) or not (i.e., $D = 1$), we use the differential operator whose output can be expressed as $b_k = a_k + a_{k+1}$ for $k = 0, \cdots, n-1$, which removes the effect of data modulation. Specifically, b_k equals to $s_k + s_{k+1}$ for both cases that $D = 1$ and $D = -1$ in the absence of noise. Moreover, b_k is equivalent to s_{k+l}, the phase-shifted version of s_k with arbitrary phase difference l ($0 \le l \le L$) due to the shift-and-add property [1]. This is the key idea of the proposed PDSE scheme.

In the loading process (i.e., the loading control switch is set to 'S'), $\{b_k\}_{k=0}^{n-1}$ are loaded into registers $\{R_k\}_{k=0}^{n-1}$ in the PN code generator with coefficients of the primitive polynomial $\{c_j\}_{j=1}^{n-1}$. After the loading process, loading control switch is set to 'P' and we can construct a PN code $\{\hat{b}_k\}_{k=0}^{L-1}$ with length of L, the l-shifted version of the PN code with correct phase. Thus, in the absence of noise, we can write $\{\hat{b}_k\}_{k=0}^{L-1}$ as

$$\{\hat{b}_k\}_{k=0}^{L-1} = \{s_k + s_{k+1}\}_{k=0}^{L-1} = \{s_{k+l}\}_{k=0}^{L-1}. \quad (2)$$

To estimate the phase difference l between $\{\hat{b}_k\}_{k=0}^{L-1}$ and $\{s_k\}_{k=0}^{L-1}$, we express $\{\hat{b}_k\}_{k=0}^{L-1}$ in a polynomial forms as $\mathrm{mod}\{(1+x), f(x)\}/f(x)$ and $\mathrm{mod}\{x^l, f(x)\}/f(x)$ from equivalent expressions $\{s_k + s_{k+1}\}_{k=0}^{L-1}$ and $\{s_{k+l}\}_{k=0}^{L-1}$, respectively, where $f(x)$ is the primitive polynomial with coefficients $\{c_j\}_{j=1}^{n-1}$. From the polynomial forms of $\{\hat{b}_k\}_{k=0}^{L-1}$, we can obtain the phase difference l, and then, delaying the phase of the output for $L-l$ by using phase-shift-network, the phase of the output PN code comes to be the same as that of the PN

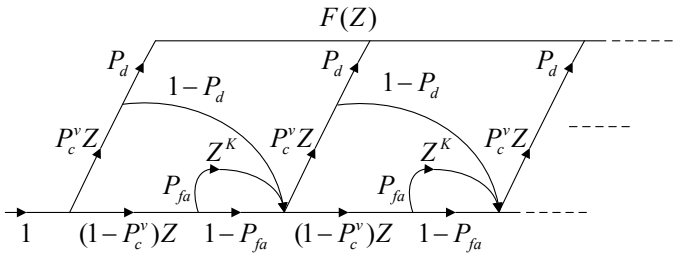


Fig. 2. A generation function flow graph of the acquisition schemes based on the sequential estimation.

code in the received signal, which can be generated by using mask polynomial and expressed as [8]

$$\begin{aligned} m(x) &= \mathrm{mod}\{x^{(L-l)}, f(x)\} \\ &= m_0 + m_1 x + \cdots + m_i x^i + \cdots + m_{n-1} x^{n-1}, \end{aligned} \quad (3)$$

where $m_i \in \{0, 1\}$ denotes the ith coefficient of the mask polynomial.

To determine whether acquisition is achieved or not, we obtain the correlation value U between $\{r_k\}_{k=0}^{n-1}$ and the output of the phase-shift-network changed into bipolar PN code by level shifter2. Then, the absolute correlation value $|U|$ is compared with a given threshold value η. If $|U|$ is smaller than η, the acquisition is determined to be failed and the receiver repeats the acquisition process. If $|U|$ is larger than η, it is determined that the acquisition is achieved. Finally, the PDSE receiver demodulate the data D by using the correlation U comparing with the threshold of zero.

III. Performance Analysis

A. Mean acquisition time

The expression of MAT can be obtained by using a generation function flow graph [9], [10]. Fig. 2 shows a generation function flow graph of the acquisition schemes based on the sequential estimation including TSE, ISE, DCSE, and PDSE, where P_c is the correct chip probability defined as the probability that the chip is estimated correctly, P_c^v is the correct chip probability of consecutive v chips (in DCSE and PDSE schemes, $v = n+1$), P_d is the detection probability defined as the probability to achieve the acquisition when all v chips are estimated correctly, and P_{fa} is the false alarm probability defined as the probability that $|U|$ is larger than η when there exist at least one error in the process of the sequential estimation. K is the penalty factor associated with the false alarm. From Fig. 2, the generation function $F(Z)$ can be derived as

$$F(Z) = \frac{P_c^v P_d Z}{1 - P_c^v(1-P_d)Z - (1-P_c^v)\{P_{fa} Z^{K+1} + (1-P_{fa})Z\}}. \quad (4)$$

From (4), the MAT can be achieved as

$$E[T_{acq}] = \frac{dF(Z)}{dZ} T_e \bigg|_{Z=1} = \frac{1 + (1-P_c^v) K P_{fa}}{P_c^v P_d} T_e, \quad (5)$$

where T_e is the estimation time spent for loading and estimation process. As we can see in (5), the MAT of the DCSE and PDSE schemes depends on the correct chip, detection, and false alarm probabilities.

Since both the hard limiters of the DCSE and PDSE schemes perform the same operation, those schemes have same correct chip probability P_c. The correct chip probability of the kth chip of the received signal is represented as

$$\begin{aligned} P_c &= 1 - \frac{1}{2}\Pr[r_k < 0 | s_k = 1] - \frac{1}{2}\Pr[r_k > 0 | s_k = 0] \\ &= 1 - \frac{1}{2}\int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left\{\frac{(x-\sqrt{P}T_c)^2}{2\sigma_n^2}\right\} dx \\ &\quad - \frac{1}{2}\int_{0}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left\{\frac{(x+\sqrt{P}T_c)^2}{2\sigma_n^2}\right\} dx \\ &= 1 - \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left\{\frac{(x-\sqrt{P}T_c)^2}{2\sigma_n^2}\right\} dx. \end{aligned} \quad (6)$$

When all consecutive $n+1$ chips are estimated correctly (i.e., correct estimation), in the PDSE scheme, the phase of the generated PN code after loading process and that of the PN code from the transmitter are same (the received PN code is non-inverted) or opposite (the received PN code is inverted). Thus, the probability density function (PDF) of the correlation value U can be written in two cases. First, when the received PN code is non-inverted, the PDF of the correlation value U of the PDSE scheme can be written as

$$f_{U|c}(x) = \frac{1}{\sqrt{2\pi}\sigma_N} \exp\left\{-\frac{(x-\sqrt{P}MT_c)^2}{2\sigma_N^2}\right\}, \quad (7)$$

where M $(= L-n-1)$, $\sqrt{P}MT_c$, and $\sigma_N^2 = N_0 MT_c/2$ are the length of the correlation, the mean, and variance of U, respectively. Next, when the received PN code is inverted, the PDF of the correlation value U of the PDSE scheme can be written as

$$f^{\text{inv}}_{U|c}(x) = \frac{1}{\sqrt{2\pi}\sigma_N} \exp\left\{-\frac{(x+\sqrt{P}MT_c)^2}{2\sigma_N^2}\right\}. \quad (8)$$

(8) has an opposite mean and the same variance compared with (7). Since the receiving probabilities of the inverted and non-inverted PN codes are the same, the detection probability of the PDSE scheme can be written as

$$\begin{aligned} P_d &= \frac{1}{2}P\{|U| \geq \eta | \text{correct estimation, non-inverted}\} \\ &\quad + \frac{1}{2}P\{|U| \geq \eta | \text{correct estimation, inverted}\} \\ &= \frac{1}{2}\left\{\int_{\eta}^{\infty} f_{U|c}(x)dx + \int_{-\infty}^{-\eta} f_{U|c}(x)dx\right\} \\ &\quad + \frac{1}{2}\left\{\int_{\eta}^{\infty} f^{\text{inv}}_{U|c}(x)dx + \int_{-\infty}^{-\eta} f^{\text{inv}}_{U|c}(x)dx\right\}. \end{aligned} \quad (9)$$

In (9), $f_{U|c}$ and $f^{\text{inv}}_{U|c}$ are symmetric. Thus, we can rewrite (9) as

$$\begin{aligned} P_d &= \frac{1}{2}\left\{\int_{\eta}^{\infty} f_{U|c}(x)dx + \int_{-\infty}^{-\eta} f_{U|c}(x)dx\right\} \\ &\quad + \frac{1}{2}\left\{\int_{-\infty}^{-\eta} f_{U|c}(x)dx + \int_{\eta}^{\infty} f_{U|c}(x)dx\right\} \\ &= \int_{\eta}^{\infty} f_{U|c}(x)dx + \int_{-\infty}^{-\eta} f_{U|c}(x)dx \\ &= 1 - \Phi\left(\frac{\eta - \sqrt{P}MT_c}{\sigma_N}\right) + \Phi\left(\frac{-\eta - \sqrt{P}MT_c}{\sigma_N}\right), \end{aligned} \quad (10)$$

where $\Phi(x)$ means the cumulative distribution function (CDF) of Gaussian distribution with zero mean and unit variance.

Since the DCSE scheme exploits the absolute correlation value and threshold same as in the PDSE scheme for acquisition, the detection probability of the DCSE scheme is the same as that of the PDSE scheme.

In the other case, if there is at least one chip which is not estimated correctly (i.e., wrong estimation), the phase of the PN code generated after loading process is different with that of the PN code from the transmitter. Let us assume that we use long PN code thus the period of the PN code L is long enough. If the PN codes are miss matched, then the correlation value during M chips can be approximated as $-1/L$. Since $L \gg 1$, the correlation value can be approximated as 0, and thus, the PDF of the correlation value of the PDSE scheme (same as that of the DCSE scheme) in the case of wrong estimation can be written as

$$f_{U|w}(x) = \frac{1}{\sqrt{2\pi}\sigma_N} \exp\left\{-\frac{x^2}{2\sigma_N^2}\right\}. \quad (11)$$

From (11), the false alarm probability can be expressed as

$$\begin{aligned} P_{fa} &= P\{|U| \geq \eta | \text{wrong estimation}\} \\ &= \int_{\eta}^{\infty} f_{U|w}(x)dx + \int_{-\infty}^{-\eta} f_{U|w}(x)dx \\ &= 1 - 2\Phi\left(\frac{\eta}{\sigma_N}\right). \end{aligned} \quad (12)$$

So far, we have analyzed the correct chip, detection, and false alarm probabilities of the PDSE and DCSE schemes. Finally, by substituting (6), (10), and (12) in (5), we can obtain the MATs of the PDSE and DCSE schemes.

We prove the analytic results by using Monte Carlo simulations. Fig. 3 shows the MAT performances of the PDSE and DCSE schemes. We use the primitive polynomials as $1 + x^2 + x^3 + x^4 + x^8$ and $1 + x^3 + x^{10}$ and assume $T_e = LT_c$; $K = 10L$. To calculate the threshold, the false alarm probability is fixed to 0.01. As the analytic results, the simulation results demonstrate that the MAT performances of both DCSE and PDSE schemes are same.

B. Complexity Comparison

In this section, we compare the hardware complexity of the PDSE scheme with that of the DCSE scheme. To compare the hardware complexity, we count the number of materials which compose each scheme as in [11]. Table I shows the number

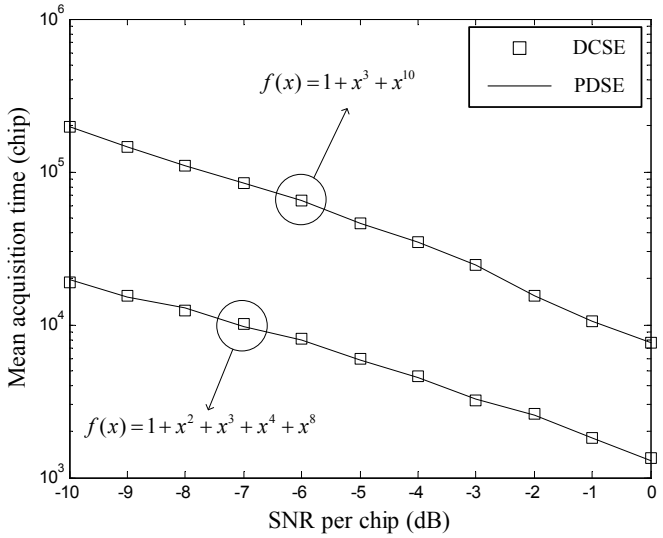


Fig. 3. The MAT performance of the proposed and conventional schemes with primitive polynomials $1 + x^2 + x^3 + x^4 + x^8$ and $1 + x^3 + x^{10}$.

TABLE I
THE NUMBER OF HARDWARES FOR THE PDSE AND DCSE SCHEMES.

Hardwares	PDSE scheme	DCSE scheme
Registers	$n + 1$	$2n + 1$
Correlators	1	2
Absolute operators	1	2
Threshold comparators	1	2
Differential operators	1	0
Inverting controllers	0	1
Combination logics	0	1

of hardwares for the PDSE and DCSE schemes. As shown in Table I, the PDSE scheme needs a half of hardware complexity of the DCSE scheme in terms of the required registers, correlators, absolute operators, and threshold comparators (generally, $n \gg 1$). Moreover, the PDSE scheme only needs a differential operator unlike DCSE scheme which requires logic blocks and an inverting control switch in order to incorporate the TSE and ISE schemes [7].

The proposed PDSE scheme can be applied to the synchronization process of the SS-based communication systems such as GPS and CDMA system, where the fast acquisition of long PN sequence with low hardware complexity is necessary for practical implementation.

IV. CONCLUSION

In this paper, we have proposed a novel sequential estimation scheme for acquisition of a long PN code based on differential estimation and the phase-shift-network. The proposed scheme has a lower complexity compared with the conventional scheme regardless of the length of PN code. Analytic and numerical results confirm that the proposed scheme, with only half complexity, has the same level of the MAT performance to that of the conventional scheme.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation (NRF) of Korea under Grants 2011-0018046 and 2011-0002915 with funding from the Ministry of Education, Science and Technology (MEST), Korea; by a Grant-in-Aid of Samsung Thales; by the Information Technology Research Center (ITRC) program of the National IT Industry Promotion Agency (NIPA) under Grant NIPA-2011-C1090-1111-0005 with funding from the Ministry of Knowledge Economy (MKE), Korea; and by National GNSS Research Center program of Defense Acquisition Program Administration and Agency for Defense Development.

REFERENCES

[1] A. W. Lam and S. Tantaratana, *Theory and Applications of Spread-Spectrum Systems: A Self-Study Course.* IEEE, 1994.

[2] H. Li, M. Lu, X. Cui, and Z. Feng, "Generalized zero-padding scheme for direct GPS P-code acquisition," *IEEE Trans. Wireless Commun.*, vol. 8, no. 6, pp. 2866-2871, June 2009.

[3] R. Kerr and J. Lodge, "Iterative signal processing for blind code phase acquisition of CDMA 1x signals for radio spectrum monitoring," *Journ. Electric. and Comput. Engin.*, vol. 2010, Article ID 282493, Aug. 2010.

[4] M. K. Simon, J. K. Omura, R. A. Scholtz, and B. K. Levitt, *Spread Spectrum Communications Handbook.* McGraw-Hill, 1994.

[5] R. B. Ward, "Acquisition of pseudonoise signals by sequential estimation," *IEEE Trans. Commun.*, vol. 13, no. 4, pp. 474-483, Dec. 1965.

[6] J. H. Chiu and L. S. Lee, "An improved sequential estimation scheme for PN acquisition," *IEEE Trans. Commun.*, vol. 36, no. 10, pp. 1182-1184. Oct. 1988.

[7] G. G. Koller and M. A. Belkerdid, "A dual correlating sequential estimator for spread spectrum PN code acquisition," in *Proc. IEEE MILCOM*, pp. 522-526, Boston, MA, Oct. 1993.

[8] J. S. Lee and L. E. Miller, *CDMA Systems Engineering Handbook.* Artech House, 1998.

[9] A. Polydoros and C. L. Weber, "A unified approach to serial search spread-spectrum code acquisition-part I: general theory," *IEEE Trans. Commun.*, vol. 32, no. 5, pp. 542-549, May 1984.

[10] S. Yoon, I. Song, and S. Y. Kim, "Seed accumulating sequential estimation for PN sequence acquisition at low signal-to-noise ratio," *Signal Process.*, vol. 82, no. 11, pp. 1795-1799, Nov. 2002.

[11] W. Kim, Y. Jung, S. Lee, and J. Kim, "Low complexity demodulation scheme for IEEE 802.15.4 LR-WPAN systems," *IEICE Electron. Express (ELEX)*, vol. 5, no. 14, pp. 490-496, July 2008.

 ISBN: 978-1-61208-174-8

Robust Integer Frequency Offset Estimation Scheme Based on Differentially Combined Correlator Outputs for DVB-T Systems

Jong In Park[1], Hyung-Weon Cho[2], Youngpo Lee[1], Sun Yong Kim[3], Gyu-In Jee[3], Jin-Mo Yang[4], and Seokho Yoon[1],[†]
[1]School of Information and Communication Engineering, Sungkyunkwan University, Suwon, Korea
[2]Samsung Thales Co. LTD., Seongnam, Korea
[3]Department of Electronics Engineering, Konkuk University, Seoul, Korea
[4]Agency for Defense Development (ADD), Daejeon, Korea
[†]Corresponding author
(E-mail: pji17@skku.edu, hyungweon.cho@samsung.com, leeyp204@skku.edu,
{kimsy, gijee}@konkuk.ac.kr, jmy1965@dreamwiz.com, and [†]syoon@skku.edu)

***Abstract*—A pilot-aided integer frequency offset (IFO) estimation scheme robust to the timing offset is proposed for orthogonal frequency division multiplexing (OFDM)-based digital video broadcasting-terrestrial (DVB-T) systems. The proposed scheme first produces correlation values between each continual pilot (CP) and a predetermined scattered pilot (SP), and then, re-correlates the correlation values in order to reduce the influence of the timing offset. Simulation results show that the proposed IFO estimation scheme is robust to the timing offset and has better estimation performance than the conventional scheme.**

***Index Terms*—DVB-T, estimation, IFO, OFDM, timing offset, emerging networks**

I. Introduction

Due to its immunity to multipath fading and high spectral efficiency, orthogonal frequency division multiplexing (OFDM) has been adopted as a modulation format in a wide variety of emerging wireless systems such as digital video broadcasting-terrestrial (DVB-T), digital audio broadcasting (DAB), and wireless local area networks (WLANs) [1], [2]. Moreover, the next generation telecommunication system including long term evolution (LTE) also employs OFDM as the physical layer implementation. However, the OFDM is very sensitive to the frequency offset (FO) caused by Doppler shift or oscillator instabilities, and thus, the FO estimation is one of the most important technical issues in OFDM-based wireless systems [3], [4]. The FO normalized to the subcarrier spacing can be divided into integer and fractional parts that bring on a cyclic shift of the OFDM subcarrier indices and intercarrier interference (ICI), respectively [5]. In this paper, we deal with the integer FO (IFO) estimation for OFDM-based DVB-T systems.

Recently, several IFO estimation schemes [6], [7] have been proposed for OFDM-based DVB-T systems. [6] estimates the IFO by using the correlation between the received pilots and reference pilots in the receiver; however, it exhibits poor performance in the multipath channel. In [7], to alleviate the influence of the multipath, the IFO is estimated based on the correlation between each continual pilot (CP) and its most adjacent scattered pilot (SP) in the received OFDM symbol; however, the scheme assumes that the timing synchronization is perfectly achieved before IFO estimation. In practical systems, only coarse timing synchronization is performed before FO estimation, and fine timing synchronization is performed after FO estimation [5]. Thus, the timing offset often remains during the IFO estimation step and affects the IFO estimation performance.

In this paper, we propose a novel pilot-aided IFO estimation scheme robust to the influence of the timing offset for OFDM-based DVB-T systems. By producing the correlation values between each CP and predetermined SP and re-correlating the correlation values, the proposed scheme reduces the influence of the timing offset on the IFO estimation. Simulation results show that the proposed scheme has more robust and better estimation performance compared with that of the conventional scheme in [7]. The rest of this paper is organized as follows. Section II describes the OFDM-based DVB-T system model. In Section III, we present the conventional IFO estimation scheme, and a novel IFO estimation scheme robust to the influence of the timing offset is proposed in Section IV. Section V demonstrates the IFO estimation performance of the two schemes in the multipath channel. Section VI concludes this paper.

II. System Model

DVB-T systems operate in 2K or 8K mode, of which the former is considered in this paper, where 1705 subcarriers among a total of 2048 subcarriers are used to transmit data and pilots of 45 CPs and 142 or 143 SPs. The pilots are used for frequency and timing synchronization and channel estimation, and have a value of either $+4/3$ or $-4/3$ depending on a pseudo random binary sequence (PRBS). Fig. 1 describes the pilot arrangement in a DVB-T system with 2K mode, where

 ISBN: 978-1-61208-174-8

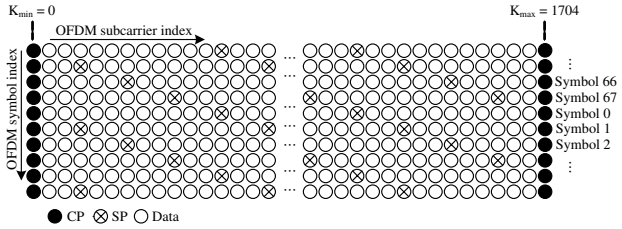


Fig. 1. Pilot arrangement in a DVB-T system with 2K mode.

SPs are periodically inserted every twelve subcarriers between the smallest and largest subcarrier indices $\mathrm{K_{min}}$ and $\mathrm{K_{max}}$ of the active subcarriers in all OFDM symbols and the insertion pattern is repeated every four OFDM symbols [8].

In the transmitter, the OFDM symbol is transmitted with the guard interval inserted at the beginning of the OFDM symbol to prevent the intersymbol interference. After passing through the channel, the n-th sample of the l-th received OFDM symbol in the receiver can be expressed as

$$y_l(n) = x_l(n+\tau)\mathrm{e}^{j2\pi\nu(lN_T+n+\tau)/N} + w_l(n),$$
$$\text{for} \quad l = 0,\ 1,\ \cdots, \quad \text{and} \quad n = 0,\ 1,\ \cdots,\ N-1, \qquad (1)$$

where τ and ν are the timing and frequency offsets normalized to the OFDM sample and subcarrier spacing, respectively, N_T is the number of the samples in an OFDM symbol including the guard interval, N is the size of the inverse fast Fourier transform (IFFT), and $w_l(n)$ is the additive white Gaussian noise (AWGN) sample with mean zero and variance $\sigma_w^2 = \mathbf{E}\{|w_l(n)|^2\}$. The signal $x_l(n)$ can be represented as

$$x_l(n) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X_l(k) H_l(k) \mathrm{e}^{j2\pi kn/N},$$
$$\text{for} \quad l = 0,\ 1, \cdots, \quad \text{and} \quad n = 0,\ 1, \cdots,\ N-1, \qquad (2)$$

where $X_l(k)$ is a pilot or data transmitted through the k-th subcarrier of the l-th OFDM symbol and $H_l(k)$ is the channel frequency response on the k-th subcarrier of the l-th OFDM symbol.

In the receiver, the estimation and compensation processes for the fractional FO (FFO) are generally performed before the IFO estimation [5], in this paper, we assume the perfect estimation and compensation for the FFO. Then the FFT output corresponding to the k-th subcarrier of the l-th received OFDM symbol is expressed as

$$Y_l(k) = \mathrm{e}^{j2\pi\Delta lN_T/N}\mathrm{e}^{j2\pi\tau k/N} \times H_l(k-\Delta)X_l(k-\Delta) + W_l(k), \qquad (3)$$

where $W_l(k)$ is the zero-mean complex AWGN sample in the frequency domain and Δ is the IFO. From (3), we can observe that the IFO and timing offset cause the cyclic shift of the subcarrier indices and the phase rotation in the received OFDM symbol, respectively.

III. Conventional Scheme

The conventional scheme presented in [7] estimates the IFO by using the template $T_m(k)$ expressed as

$$T_m(k) = \frac{Z_m(k')}{Z_m(k)},$$
$$\text{for} \quad k \in C_{cp} \quad \text{and} \quad m \in \{0,1,2,3\}, \qquad (4)$$

where m is the index of the OFDM symbol to distinguish the four different pilot patterns, $Z_m(k)$ denotes the CP with the subcarrier index k in the m-th pilot pattern, and $Z_m(k')$ is the most adjacent SP of $Z_m(k)$, respectively, C_{cp} is the set of the subcarrier indices of the CPs.

In the conventional scheme, the following metric is first performed for all trial values for the IFO estimation in order to recognize the pilot pattern of the received OFDM symbol.

$$m_0 = \underset{m\in\{0,1,2,3\}}{\arg\max} \Big\{\mathbf{Re}\big(\Psi(f,m)\big)\Big\}, \quad \text{for} \quad |f| \le N/2, \quad (5)$$

where $\Psi(f,m) = \sum_{k_m\in C_{cp}} Y_0(k_m+f)Y_0^*(k'_m+f)T_m(k_m)$, f is a trial value, and k_m and k'_m are the indices of the CP and its most adjacent SP in the m-th pilot pattern, respectively. Then, the most reliable α trial values in all trial values are chosen as follows

$$\{f_1,\ldots,f_\alpha\} = \underset{|f|\le N/2}{\arg\max_\alpha} \Big\{\mathbf{Re}\big(\Psi(f,m_0)\big)\Big\}, \qquad (6)$$

where $\underset{|f|\le N/2}{\arg\max_\alpha}\{b(f)\}$ selects the α largest values among f according to the results of $b(f)$. By exploiting the selected α trial values and correlation values of the pilots in the D consecutive OFDM symbols, the estimate $\widehat{\Delta}$ of IFO is obtained as follows

$$\Omega(\overline{f}) = \sum_{k_{m_0\oplus l}\in C_{cp}} \sum_{l=0}^{D-1} Y_l(k_{m_0\oplus l}+\overline{f}) \times Y_l^*(k'_{m_0\oplus l}+\overline{f})T_{m_0\oplus l}(k_{m_0\oplus l}), \qquad (7)$$

$$\widehat{\Delta} = \underset{\overline{f}\in\{f_1,f_2,\ldots,f_\alpha\}}{\arg\max} \Big\{\mathbf{Re}\big(\Omega(\overline{f})\big)\Big\}, \qquad (8)$$

where $\overline{f}$ is an element of the set of trial values selected in (6), D is the number of consecutive OFDM symbols used for IFO estimation, and $(m_0 \oplus l)$ is the remainder operation when the sum of the m_0 and l is divided by four, respectively. If α is set to be 1, (7) and (8) become meaningless. Due to the narrow subcarrier spacing in the OFDM-based DVB-T systems, it can be assumed that the channel frequency responses of a CP and its most adjacent SP are the same. Thus, when there exists the timing offset, (7) can be rewritten as

$$\Omega(\overline{f}) = \sum_{k_{m_0\oplus l}\in C_{cp}} \sum_{l=0}^{D-1} \mathrm{e}^{j2\pi\tau(k_{m_0\oplus l}-k'_{m_0\oplus l})/N} \times \Big|H_l(k_{m_0\oplus l}+\overline{f}-\Delta)\Big|^2 X_l(k_{m_0\oplus l}+\overline{f}-\Delta) \times X_l^*(k'_{m_0\oplus l}+\overline{f}-\Delta)\frac{Z_{m_0\oplus l}(k'_{m_0\oplus l})}{Z_{m_0\oplus l}(k_{m_0\oplus l})} + \widehat{W_l}(k_{m_0\oplus l}), \qquad (9)$$

where $\widehat{W}_l(k_{m_0\oplus l})$ represents the noise term. As shown in (9), the phase rotation caused by the timing offset affects the estimation of the IFO in the conventional scheme.

IV. Proposed Scheme

In this section, we propose a novel IFO estimation scheme robust to the timing offset. In the proposed scheme, the correlation value between each CP and a predetermined SP is calculated for all CPs, and then, the correlation values are classified into several groups according to the predetermined sample distances between the CP and predetermined SP. Since the correlation values with the same sample distance are equally affected by the timing offset, the influence of the timing offset can be removed by re-correlating the classified correlation values.

If the subcarrier indices of the CP and SP are the same, the pilot on the index is considered as the CP, and the most adjacent SP to the CP is called the predetermined SP. According to the DVB-T standard document, the sample distances between a CP and its predetermined SP in an OFDM symbol is among ± 3, ± 6, ± 9, and ± 12 [8], and thus, the pilot pattern of the received OFDM symbol is recognized as follows

$$m_0 = \underset{m\in\{0,1,2,3\}}{\arg\max} \Big\{ \mathbf{Re}\big(\Lambda(f,m)\big)\Big\}, \quad \text{for} \quad |f| \le N/2, \quad (10)$$

where $\Lambda(f,m)$ is given by

$$\begin{aligned}\Lambda(f,m) = &\sum_{g\in G}\sum_{i=1}^{G_n(m,g)-1}\sum_{j=i+1}^{G_n(m,g)} Y_0(I_{g,m}(i)+f)\\ &\times Y_0^*(I_{g,m}(i)+g+f)T_m(I_{g,m}(i))\\ &\times \Big\{Y_0(I_{g,m}(j)+f)Y_0^*(I_{g,m}(j)+g+f)\\ &\times T_m(I_{g,m}(j))\Big\}^*, \end{aligned} \quad (11)$$

where g is the sample distance between the CP and predetermined SP, G is the set of the sample distances between the CPs and predetermined SPs, and $G_n(m,g)$ is the number of the CPs with the sample distance g in the m-th pilot pattern. $I_{g,m}(i)$ is the subcarrier index of the i-th CP included in the group with the sample distance g in the m-th pilot pattern. The template of the proposed scheme is expressed as

$$\begin{aligned} T_m\big(I_{g,m}(k)\big) &= \frac{Z_m(I_{g,m}(k)+g)}{Z_m(I_{g,m}(k))},\\ \text{for} \quad k\in C_{cp} \quad &\text{and} \quad m\in\{0,1,2,3\}. \end{aligned} \quad (12)$$

After recognizing the pilot pattern of the received OFDM symbol, we select the most reliable α trial values as follows:

$$\{f_1,\ldots,f_\alpha\} = \underset{|f|\le N/2}{\arg\max_\alpha}\Big\{\mathbf{Re}\big(\Lambda(f,m_0)\big)\Big\}. \quad (13)$$

By exploiting the selected α trial values and correlation values of the pilots in the D consecutive OFDM symbols, the proposed scheme estimates the IFO as follows:

$$\begin{aligned}\Gamma(\overline{f}) = &\sum_{l=0}^{D-1}\sum_{g\in G}\sum_{i=1}^{G_n(m,g)-1}\sum_{j=i+1}^{G_n(m,g)} Y_l(I_{g,m_0\oplus l}(i)+\overline{f})\\ &\times Y_l^*(I_{g,m_0\oplus l}(i)+g+\overline{f})T_{m_0\oplus l}(I_{g,m_0\oplus l}(i))\\ &\times \Big\{Y_l(I_{g,m_0\oplus l}(j)+\overline{f})Y_l^*(I_{g,m_0\oplus l}(j)+g+\overline{f})\\ &\times T_{m_0\oplus l}(I_{g,m_0\oplus l}(j))\Big\}^* \end{aligned} \quad (14)$$

and

$$\widehat{\Delta} = \underset{\overline{f}\in\{f_1,f_2,\ldots,f_\alpha\}}{\arg\max} \Big\{\mathbf{Re}\big(\Gamma(\overline{f})\big)\Big\}, \quad (15)$$

where $\overline{f}$ is an element of the set of trial values selected in (13).

Assuming that the channel frequency responses of the CP and predetermined SP are the same, when there exists a timing offset, we can re-write (14) as

$$\begin{aligned}\Gamma(\overline{f}) = &\sum_{l=0}^{D-1}\sum_{g\in G}\sum_{i=1}^{G_n(m,g)-1}\sum_{j=i+1}^{G_n(m,g)} \Big|H_l(A)\Big|^2 X_l(A)\\ &\times X_l^*(A+g)\frac{Z_{m_0\oplus l}(I_{g,m_0\oplus l}(i)+g)}{Z_{m_0\oplus l}(I_{g,m_0\oplus l}(i))}\Big|H_l(B)\Big|^2\\ &\times X_l^*(B)X_l(B+g)\\ &\times \Big\{\frac{Z_{m_0\oplus l}(I_{g,m_0\oplus l}(j)+g)}{Z_{m_0\oplus l}(I_{g,m_0\oplus l}(j))}\Big\}^*\\ &+ \widehat{W}_l(I_{g,m_0\oplus l}(j)), \end{aligned} \quad (16)$$

where $A = I_{g,m_0\oplus l}(i)+\overline{f}-\Delta$, $B = I_{g,m_0\oplus l}(j)+\overline{f}-\Delta$, and $\widehat{W}_l(I_{g,m_0\oplus l}(j))$ represents the noise term.

V. Simulation Results

In this section, the proposed scheme is compared with the conventional scheme in terms of the IFO detection probability in the multipath channel. We consider a DVB-T system with 2K mode and quadrature amplitude modulation. The parameters used in the simulation are as follows: the guard interval size of 128, $D = 1$ and 2, $N = 2048$, $\Delta = 1$, and $\alpha = N/4$. The signal-to-noise ratio (SNR) is defined as σ_x^2/σ_w^2 where $\sigma_x^2 = \mathbf{E}\{|x_l(n)|^2\}$. We consider a ten-path Rayleigh fading channel with path delays of $0, 10, \cdots, 90$ samples and an exponential power delay profile where the power ratio of the first path to the last path is 20 dB. The phase of each path is assumed to be distributed uniformly in $(-\pi,\pi]$ and the maximum Doppler frequency is set to be 100 Hz.

Fig. 2 shows the IFO detection probabilities of the proposed and conventional schemes as a function of the timing offset normalized to the length of one OFDM symbol when the SNR is 5 dB. As shown in Fig. 2, the IFO detection probability of the conventional scheme is severely degraded as the timing offset is increased, whereas that of the proposed scheme is almost constant regardless of the timing offset value.

Figs. 3-5 show the IFO detection probabilities of the proposed and conventional schemes as a function of the SNR when the timing offset is $1/64$, $2/64$, and $3/64$, respectively.

 ISBN: 978-1-61208-174-8

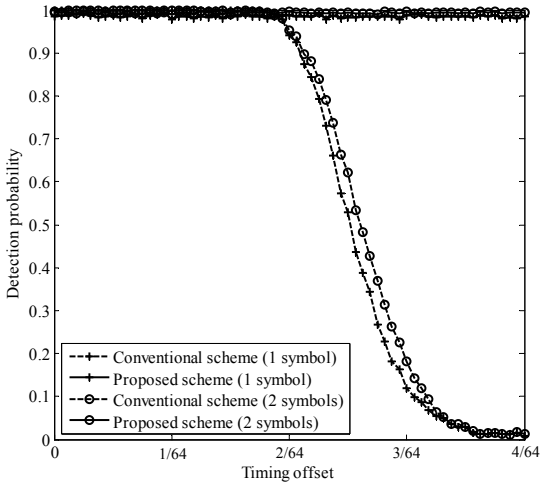


Fig. 2. IFO detection probabilities of the conventional and proposed schemes when the SNR is 5 dB.

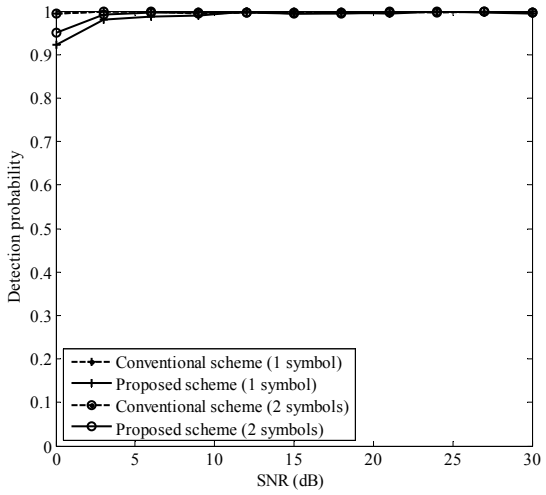


Fig. 3. IFO detection probabilities of the conventional and proposed schemes when the timing offset is $1/64$.

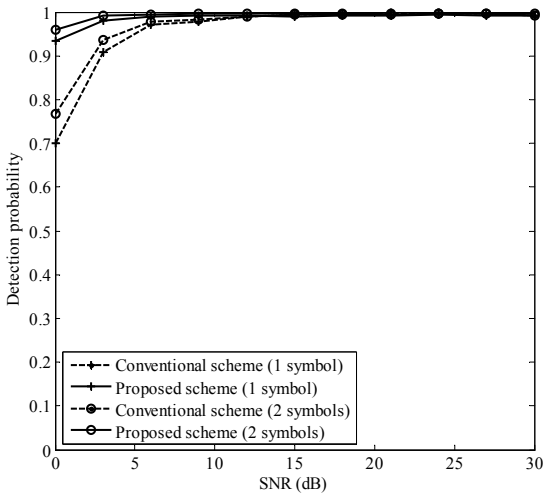


Fig. 4. IFO detection probabilities of the conventional and proposed schemes when the timing offset is $2/64$.

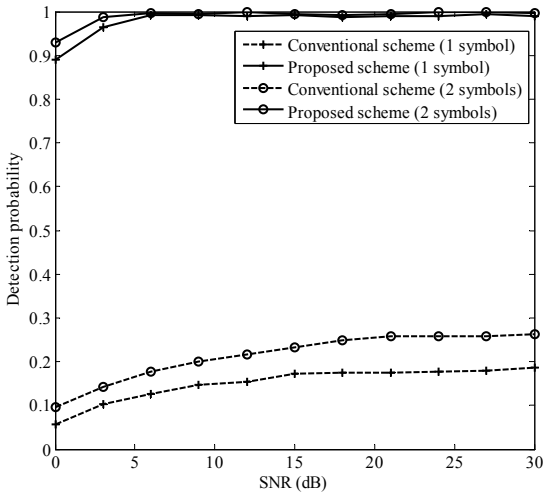


Fig. 5. IFO detection probabilities of the conventional and proposed schemes when the timing offset is $3/64$.

From the figures, it is found that the difference in IFO estimation performance between the proposed and conventional schemes becomes larger as the timing offset is increased.

The proposed scheme has a better performance under the influence of timing offset, however, it requires higher computational complexity than the conventional scheme. Thus, the robust scheme to the timing offset with a low complexity will be studied in our future work.

VI. Conclusion

By first producing the correlation values between CPs and their own predetermined SPs and then re-correlating the correlation values, we have proposed a novel IFO estimation scheme robust to the influence of the timing offset. From the simulation results, it has been confirmed that the proposed scheme has more robust and better estimation performance compared with that of the conventional scheme.

Acknowledgment

This work was supported by the National Research Foundation (NRF) of Korea under Grants 2011-0018046 and 2011-0002915 with funding from the Ministry of Education, Science and Technology (MEST), Korea; by a Grant-in-Aid of Samsung Thales; by the Information Technology Research Center (ITRC) program of the National IT Industry Promotion Agency (NIPA) under Grant NIPA-2011-C1090-1111-0005 with funding from the Ministry of Knowledge Economy (MKE), Korea; and by National GNSS Research Center program of Defense Acquisition Program Administration and Agency for Defense Development.

References

[1] A. Filippi and S. Serbetli, "OFDM symbol synchronization using frequency domain pilots in time domain," *IEEE Trans. Wireless Commun.*, vol. 8, no. 6, pp. 3240-3248, June 2009.

 ISBN: 978-1-61208-174-8

[2] M. M. Wang, L. Xiao, T. Brown, and M. Dong, “Optimal symbol timing for OFDM wireless communications,” *IEEE Trans. Wireless Commun.*, vol. 8, no. 10, pp. 5328-5337, Oct. 2009.
[3] M. Morelli and M. Moretti, “Integer frequency offset recovery in OFDM transmissions over selective channels,” *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5220-5226, Dec. 2008.
[4] P. Dharmawansa, N. Rajatheva, and H. Minn, “An exact error probability analysis of OFDM systems with frequency offset,” *IEEE Trans. Commun.*, vol. 57, no. 1, pp. 26-31, Jan. 2009.
[5] M. Speth, S. Fechtel, G. Fock, and H. Meyr, “Optimum receiver design for OFDM-based broadband transmission - part II: a case study,” *IEEE Trans. Commun.*, vol. 49, no. 4, pp. 571-578, Apr. 2001.
[6] P. Liu, B.-B. Li, Z.-Y. Lu, and F.-K. Gong, “A new frequency synchronization scheme for OFDM,” *IEEE Trans. Consumer Electron.*, vol. 50, no. 3, pp. 823-828, Aug. 2004.
[7] K.-T. Lee and J.-S. Seo, “Pilot-aided frequency offset estimation for digital video broadcasting systems,” *IEICE Trans. Commun.*, vol. E90-B, no. 11, pp. 3327-3329, Nov. 2007.
[8] ETSI EN 300 744, “Digital video broadcasting (DVB); framing structure, channel coding and modulation for digital terrestrial television,” *ETSI, Tech. Rep.*, Jan. 2001.

 ISBN: 978-1-61208-174-8

Investigating the Robustness of Detection vis-à-vis the Detector Threshold in WSN with Fading MAC and Differing Sensor SNRs – Optimal Sensor Gains vs. Uniform Sensor Gains

R. Muralishankar*, H. N. Shankar†, Manisha Sinha‡ and Aniketh Venkat§
*CMR Inst. of Technology, Bangalore, INDIA.
muralishankar@cmrit.ac.in
†Dean – Academics and Research,
CMR Inst. of Technology, Bangalore, INDIA.
hnshankar@cmrit.ac.in
‡Indian Institute of Science, Bangalore, INDIA.
manisha@mbu.iisc.ernet.in
§Banashankari 3rd Stage, Bangalore, INDIA.
anikethvenkat@gmail.com

***Abstract*—In Wireless Sensor Networks, if the Multiple Access Channel between distributed sensors and multiple antennas is fading and the envelope of the channel gain distribution is unknown and time-varying, fusion at the antennas is usually incoherent. Often, the overall sensor power is upper bounded by a constraint on the onboard battery power. Then, the optimal sensor power allocation scheme which minimizes the probability of missed detection is kno-wn to outperform uniform sensor power allocation scheme. Further, if the observation signal-to-noise ratios at the sensors are non-identical, optimizing the probability of detection must take into account the combined effect of the differing sensor signal-to-noise ratios and the fading nature of the channel *as seen by the sensors*. Neyman-Pearson formulation of this problem sets out by setting an upper bound on the permissible probability of false alarm. Consequently, the detector threshold is governed by the power allocation scheme – uniform or optimal. We examine here the inter-dependencies between the probability of false alarm, the probability of detection and the detector threshold. We demonstrate that for robust detection vis-à-vis variations in detector threshold, there is an additional compelling case for optimal power allocation over uniform power allocation.**

***Keywords*-Wireless Sensor Networks; Multiple Access Fading Channel; Optimal Power Allocation; Detector Threshold; Robustness.**

I. INTRODUCTION

We address the problem of distributed detection over a resource constrained Wireless Sensor Network (WSN). The schematic of the system taken from [1] is shown in Fig 1. On-board batteries with limited power drive the sensors. The sensed signal received by the sensors is corrupted by additive noise, amplified by the sensor gain and transmitted over a fading channel to the Fusion Centre (FC). To detect the sensed parameter/event at the FC, we employ the Neyman-Pearson (NP) formulation.

A. State of the Art

Uniform Power Allocation (UPA) to the sensors is shown to be sub-optimal when the Multiple Access Channel (MAC) is fading [1]. The authors there show that Optimal Power Allocation (OPA) is superior to UPA under the following conditions: (i) the channel is fading; (ii) the sensor observation noise is i.i.d.; (iii) the sensor observation Signal-to-Noise Ratio (SNR) is time-invariant; (iv) there is an overall sensor power constraint; and (v) the False Alarm (FA) rate has a fixed acceptable upper bound. Thus, there is a saving in onboard power even with i.i.d. sensor observation noise and time-invariant sensor observation SNR. However, with non-identical sensor observation SNRs, the OPA of [1] may lead to wastage of system resources. OPA for the case of sensor noise with different SNRs is addressed in [2].

B. Motivation for this Work

It was seen in Section I-A that UPA is sub-optimal when the MAC is fading and that OPA is superior to UPA under some conditions [1]. Suppose the sensor observation SNRs are non-identical. Then, the OPA of [1] which does not take into account the combined effect on the overall performance of (i) the differing sensor SNRs; and (ii) the particular fading characteristic of the channel path seen by the individual sensors, has been shown to result in wastage of system resources [2].

Within a fixed permissible probability of FA, P_{FA}, the NP scheme admits a choice of the detector threshold. Of particular concern that we focus upon here is from a designer's perspective. It lies in examining the interval allowable to choose the detector threshold, τ, within the constraints placed by the tolerable P_{FA} concurrent with the desired probability of detection, P_D, albeit for a given total sensor power constraint, P_T.

 ISBN: 978-1-61208-174-8

Moreover, as τ varies within the admissible interval of τ, it is desired to study the nature of degradation of performance. In relation to P_{FA}, the detector threshold, τ, impacts P_D. The natural question that arises here is therefore the 'goodness' of the choice of τ for 'enhanced performance'. To answer this question, we study the nature of inter-relationships between P_{FA}, τ and P_D. The comparison here is across OPA and UPA. Thus, equipped with a priori knowledge of the nature of impact of τ on system performance, the user can choose τ to balance conflicting criteria, dictated by the demands of the specific application.

The rest of the paper is organized as follows. Section II describes the system model with the power constraint. The detection algorithm and its formulation comprise Section III. Power allocation schemes are discussed in Section IV. Simulation set up and results are presented and analyzed in Section V. Concluding remarks form Section VI.

II. System Description and Model

To facilitate readability, we briefly describe the system setup on the lines of the formulation in [2].

A. The Sensor

The sensor network comprises L sensors transmitting to N antennas over NL channels as in Fig. 1 [3]. The sensed parameter/event is $\Theta \in \{0, \theta\}$. Let H_0 and H_1 be the hypotheses corresponding respectively to $\Theta = 0$ and $\Theta = \theta$. Further, in terms of the priors p_0 and p_1, let

$$\Theta \triangleq \begin{cases} H_0 : 0 & w.p.\ p_0; \\ H_1 : \theta & w.p.\ p_1. \end{cases}$$

At the ℓth sensor, the additive noise, η_ℓ, is characterized as $\eta_\ell \sim \mathcal{CN}(0, {\sigma_\ell}^2)$. Thus the sensor SNRs are not identical. The gain of the ℓth sensor is $\alpha_\ell \in \mathbb{C}$. Then, the sensor output is

$$\alpha_\ell(\Theta + \eta_\ell);\ \ell = 1, 2, \cdots, L.$$

The constraint on the overall sensor power, P_T, is given by

$$\begin{aligned} P_T &= E\left[\sum_{\ell=1}^{L} \mid \alpha_\ell(\Theta + n_\ell) \mid^2\right] && (1) \\ &= \sum_{\ell=1}^{L} \mid \alpha_\ell \mid^2 (p_1\theta^2 + \sigma_\ell^2) \\ &= \alpha^H \left[p_1\theta^2 \mathbf{I}_L + \mathbf{D}(\sigma)\right] \alpha, && (2) \end{aligned}$$

in matrix notations. Here, $E[\cdot]$ is the expectation operator; $\alpha,\ \sigma \in \mathbb{C}_{L\times 1}$; x^H is the Hermitian of x; $\mathbf{I}_L$ is the identity matrix of size L; and $\mathbf{D}(u) \in \mathbb{C}_{m\times m}$ is the diagonal matrix with the entries of the vector $u \in \mathbb{C}_{m\times 1}$ on the diagonal.

B. Multiple Access Channel (MAC)

The sensors feed into a multiple access fading channel. The random gain from the ℓth sensor to the nth antenna is $h(n, \ell)$.

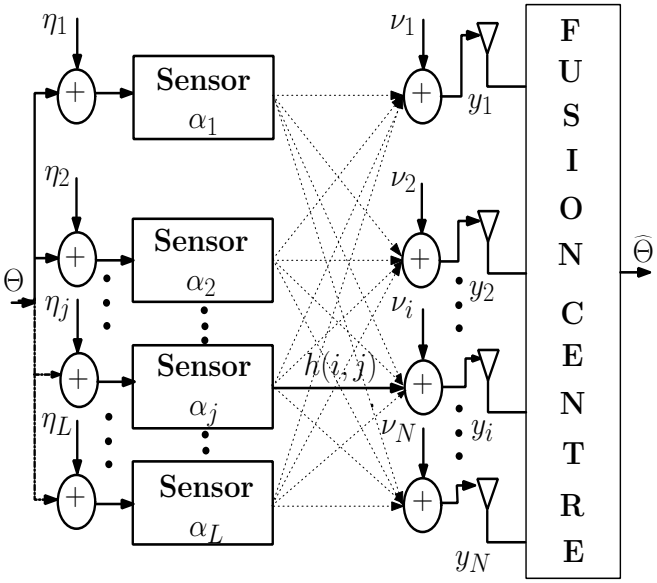


Fig. 1. Schematic of the Setup

The channel gain matrix is thus $\mathbf{H} = [h(n, \ell)] \in \mathbb{C}_{N\times L}$.

C. Antennas and the Fusion Center

There are N antennas at the receiving end. The additive noise, $v_n \sim \mathcal{CN}(0, {\sigma_v}^2)$, at the nth antenna, $n = 1, 2, \cdots, N$, is taken to be i.i.d. For simplicity, we assume the antenna noise to be of unit variance. The output of the N antennas received by the FC is $y \in \mathbb{C}_{N\times 1}$. Thus,

$$y = \mathbf{H}\alpha\Theta + \mathbf{HD}(\alpha)\eta + v, \quad (3)$$

where $\eta \in \mathbb{C}_{L\times 1}$ and $v \in \mathbb{C}_{N\times 1}$. From the observed output, $\widehat{\Theta} \in \{0, \hat{\theta}\}$, at the FC, the problem is to detect the parameter, $\Theta \in \{0, \theta\}$, emitted by the source and to analyze the system performance.

III. Detection Algorithm

For detection, we assume y to be Gaussian. Thus,

$$\begin{aligned} H_0 &: y \sim \mathcal{CN}(\mathbf{0}_N, \mathbf{R}); \\ H_1 &: y \sim \mathcal{CN}(\theta\mathbf{H}\alpha, \mathbf{R}). \quad (4) \end{aligned}$$

Here, $\mathbf{0}_N$ is the $N \times 1$ zero vector and $\mathbf{R}$ is the $N \times N$ covariance matrix of the received signal given by

$$\mathbf{R} = \mathbf{HD}(\alpha)\mathbf{D}(\sigma)\mathbf{D}(\alpha)^H\mathbf{H}^H + \mathbf{I}_N. \quad (5)$$

Define

$$\delta \triangleq \alpha^H\mathbf{H}^H\mathbf{R}^{-1}\mathbf{H}\alpha \quad (6)$$

and

$$Q(x) \triangleq (1/\sqrt{2\pi})\int_x^\infty \mathrm{e}^{-t^2/2}\mathrm{d}t.$$

Then, the probability of false-alarm, P_{FA}, becomes

$$P_{FA} \leq Q\left(\frac{\theta\sqrt{\delta}}{2} + \frac{\tau}{\theta\sqrt{\delta}}\right). \quad (7)$$

In (7), τ is the detector threshold which is a consequence of the likelihood ratio,

$$\frac{\Pr\{y \mid H_1\}}{\Pr\{y \mid H_0\}}.$$

Specifically, τ influences detection in accordance with

$$\left(\theta y^H \mathbf{R}^{-1}\mathbf{H}\alpha\right) \underset{H_0}{\overset{H_1}{\gtrless}} \left(\frac{1}{2}\theta^2\alpha^H\mathbf{H}^H\mathbf{R}^{-1}\mathbf{H}\alpha + \tau\right) = \frac{\theta^2\delta}{2} + \tau. \quad (8)$$

Finally, the probability of missed detection, P_{MD}, and hence, the probability of detection, P_D, are given by

$$P_{MD} = 1 - P_D \leq \left[1 - Q\left(Q^{-1}(P_{FA}) - \theta\sqrt{\delta}\right)\right]. \quad (9)$$

IV. Power Allocation Algorithms

We discuss here two schemes, viz., Uniform PA and Optimal PA. The relations of Section II and Section III are valid for both these schemes, although they represent different quantities in the two schemes.

A. Uniform Power Allocation (UPA)

The total sensor power with UPA is equally distributed among L sensors as P_T/L, thus giving the sensor gains as

$$\alpha_{uni,\ell} = \sqrt{P_T/L},\ \ell \in \{1, 2, \cdots, L\}.$$

Setting $\alpha = \alpha_{uni}$ in (5) and (6), we obtain R and δ. We stipulate the maximum allowable false alarm P_{FA}. Then, taking the equality in (7), we solve for the corresponding limiting detector threshold, $\tau = \tau_{uni}$. Similarly, taking the equality in (9), we solve for the corresponding limiting probability of detection, $P_D = P_{Duni}$.

It is shown in [1] that this UPA results in wastage of system resources if the channel is fading and/or if the sensor observation SNRs are not identical [2], albeit time-invariant. This brings us to the optimal PA scheme.

B. Optimal Power Allocation (OPA)

In the context of a fading MAC, if the sensor observation SNRs are time-varying and/or non-identical, the optimal PA scheme proposed in the setting of [1] is indeed non-optimal. Even though the channel considered there is fading, the sensor noise is i.i.d. Hence, if the sensor observation noise is not identical, due to very poor SNR of a certain sensor, amplified noise transmitted by it over even a noise-free channel may lead to (i) missed detection, (ii) false alarm and (iii) wastage of resources. A comprehensive optimization algorithm must therefore consider the combined effect on detection of the differing sensor SNRs and the fading channel, subject to a total power constraint [2].

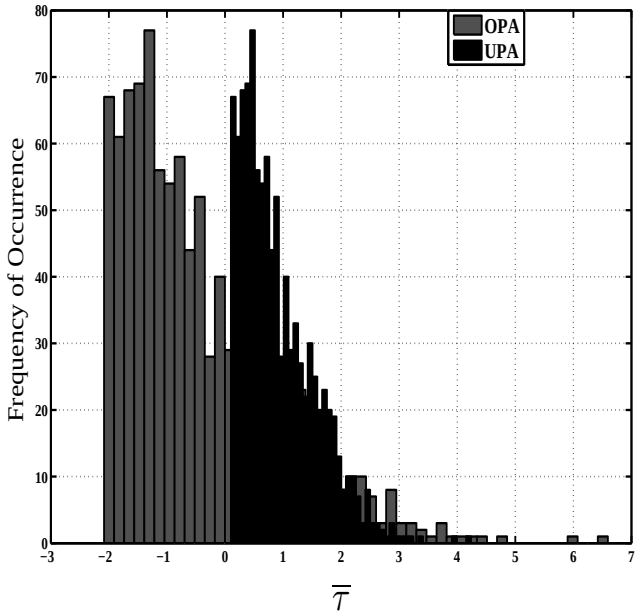


Fig. 2. Distribution of mean detector threshold for UPA and OPA

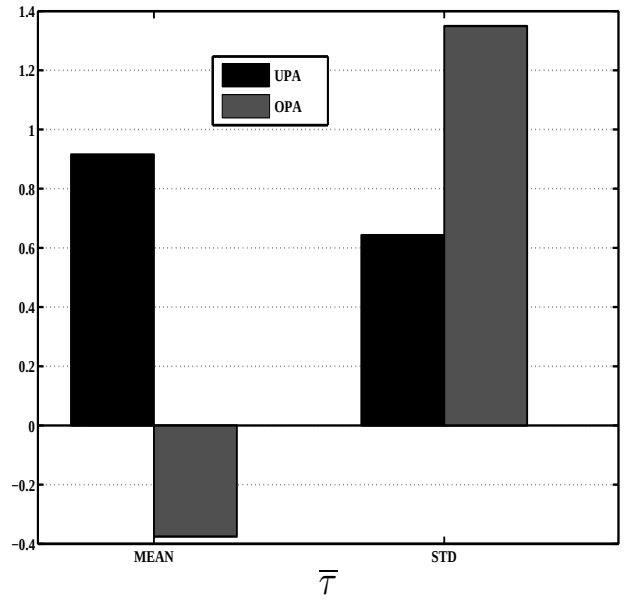


Fig. 3. Statistics of mean detector threshold

In this backdrop, (9) shows that for a fixed P_{FA}, maximizing δ is equivalent to minimizing P_{MD}. It is clear from (6) that this requires choosing α that maximizes δ. Thus, the problem of maximizing P_D reduces to finding that optimal sensor gain, α_{opt}, such that

$$\alpha_{opt} = \underset{\alpha}{\mathbf{argmax}}\left[\alpha^H\mathbf{H}^H\mathbf{R}^{-1}\mathbf{H}\alpha\right], \quad (10)$$

 ISBN: 978-1-61208-174-8

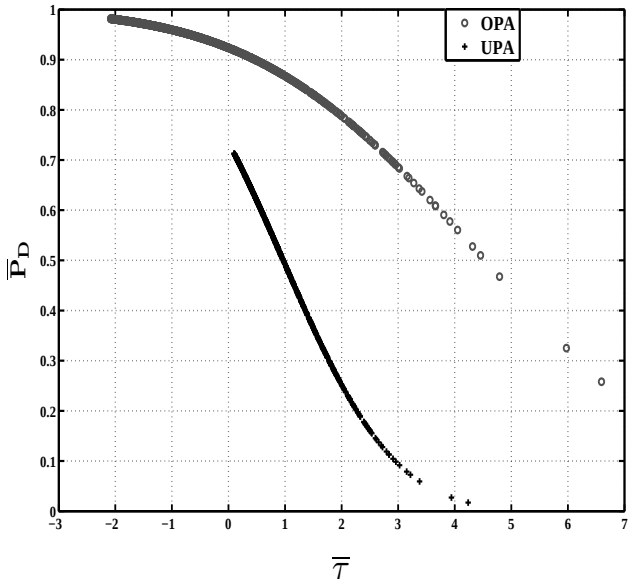


Fig. 4. Mean probability of detection vs. Mean detector threshold

subject to the power constraint of (2) rewritten as

$$P_T = \sum_{\ell=1}^{L} \left[|\alpha_\ell|^2 \left(p_1\theta^2 + {\sigma_\ell}^2 \right) \right] \triangleq \sum_{\ell=1}^{L} |\alpha_\ell \beta_\ell|^2 \triangleq \sum_{\ell=1}^{L} |\gamma_\ell|^2. \quad (11)$$

Here, $\beta_\ell = \sqrt{(p_1\theta^2 + \sigma_\ell^2)}$. By sampling on a grid on the surface of the sphere of radius $\sqrt{P_T}$ and centered at the origin of the L dimensional complex space, we obtain candidates for γ. Since p_1, θ and σ_ℓ are known apriori, β_ℓ can be calculated $\forall \ell$ and hence, α_ℓ also using $\alpha_\ell = \gamma_\ell / \beta_\ell$ from (11). The candidates of α thus derived are used in (10) yielding α_{opt}. Now, we may write

$$\alpha_{opt,\ell} = \frac{\gamma_{opt,\ell}}{\beta_\ell} = \frac{\gamma_{opt,\ell}}{\sqrt{(p_1\theta^2 + \sigma_\ell^2)}} \ . \quad (12)$$

Clearly, $\alpha_{opt,\ell}$ from (12) depends on σ_ℓ. Hence, for sensors with different observation SNRs and for different realizations of the random channel matrix, H, we run the optimization algorithm again to find a new α_{opt}. After substituting $\alpha = \alpha_{opt}$ in (5) and (6), we obtain the corresponding R and δ respectively. Hereafter, similar to the procedure in Section IV-A, we specify the desired maximum allowable false alarm P_{FA}, the same as was stipulated in Section IV-A. Considering the equality in (7), we solve for the corresponding limiting detector threshold, $\tau = \tau_{opt}$. Likewise, taking the equality in (9), we solve for the corresponding probability of detection, $P_D = P_{Dopt}$, the desired optimum.

It was seen in Section I-B that the question we seek to answer concerns the 'goodness' of the choice of τ for 'enhanced performance'. Towards this end, we study the nature of inter-relationships between P_{FA}, τ and P_D, across both OPA and UPA. The relative computational burden with OPA over UPA must indeed be justified by commensurate profit in performance.

V. Simulation Details, Results and Discussions

The dependence that we seek to study in Section I-B is through simulations over a slew of P_{FA}. Thus, for each of the several different values of fixed P_{FA}, the goal here is to compare and analyze the behavior of τ_{uni} and P_{Duni} from UPA of Section IV-A vis-à-vis the behavior of τ_{opt} and P_{Dopt} from OPA of Section IV-B, for a given power constraint, P_T. We first present the simulation settings.

A. Simulation Details

We simulate with the parameter being sensed, $\theta = 1$, no. of sensors, $L = 5$, no. of antennas, $N = 3$, probability of the null hypothesis, $p_0 = 0.4$ and the observation SNRs, θ^2/σ_ℓ^2 being 5 dB, 10 dB, 0 dB, 15 dB and 20 dB respectively for the sensors $\ell = 1, \cdots, L$. The power constraint is $P_T = 1$. We run Monte Carlo simulations with 1000 realizations of the random channel envelope, H, corresponding to a fixed P_{FA}. For each of these 1000 realizations, we implement the following.

1. By the UPA scheme of Section IV-A, obtain τ_{uni} and P_{Duni} through α_{uni}.
2. By the OPA scheme of Section IV-B, obtain τ_{opt} and P_{Dopt} through α_{opt}.

We run 1000 such epochs after taking 1000 samples of P_{FA} drawn from a uniform distribution supported on $(0, 0.2)$. The choice of this upper bound of 0.2 for P_{FA} is such that it is 50% of p_0 which has been set to 0.4.

B. Results and Discussions

As stated in Section I, we seek to study of the nature of inter-relationships between P_{FA}, τ and P_D. Specifically, we report here the initial results of an ongoing investigation into the influence of the detector threshold on P_{FA} and P_D. Towards this end, for each of the 1000 instances of $P_{FA} \in (0, 0.2)$, we average τ and P_D over 1000 realizations of the random channel matrix, H. Let the averaged values for the UPA scheme be $\overline{\tau}_{uni}$ and $\overline{P}_{Duni}$, and the corresponding quantities for the OPA scheme be $\overline{\tau}_{opt}$ and $\overline{P}_{Dopt}$. Thus, we get 1000 each of the above four averaged quantities.

Fig. 2 shows the histograms of $\overline{\tau}_{uni}$ and $\overline{\tau}_{opt}$. The standard deviation of the mean of the detector threshold, $\overline{\tau}_{uni}$, is 0.643 with the UPA scheme, whereas the corresponding standard deviation for the OPA scheme is 1.35. This comparison along with the mean values of $\overline{\tau}_{uni}$ and $\overline{\tau}_{opt}$ is depicted in Fig. 3. The implication is that in comparison to the UPA scheme, the OPA scheme admits a greater leeway in the choice of the

 ISBN: 978-1-61208-174-8

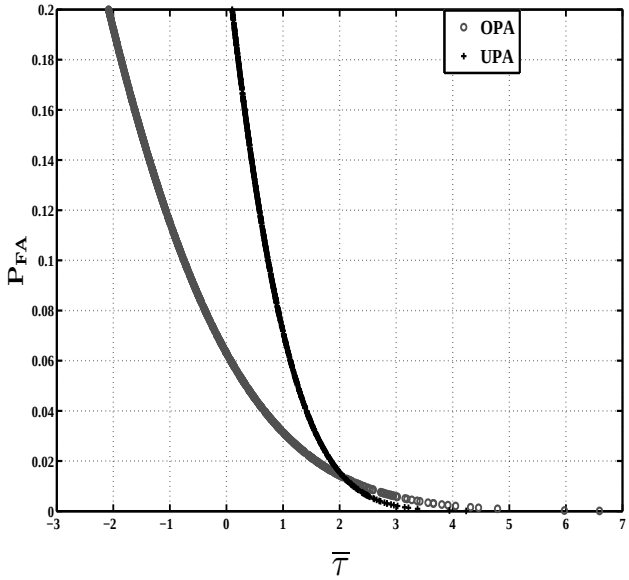


Fig. 5. Probability of false alarm vs. Mean detector threshold

detector threshold. This is empirical evidence in support of the claim that in relation to the detector threshold, detection with the OPA scheme is more robust than that in the UPA scheme.

The foregoing result has credibility only if the P_{Dopt} compares favorably with P_{Duni}. That it is indeed true is borne out by Fig. 4. It shows that $\overline{P}_{Duni}$ and $\overline{P}_{Dopt}$ vs. their respective $\overline{\tau}$. We observe four significant features here.

1) The lowest value of $\overline{P}_D$ for the UPA and OPA schemes are respectively 0.0172 and 0.2581. That is, the OPA scheme outperforms the UPA scheme in respect of the empirical worst case detection. This is after taking even the outliers into consideration, without which, the disparity is further pronounced.
2) The highest value of $\overline{P}_D$ for the UPA and OPA schemes are respectively 0.7127 and 0.9816. That is, the OPA scheme improves upon the UPA scheme in respect of the empirical best case detection.
3) The mean of $\overline{P}_D$ for the UPA scheme which is 0.5166 is considerably lower than that for the OPA scheme which is 0.9216.
4) Finally, on any interval of common support of the detector threshold, the rate of fall of $\overline{P}_D$ with the OPA scheme is smaller than that with the UPA scheme.

Often, the price for a higher P_D is a higher P_{FA}. Hence, in laying claims to a higher P_D, one must bring into perspective the associated P_{FA}. Fig. 5 shows P_{FA} vs. $\overline{\tau}_{uni}$ and $\overline{\tau}_{opt}$. Here too, we may note these characteristics.

1) For the same distribution of P_{FA}, the range of $\overline{\tau}$ over which P_D may be optimized is larger for OPA than it is for UPA.
2) The probability of false alarm falls more rapidly with the threshold in the case of UPA than in the case of OPA. This must be expected in view of the fact that for optimal detection, the OPA offers a larger interval for the detector threshold in comparison to the UPA as borne out by Fig. 4.
3) For the same value of $\overline{\tau}$, compared with the UPA scheme, the OPA scheme yields not merely a higher probability of detection (vide Fig. 4), but concurrently operates at a lower probability of false alarm, except for $\overline{\tau} \in (2, 4.5)$. Moreover, even for this interval of $\overline{\tau}$, it is noteworthty that $P_{Dopt} - P_{Duni} > 0.5$.

In essence, the investigation here makes a strong prima facie case for the OPA scheme over the UPA scheme in terms of robustness of detection w.r.t. the detector threshold when operating under an overall sensor power constraint. In fact, the enhancement in performance is concurrently over conflicting requirements.

VI. Conclusion

In Wireless Sensor Networks, we relaxed the AWGN condition on the Multiple Access Channel and considered the envelope of the channel gain distribution to be unknown and time-varying. Moreover, the observation SNRs at the multiple sensors were take to be non-identical. The detector threshold in the Neyman-Pearson formulation holds a key to the detection probability in relation to the probability of false alarm. With an overall sensor power constraint, we used an optimal detection scheme which takes into consideration the combined effect of the sensor noise and the fading MAC on detection. We examined the impact of the detector threshold on the probability of detection and the probability of false alarm under the uniform and the optimal PA schemes. Optimizing the probability of detection independently in each of the power allocation schemes was the common basis. For the case of a single power constraint, we showed through simulations that the optimal PA scheme outperforms the uniform PA scheme concurrently on three counts: (a) Relative robustness of the probability of detection vis-à-vis the detector threshold; (b) Comparatively high probability of detection; and notwithstanding this, (c) relatively low probability of false alarm.

Acknowledgements

Our thanks are due to Dr. Mahesh K Banavar, Arizona State University, USA, for his critical comments which have contributed to improvements in this paper. We acknowledge with thanks the support we have received from CMR Institute of Technology, Bangalore, India, throughout this work.

References

[1] A. Venkat, H. N. Shankar, and R. Muralishankar, "Optimal Sensor Gain Design over Fading MAC using Distributed Sensing on a Wireless Sensor Network with False Alarm Rate Constraint," *Proc. APCC 2010*, pp. 311–315, Oct. 31-Nov. 4, 2010.

[2] R. Muralishankar, H. N. Shankar, A. Venkat, , and M. Sinha, "Optimal Power Allocation over a Fading MAC with Varying Observation SNRs in Resource Constrained Wireless Sensor Network," *Proc. IEEE ICC-2011*, Jun. 05-09, 2011, Kyoto, Japan.

[3] M. K. Banavar, A. D. Smith, C. Tepedelenlioglu, and A. Spanias, "Distrubuted Detection over Fading MACs with Multiple Antennas at the Fusion Center," *Proc. ICASSP'10*, pp. 2894 – 2897, March 2010.

 ISBN: 978-1-61208-174-8

A New Telesupervision System Integrated in an Intelligent Networked Operating Room

Marcus Koeny, Marian Walter, Steffen Leonhardt
Chair for Medical Information Technology
RWTH Aachen University
Aachen, Germany
koeny@hia.rwth-aachen.de, walter@hia.rwth-aachen.de,
leonhardt@hia.rwth-aachen.de

Michael Czaplik, Rolf Rossaint
Department of Anaesthesiology
University Hospital Aachen
Aachen, Germany
mczaplik@ukaachen.de, rrossaint@ukaachen.de

***Abstract*—In this article, a telesupervision system for the anesthesiologist is presented. It makes use of a manufacturer independent standard for a networked operating room, which is currently under development in a research project called smartOR. The telesupervision system itself is part of the smartOR network and consists of a workstation at the anesthesiologist's workplace and a remote tablet PC, a supervisor can use. Besides a short description of the smartOR network, this work shows the technical principles of the anesthesiology workstation and telesupervision system called SomnuCare. Furthermore, some exemplary opportunities a networked operating room can offer are presented, based on a data acquisition during surgery in the University Hospital Aachen.**

Keywords-telesupervision; anesthesiology; smart operating room; network

I. INTRODUCTION

Currently, most of the devices in an operating room are stand alone, each having its own display and control panel for human interaction. In Figure 1 such a scenario from the anesthesiologists view can be seen. It is obviously challanging to observe all these displays, operate the devices and locate possible acoustic alarms, especially in critical situations. A further issue is the communication between the anesthesiologistical and the surgical team. Due to hygienic reasons there is often a barrier between the patients head, respectively the anesthesiologist's workplace, and the surgical team, which complicates the communication between the anesthesiologist and the surgical team.

Networking all devices in the operating room can improve the efficiency of the operating room staff and therefore improve the patients safety [2]. It can help breaking the barrier between the surgeon and the anesthesiologist using central displays, which shows individually optimized information for every side of the barrier. For several special procedures like cardiac surgery, shared, central monitors for vital data are already common. Telesupervision systems [3] would even benefit from such a network. They can easily acquire information from the whole network without effort and are able to support the patients treatment in critical situations by an experienced supervisor. Furthermore, intelligent patient alarms can be generated and displayed using all information available in the network.

Figure 1. Operating room scenario [1]

A. State of the art

As above mentioned, most of the devices are stand alone. The patient monitor and the anesthetic machine are linked in most cases, via proprietary or semi proprietary protocols like Draeger Medibus [4]. In the meantime, first approaches to an integrated networked operating room were made for example by Olympus [5] or Brainlab [6]. A further restriction is that only manufacturers which are able to implement the proprietary standard are able to integrate their own devices. This certification process is expensive and time consuming, so new innovations can enter slowly into the operating room. As telesupervision systems for the operating room, especially for anesthesia purposes, only prototypes like [3] exist. One is the previous version of the SomnuCare system, which was mentioned before. Patient alarms are, in most cases, generated by each device on its own. They are primarily triggered by exceeding pre-defined limits. There are other concepts like the medical device plug and play [7]. But, there are currently no systems which are able to generate alarms

 ISBN: 978-1-61208-174-8

regarding the anesthesiology with an integrated supervision system.

B. The vision

Assuming that all devices in an operating room are networked and able to exchange information, it is possible to build a central display for both sides of the barrier, optimized for their special requirements. The display on the surgical side for example could show only the relevant vital data for the surgeons. On the anesthesia side, the monitor shows vital signs, ventilator settings, respiration parameters. All of these displays can be further optimized to adapt displayed information to the current workflow. For example alarm limits can be adopted or the anesthesiologist can be warned if the depth of the anesthesia is not adequate, depending on the current state of the surgical intervention. Additionally, the anesthesiologist can consult a colleague in critical situations using the integrated telesupervision system. Smart alarms could support the anesthesiologist and the supervisor to get a quick overview of the current patient's situation.

II. Materials and Methods

A. The smartOR network standard

The smartOR network [8] standard is still in development, hence the current state is described. As a physical communication layer, Ethernet will be used, as it is a commonly used standard and it is already used and approved in many medical applications like MDPNP [7] or Draeger Infinity network [9]. To solve problems with the Ethernet CSMA/CD protocol in time critical networks each operating room has its own network and is connected to the hospital IT via a special gateway. This reduces network load, because the gateway filters irrelevant traffic from other networks. As upper layer IP with UDP or TCP will be used, with a SOAP protocol as a web service architecture [10]. The complete context of the smartOR network and the developed anesthesia system is described in Figure 2. The gateway only exchanges relevant data between the networks. For example vital signs from the smartOR network are not transmitted to the clinic IT. Otherwise DICOM images can be send to the smartOR network for analysis during the surgery, whereas other information from the clinic network is filtered out.

B. Network structure

As shown in Figure 2, the smartOR network is the central component. All devices supporting the smartOR protocol in the operating room are connected via Ethernet. For the connection to the clinic IT a gateway is used, as described in Section II-A. This ensures an isolation of the smartOR network of each operating room and the clinic network. The anesthesia workstation is connected to the smartOR network, too. As central processing and viewing component of the anesthesiologist's workplace, it has much more features. It

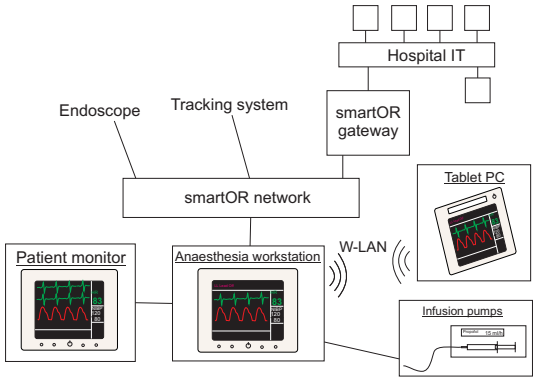


Figure 2. Network structure

integrates devices of the anesthesiologist's workplace and therefore all of the devices are directly connected to the workstation. Necessary data of the anesthesiologist's devices can be converted to the smartOR protocol and be provided to other smartOR devices. Ideally, all devices should be connected to the smartOR network, but especially the anesthesiologist's devices like patient monitor, respirator and perfusion pumps are critical. The necessary modifications of such devices could not be made during the research project. Due to high requirements in data processing for vital signs, it is necessary to process the data stream without any interruption at data rates of up to 200 Hz with up to 50 channels. Furthermore, a continuously transmission must be ensured without interruptions. Hence, the remote tablet PC for the supervisor is directly connected to the anesthesia workstation and not to the smartOR network. Additionally the direct connection allows the workstation a pre-processing of vital signs and enables it to send these directly to the tablet pc without diversion over the smartOR network. The anesthesiologist's workstation and telesupervision system is described in the following part.

C. The anaesthesia workstation

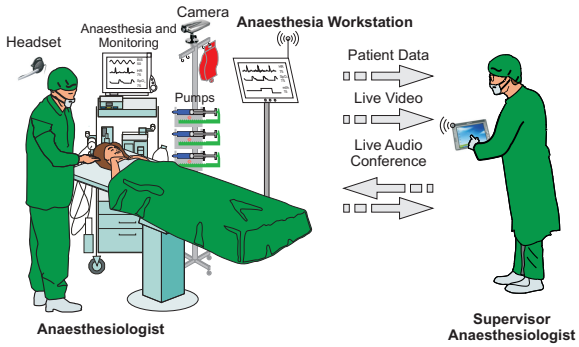


Figure 3. Telesupervision system. Modified after [3]

The essential parts of the anesthesiologist's workplace are shown in Figure 3. The workstation supplies the smartOR

components with the necessary physiologic vital signs and retrieves the relevant data from the smartOR network. This can be events like start of the surgical intervention or other critical procedures. All of the data processed by the workstation can be transmitted to the supervisor via wireless LAN. Additionally to the vital signs, events and alarms, a live video stream from the operating room and a live audio conference can be transmitted.

D. Description of the SomnuCare software

An overview of the important elements of the SomnuCare software can be seen in Figure 4. The central element is the memory mapped file engine. It stores and caches all data from the interfaces, where each memory mapped file represents one vital sign. Data acquisition is done by the interfaces described in Section II-D1. All interfaces in SomnuCare have a specific API, which is equal to all interfaces. This concept enables the programmer to add further interfaces for data acquisition without change of other parts of SomnuCare. Furthermore, the telesupervision server emulates an interface for the supervisors tablet PC and mirrors all data stored in the memory mapped files to a special interface on the tablet PC.

1) SCInterface: All interfaces in SomnuCare have a specific API. This makes it easy to implement and integrate new devices into the workstation. The complete communication and handling with the connected device is done by the interface, so independently from the physical connection each device can be integrated in SomnuCare. The interface only has to implement functions like configuration, starting, stopping and resetting the interface. Every interface holds an internal state machine, which is controlled by these functions. This enables SomnuCare to automatically handle different types of connected devices the same way. Received data are directly written by the interface to the memory mapped file engine. For each data stream or vital sign one file is used.

2) Memory Mapped File engine: The memory mapped file engine makes use of the operating system's memory mapped file API and supports reading and writing data to a segmented memory mapped file. So, writing and reading data is cached via operating system functions on the one hand. On the other hand, the memory mapped file is a complete log file of all inserted data. The engine supports only one writer which can append data to the memory mapped file. This is sufficient because the interfaces are the only writers to the memory mapped file and they only need to append data, because the vital signs are time continuous. As it can be seen in Figure 5, more readers are allowed. This is necessary because the data from the interfaces are needed in parts of the software. The GUI as viewing element needs access to the last appended piece of data, the telesupervision server must send the last inserted piece of data to the client and the smartOR interface needs to send data on request. To save memory not the whole file is mapped into

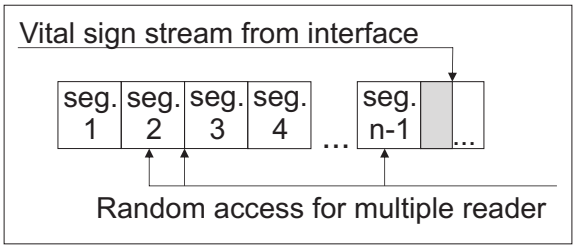


Figure 5. Memory Mapped File

memory. For writing, only the last segment is allocated. If the last segment is full, the file will be expanded with the segments size and this new segment will be allocated. The reading function is able to randomly access segments. To improve the performance of multiple readers, appended data are tagged as new, so with a special read function only new recently appended data can be retrieved. The control of the memory mapped file is done via a special control segment. This is held in a separate memory mapped file, the control file. It stores information for the segment handler and tagging new data. Furthermore individual information about the data held in the memory mapped file can be stored in the control segment.

3) Telesupervision server: The telesupervision server mirrors all data in all memory mapped files over the wireless network to a client. It makes use of the memory mapped file engine and sends all new data over the network. On the client side the receiver is implemented as a regular interface (IfNetwork) so there is no modification needed for the clients except activating the IfNetwork interface and setting the IP address of the server. It acts as a normal interface, but additionally renames the vital signs and therefore emulates the interfaces from the server. Like all other interfaces the IfNetwork interface must implement the state machine for the interface state. This enables the network interface to automatically reconnect after a WLAN disconnect or any other failure. After such a disconnect the IfNetwork tries, as fast as possible, to reconnect and load missing data and then continuous with normal operation. The vital signs are automatically cached by the memory mapped file engine described in Section II-D2. Due to automatic caching and the consistency of the memory mapped files, no data will be lost, so a resynchronization is not necessary.

4) Graphical user interface: The graphical user interface is implemented as a QT grid [11] layout. It can be customized via a configuration stored in a SQLite database, but the standard view is similar to the one of a patient monitor.

5) Simulation interface: The simulation interface is used to evaluating the new alarming concepts. It is able to load saved memory mapped files and CDF data. These data are send to SomnuCare in defined time steps to simulate a conventional interface. This enables the user to replay a

 ISBN: 978-1-61208-174-8

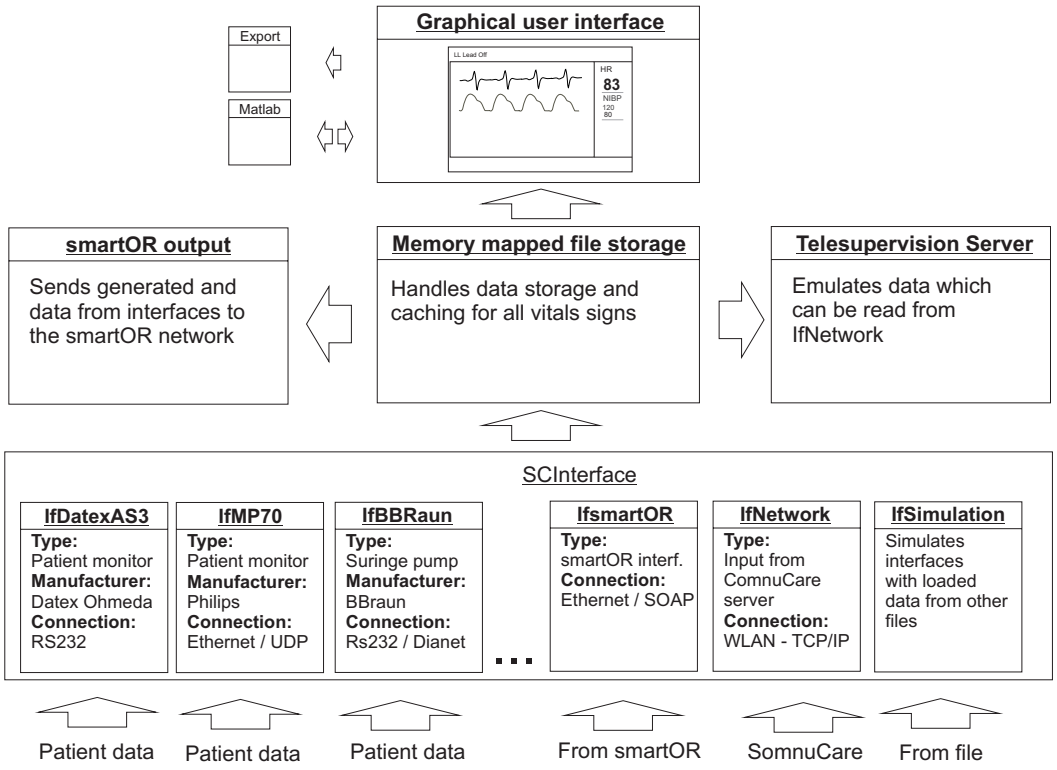


Figure 4. Functional diagramm of SomnuCare

scenario faster than normal time for studies evaluating the new smart alarms.

E. Data acquisition during surgery

The described data acquisition during surgery is the foundation of the alarm system integrated in SomnuCare. First the realization of this data acquisition is described. After that the first approach to the smart alarms is described in Section II-E2.

1) Realization of the data acquisition: In order to improve comparability of the collected data, similar surgical interventions, most of them laparoscopic, were selected for recording. Anonymous data acquisition took place at the University Hospital Aachen (UKA) after approval by the local ethics committee within a time period of three weeks. Generally, the most important steps were pointed out as milestones:

- Start of the presence of the anesthesiologist
- Start of anesthetization
- Approval for surgery
- Start of the surgical preparation
- Start of the surgical intervention
- End of the surgical intervention
- End of the surgical wrap-up
- End of anesthesia
- End of presence of the anesthesiologist

Furthermore the following events were recorded:

- Anesthetic events, like intubation or inserting of a stomach tube
- Surgical events like skin incision, intra-operative relocation
- Intravenous drug injection

All vital signs and information from the following devices were recorded:

- Datex Ohmeda AS/3 patient monitor
- Draeger Cicero and Cato anesthetic machine connected over Datex monitor
- Up to four BBraun perfusor infusion pumps

This represents the standard setup in the UKA for these surgical interventions and resulted in the following recorded vital signs:

- Heart rate, non-invasive and/or invasive blood pressure, oxygen saturation
- Respiratory rate, tidal volumes, pressures, fractions of end-tidal CO2, O2 and anesthetic gases
- Anesthesia agents via perfusion pumps or and/or anesthetic gas concentration via the anesthetic machine

For recording all the vital signs and events, a special software has been developed. This software is a prerelease of SomnuCare and is able to capture the data from Datex and BBraun serial interfaces and store them as comma separated

 ISBN: 978-1-61208-174-8

text files. The events and milestones were recorded using the software as well, so all timestamps have the same time basis. Furthermore the patients sex, weight, age and size were recorded.

In total, data from 17 surgical interventions were recorded. (8 female, 9 male patients) A balanced anesthesia using anesthetic gases was carried out in 8 cases, the remaining 9 received a total intravenous anesthesia using propofol and remifentanil.

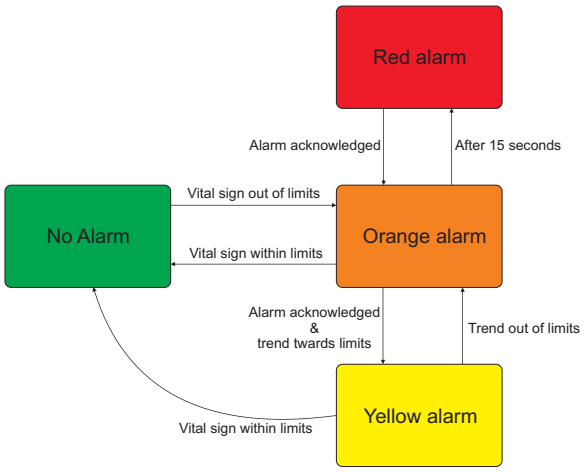


Figure 6. Stateflow

2) First approach to smart alarms: As a first approach to intelligent alarms, the following state machine was implemented for every important vital sign, for example like heard rate, non-invasive blood pressure and oxygen saturation. Compared to conventional alarms, which are triggered by exceeding pre-configured but fixed limits, the state machine considers the change of the vital sign after exceeding the limits. So, the concept of classical limits is kept, but supplemented with the state machine after the alarm rises. The alarm is rated into four conditions, similar to a traffic light.

- RED for a serious danger for the patient
- ORANGE for a situation with a potential danger for the patient
- YELLOW for the phase after a RED or ORANGE alarm is cleared
- GREEN for no alarm

The resulting state machine and flow diagram can be seen in Figure 6. For evaluation the state machine was implemented in Matlab/Simulink/Stateflow [12]. The results are presented in Section III-B.

III. Results

Some parts described in this paper like the smartOR protocol and the integration of the anesthesia workstation into the smartOR network are work in progress. Therefore, the authors cannot present any results regarding these parts.

A. SomnuCare software

The SomnuCare software is, except the smartOR part, fully functional. It will be improved continuously as well as its design.

1) Memory mapped file: The memory mapped file engine has been tested and verified by regular software operation. By only allocating needed segments, we have a much better memory efficient use. As tested under Microsoft Windows 7 the allocation granularity of the segments is 64 KB. So only 128 KB are reserved for one reading and one writing segment for each memory mapped file. Compared to a 12 channel ECG signal sampled with 100 Hz over one hour and each value stored as double with 8 Byte timestamp which needs 66 MB this signal stored under the same condition with the memory mapped file. Regarding a surgical intervention lasting 5 hours with multiple other vital signs recorded, the benefit is obvious. Because of the pointer exchange function the MS Windows API offers there is nearly no overhead due to the allocation algorithm.

2) Wireless LAN connection to the client: The WLAN connection is currently in implementation. First tests showed a good performance. Disconnects are handled directly by the interface manager, which reconnects the interface as soon as possible with the device, in this case with the server. It should be noted that for the network interface a faster handling should be implemented by the manager. For short disconnects, which can depend on the network architecture, it is not sufficient to wait 2 seconds for the reconnect. Exact timing measurements have not yet be done, but will follow.

B. Data acquisition during surgery

The benefit of this concept can be demonstrated by the following example in Figure 7: The patient shows a strong reaction to the anesthesia agents. The blood pressure and heart rate decreases. As reaction to this, the anesthesiologist decreases the anesthesia depth. Once the skin incision occurs the patient reacts heavily to this pain stimulus. The heart rate and blood pressure increase rapidly and on the O2 curve spontaneous breathing can be seen, due to an inadequate depth of analgesia. The above described state machine is currently not able to prevent the situation, but can relieve the anesthesiologist from unnecessary raised alarms. Possibly advising the physician of a trend (like an increasing blood pressure), at an early stage, would result in a more appropriate behavior of the anesthesiologist. Compared with the state diagram in Figure 6, the blood pressure alarm will raise directly after the skin incision, because the alarm limits are exceeded. For the first 15 seconds it will turn to an orange alarm. If nobody acknowledges the alarm it will turn into a red one. If the alarm is acknowledged, it will turn into an orange one and as long as the blood pressure shows a falling trend it will turn into a yellow alarm with no acoustic warnings. If the vital sign is back in normal range, the alarm is switched off and reaches the green state. Compared to

 ISBN: 978-1-61208-174-8

conventional alarms, the alarms rises every 2 minutes with an acoustic sound.

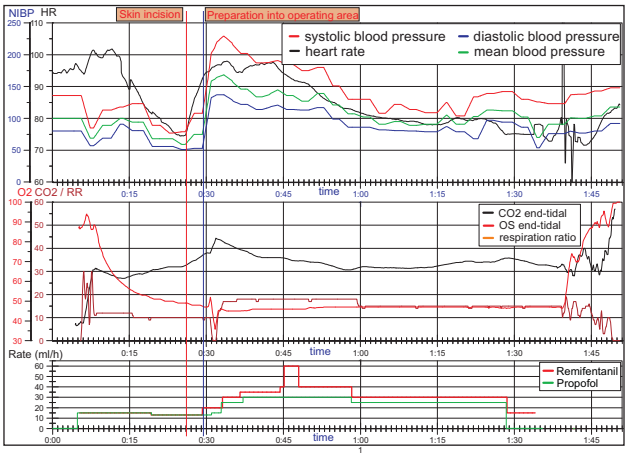


Figure 7. Recorded data of a patient during a surgery

So, at this point the alarm state machine is based only on the information of one vital sign, as described in Section II-E2. This reduces the number of unnecessary alarms essentially and helps the anesthesiologist to keep an overview in critical situations. For example a conventional alarm which is generated by exceeded limits can only be switched off for 2 minutes and then raises again. The described concept is only a basic example and it is still in development.

IV. Discussion and Outlook

A. Multi operating room ability

As described in Section II-A, the smartOR network is an isolated network for each operating room, which is only connected to the hospital IT via a special gateway, due to security reasons. Hence exchanging data between different operating rooms is difficult, but usually not necessary, the telesupervision system can only be connected to one operating room. However in reality the supervisor must be able to supervise more than one operating room. One possible realization could be a network between the gateways of each operating room connected to a central Wireless LAN access point, which is accessed by the supervisors tablet PC. This is more efficient than realizing a Wireless LAN for every operating room and switching between them, because the tablet PC is only able to login to one WLAN at the same time.

B. Data acquisition during surgery

With the basic concept described above it is only possible to reduce the number of unnecessary acoustic alarms. As mentioned in Section II-E2 the system cannot prevent the above described situation, because the actual context of the surgical intervention is not considered and the state machine analyses only one vital sign. Furthermore the state machine should consider multiple vital signs for a smart alarm.

In the future the state machine will be developed further to analyze heart rate, blood pressure, CO2/O2 concentration and depleted anesthesia agents at the same time. Furthermore the state of the surgical intervention will be considered. This will enable the system to recognize the trend in future and support the anesthesiologist during the anesthesia.

Acknowledgment

The project is supported by the Federal Ministry of Economics and Technology. Furthermore the authors would thank Christoph Schorn and Paul Voigtlaender for the great programming support.

References

[1] M. Walter, “Telesupervision und automatisierung in der anasthesie,” *VDE Kongress 2006 Aachen*.

[2] J. Goldman, “Advancing the adoption of medical device plug-and-play interoperability to improve patient safety and healthcare efficiency,” Medical Device ”Plug-and-Play” Interoperability Program, Tech. Rep., 2000.

[3] M. Walter, A. Kanert, S. Macko, J. Schnoor, R. Rossaint, and S. Leonhardt, “Tele-assist system for anaesthesia,” *European Society for Computing and Technology in Anesthesia and Intensive Care (ESTIAC)*, 2006.

[4] “Draeger medical company website (last visited: June 2011) http://www.draeger.com.”

[5] “Olympus company website (last visited: June 2011) http://www.olympus.com.”

[6] “Brainlab company website (last visited: June 2011) http://www.brainlab.com.”

[7] “Medical device ”plug-and-play” website (last visited: June 2011) http://mdpnp.org.”

[8] “Website smartor project (last visited: June 2011) http://www.smartor.de.”

[9] (2011, Jun.) Draeger infinity protocol. [Online]. Available: http://www.draeger.de

[10] J. Benzko, B. Ibach, and K. Radermacher, “Der traum vom plug u. play im op,” *MED engineering*, vol. 03-04, pp. 76–79, 2011.

[11] “Qt website / layout management (last visited: June 2011) http://doc.qt.nokia.com/latest/layout.html.”

[12] “The mathworks company website (last visited: June 2011) http://www.mathworks.com.”

 ISBN: 978-1-61208-174-8

Multi-Objective Optimization for Virtual Machine Migration on LANs for Opportunistic Grid Infrastructures

Nathalia Garcés, Nicolás Ortiz, David Mendez, Yezid Donoso

Departamento de Ingeniería de Sistemas y Computación
Universidad de los Andes
Bogotá, Colombia
{n.garces26, dg.mendez67, n.ortiz980, ydonoso}@uniandes.edu.co

Abstract—**This paper illustrates how to apply a solution for a multiple objective problem in a simple and efficient way through the case study of an example where we must copy a single file, in this case a virtual machine, on to all computers of a LAN. Our solution is intended to be used for the creation of Virtual Clusters, which are clusters composed of virtual machines that execute on opportunistic grid infrastructures. We specify the restrictions through a mathematical model and then proceed to implement a two-part solution: First, we use the solver CONOPT to determine the Pareto frontier; then we implement an evolutionary algorithm to generate possible solutions, and match them to the Pareto frontier. Finally, we evaluate our solution as an efficient way of solving the problem through the result's attributes and conclude which are the advantages of using evolutionary algorithms to find an answer for a multiple objective problem (time, possible solutions and variability between them).**

Keywords- MOP; SPEA; resource sharing; Grid Computing.

I. INTRODUCTION

Often, we find that there are problems for which no unique answer is clear and that there are many factors to consider before a decision is made. Take for example an archive copying problem where we have to give a single file to many computers connected to a Switch-based LAN network in the least amount of time. But, if we also say that we have to do it while other people are using those computers, so we cannot occupy the entire span of the bandwidth, then it becomes a bit more difficult to nail down an answer. This type of problem is called a multiple objective problem.

This problem is found on opportunistic infrastructures, where the goal is to take advantage of the unused capabilities of desktop computers. On a university campus, there are computer labs in which students can develop their daily activities. These daily activities do not fully use the processing capability of the computer. By the use of virtual machines, it is possible to create Virtual Clusters (VCs) running on desktop computers to be used as a part of a grid infrastructure. This is the goal of the infrastructure UnaGrid [1], which allows the creation of VCs and their execution on desktop computer labs at the Universidad de Los Andes. However, the Virtual Machines (VM) must be copied to each desktop computer.

In order to deploy a new cluster it is needed to transfer the VM from a source computer to the rest. This can be seen as a file transfer from one node to every other node in a network. Because the transfer is on an opportunistic infrastructure, it must be designed to not disturb the users of the computer lab.

In this paper, this case study is analyzed. The article begins by addressing other solutions for similar problems and then describes the specifics of this problem. After that, it evaluates the many components of its answer, including the SPEA-based algorithm that is used (its chromosome design, its population selection process, its genetic operators and its exit point). Finally, the results to our solution are explained and it concludes by specifying how it pertains to multiple objective problems.

II. RELATED WORK

Efficient data management has been always a challenge on large scale infrastructures. On Computational Grids, it is required to have a flexible, fast, reliable, and secure way of sharing resources across sites. Several solutions have been developed while looking for an efficient way of file sharing between numerous nodes.

On Grid environments, Grid FTP [2] provides efficient and reliable data transfer between computing nodes located among different sites. Some transfer services have been built on top of GridFTP and added to the Globus Toolkit in order to provide fault recovery mechanisms. Nevertheless, GridFTP is designed to transfer large amounts of data across different networks. Another work based on FTP, designed for parallel transfer sessions is P-FTP. This protocol is also intended to be used across different networks. Since our work aims to analyze the transfer inside a single LAN, it requires a different approach.

On Data Grids, several efforts have been made to manage the data transfer between nodes [3] [4] [5]. On these infrastructures, the data is distributed on replicas among them. There are applications that analyze the bandwidth status between connections, in order to adjust the workloads and to reduce the file transfer time. They try to adapt the file transfer to network links which do not have a predictable or

 ISBN: 978-1-61208-174-8

stable bandwidth. These algorithms also seek to reduce idle time wasted waiting for the slowest server. This is the kind of approach that is useful in this work's solution. However, our context does not have replicas for file transfer; the file must be transferred from one node to every other node on a LAN.

In other words, it is needed to transfer one virtual machine from one node to every other node on a LAN of interconnected physical desktop computers. The solution proposed by the Ohio State University [6] migrates virtual machines by using RDMA (Remote Direct Memory Access), which allows a computer to access the main memory of another without involving the operating system. This schema permits a very high-throughput data transfer between the nodes; it also reduces the transmission overhead up to 80%.

Furthermore, this solution requires a special configuration of each computer involved on the cluster. In the UnaGrid infrastructure the VMs that constitute the VCs are deployed on common computer labs, in which the computers are not controlled by the administrators of UnaGrid. This restriction makes it unfeasible to conFigure the physical machines to support RDMA.

On the other hand, multicast can save great amounts of bandwidth if it is used for file transfer in a LAN. Earlier, UnaGrid had been extended to copy VMs using reliable multicast. However, the firewall configuration of the computer labs blocks any multicast transfer when it involves large files. Generally, the VMs copied on to the computer labs consist of files greater than 5 GB. So another approach is needed, rather than multicast based schemas.

Some solutions involve the use of multicast-based schemas such as overlay multicast [7]. That particular solution describes a method for reliable data transfer based on this protocol, achieved by the usage of the application layer. They create a binary distribution tree, in which each node acts both as a sender and a receiver of packets using TCP, and it changes its structure according to the network condition. This solution is not in the scope of this paper but will be explored as future work.

III. PROBLEM STATEMENT

The situation starts off with a topology where *n* terminals are connected to a switch with *Cs* capacity and all of the connections are Symmetric DSL with the same amount of Bandwidth. (For practical purposes, assume that *Cs* is greater than or equal to the sum of the connections of the terminals, establishing that the network will not collapse if all of the terminals are using their connection to their full extent.)

In one of these terminals, there is a large archive (a VM) that requires to be transfered to other computers throughout the LAN. UDP cannot be used because of firewalls and it must be done without the users of the computers knowing about it, so the Switch's capacity cannot be consumed too much or else the users will start to notice a lag in their connection.

The objective is to be able to copy the large archive onto all of the computers in the least amount of time and having all of the computers finish downloading the archive at approximately the same time.

A. Mathematical Formulation

1) Objective Functions

a) Min (Bwmax – Bwmin)

Minimize the waste. This means we want all the nodes to end at approximately the same time.

b) Min (Tf)

Minimize the time that it takes to copy the information to the last node.

2) Constraints

a) $C_{ij} - U_{ij} = R_{ij} \quad , \quad \forall i,j \in E$

The remnant R_{ij} that can be used, from the connection of the ith terminal, to transmit the file is equal to the capacity C_{ij} of the connection minus the capacity U_{ij} being consumed by other functions or the user.

b) $\alpha_{ij} = \%R_{ij}$

The percentage of use α_{ij} is equal to a defined percentage that will be used from the remnant of R_{ij}. Note that it is a percentage, not a capacity.

c) $Bw_{min} = min(\alpha_{ij}R_{ij})_{j \in E} \quad , \quad i = 0$ *(download)*

The minimum bandwidth of the configuration is equal to the minimum actual usage of the connection, which is the percentage of use α_{ij} times the remnant of the connection R_{ij}. Note that the usage is only from the central switch ($i = 0$) to a node.

d) $Bw_{max} = max(\alpha_{ij}R_{ij})_{j \in E} \quad , \quad i = 0$ *(download)*

The maximum bandwidth of the configuration is equal to the maximum actual usage of the connection, which is the percentage of use α_{ij} times the remnant of the connection R_{ij}. Note that the usage is only from the central switch ($i = 0$) to a node.

e) $T_f = \frac{A}{Bw_{min}} + t * (n-1)$

The time it takes for the final node to finish downloading the file is equal to the size of the file A divided by the minimum Bandwidth Bw_{min}, plus the time it takes to establish the connection and start sending the information to the next node (t) times the number of nodes n minus one.

f) $\sum_{ij \in E} \alpha_{ij}R_{ij} \leq k * Cs$

The sum of the actual usage of the connections must be less than or equal to the Switch's capacity Cs times a defined percentage k.

 ISBN: 978-1-61208-174-8

g) $0.1 \leq \alpha_{ij} \leq 0.9 \quad , \quad i = 0$

The percentage of use α_{ij} must be less than or equal to 0.9, but more than or equal to 0.1 in order for the program to work.

IV. NETWORK

The network is a star graph due to the fact that none of the terminals has a direct connection between them, but may communicate to each other through a central switch. In this particular problem a specific configuration has been determined for the simulation:

- A = 4 GB
  The size of the file that is to be transmitted is equal to 4 GB.
- Cs = 500 Mb/s
  The capacity of the Switch, which is the total amount of bandwidth (being used at the same time) it can sustain without running into problems, is 500 Mb/s.
- k = 0.5
  In this case, half of the total capacity of the switch is used. Thus, k is the percentage restrain over the capacity of the switch usable by the solution.
- t = 1 s
  Assume that the delay of establishing a connection and to start sending the file between terminals is 1 second long.
- n = 8
  The total nodes in the network will be eight. Remember than in all of the cases, one of them already starts with the file.

V. SOLUTION

Once the topology to test the MOP has been established, two methods have been selected to find the best solutions. One approach is to convert the MOP in a single objective problem and the other is to optimize both of the objective functions simultaneously.

For the mathematical problem the first approach was selected: a classic method called weighted sum. This method would provide the real Pareto frontier of the problem, or at least some of the solutions. Since the problem was not established as convex or not, it can't be guaranteed that this method would be the best approach to discover all the solutions of the Pareto frontier.

Therefore, the second approach was necessary: a meta-heuristic method called Evolutionary Algorithms (EAs). EAs are not entirely probabilistic because they have some intelligence that they use to find the correct solutions; also their computational time is lower than a probabilistic method. In addition this method breaks the convexity problem, which means that it can be used to find the entire Pareto frontier.

Multi objective evolutionary algorithms have a lot of implementations, but SPEA algorithm [8] was deemed appropriate enough for the problem.

VI. WEIGHTED SUM METHOD

To find the solution of the mathematical model the classic method was implemented with the solver CONOPT. Since this approach for MOP uses only one objective function (F), the weights assigned to each sub-function (f1, f2) were varied by increasing or decreasing 0.1.

$$F = w1 * f1 + w2 * f2$$

When executing the solver the first objective function was always converging to zero. It was then necessary to impose a new restriction to the mathematical model so that the value could vary.

$$Bw_{min} * 0.1 \leq (Bw_{max} - Bw_{min})$$

In other words, this restriction means that the value of F1 must be at least ten percent of Bwmin.

This variation of the model resolved the difficulty. At last, the solver was executed ten times and yielded 8 points of the Pareto frontier. These points are the red ones shown in Figure 4.

VII. SPEA

The solutions granted by this method are the blue ones shown in Figure 4.

The overall algorithm is shown in Figure 1. Next, the implementation of each step of the algorithm is described.

Begin
Generate initial population *P* randomly. Make *P* feasible according to constrains
Generation $G = 1$
WHILE $G < G_{max}$
Calculate objective values for every solution in *P*
Find non-dominated solutions in *P*
Incorporate non-dominated solutions of *P* to *P'*
IF number of solutions in *P'* exceeds maximum number allowed
Purge *P'* using *clustering*
End IF
Calculate *fitness* for every solution in *P* and *P'*
Apply *selection* to *P* + *P'*
Apply *cross over* and *mutation* to generate new population *P*
Make *P* feasible according to constrains
$G = G + 1$
End WHILE
Print results in output file
End

Figure 1. SPEA algorithm structure.

A. Individual (chromosome)

A chromosome represents a solution to an EA problem. This solution is represented as a vector of genes in which each one contains a node of the topology followed by the alpha and it's respective Bw.

Node 1 $\|\alpha_{0,1}\|Bw_{0,1}$	...	Node i $\|\alpha_{0,i}\|Bw_{0,i}$	...	Node 8 $\|\alpha_{0,8}\|Bw_{0,8}$

Figure 2. Coding of the chromosome.

B. Initial population

A population is an array of chromosomes. As stated in Figure 1, this initial population (P) is generated randomly. In this specific implementation it starts at 100 chromosomes

with random alpha values. In addition it's guaranteed that each chromosome is a feasible solution, which means that it takes into account every constraint.

C. Fitness calculation

The higher the fitness of a chromosome, the better the solution is. This means that the values for the non-dominated chromosomes are higher than the dominated ones. To achieve this each chromosome is compared with the entire population P, and for every chromosome that it dominated we added one to its fitness value. At the end of the comparisons the chromosomes with higher fitness were deemed the best solutions so far, so the 10% of the population with the highest fitness where moved to the elitist population (P').

D. Clustering

Although clustering is a very important process to maintain P' small, it wasn't implemented. Instead, every 150 generations, 60% of the chromosomes with the lowest fitness were deleted.

E. Genetic Operators

1) Selection

It was unknown if the values of the fitness of the chromosomes were very far apart. Which is why the Ranking selection method was applied. This way each of the chromosomes had a guaranteed chance to be chosen to take part in the combinatorial process.

2) Crossover

For this operation to take place a 70% possibility was determined.

For its implementation the simple crossover, with two parents/chromosomes involved, was chosen. A number between 1 and 7 was picked at random and the result lead us to the crossover point.

At the end of this process the offspring was verified as a feasible solution, if it wasn't it was discarded.

3) Mutation

For this operation to take place a 30% chance was stipulated.

Also, two types for this algorithm were implemented, each one with 50% chance. The first one is a permutation and the second one is changing the value of a random alpha value. This operation uses one parent and produces another one. This new chromosome was also checked as a feasible solution and discarded if it wasn't.

4) Exit point

It was concluded that 4000 generations where enough for the chromosomes to converge into optimal solutions.

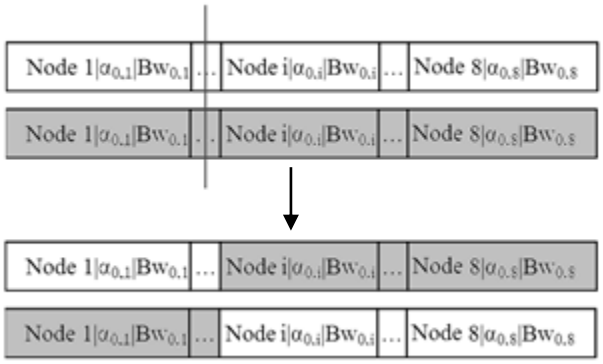


Figure 3. Crossover Operation.

VIII. RESULTS

A. Pareto Optimality

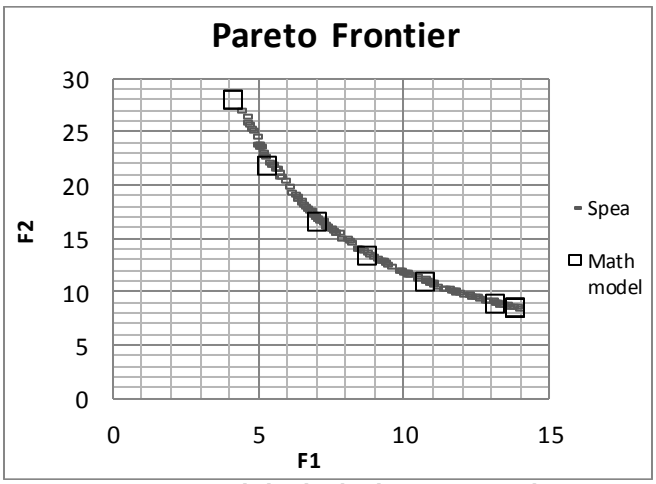


Figure 4. Pareto Frontier calculated with solver CONOPT and SPEA.

B. Evaluation

The solution is evaluated by providing 6 metrics [7]: the first three metrics indicate quantity and the last three indicate quality. It is important to consider more than one metric because only one doesn't take into account all of the performance of our SPEA algorithm.

1) GVND

The non-dominated points that present at the end of the SPEA execution are 135.

2) RGVND

$$\frac{135}{8} = 16{,}875$$

Where 8 is the number of the points found by the mathematical solver.

3) GRVND

The sum of the points found by SPEA and by the math model is 143.

4) Error

$$\frac{135 - 143}{135} = -0{,}059$$

 ISBN: 978-1-61208-174-8

5) Generational Distance

By applying the formula, the value of this metric is 0,102.

6) Spacing

$$s = 0{,}12$$

The ideal value of the last three metrics is zero, however our values are not that far apart. This means that we have a good behavior of our algorithm. In other word, the distribution of our solution is adequate, the distance between our solutions and the mathematical model is minimal and our error (our solution compared with the solution given by the solver CONOPT) is nearly insignificant.

In addition, the first three metrics indicate that we have a considerable amount of solutions, which is something good.

IX. CONCLUSION

Our model can find an efficient way of transferring a file from one source node to every other node on a LAN. This transfer is done in an opportunistic way, it is intended to adapt to the use-percentage of each link of the network. So our algorithm can find a solution in which the user does not perceive any change on the quality of service of the network due to the transfer.

From both our solutions calculated from the evolutionary algorithm and CONOPT we obtain practically the same Pareto Frontier. This can be proven by our result for generational distance. This tells us that our evolutionary algorithm is not obtaining only local minimums, and that our mathematical model is well defined for our problem.

We obtain more diversity of solutions from our evolutionary algorithm. This shows that our evolutionary algorithm can provide more information in case we wanted to develop software to calculate the best way of migrating virtual machines on a LAN.

X. FUTURE WORK

Our algorithm has proven a good performance and efficiency in small networks. However it is necessarily to test it in bigger networks because in these tests is where it can evaluated the scalability of the proposed solution and the complexity of the algorithm.

Also in our algorithm, we simplified the way in which we reduce the size of the elitist population. The SPEA algorithm proposes that to control the size of the elitist population some solutions have to be removed by using Clustering [8]. We simplified this calculation by using a different algorithm to reduce the population. When our elitist population reaches a size S greater than a defined maximum M, what we do is to calculate the fitness inside the elitist population and then we eliminate the last M-S solutions. We suggest as a future work to implement clustering for reducing the size of the elitist population.

In our solution we propose a way of transferring the file by using a pipeline. A future work could consider a different schema, for example binary trees, or n-ary trees where the root is the node that is the source of the file, and the tree defines how to organize the transfer. It must be considered that these kinds of solutions require a different mathematical model.

REFERENCES

[1] H. Castro, E. Rosales, M. Villamizar and A. Jiménez, "UnaGrid: On Demand Opportunistic Desktop Grid" 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing (VIC 2010), May. 2010, pp. 661 - 666, doi: 10.1109/CCGRID.2010.79.

[2] Globus. GridFTP. [Online] 5 2011. http://globus.org/toolkit/docs/3.2/gridftp/, retireved September, 2011

[3] R. Madduri, C. Hood, W. Allcock and W. E, "Reliable file transfer in Grid environments" Proc. Local 27th Annual IEEE Conference on Computer Networks, (LCN 2002), pp. 737-738, doi: 10.1109/LCN.2002.1181855.

[4] V. Velusamy, A. Skjiellum, and A. Kanevski, "Employing an RDMA-based file system for high performance computing" Proc. 12th IEEE International Conference on Networks (ICON 2004), pp. 66-70, doi: 10.1109/ICON.2004.1409089.

[5] Y. Chao-Tung, C. Yao-Chun, Y. Ming-Feng and H. Ching-Hsien, "An Anticipative Recursively-Adjusting Mechanism for Redundant Parallel file transfer in data grids" 13th Asia-Pacific. Computer Systems Architecture Conference (ACSAC 2008), Aug 2008, pp. 1-8, doi: 10.1109/APCSAC.2008.4625456

[6] H. Wei, G. Qi, and L. Jiuxing, "High performance virtual machine migration with RDMA over modern interconnects" IEEE International Conference on Cluster Computing (CLUSTR 2007), Sep 2007 pp. 11-20, doi: 10.1109/CLUSTR.2007.4629212

[7] E. Kwon, J. Park, and S. Kang "Reliable data transfer mechanism on dynamic nodes based overlay multicast" The 7th International Conference on Advanced Communication Technology (ICACT 2005), Feb 2005, pp. 1349 – 1352, doi: 10.1109/ICACT.2005.246199

[8] Deb, Kalyanmoy. Mulit-Objectctive Optimization using Evolutionary Algorithms. Chichester : Wiley, 2004.

 ISBN: 978-1-61208-174-8

Restoring CSCF by Leveraging Feature of Retransmission Mechanism in Session Initiation Protocol

Takeshi Usui and Nozomu Nishinaga
New Generation Network Laboratory
National Institute of Information and Communications Technology
Koganei, Tokyo
e-mail: ta-usui@nict.go.jp and nisinaga@nict.go.jp

Yoshinori Kitatsuji and Hidetoshi Yokota
Mobile Network Laboratory
KDDI R&D Laboratories, Inc.
Fujimino, Saitama
e-mail: kitaji@kddilabs.jp and yokota@nict.go.jp

***Abstract*— The IP multimedia subsystem (IMS) is a key technology for providing various services over IP-based networks. IMS enables network service providers to collect information relating to the communications of customers, such as an accounting through call/session control function (CSCF) which is used to establish a session using the session initiation protocol (SIP). Therefore, the availability of CSCF has become important. We propose a system to recover the session states maintained by the CSCF in the alternate CSCFs. This is achieved in a low cost solution by leveraging the features of retransmission mechanism in SIP. Our proposed system selectively saves the session state in order to reduce the saved data and recovers the saved session state in the alternate CSCFs rapidly when the faults occur in CSCFs. We show that our proposed system can achieve 60% reduction of the backup servers and that the overhead of our proposed system is not large.**

Keywords-component; Restoration system; IMS; SIP; Network Operation

I. INTRODUCTION

Many network service providers (NSPs) are now promoting convergence towards providing various services over IP-based networks. IP Multimedia Subsystem (IMS) [1] is supposed as the service management and control function over IP-based networks. In the IMS, a call/session control function (CSCF) [2], that is a session initiation protocol (SIP) [3] proxy server, is used to establish a session, that accompanies the information to start the application between user equipments (UEs), through the SIP. When the UE starts an application managed by IMS, such as IP telephony, a series of SIP messages are exchanged through multiple types of CSCFs that is referred to as the SIP signaling call flow.

The availability of the CSCFs is essential for session management, i.e., service control, administration, and accounting. If the CSCFs are not available, the UEs cannot start the application. In the worst case, the services managed by IMS may go down because the application initiation and the communications of the UE are broken off. Therefore, the availability of the CSCFs is of great importance.

For continuing services, the service status need to be recovered to an available server and the service execution is succeeded with this server when the original server becomes unavailable. E.g., in the case of the CSCFs, it is required to recover the session states which the CSCF keeps for each UE. A session state is the information which is used for establishing the session (i.e., transaction and dialog, details are described in Section III-B). For restoring the CSCF, the NSP may adopt a fault tolerant (FT) server [4] which consists of a pair of active and backup hardware and copies all the data from the active server to the backup server. In this case, the NSPs must prepare the same number of backup servers as active servers. As the other methods, Peer-to-Peer (P2P) SIP [5] can be used to improve the redundancy of SIP proxy servers. However, P2P SIP cannot recover a session state when the SIP signaling call flow is not completed and a session is not yet established (hereafter, this incomplete session is termed “pre-session”). Here, we propose a system that recovers even the pre-session as well as the established sessions from the aggregated backup servers.

For completely restore the CSCF that halted because of faults (e.g., hardware fault) (hereafter, termed “fault CSCF”), the session states must be saved in the backup server which is positioned as the backup hardware in the case of FT and the alternate CSCF needs to take over the session state of the fault CSCF as described in a literature [6]. However, this process has two issues. The first is that the NSP encounters to maintain a number of the backup servers because of the high volume of the data to be saved. Now, NSPs aim to reduce the number of the servers for cost saving by aggregating multiple servers to fewer servers.The second is that the NSP needs to restore the CSCF within a limited period. This is because the NSP never want to lose the information related to the accounting.

To solve the first issue, we reduce the number of the backup server by selectively saving the session states. The SIP application retransmits SIP messages if they are lost. Our proposed system does not save the session states that can be restored by the retransmitted SIP messages, but save the session states based on the relationships between the retransmitted SIP messages and the session state that is necessary for the CSCFs to handle the retransmitted SIP messages. Our proposed system enables few backup servers to save the session states from a large number of CSCFs because the volume of the saved data is sufficiently reduced.

To solve the second issue, we propose that the specific session state be recovered rapidly in turn based on the type of each session state. We defined priorities from the session states which are required from the service. At first our proposed system recovers the pre-session in order to

 ISBN: 978-1-61208-174-8

continue the SIP signaling call flow and restricts the influence of the CSCF faults at the minimum level.

The rest of the paper is organized as follows. Section II and III describe the requirement for CSCF restoration system and the design of our proposed system. Section IV evaluates the overhead of our proposed system and how many backup servers are reduced. Section V concludes the paper.

II. REQUIREMENTS FOR CSCF RESTORATION SYSTEM

A. Overview of IMS Architecture

One of the features in the IMS is that three different SIP proxy servers are used. The core functional components of the IMS are these three different kinds of CSCFs: Proxy-CSCF (P-CSCF), Interrogating-CSCF (I-CSCF), and Serving-CSCF (S-CSCF). These components are responsible for the management of SIP signaling call flow, such as routing a SIP message, and authentication and registration for UEs.

The P-CSCF is the first functional point of contact between UE and IMS components. It maintains IPSec security associations (SA) with UE and applies integrity and confidentiality protection for the SIP message. The I-CSCF is responsible for assigning a suitable S-CSCF to the UE and routes incoming requests to S-CSCF. The S-CSCF handles the registration process and downloads authentication data from the Home Subscriber Server (HSS), which is a database containing the information base of the subscribed users, such as the users identifiers, location information, and so on.

In this paper, we focus on the P-CSCF and the S-CSCF for the restoration system because these servers maintain the sessions in a stateful mannr. Because the I-CSCF deals with the SIP messages in a stateless manner, the UE and CSCFs does not need to execute registration process from the beginning when the alternate I-CSCF is prepared. Therefore, recovering the session states of the I-CSCF is not necessary.

B. Functions of CSCF Restoration System

In this paper, we assume the CSCFs and UE use UDP as the transport protocol for exchanging SIP messages because many NSPs do so. Therefore, the CSCFs and UE use the retransmission mechanism in SIP. (When adopting TCP, our proposed system cannot be used because the CSCFs do not use the retransmission mechanism)

The following three functions are considered necessary for the restoration system.

(1) Monitoring function to detect CSCFs failure quickly
(2) Saving function to save the minimized data of the session states from the CSCFs before their faults occurred
(3) Restoring function to rapidly recover the session state in the alternate CSCF without requiring any action of UE or without losing session management at the UE due to the faults.

After function (1) detects faults of CSCF, we can execute the procedures to recover the session states. In this paper, we assume that the existing tools and method (e.g., [7] [8]) are used to detect the faults of CSCF. We thus focus instead on the design and implementations of functions (2) and (3) in a low cost solution.

In function (2), a key is reducing the saved data of the session state. To implement the low-cost restoration system, it is desirable that the volume of the saved data is low. Saving all the change of session state as the CSCF receives any types of SIP messages leads to a large amount of data transferred to and stored in the storage. The reduction of the saved data allows the number of backup servers maintained in the NSP to be reduced. We distinguish SIP message (INVITE, UPDATE, ACK, and so on) based on the priorities in recover the session state just after the CSCF is halt, and let the session state be saved when the important SIP message is received by the CSCF. Thus, the CSCF saves the limited session states instead of saving every change of the session state. The definition of the importance of SIP message and the other detail are described in Section IV.

Function (3) needs to be executed as rapidly as possible in order to not affect the recovered sessions maintained in the alternate CSCF server after the failure of the original CSCF server. No particular functions in the implementation of the UE should be additionally required for function (3).

Function (3) is also required to recover the pre-session. If the pre-session is not recovered, the UE fail to start the application and needs to execute the application initiation procedures from the beginning. In the SIP, there is no procedure to recover the broken procedure of the application initiation. In order that the UE start to use the application, that procedure is need to be defined.

C. Related Work

The FT server [4] requires the pairs of active and backup hardware resources because it saves data from active server to backup server whenever the CPU executes any commands. Therefore the FT server requires the specialized hardware is required. The FT server has function (3), but does not satisfy function (2).

The redundancy techniques for the web servers [9] [10] are applicable for CSCFs. However, these technologies do not achieve minimizing the saved data of the session states because these technologies do not consider the feature of the SIP. These technologies do not satisfy function (2).

The system for the replication of the SIP proxy server as described in [6] has been proposed. In this system, a pair of an active and backup server is prepared. The replication is executed every time SIP messages are exchanged between the SIP proxy and the UE. This system satisfies the function (3). But, this system does not consider the reduction of the data to be saved in the backup servers. Therefore, this system does not satisfy the function (2).

P2P SIP is also used to improve the redundancy of the servers, but it cannot recover the pre-session because there is no mechanism that shares the session states that the SIP proxy servers keep. For P2P SIP, it is necessary for the UE to re-register with the SIP proxy server so that a different SIP proxy server manages the session state of the UE. P2P SIP does not satisfy the function (3) because the data for restoring the CSCF is not saved anywhere.

 ISBN: 978-1-61208-174-8

III. PROPOSED RESTORATION SYSTEM

This section presents the design of our proposed system which leverages the features of the retransmission mechanism in SIP.

A. Overview of our proposed system

An overview of our proposed system is shown in Figure 1. The restoration system consists of multiple active and backup servers and does not include the monitoring function. The backup servers save the session states from multiple CSCFs through the out-band line (bidirectional arrowed lines in Figure 1). When the monitoring system detects a fault, the restoration system is commanded to set up an alternate CSCF. The alternate CSCF is selected from one of standby servers and takes over the IP address and configuration of the fault CSCF. The alternate CSCS is prepared before the migration of the transactions and dialogs.

Our proposed system takes the following steps if the CSCF halts with the fault.

1. The monitoring system or the operator order the restoration system to restore the CSCF
2. The alternate CSCF is set up and the saved transactions and dialogs are recovered from the backup server in order to take over the session states.

B. Analsys of SIP Signaling Call Flow

The registration procedure from UE-A is shown in Figure 2. The procedure is normally executed when UE is turned on. In this call flow, authentication and key agreement (AKA) [12] that is necessary for the UE to get access the IMS-based services is also executed, and the IPSec is established between the UE and the P-CSCF. If any SIP messages are lost, the first Register message is retransmitted.

We assumed the application initiation procedures of IMS as shown in Figure 3. Each SIP message in the Figure is numbered. The INVITE message is used to initiate a session between UE-A and UE-B. Here, UE-A and UE-B register with the different P-CSCFs and S-CSCFs. We next summarize the retransmission behavior of the SIP messages by separating the SIP signaling call flow into seven phases. The SIP message numbers are in parentheses.

In Figure 3, there is a Request-Response relationship among the SIP messages, such as Bye and 200OK. In the case that a SIP message is lost, the SIP signaling call flow always restarts from the SIP messages that corresponds to the request message. The CSCF forwards an incoming request message to the UE or the other CSCF and keeps waiting for the response message until either the response message is received for the request message or a set timer expires. This relationship among the SIP messages is called a “transaction” in SIP. A “dialog” means the state that the CSCFs create about the communications between the UEs through the SIP signaling call flow. The dialog is kept in the CSCFs while the UEs communicate, and erased when the communication is terminated.

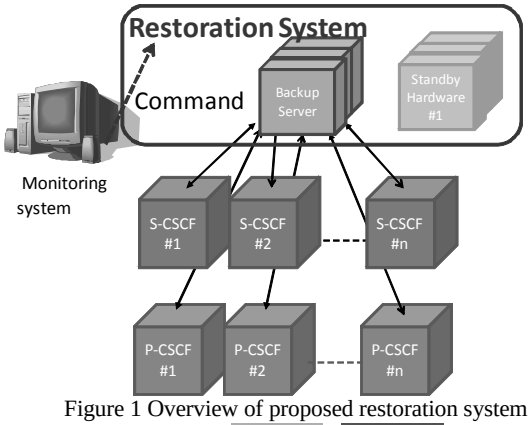


Figure 1 Overview of proposed restoration system

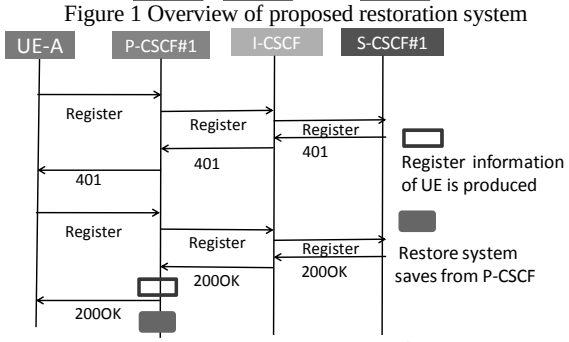


Figure 2 AKA Register procedure

In this paper, we define that the session state consists of the transaction and dialog. A transaction and dialog are updated whenever SIP messages are exchanged through the SIP signaling call flow. The specific transaction and dialog during the SIP signaling call flow are required in order to recover the pre-session in the alternate CSCF. When the CSCF does not have the transaction and dialog related to that SIP message, the CSCF cannot handle that SIP message. Our proposed system save the session states that are required to recover the pre-session.

1) INVITE/100Trying/183Session Progress (1–17)

This phase is initiated with the sending of an INVITE message and is terminated when UE-B receives 100Trying or 200OK message. This phase actually consists of multiple parts: part of INVITE and 100Trying message and part of INVITE and 183Session Progress. Once the P-CSCF#1 receives the INVITE message from UE-A, it replies with 100Trying message to UE-A. The same exchange of the SIP messages applies to each hop until the INVITE message reaches UE-B. If the INVITE or 100Trying message is lost, the UE will retransmit the request after the expiration of a timeout value called T1. If the INVITE message sent by the P-CSCF to the S-CSCF or the 100Trying message sent by the S-CSCF to the P-CSCF is lost, the P-CSCF will retransmit the INVITE message after T1 seconds. The same applies to all hops until UE-B. Note that the INVITE message is only the case in which the reply message are sent

 ISBN: 978-1-61208-174-8

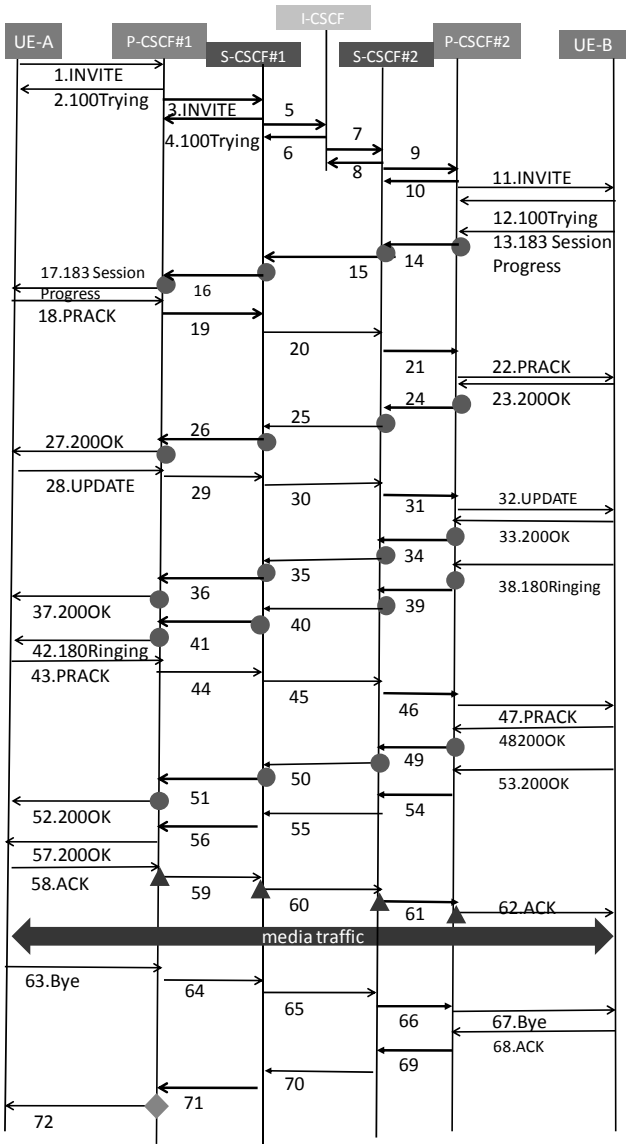


Figure 3 Application initiation procedures of IMS

hop-by-hop. In all the other phases 2) through 7), the UE is responsible for generating appropriate acknowledgements in an end-to-end manner. In the rest of this phase, UE-A and UE-B exchange the information of the audio and video codes to be used as well as the quality of service (QoS) criteria. The 183Session Progress message is sent reliably. If this message is lost, the INVITE message is retransmitted.

2) PRACK/200OK (18–27)

UE-B acknowledges receipt of the 183Session Progress message. The PRACK message will be retransmitted by UE-A until a 200OK message is received. Note that the transaction related to the PRACK message starts its retransmission timer just after it receives the 183Session Progress message. Furthermore, the PRACK message will not be retransmitted each time the 183Session Progress message is retransmitted but only when a retransmission timer is triggered (i.e., T1, 2T1, etc.).

3) UPDATE/200OK (28–37)

UE-A and UE-B complete the exchange of the code and QoS information. If the UPDATE message or the 200OK message is lost, the caller will retransmit the UPDATE message until the 200OK message is received.

4) 180Ringing/PRACK (38–47)

UE-A informs the caller that the user is being alerted about a call. The 180Ringing message is sent reliably. That is, UE-A retransmit it until the PRACK message is received.

5) PRACK/200OK (43–52)

UE-B acknowledges receipt of the 180Ringing message. The PRACK message will be retransmitted by UE-A until a 200OK message is received.

6) 200OK/ACK (53–62)

The callee informs the caller that the call was accepted. If the response or an ACK message is lost, the callee will retransmit the 200OK message until the ACK message is received.

7) BYE/200OK (63–72)

One UE terminates the call. If the BYE or acknowledging 200OK message is lost, the sender of the BYE message will retransmit that message until a 200OK message is received.

C. Saving Session States from CSCFs

Saving all the transactions and dialogs in the SIP signaling call flow generates a number of request messages and its acknowledgement messages for saving the session states and makes the load of the backup servers high. Therefore, a certain number of backup servers are necessary for the restoration system. When the alternate CSCF is set up and the UE sent one SIP message in the SIP signaling call flow, the CSCF cannot handle the SIP message if the CSCF does not have the transaction and dialog related to the SIP messages that UE sent.

The backup servers do not need to save the session states whenever the CSCF sends the SIP messages. Saving the specific session states that is necessary for the CSCF to handle the retransmitted SIP messages is enough. To reduce the load of the backup servers, we propose that the restoration system not save the transactions and dialogs that are re-generated after the retransmission by the UEs. We select the specific session states by leveraging the features of SIP retransmission mechanism.

Our proposed system save the session states which the CSCFs need in order to handle the retransmitted SIP messages. To recover the session states completely, the SIP signaling call flow is made to wait for the completion of saving from the CSCFs to the backup server. Only after the session states are saved, the CSCF can send the next SIP message in the SIP signaling call flow. If the saving process is not completed and the next SIP messages have been sent before a fault occurs, the UE cannot complete the SIP signaling call flow because the CSCF cannot handle the SIP messages that the UE sent.

The request messages for saving the session states are sent by the CSCF to the backup server and the acknowledgement message for informing the completion of saving is returned by the backup sever to the CSCF. If much time is spent on

 ISBN: 978-1-61208-174-8

treating the request messages from the CSCFs, the SIP signaling call flow is delayed. To prevent this situation, it is necessary to increase the number of backup server or to reduce the number of the total durations for saving the session states. In Section IV, we evaluate about the saving process in the backup server.

For the registration procedure (Figure 2), the restoration system saves the register information (e.g., UE IP address and URI) and the key information for the IPSec. The rectangle in Figure 2 indicates the points at which the restoration system saves the register and key information.

For the application initiation procedure (Figure 3), the transactions and dialogs which the CSCFs need in order to handle the retransmitted SIP messages are summarized in Table 1. We assign phase numbers for the sequenced SIP messages. The circles in Figure 3 indicate the points at which the restoration system need to save the transactions and dialogs before the CSCFs send the next SIP messages to the UE or the other CSCF.

If the 183Session Progress message does not arrive, or the other SIP messages are lost, the UE retransmits the INVITE messages. For the INVITE, 100Trying, and 183Session Progress messages (phase 1), the restoration system does not need to save any session states. After the 183Session Progress message is sent, the transaction and dialog need to be saved for when the PRACK message or the 200OK message is retransmitted (phase 2). After the 200OK message is sent from UE-B, the transaction and dialog need to be saved for when the UPDATE message or the 200OK message is retransmitted (phase 3). To handle the retransmitted 180Ringing message (phase 4), the CSCFs need to save the transaction and dialog after sending the 200OK message. After the CSCFs send the 180Ringing message, the restoration system needs to save the transaction and dialog for when the PRACK or 200OK message is retransmitted (phase 5).

After the ACK messages are received by the CSCFs (phase 6), only the dialogs are saved from the CSCFs because. The triangles on the SIP signaling call flow in Figure 3 indicate the points at which the restoration system saves only the dialog because the transaction is erased in the CSCF. The session is then established, and the keep-alive messages are exchanged in some periods until one UE sends the Bye message. After a few minutes, the restoration system erases the transaction because the session is established and no retransmission is generated from now. As for the dialog of the Bye and 200OK messages (phase 7), the restoration system erases the dialogs of the UE. The diamond on the SIP signaling call flow in Figure 3 indicates the point at which the restoration system erases the dialogs of P-CSCF#1, S-CSCF#1, S-CSCF#2 and P-CSCF#2.

D. Recovering Session States

We propose that the restoration system puts priorities on the session states in order to recover the session states which will be used by the CSCF immediately because the SIP signaling call flow is processed (not completed) and the UE retransmits the SIP messages. For this prioritization, we classify three types of session states. The session state at which the SIP signaling call flow is not completed between the UEs is “pre-session”, the session state in which the final response 200OK message and ACK message are sent and received is “active session”, and the session state in which the UE only registers with the CSCFs is “register session”.

Table 1 Relationship between retransmitted SIP messages and transactions and dialogs necessary for CSCF to handle the retransmitted SIP messages

Phase No.	SIP message	Transactions and dialogs are saved before SIP proxy server send SIP message
1	1-17	None
2	18-27	14,15,16,17
3	28-37	24,25,26,27
4	38-42	34,35,36,37
5	43-52	39,40,41,42
6	53–62	49,50,51,52

In our proposed system, first the pre-sessions must be recovered because the application initiation by the UE will halt if the CSCFs do not return the SIP messages. Second, the active sessions are migrated and restored. Even if the restoration of the active-sessions is delayed, there is not severely influence on the communication between the UEs. But it is important to manage the termination of the communication for the accounting. Finally, the register sessions are recovered

The number of pre-sessions in the CSCFs is small compared to the register and active sessions, although the CSCFs accommodate a large amount of UE. Therefore, we consider that our proposed system enables the transactions and dialogs to be migrated into the alternate CSCF and the alternate CSCF to conduct the session management on them at an earlier time in the entire session restoration.

Because the active sessions are migrated second, a probability exists that the Bye messages are sent by the UE before migration of the active sessions is completed. In this case, the CSCFs normally return error code messages and the some applications of the UE stop retransmitting the Bye messages and terminated, which depend on the implementation. As a result, the NSPs then cannot obtain the accounting information from the UE. To solve this problem, we append a function, that is, the CSCF discards the Bye message and does not return any error messages before the completion of restoring the session state. When the UE retransmits the Bye messages, the session is terminated as normal by the CSCF, which takes over the session state.

IV. EVALUATION

A. Model of Experiment

Our proposed system aims to reduce the number of backup servers. The number of CSCFs treated by the single backup server increases as the number of backup servers decreases. In the case that the process for saving session state congests in the backup server, the slow responses from the backup server to the CSCF affects the completion time of the SIP signaling call flow. We define the queuing delay as the duration taken by the backup servers to finish the process for

 ISBN: 978-1-61208-174-8

Table 2 Parameters employed in experiments

Propagation Delay	Value (millisecond)
Between UE and P-CSCF	10
Between CSCFs (P-CSCF#1 and S-CSCF#1,S-CSCF#1 and S-CSCF#2, S-CSCF#2 and P-CSCF#1)	1
Between backup server and CSCF	10

saving the session states. If a large queuing delay for treating the request messages that arrive continuously from the CSCFs occurred in the backup server, the duration for establishing the session between the UEs is delayed because the CSCF can send the next SIP message in the SIP signaling call flow only after the session states are saved. The approximate propagation delay is easy to estimate beforehand between the backup servers and the CSCFs because NSP can constructs dedicated network for the transaction between CSCFs and backup servers and the sufficient network capacity can be prepared. We evaluated how long the queuing delay in the backup servers will affect on the duration for establishing the session between the UEs with an event-based simulator.

The CDMA 2000 defines about 5 seconds as the suggested time of paging process [13]. This indicates that the additional duration for establishing the sessions between the UEs is sometimes required when the location of the UE is searched for. We adopt 5 seconds as the criterion in order to evaluate the delayed time of the application initiation procedures by our proposed system. We believe that there is no problem if the overhead of saving session state reached to the same as the duration of the paging process and that total duration for establishing the session is not beyond 20 seconds even if the paging process spends about 5 seconds.

We compared our proposed system to the case where the backup servers save the session states every time the P-CSCFs and S-CSCFs send SIP messages (hereafter, termed "Allcopy"), that is same with the approach reported in the [6]. In this evaluation, the number of backup servers (= *b*) was changed and Figure 2 is used as the SIP signaling call flow where 2 set of P-CSCFs and S-CSCFs exist.

We also assumed that 100 million UEs are accommodated by multiple CSCFs. The call arrival rate follows the Poisson process for that number of UE. Additionally, to represent an on-peak period, the offered calls were 20 times larger than the 1-year average call arrival rate (1.7 calls per user a day) [13]. The duration of call ringing was a uniform random numbers between 1 and 5 seconds. The call duration followed the exponential distribution with 120 seconds as the average [13].

The parameters for the for the propagation delay of links are shown in Table 2. We used 10 and 20 microseconds, respectively as the time *c* for executing saving the session states in the backup server after the backup servers receive the request messages for saving from the CSCFs.

B. *Evaluation Result*

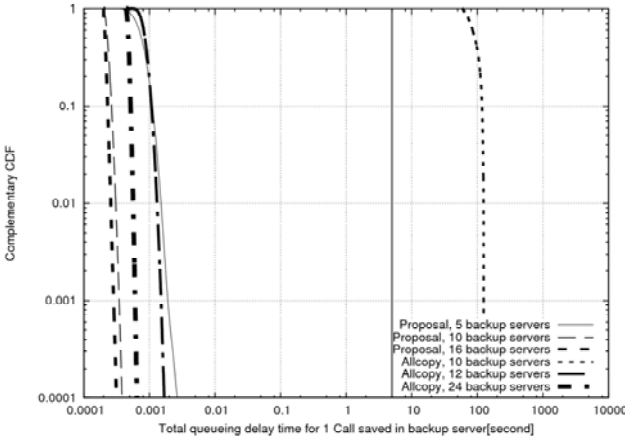


Figure 4 Total queuing delay time for 1 call saved in backup servers when *c*=10 microseconds

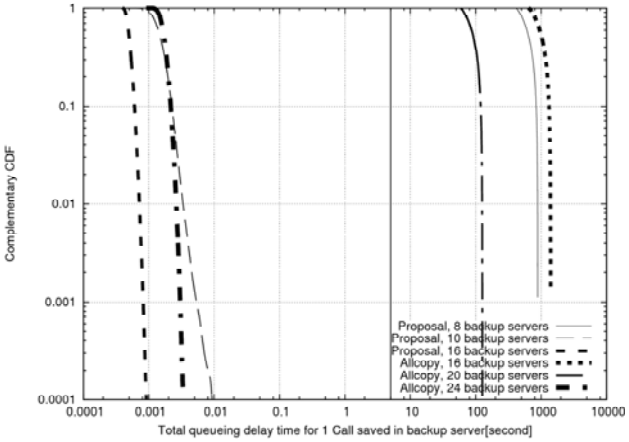


Figure 5 Total queuing delay time for 1 call saved in backup servers when *c*=20 microseconds

As the total number, 72 messages in Figure 3 were processed through the CSCFs, 12 of which are sent by the UEs. In our proposed system, the session states are saved at 20 entry points, at which the CSCFs are made to wait for the completion of saving. The total duration of the propagation delay for sending the session states to the backup server and returning the acknowledgement from the backup server was 400 milliseconds.

In the all copy case, the session states are saved at 48 times from the P-CSCF and S-CSCF (except the cases where the I-CSCF sends the SIP message and where the UE sends the Bye message). In this case, the total time of the propagation delay for sending the session states to the backup server and returning the acknowledgement from the backup server was 960 milliseconds. We consider neither value had a large impact on the UE.

The total queuing delay time for one call (from INVITE message to ACK message) saved in the backup servers for *c* = 10 and 20 microseconds is shown in Figures 4 and 5, respectively. The x-axis represents that the queuing delay in the case *c*= 10 or 20 microseconds but does not include the duration for sending the session states to the backup server

 ISBN: 978-1-61208-174-8

and returning the acknowledgement from the backup server. The y-axis represents the complementary cumulative distribution function (CCDF). Both axes use a log scale. We drew the line which means 5 seconds in Figures 4 and 5.

As Figure 4 shows, our proposed method did not accumulate large queuing delay when b = 6, 10 and 16, and c= 10 microseconds and gave less impact on the duration for establishing the sessions between the UEs. However, the All copy case accumulated the large queuing delay (more than 5 seconds) when b = 10 because the treatment for the request messages in the backup servers was congested. When b= 10, the quality of the IMS-based services is degraded because the duration for establishing the session between the UEs is beyond 5 seconds. In this situation, in order not to delay the duration for establishing the sessions, 12 backup servers are necessary for the Allcopy, however 5 backup servers are necessary for our proposed system. Our proposed system achieves about 60% reduction of the backup servers.

As Figure 5 shows, our proposed system did not accumulate large queuing delay when b = 10 and 16, and still gave less impact on the duration for establishing the sessions between the UEs. However, our proposed system generated a large queuing delay when b = 8. The Allcopy also accumulated a large queuing delay when b=16, and 20. This indicates that more backup servers are necessary to shorten the duration for establishing the sessions between the UEs. In this situation, in order not to delay the duration for establishing the sessions, 24 backup servers are necessary for the Allcopy, however 10 backup servers are necessary for our proposed system. Our proposed system also achieves about 60% reduction of the backup servers.

Compared the result in Figure 4 to the one in Figure 5 when b =10, 16 and 24, each increased time of queuing delay in the Allcopy is much more than in our proposed system. When the duration for executing saving the session states becomes longer, the effect of our proposed system becomes larger.

V. CONCLUSION

This paper introduced an approach to save and recover the session states of the CSCFs in a low cost solution. In this paper, we aim to reduce the number of the times to be saved by leveraging the feature of the retransmission in SIP. Our proposed system contributes to the recovery of even the pre-session and, at the same time, the reduction of the number of the backup servers. We presented selectively saving and restoring in turn based on the type of the session state. We highlighted the relationship between the retransmitted SIP messages and the transactions and dialogs necessary for CSCF to handle the retransmitted SIP messages. The existing system unnecessarily saves the data from the SIP proxy servers for restoring the servers.

We evaluated the queuing delay which the backup servers spend for saving the session state as the overhead of our proposed system. The overhead of our proposed system was shown with simulation experiments compared to the case where the backup servers save the session states every time the P-CSCFs and S-CSCFs send SIP messages The experiments showed that the overhead of the selectively-saving in our proposed system did not affect much more on the completion of the SIP signaling call flow than the system described in [6]. Our proposed system can achieve about 60% reduction of the backup servers. Our proposed system contributes to cost saving because it succeeds in saving the session states in fewer backup servers.

As the future work, we need to evaluate how long it will take to recover the session states in the alternate CSCF with the implementation of our proposed system.

REFERENCES

[1] ETSI “Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN)”, ETSI E2 282 001 V1.1.1 NGN Functional Architecture Release 1.

[2] ETSI, “Telecommunications and Internet Converged Services and Protocols for Advanced Networking (TISPAN); IP Multimedia Call Control Protocol based on Session Initiation Protocol (SIP) and Session Description Protocol (SDP) Stage 3”, ETSI ES 283 003, V1.1.1, April 2006.

[3] M. Handley, H. Schulzrinne, E. Schooler and J. Rosenberg, "SIP: Session Initiation Protocol", IETF RFC 3261, June 2002

[4] D.Jewett: “Integrity S2: A Fault-Tolerant Unix Platform”, IEEE 21th FTCS International Symposium 1991, pp.512-519

[5] D.Bryan, P.Matthews, E.Shim, D.Willis, and S.Dawkins, “Concepts and Terminology for Peer to Peer SIP”, draft-ietf-p2psip-concepts-03, Oct, 2010

[6] G.Kambourakis, D.Geneiatakis, S.Gritzalis, C.Lambrinoudakis, T.Dagiuklas, and J. Fiedler, “High Availability for SIP: Solutions and Real-Time Measurement Performance Evaluation”, International Journal of Dissaster Recovery and Business Continuity, Vol. 1, No.1, Feburary,2010

[7] A.Kobayashi, H.Ishizuka, N.Tomoed, T.Sone, E.Kosugi, and A.Iwatani, “VoIP network monitoring solution for IP telephony service as public communication infrastructure”, Mitsubishi Electoronics Technical Paper, April, 2006

[8] T.Usui, T.Kubo, Y.Kitaji and H.Yokota, “A Study on Locating Lossy Links of Signaling Messages in SIP-based Services”, IEICE trans-b, vol.E94-B, No.1, January 2011

[9] O. Damani, P. Chung, Y. Huang, C. Kintala, and Y. Wang,"ONE-IP: techniques for hosting a service on a cluster of machines", Computer Networks, vol. 29, pp. 1019-1027, September, 1997

[10] D. Oppenheimer, A. Ganapathi, and D. Patterson,"Why do internet services fail, and what can be done about it?", in 4th USENIX Symposium on Internet Technologies and Systems (USITS '03), (Seattle, WA), March, 2003

[11] 3GPP TS 33.203 (v7.6.0), "3G security; Access security for IP based services", Release 7, June, 2006

[12] V. Ramaswamy and J.Chung, “Performance Analysis of the Quick Idle State Protocol of CDMA 1xEV-Do Rev.B systems”, IEEE Globecom 2010

[13] Telecom Data Book 2010, Telecommuncations Carries Association in Japan, http://www.tca.or.jp/databook/index.html

 ISBN: 978-1-61208-174-8

Mobility Aware Routing for Multihomed Wireless Networks Under Interference Constraints

Preetha Thulasiraman
Department of Electrical and Computer Engineering
Naval Postgraduate School
Monterey, CA, USA
pthulas1@nps.edu

Abstract— **In this paper, the problem of interference aware routing using local mobility management (LMM) is addressed in multihomed wireless networks in which multiple fixed relay nodes are deployed to locally maintain and deliver mobility information collected from the surrounding mobile users. We present a new interference aware routing algorithm that uses the signal to interference noise ratio (SINR) value as the routing metric. The LMM model, based on the Hidden Markov Model (HMM), is implemented to calculate the SINR value of a specific link at particular time instances. This information is used to proactively perform route construction based on least interference. We minimize the total cost of routing as a cost function of SINR while guaranteeing that the load on each link does not exceed its capacity. We compare our LMM and SINR based routing algorithms with conventional counterparts in the literature and show that our algorithms have better prediction accuracy while reinforcing routing paths with high link quality and low latency.**

Keywords – **Interference; hidden markov model; SINR routing; mobility prediction**

I. INTRODUCTION

In recent years, services supported by mobile communications have expanded from simple voice traffic to various multimedia applications, resulting in the rise of 4G systems. These 4G cellular systems are required to provide high and homogeneous data rates over the complete cell coverage area while assuring a level of quality of service (QoS). Traditional cellular architectures, where each Mobile Station (MS) directly communicates with the Base Station (BS), are not capable to provide such homogeneous high bit rates due to the signal attenuation with increasing distance. Achieving the defined 4G objectives requires installing either a higher number of base stations, or integrating cellular and ad-hoc networking technologies. The integration of cellular and ad-hoc technologies, also referred to as Multi-hop Cellular Networks (MCN) [1], has gained significant research attention given its capacity to achieve the 4G objectives by substituting a direct MS-BS link by multi-hop links using intermediate nodes (relays) to retransmit the information from source to destination. Various architectures are available to MCNs [2], including both fixed and mobile relays. In this paper we focus on MCNs with fixed relay nodes where the base station communicates directly with fixed relay nodes which in turn cooperatively relay information in an ad hoc fashion to other users in connectivity range. In this architecture, each fixed relay behaves as a "pseudo-base station" or "home" for the mobile users by providing services (i.e., routing and mobility management) that would normally be taken care of by the base station in a centralized manner. This is termed a *multihomed* MCN. The concept of multihoming has been extensively discussed in the context of Mobile IP [3] to improve network connectivity and manage mobility. Multihomed architectures have also been predominantly used to develop fault-tolerant routing protocols by ensuring that user nodes have multiple connection opportunities in the event that one home relay fails [4], [5].

A. Motivations and Related Work

The cooperation between fixed relays and the base station is the cornerstone for efficient communication at the network layer. A mobile user, MS, is served by a nearby relay node that forwards packets (potentially over multiple wireless hops) to the BS. In addition to traffic forwarding and route decision making, the relays also have the responsibility of managing user mobility by collecting information regarding user movements from one home relay to another. This essentially reduces the burden on the base station by localizing mobility management.

A consequence of the increased use of cellular networks is the inherent interference that is induced. Wireless interference is influenced by node mobility and can lead to performance degradation. The time varying mobility patterns of the users (i.e., mobility patterns, speed, direction etc.) can cause new interference to be induced at neighboring nodes [6]. Interference can be controlled/mitigated in the network layer i.e., with routing. In order to design an effective routing algorithm that mitigates the interference experiences of the wireless links, the mobility of the users must be considered. Mobility assisted routing has been studied in the literature for several years, more recently focusing on ad hoc and delay tolerant networks [7], [8]. However, none of these works discuss the direct impact of interference on the routing protocols. More recently, in [6], mobility aware routing using interference constraints was developed. However, the interference is modeled using the protocol model which induces binary conflicts (either two links interfere or they do not despite neighboring simultaneous transmissions) which is not true in practice. Our focus is on the use of the signal to interference noise ratio (SINR) interference

 ISBN: 978-1-61208-174-8

model (also known as the physical interference model), which is based on practical transceiver designs of communication systems that treat interference as noise. Under the SINR model, a transmission is successful if and only if the SINR at the intended receiver exceeds a threshold such that the signal can be decoded with acceptable bit error probability. Although the SINR model has been shown to be more computationally complex than the protocol model, it also provides a more practical and realistic assessment of wireless interference [9]. Routing protocols using SINR to model interference have been studied in both static networks [10], [11], [12] and mobile networks [13]. However, although the work of [13] uses SINR for route selection, the mobility modeling is based on the random waypoint model, and therefore no specific mobility prediction is introduced. In addition, [13] does not correlate wireless interference with mobility.

Our objective in this paper is to study SINR and its relationship to interference based routing using localized mobility management information.

B. Contributions and Organization

The contributions of this paper are two-fold. First, we propose a localized mobility management (LMM) model based on the Hidden Markov Model (HMM) where the mobility information (i.e., location) of each user is collected by the corresponding home relay node for movement prediction purposes. Second, we develop a SINR based routing algorithm which uses the location of a mobile user at time t to determine least interfering paths. Specifically, we develop the routing algorithm such that the link costs are derived from the SINR values and the chosen routes have minimum cost (minimum interference). In addition, we ensure that the capacity of each link is not violated when the traffic is routed.

The rest of the paper is organized as follows: Section II describes the system model. In Section III, we discuss the LMM model used in this paper while in Section IV the SINR based routing algorithm is developed. The performance evaluation of the LMM model and SINR routing algorithm is discussed in Section V. We conclude the paper in Section VI.

II. System Model

The multihomed MCN that is the focus of this paper is shown in Fig. 1. Each home relay interacts with a set of mobile users as well as with each other. In addition, as in traditional MCNs, the various MS nodes can also interact with each other. Thus a MS node may use other MS nodes to relay information to a home relay or to the BS. It must be noted that a MS can directly interact with a BS rather than a home relay if it is closer to the BS than to the home relay. The BS is connected to the wired infrastructure and behaves as a gateway to the Internet. The LMM model that is used to predict the next location of each user node is handled by the individual home relays. Each home relay collects and maintains information regarding the movement of the mobile users connected to it.

To understand the interaction between the various components of our framework, we provide a block diagram given in Fig. 2. The block diagram shows the LMM model and its relationship to the SINR based routing algorithm. The prediction of the user's movement is driven locally by a HMM that is performed by each home relay. The current mobility information and the history of the user's past movements is used to make predictions. This information is maintained in the mobility database of each home relay which keeps track of users that are connected, were connected or will be connected (prediction) to the home relay. The next predicted location of the mobile user, as determined by the home relay, is broadcast to other home relays within transmission range so that they may update their databases accordingly. This updated information is then used to calculate the induced SINR interference at the receiver to proactively construct paths with least interference. The calculation of the SINR value at a time t in a mobile setting must be computed instantaneously. To facilitate the SINR calculation and the execution of the LMM and routing algorithms, it is assumed that the user nodes are quasi-mobile [14]; each user moves with a certain velocity and for a time T stays at one location before moving to a new random location.

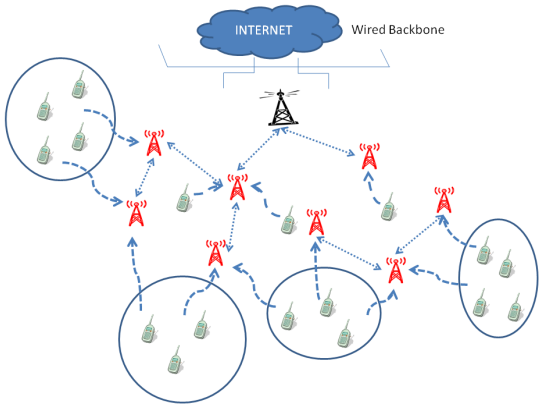


Fig. 1. Multihomed MCN where sets of user nodes are connected to a home relay and home relays communicate with other home relays in its transmission range to transmit information to the base station

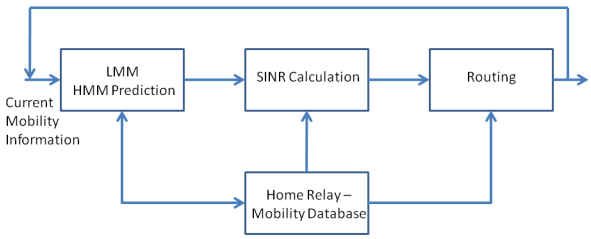


Fig. 2. Block diagram that illustrates the interaction between the LMM model and the interference aware routing algorithm

III. LMM Model

A HMM is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (hidden) states. In a regular Markov model, the state is directly visible to the observer, and therefore the state

 ISBN: 978-1-61208-174-8

transition probabilities are the only parameters. In a hidden Markov model, the state is not directly visible, but output, dependent on the state, is visible. A HMM has two kinds of stochastic variables: state variables (hidden variable) and the output variables (observable variable). A HMM can be defined as follows:
$S : \{s_1 s_2 ... s_N\}$ are the N hidden states of the system
$O : \{o_1 o_2 ... o_N\}$ are the values of the observed sequences
$\Pi : \{\pi\}$ is the initial state probabilities. π_i indicates the probability of starting in state i
$A = \{a_{ij}\}$ are the state transition probabilities where a_{ij} denotes the probability of moving from state i to j

$$a_{ij} = P(t_k = s_j | t_{k-1} = s_i)$$

$B = \{b_{ik}\}$ are the observation state probabilities where b_{ik} is the probability of emitting symbol k at state i

$$b_{ik} = P(o_k | t_k = s_j)$$

The 3-tuple (A, B, π) provides a complete specification of the HMM for the system considered in this paper.

A. Mobility Model Using HMM

To track the state of a mobile user we apply two approaches: 1) forward-backward algorithm and 2) re-estimation algorithm for the HMM parameters discussed above. The main steps of the tracking algorithm can be summarized as follows:

1) Apply HMM re-estimation algorithm to obtain initial estimates of (A, B, π) of the HMM.
2) Apply the HMM forward-backward estimation algorithm to predict at time t the next state of a user.
3) Obtain refined estimates of (A, B, π) by again applying the HMM re-estimation algorithm to the given observation sequences.

These steps are performed at each home relay node during each observation interval. We define the observation interval as the time intervals during which observations (mobility information is collected) occur. The observation interval is assumed to be segmented into T subintervals indexed by $1, 2, ..., T$. T is defined as the time during which the mobile user remains stationary. Thus, the time during which the node remains stationary is the predicted state of the mobile network in the HMM.

1) Forward-Backward Algorithm: A forward-backward algorithm is an algorithm for computing the probability of a particular observation sequence in the context of hidden Markov models [15]. The algorithm first computes a set of forward probabilities which provide the probability of observing the first k observations in the sequence and ending in each of the possible Markov model states. The algorithm also computes a set of backward probabilities which provide the probability of observing the remaining observations given an initial state. For our model, we define the following forward and backward variables:

Forward variables:
$\alpha_t(n) = P[o_1^t, \text{state } n \text{ sojourn ends at } t], t \geq 1$
$\alpha_t^*(n) = P[o_1^t, \text{state } n \text{ sojourn begins at } t+1], t \geq 1$

Backward variables:
$\beta_t(n) = P[o_t^T | \text{sojourn in state } n \text{ begins at } t], t \leq T$
$\beta_t^*(n) = P[o_t^T | \text{sojourn in state } n \text{ ends at } t-1\, t \leq T$

The forward variables are then computed inductively for $t = 1, 2, ..., T$. Similarly, the backward variables are computed inductively for $t = T, T-1, ..., 1$. After computing the forward and backward variables, a state estimate can be found. Let us define,

$$\gamma_t(n) = P[o_1^T; s_t = n]$$

as the probability that s is observed to be in state n at time t. Then the estimate of s_t is given by

$$\hat{s}_t = \arg\max_{1 \leq n \leq N} \frac{\gamma_t(n)}{P[o_1^T]}, t = T, T-1, ..., 1$$

2) Re-estimation Algorithm: A simple iterative procedure for re-estimating the HMM parameters each time a node moves is reported in [15] and implemented in this paper.

IV. SINR Based Routing Using LMM Model

This section will discuss the formulation of the SINR routing algorithm using the developed LMM model.

A. Challenge of Routing with Interference and Mobility

Using the LMM model based on the HMM, we are able to track the movement of the users to determine which relay it is connected to. Interference depends on the existence of other sources/intermediate relays and their spatial separation. Thus the routing decision of a given source-BS pair becomes coupled to the routing decision of other source-BS pairs. To determine appropriate routing paths from the relay to the BS that are cognizant of interference, we use SINR as a routing metric.

B. Problem Formulation

For our analysis, we model the multihomed MCN as a graph, $G(V, E)$, where V is the set of nodes (relays, mobile users and base station inclusive) and E is the set of links. Let V_N be the set of mobile users and let V_M be the set of home relays. Note that the network has only one base station. The successful reception of a packet depends on the received signal strength, the interference caused by the simultaneously transmitting nodes, and the ambient noise level η. The SINR of a link (i, j) is given as follows

$$SINR_{ij} = \frac{P_j(i)}{\eta + \sum_{k \in V'} P_j(k)} \geq \beta \quad (1)$$

where $P_j(i)$ is the received power at node j due to node i, V' is the subset of nodes in the network that are transmitting simultaneously, and β is the SINR threshold. Our proposed routing protocol is implemented to route data using the least interfering path out of all path possibilities. If a link has a high SINR, it is an indication that it is experiencing low interference.

 ISBN: 978-1-61208-174-8

Each link (i, j) has an associated cost which is derived from the SINR value calculation. Each link also has an associated capacity denoted u_{ij}. The capacity is formulated using Shannon's formula, given in Eq.2.

$$u_{ij} = log_2(1 + SINR_{ij}) \quad (2)$$

In addition, the flow of packets from node i to its neighbor j over wireless link (i, j) is represented by f_{ij}.

C. SINR Based Routing

The position of each user node at time t affects the cumulative SINR on each link. The SINR is also affected by the path loss model and channel gain. The SINR at time t on link (i, j) is given by Eq.3,

$$SINR(t)_{ij} = \frac{G_{ij}P_j(i)(t)}{\eta + \sum_{k \in V'} G_{kj}P_j(k)(t)} \geq \beta \quad (3)$$

where G_{ij} is the channel gain on link (i, j) (in the simulations, the channel gain of each link is calculated using a Rayleigh fading model and an appropriate path loss factor), $P_j(i)(t)$ is the received power at node j due to node i at time t, and k is a simultaneously transmitting node. The corresponding capacity u_{ij} is then modified to be

$$u_{ij}(t) = log_2(1 + SINR_{ij}(t)) \quad (4)$$

The SINR is calculated during each observation interval, $t \in T$.

In order to determine the least cost (least interfering) paths, we use the minimum cost flow optimization technique. In our case, the cost of a link is motivated by the amount of interference on that link due to neighboring transmissions and/or noise. As we are using SINR as the routing metric, the higher the SINR, the better the link quality. Therefore, we want to minimize the *inverse* of the SINR value.

The objective of the SINR routing problem is to deliver all the data packets generated by the user nodes to the base station in the most cost-effective (least interfering) manner without exceeding the link capacities. Formally, the problem can be stated as follows.

$$\text{minimize} \quad \sum_{(i,j) \in E} SINR(t)^{-1} f_{ij}(t) \quad (5)$$

subject to

$$\sum_{j:(i,j) \in E} f_{ij}(t) - \sum_{j:(j,i) \in E} f_{ji}(t) = d_i(t), \forall i \in V_N \quad (6)$$

$$\sum_{k:k \in V_M \cup BS} \Big(\sum_{j:(k,j) \in E} f_{kj}(t) - \sum_{j:(j,k) \in E} f_{jk}(t) \Big) = - \sum_{i:i \in V_N} d_i(t) \quad (7)$$

$$0 \leq f_{ij}(t) \leq u_{ij}(t) \quad (8)$$

In the above formulation, d_i represents the rate at which the data packets are generated at user node i per unit time. The first constraint (Eq. 6) ensures flow conservation at each node. The second constraint (Eq. 7) ensures that the base station receives all the packets generated by all the nodes. The flow of packets on a link must not exceed its capacity and this is ensured by the third constraint (Eq. 8).

1) Solution: The above defined problem is similar to the minimum-cost flow problem, known in the operations research literature [16]. We will convert the above problem into the minimum-cost circulation problem as follows.

1) Add a super source x, and a super base station node y, to the graph $G(V, E)$.
2) Add directed links (x, i), connecting the super source x to node i, for all $i \in V_M \cup V_N$. Set costs of these links to 0 and the capacities to d_i.
3) Add directed links (j, y) connecting the base station and relay nodes to the super base station y. Set costs of these links to 0 and the capacities to infinity.
4) Add a directed link (y, x) connecting the super base station y to the super source x. Set the cost of the link (y, x) to $-|V|\beta$ and the capacity to infinity, where β is the minimum of any link cost (lower bound of SINR).
5) The modified graph is defined as $G'(V \cup \{x, y\}, E \cup E')$, where $E' = \{(x, i) : i \in V_N\} \cup \{(j, y) : j \in V_M \cup BS\} \cup \{(y, x)\}$.

2) Analysis of the Solution: Pushing more flow from x to y will decrease the overall cost of the flow due to the fact that the link from y back to x has sufficiently large negative cost. It is clear that the maximum flow is bounded from above by $F = d_1 + d_2 + ... + d_{|V_N|}$ because F is the maximum possible flow going out of x, the super source. There are two possibilities that have to be analyzed.

Case 1: $\sum_{i:i \in V_N} f_{xi} = \sum_{i:i \in V_N} d_i$

In this case, all the links of the form (x, i), $i \in V_N$ are saturated. The maximum-flow is restricted by the capacities of these links. Consider a link $(x, 1)$ having the capacity d_1. Since all the (x, i) links are saturated, the input flow at node 1 must be d_1 + $\sum_{j:(j,1) \in E} f_{j1}$ and the output flow must be equal to the input flow (flow conservation). There must be paths from node 1 to base stations which carry the flow d_1 + $\sum_{j:(j,1) \in E} f_{j1}$. The same argument holds for other nodes.

Case 2: $\sum_{i:i \in V_N} f_{xi} < \sum_{i:i \in V_N} d_i$

In this case the maximum flow is restricted by the capacities on the actual links ($(i, j) \in A$) of the network. The minimum cost flow algorithm will identify the paths from the user node i to the base stations which carry the flow d'_i where $0 \leq d'_i \leq d_i$, $\forall i \in V_N$. The flow on the links (x, i) would be d'_i, $\forall i \in V_N$.

V. Performance Evaluation

We first evaluate the LMM model separately to gauge its effectiveness in prediction accuracy. The initial parameters of the HMM are randomly generated using a uniform distribution (the number and locations of users and relays, relay-user associations and the initial transition probabilities are randomly generated). Once the users begin to move, its movement history is tracked and stored in the databases of each home relay for prediction.

We evaluate the SINR based routing algorithm using the following performance metrics: packet delivery ratio and end-to-end delay. We use NS-2 to simulate our evaluations and use CPLEX to solve the optimization formulation for the minimum cost SINR based routing algorithm.

The simulation environment is based on a 2250m x 2250m Manhattan type scenario, emulated with the NS-2 software

 ISBN: 978-1-61208-174-8

platform, with the BS located at the centre of the environment. The propagation loss is modeled using the Rayleigh fading model. The traffic is constant bit rate (CBR) with UDP based traffic at 4 packets per second and payload of 512 bytes. The data transmission rate is homogeneous among all the nodes and is set to 12Mbps. The radio transmission range of each node is 130m. The speed of the user nodes ranges from 1.5m/s to 5m/s and the simulation time is 1000 seconds. The simulated networks have 256 subcarriers with a system bandwidth of 2MHz. We also use different observation interval times, T. All results shown are an average of 20 different simulations.

A. Simulation Results for LMM Model

We first evaluate the prediction accuracy of our LMM model. Prediction accuracy is defined as the ratio of the number of times a user node moves to different relay nodes to the ability of the system to predict the location. For example if node n moves to relay node A and then to relay node B, and our prediction model predicts correctly that it moved to A but not B, then the prediction accuracy is 50%. Fig. 3 and Fig. 4 show the prediction accuracy in percentages for two user nodes in the network. We compare our LMM model with prevalent prediction models, specifically a generic Markov chain and a second-order Markov chain. When the user nodes make first contact with a relay node, the initial, randomly generated parameters of the HMM are used. Once the user nodes begin to move, its movement history is tracked and stored in the databases of each relay node for prediction. Each network that is simulated has relay nodes varying from 2 to 14 and the number of users range from 10 to 120. From Figs. 3 and 4, we can conclude that the LMM has an advantage in prediction accuracy compared to the Markov and second-order Markov chains. The results also show that the LMM can better adapt to a user node's change in movement. In other words, the LMM learns faster than the generic Markov based approaches.

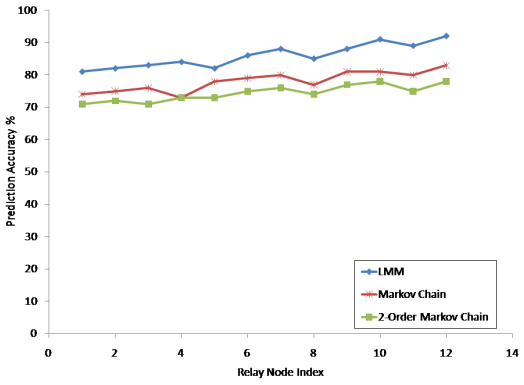


Fig. 3. Comparison of prediction accuracy for the proposed LMM model, generic Markov chain and second-order Markov chain for User Node 1 in networks with 120 users

B. Simulation Results of SINR Based Routing Algorithm

The performance of the SINR routing algorithm is evaluated compared to two SINR based routing algorithms given in [10] and [13]. In [10], an algorithm, 2-HEAR, is developed in which a routing metric is used such that a node calculates the SINR to its neighboring links based on a 2-hop interference range only. In [13], a modified version of the AODV routing algorithm is proposed in which SINR is used to calculate the route quality while using a random waypoint mobility model. We denote the above approaches as 2-HEAR and AODV-INT, respectively, in the simulation graphs. To calculate the SINR, we take the following steps. The received power, $P_j(i)(t)$, is calculated according to the radio propagation model at the receiver. The noise, η, is calculated as additive white Gaussian noise (AWGN) that is modeled as a Gaussian random variable. The pathloss exponent (LOS/NLOS) is set to 2.35/3.76. The same networks used in the LMM simulations of Section V-A are used in the simulations of the SINR routing algorithm.

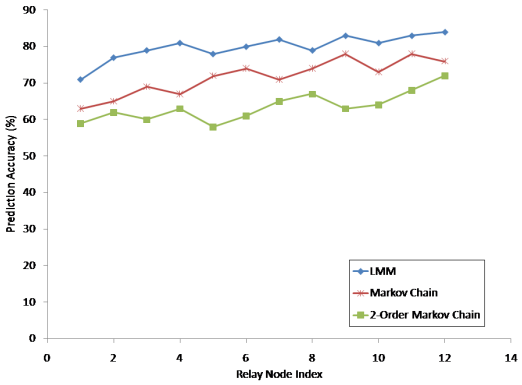


Fig. 4. Comparison of prediction accuracy for the proposed LMM model, generic Markov chain and second-order Markov chain for User Node 2 in networks with 120 users

We first evaluate the packet delivery ratio for our SINR based routing algorithm and its two relevant counterparts in the literature. In Fig. 5 and Fig. 6, the results of the packet delivery ratio for varying node speed and observation intervals ($T = 10ms$, $T = 1ms$) are shown. From the results it can be seen that our algorithm provides better packet delivery ratios when compared to the other approaches. We can justify the better performance of our results as follows: In 2-HEAR the SINR calculated by each node only includes those nodes within a 2-hop range which means that even if interference beyond this range occurs, it is not captured in the routing metric. If the interference level is high beyond the 2-hop range, packets drops may occur, requiring retransmissions.The results of the algorithm from AODV-INT are better than 2-HEAR but because it does not use a specific mobility prediction model, it fails to capture precise interference information as is done in our proposed routing algorithm.

We next evaluate the end-to-end delay of our algorithm for varying node speeds and $T = 1ms$. The results are shown in Fig. 7. The average end-to-end delay is improved compared to 2-HEAR and AODV-INT mainly due to more robust routes and less route discoveries. For the LMM model and the SINR routing algorithm, the density of the networks impacts the network performance. Simulations were performed that showed a decrease in the routing performance when the

 ISBN: 978-1-61208-174-8

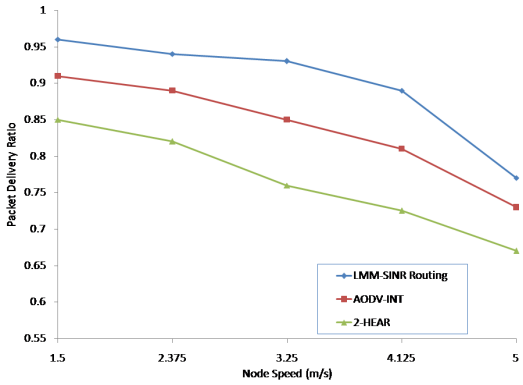


Fig. 5. Packet delivery ratio versus varying node speeds for $T = 10ms$

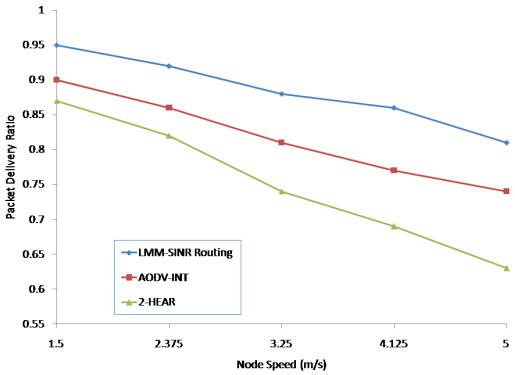


Fig. 6. Packet delivery ratio versus varying node speeds for $T = 1ms$

density increased (i.e., routing overhead due to prediction increased). Therefore, the routing and prediction algorithms are limited to an extent because of scalability. Due to space constraints, these simulations are not presented in this paper.

VI. Conclusion

In this paper we develop a minimum interference aware routing algorithm for multihomed wireless networks where link costs are derived from the SINR values. The mobility of each user is captured by a localized mobility management model based on HMM where home relays locally collect mobility information. We show that our LMM model has better prediction accuracy than other generic Markov based mobility predictors. We also show that our SINR based routing algorithm guarantees minimum interference paths by increasing the packet delivery ratio and reducing latency compared to established SINR based routing approaches in the literature. In our future work, we plan to integrate the mobility of relay nodes to analyze the impact of SINR induced interference on routing and overall network performance.

Acknowledgement

This work was funded by the Research Initiation Program (RIP) at the Naval Postgraduate School, Monterey, CA, USA.

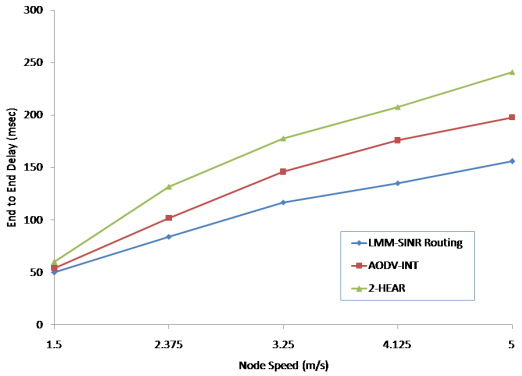


Fig. 7. End-to-end delay for $T = 1ms$ and varying node speed

References

[1] Y.-D. Lin and Y.-C. Hsu, "Multihop cellular: a new architecture for wireless communications," in *Proceedings of IEEE INFOCOM*, 2000, pp. 1273–1282.

[2] X.J. Li, B.-C. Seet, and P.H.J. Chong, "Multihop cellular networks: Technology and economics," *Computer Networks (Elsevier)*, vol. 52, no. 9, pp. 1825–1837, June 2008.

[3] Y. Li, D.-W. Kum, W.-K. Seo, and Y.-Z Cho, "A multihoming support scheme with localized shim protocol in proxy mobile ipv6," in *Proceedings of IEEE ICC*, 2009, pp. 1–5.

[4] P. Thulasiraman, S. Ramasubramanian, and M. Krunz, "Disjoint multipath routing to two distinct drains in a multi-drain sensor network," in *Proceedings of IEEE INFOCOM*, 2007, pp. 643–651.

[5] Y. Amir, C. Danilov, R. Musaloiu-Elefteri, and N. Rivera, "An interdomain routing protocol for multi-homed wireless mesh networks," in *Proceedings of IEEE WoWMoM*, 2007, pp. 1–10.

[6] R. Langer, N. Bouabdallah, and R. Boutaba, "Mobility-aware clustering algorithms with interference constraints in wireless mesh networks," *Computer Networks*, vol. 53, no. 1, pp. 25–44, January 2009.

[7] L. Badia, N. Bui, M. Miozzo, M. Rossi, and M. Zorzi, "Mobility-aided routing in multi-hop heterogeneous networks with group mobility," in *Proceedings of IEEE GLOBECOM*, 2007, pp. 4915–4919.

[8] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Efficient routing in intermittently connected mobile networks: the multiple-copy case," *IEEE/ACM Transactions on Networking*, vol. 16, no. 1, pp. 77–90, February 2008.

[9] A. Iyer, C. Rosenberg, and A. Karnik, "What is the right model for wireless channel interference?," *IEEE Transactions on Wireless Communications*, vol. 8, no. 5, pp. 2662–2671, May 2009.

[10] R.M. Kortebi, Y. Gourhant, and N. Agoulmine, "On the use of sinr for interference-aware routing in wireless multi-hop networks," in *Proceedings of ACM MSWiM*, 2007, pp. 395–399.

[11] S. Kwon and N.B. Schroff, "Energy-efficient sinr-based routing for multihop wireless networks," *IEEE Transactions on Mobile Computing*, vol. 8, no. 5, May 2009.

[12] P. Thulasiraman, J. Chen, and X. Shen, "Multipath routing and max-min fair qos provisioning under interference constraints in wireless multihop networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 22, no. 5, pp. 716–728, March 2011.

[13] J. Park, S. Moh, and I. Chung, "A multipath aodv routing protocol in mobile ad hoc networks with sinr-based route selection," in *Proceedings of IEEE International Symposium on Wireless Communication Systems (ISWCS)*, 2008, pp. 682–686.

[14] R.C. Ramos and L.F.G. Perez, "Quasi mobile ip-based architecture for seamless interworking between wlan and gprs networks," in *Proceedings of IEEE Conferences on Electrical and Electronics Engineering (CIE)*, 2005, pp. 455–458.

[15] S.-Z. Yu and H. Kobayashi, "Practical implementation of an efficient forward-backward algorithm for an explicit-duration hidden markov model," *IEEE Transactions on Signal Processing*, vol. 54, no. 5, pp. 1947–1951, May 2006.

[16] R. Ahuja, T. Magnanti, and J. Orlin, *Network Flows*, Prentice Hall, 1993.

 ISBN: 978-1-61208-174-8

Priority-based Packet Scheduling in Internet Protocol Television

Mehmet Deniz Demirci
Computer Science Department
Istanbul University
İstanbul, Turkey
e-mail:demircid@istanbul.edu.tr

Abdul Halim Zaim
Institute of Science and Engineering
Istanbul Commerce University
Istanbul, Turkey
e-mail:azaim@iticu.edu.tr

***Abstract*—Techniques to provide the quality of service for policing and scheduling IPTV are discussed (Weighted Fair Queuing and Alpha-Beta Virtual Clock). Also, a new scheduling algorithm is proposed for prioritized services over IPTV. Simulation results for the proposed algorithm is presented. In the proposed algorithm, the amount of packets that belongs each of the five priority classes is broadcast to every switcher. A series of estimated values is obtained by calculating the status of the amount of received packets. Then, these estimation values are used to calculate the credit of each packet received. Switcher selects the appropriate packet to forward to the next switcher by its credit value. The results are discussed as epilogue with objective comments.**

Keywords-IPTV; Priority; Scheduling; QoS; Class of Service; Priority Class

I. INTRODUCTION

Internet Protocol Television known as IPTV consists of several services that provide triple play entertainment. IPTV has 3 different services like television broadcasting, voice over IP [1] (Internet Protocol; RFC791 [2]) and data services bundled together to the subscribers. Providing all of these services at the same time to the subscribers is still a big challenge for researchers all around the world.

In IPTV, there have to be more than one types of packets. If one tries to transmit a real-time live soccer match with normal best-effort traffic then the packets belonging to the match's flow will suffer severe packet loss, delay and jitter.

Packet loss, delay and jitter are the most important aspects of Quality of Service (QoS). That influence the performance and Quality of Experience (QoE) received by the subscriber.

In multimedia streams, packet loss between 1% to 20%, and end-to-end delays until 100 ms are optimum but delays between 100 ms to even 1000 ms (with additional set top box buffering) are acceptable [3]. So, that traffic scheduling for IPTV must respect the difference of class for all kinds of triple-play packets.

The main challenge in IPTV implementation is distributing fairly the resources to each class of service. The Internet, is a best effort service which is neutral with respect to different services. The goal of this work is adjusting the flow of multi-serviced packets in terms of QoS.

Even though IP is a best effort service, the end-to-end delay and jitter are reduced using the proposed algorithm.

In Section 2, how the services are classified is presented. In Section 3, the proposed algorithm is compared to other work. In Section 4, the details of the proposed algorithm are given. In Section 5, the results, and in Section 6, the conclusion, are shown, respectively.

II. CLASSIFICATION OF SERVICES

IPTV packets can be grouped into five priority classes shown as in Figure 1. Priority class 1 is the real time video broadcasting, i.e. a live Champions League Match. This is the most important class of service because a live stream never stops and every seconds count. Priority class 2 is video on demand (VoD) an IPTV class of service which means, a multi million dollar budget new movie is rented by the user and sometimes user stops or pauses the movie even rewind or forward. Voice over IP applications such as internet telephony, belong to Priority Class 3. Best effort services such as web surfing, e-mail or ftp are members of the priority class 4. The last one is the priority class 5 that possess the signalization data. The importance of the class of services is given in descending order.

 ISBN: 978-1-61208-174-8

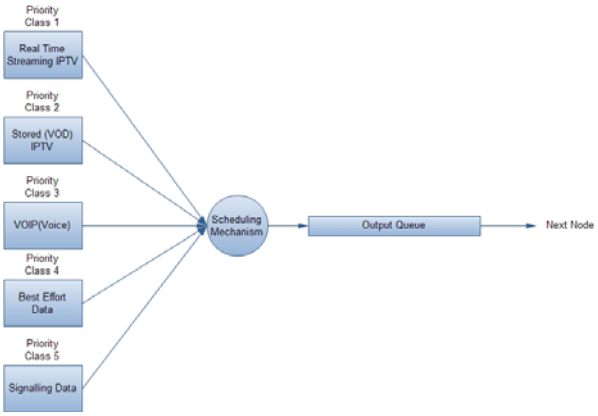


Figure 1. Priority Based Classified Scheduling Node Structure

III. RELATED WORK

The proposed algorithm is compared with two well-known algorithms which are Weighted Fair Queueing (WFQ) [4] and Alpha-Beta Virtual Clock (ABVC) [5].

In WFQ, incoming packets are classified by their type of service and placed into the appropriate queue. The packets are serviced by the node's scheduler in circular form. First of all, the first queue is serviced, then second, and so on. After serving the queue with the least priority, cycle restarts with the highest priority queue. During the cycle, an empty queue is skipped. WFQ is a little bit different than Round Robin [3] because every queue has a weight symbolized as w_i. Thanks to the weights assigned to each queue it guarantees (1)

$$w_i / \sum w_j \tag{1}$$

of the total bandwidth. Thus every time, for the transmission rate R, the class i has (2)

$$Rw_i / \sum w_j \tag{2}$$

of guaranteed rate. WFQ is depicted in Figure 2 [3].

The other algorithm, Alpha-Beta Virtual Clock (ABVC), is an enhanced version of Zhang's Virtual Clock [6]. ABVC forward packets to output queue of the node by inspecting which flow is sending packets and which not.

Initially, the number of active flows is n. The Virtual-Tick value of active flows is:

$$VT_i = \frac{1}{r_i}; \sum_i r_i = 1 \tag{3}$$

In any interval of time (n-j), flows start to send packets then the total bandwidth of all passive flows, δ becomes (4)

$$\delta = \sum_{i=j}^{n} r_i \tag{4}$$

δ is divided equally to j active flows. Then the virtual-tick value of active flows is calculated as (5)

$$VT_i = \frac{1}{(r_i + \frac{\delta}{j})} \tag{5}$$

Whenever a new flow starts to send packets, it takes equal amount of bandwidth from each flow without affecting the other flows. Virtual-tick value is updated as (6)

$$VT_i = \frac{1}{(r_i - \frac{r_{n+1}}{n})} \tag{6}$$

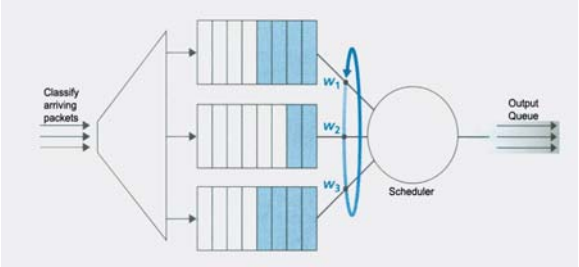


Figure 2. Schema of Weighted Fair Queuing

Each flow has an auxVC value which is equal with real time, after receiving new packets from a passive flow, the auxVC value is updating as (7)

$$auxVC_i = auxVC_i + VT_i \tag{7}$$

IV. PROPOSED WORK

Our scheduling algorithm is implemented on the topology depicted in Figure 3. Broadcasting is made by n units of IPTV service provider as 5 classes of service. Each node is an onboard switcher situated in IRIDIUM satellites. The proposed algorithm can be used in several different wireless and wired network topologies.

In simulations, for a scheduling task, packets generated by a single service provider are routed to a subscriber over a predefined route. In order to simulate a real time IPTV traffic other service providers are also included. The service provider sends an estimated value of packets that belongs to each of the priority classes to all switchers of how many packets from each priority class will be sent. The credit of each packet is calculated while the packet placed in the corresponding queue. An iterative estimation value is obtained by (8) whenever a packet is received from any priority class. (8) is a slightly specialized form of exponential smoothing or exponential averaging [7].

 ISBN: 978-1-61208-174-8

$$Est_i^{n+1} = \alpha(Est_i^n) + (1-\alpha)RT_i \text{ ,n=0,1,2,..} \quad (8)$$

Est_i^n : estimated packet count of i.th service class' n.th packet

α : alpha smoothing constant. $0 < \alpha < 1$. In the proposed algorithm the value 0.125 is selected.

RT_i : the number of received packets of i.th class

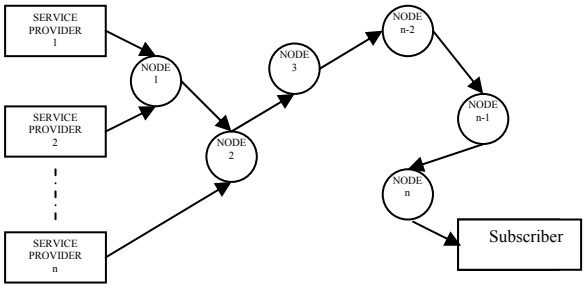


Figure 3. Exemplary Network Topology

After the estimated value is obtained, the credit of nth packet is computed as following:

$$Credit_i^n = \frac{Est_i^{n-1}}{PC} + SubCredit_i x Threshold_i \quad (9)$$

n=1,2,….

Est_i^{n-1} : estimation value of (n-1)th packet belonging ith service class.

$SubCredit_i$: momentary SubCredit value of ith service class

$Threshold_i$: momentary threshold value of ith service class

Table 1 shows the initial values of the SubCredit for each service class.

Threshold values starts all from zero and are updated when a packet is served by the scheduler from any of the service classes (SC). The value of and the other SC values are augmented as shown in Table 2, which indicates the values added to the previous value of the Threshold in each class. n is a natural number. Whenever the threshold of a SC reaches 35n, the first packet in the class queue is served automatically without checking the maximum credit.

If any of the threshold values reach 35n then the credit of each leading packet in the class queues are checked by the scheduler and the maximum one is routed to the output queue of the active switch.

All packets driven to the output queue of the current switcher are released to the link in the FIFO (First in First Out) order. At the next switcher all of the process is executed again.

The other packets coming from the other service providers through other switchers are also competing with our subscriber's packets.

TABLE I. SUBCREDIT INITIAL VALUES

SubCredit Initial Values	
Service Class 1 (PC1)	10
Service Class 2 (PC2)	8
Service Class 3 (PC3)	6
Service Class 4 (PC4)	4
Service Class 5 (PC5)	2

TABLE II. AUGMENTING THRESHOLD VALUES

Augmenting Threshold Values	
Service Class 1 (PC1)	+11n
Service Class 2 (PC2)	+7n
Service Class 3 (PC3)	+4n
Service Class 4 (PC4)	+2n
Service Class 5 (PC5)	+n

Table 3 shows how SubCredit of each SC is changed whenever a packet is selected and sent to the output queue. Figure 4 summarizes the algorithm by showing the flow diagram of the process.

 ISBN: 978-1-61208-174-8

TABLE III. SUBCREDIT UPDATE TABLE

SubCredit Update Table					
	PC1	***PC2***	***PC3***	***PC4***	***PC5***
A packet of PC1 is selected	$\frac{(\frac{PC2+PC3}{2}+10)xPS1}{50.Max(SubCredit)}$	PC2+20	PC3+30	PC4+40	PC5+50
A packet of PC2 is selected	PC1+10	$\frac{(\frac{PC3+PC4}{2}+20)xPS2}{50.Max(SubCredit)}$	PC3+30	PC4+40	PC5+50
A packet of PC3 is selected	PC1+10	PC2+20	$\frac{(\frac{PC4+PC5}{2}+30)xPS3}{50.Max(SubCredit)}$	PC4+40	PC5+50
A packet of PC4 is selected	PC1+10	PC2+20	PC3+30	$\frac{(\frac{PC3+PC5}{2}+40)xPS4}{50.Max(SubCredit)}$	PC5+50
A packet of PC5 is selected	PC1+10	PC2+20	PC3+30	PC4+40	$\frac{(\frac{PC3+PC4}{2}+50)xPS5}{50.Max(SubCredit)}$

PCi : the initial estimation value of ith service class

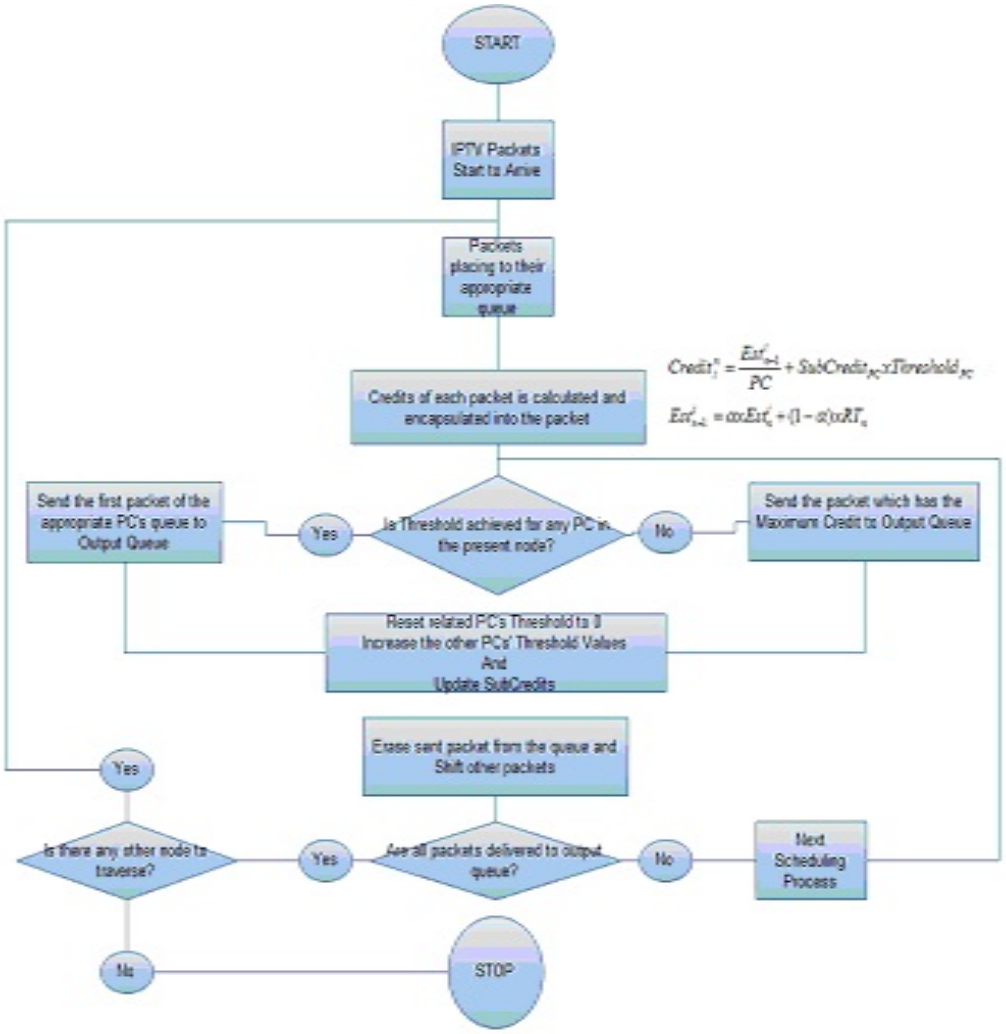


Figure 4. Flowchart of the Proposed Algorithm

V. RESULTS

The proposed algorithm is simulated versus WFQ and Alpha-Beta Virtual Clock algorithms in a framework written in MATLAB, at 100% load. In a flow started by the first service provider, packet distribution for each priority class is selected as:

PC1:2421
PC2: 2902
PC3: 3622
PC4: 8027
PC5: 815

Total duration of the real-time simulation is: 30955 ms

Queue length is chosen as 20 packets (all packets are at same size. Each node has a queue to buffer the arriving packets before forwarding occurs)

Mean packet delay variation and end-to-end mean delay for PC1 packets using one to nine switchers are shown in Figure 5 and 6 respectively.

VI. CONCLUSION

The proposed algorithm successfully staying below 30 ms through even 9 nodes offers a better solution against these two well-known algorithms. The advantage of the proposed algorithm is lied on its relative fairness by letting high priority class's class number in the denominator of the left side of the addition in the credit calculation equation number 9, giving advantage to itself. The other way around, if the right side of the equation is inspected, it is evident that SubCredit value favors the low priority classes. Threshold value itself favors all classes to reach good performance of packet delay variation.

In the future, the delay and the delay variation will be investigated thoroughly. It may be possible to be improved by changing the parameters and the algorithm formula.

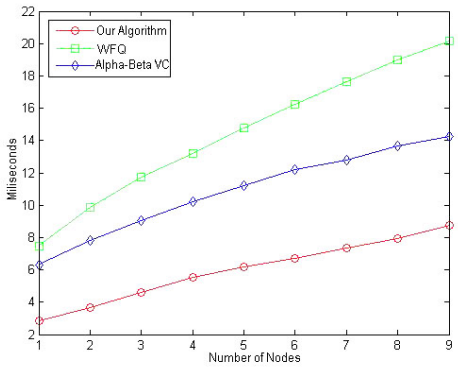


Figure 5. End-to-end mean packet delay variation for 3 algorithms

Furthermore, the performance of the proposed algorithm will be investigated on different wireless network topologies such as mesh networks.

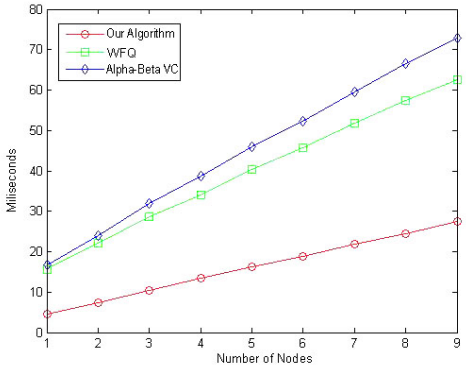


Figure 6. End-to-end mean delay for 3 algorithms

VII. REFERENCES

[1] VoIP Specifications, "http://www.ipdr.org/public/Service_Specifications/2.X/VoIP/VoIP2.5-A.0.pdf"

[2] RFC791, "http:// tools.ietf.org/html/rfc791", Sep. 1981

[3] J.F Kurose, K.W. Ross, Computer Networking:A Top Down Approach Fifth Ed., PEARSON, 2010

[4] A. Demers, S. Keshav, S. Shenker, "Analysis and Simulation of a Fair Queuing Algorithm", Internetworking: research and Experience, Vol 1, No1, 1990,pp 3-26

[5] M. Hosaagrahara, H. Sethu, 2001, Simulation-Based Analysis of a Novel Enhancement of Virtual Clock, *Proceedings of the Applied Telecommunication Symposium*

[6] L., Zhang, 1990, Virtual Clock: A New Traffic Control Algorithm for Packet Switched Networks, *Proceedings of ACM SIGCOMM'90*, Philadelphia

[7] Holt, C.C., "Forecasting seasonals and trends by exponentially weighted moving averages", International Journal of Forecasting Volume 20, Issue 1, January-March 2004, Pages 5-10

 ISBN: 978-1-61208-174-8

Optical CDMA Using Dual Encoding with Different Optical Power

Shusaku Hata, Hiroyuki Yashima
Department of Management Science, Faculty of Engineering
Tokyo University of Science
1-3, Kagurazaka, Shinjuku-ku, Tokyo, Japan
E-mail: hatah@ms.kagu.tus.ac.jp, yashima@ms.kagu.tus.ac.jp

***Abstract*— In this paper, we propose optical code division multiple access (CDMA) systems using dual encoding with different optical power to improve system performance. In the proposed system, each user has two signature sequences, and an information bit is modulated by the sequences with the different power. In the receiver, the received signal is fed into optical hard limiter (OHL). The reflected signal and the transmitted signal of OHL are decoded by corresponding decoders of the sequences, respectively. At each decoder, correlation between received sequence and an assigned sequence is calculated and then information bit is detected. We theoretically derive bit error rate (BER), and show that our proposed system can significantly improve BER.**

Keywords-optical code division multiple access(CDMA); power control; optical orthogonal codes.

I. INTRODUCTION

Recently, Optical Code Division Multiple Access (CDMA) systems attract much attention particularly in the field of fiber optic networks. In optical CDMA systems, each user is assigned a unique signature sequence to allow multiple accesses. Information bit is modulated by On-Off Keying (OOK) signaling, then coded by each encoder, which has own signature sequence [1]-[13]. Signals of all users are coupled and send to optical fiber network. At the receiver, multiplexed signal is fed into a decoder and correlated with its sequence. When the correlation value is larger than decision threshold, received bit is determined as "1", otherwise as "0." Generally, optical CDMA systems suffer from multiple access interferences (MAI) from other simultaneous users. In order to decrease MAI, signature sequences are constructed so that cross correlation is small. Optical Orthogonal Code (OOC) with $\lambda_a = \lambda_c = 1$ is frequently used, and is discussed on its construction methods and its property [1]-[5]. Here, λ_a is the maximum off-peak of auto correlation, and λ_c is the maximum cross correlation. The code length must be long to improve the BER, but it reduces bit rate. Increasing weight of sequence also improve the BER, however, the available number of user decrease. Since OOC with $\lambda_c = 1$ is restricted in the number of sequence for a given code length, OOC with $\lambda_c \geq 1$ is also discussed to increase the number of sequence [6].

In order to improve the BER, various methods have been reported [7]-[9]. In [1], an optical hard limiter (OHL), where nonlinear optical effect is used for limiting power of optical signal, is placed at the front of the receiver to reduce the effect of interference.

On the other hand, optical CDMA systems using different optical power have been also proposed [10]-[12]. In [11], users are divided into some groups, and each user uses the optical power level assigned to own group, then users of different groups can obtain different BER. In conventional researches, different optical power is used to achieve different requirement on BER in multimedia communications or increase the number of user.

In this paper, we propose an optical CDMA using dual encoding with different optical power to improve BER by using two levels of optical power by each user. In the proposed system, each user has two signature sequences, and an information bit is modulated by the two sequences with different power levels. In the receiver, the received signal is passed through OHL, and the reflected signal and the transmitted signal of OHL are decoded by corresponding decoders, respectively. The proposed transmitter and receiver can remove MAI between signals of different power. Since we can increase the weight of sequence without increasing MAI in the same code length, the BER is improved. Moreover, we discuss how to assign two sequences to each user. By using cyclic shifted sequence, we can assign sequences without decreasing the number of user. We also derive BER theoretically, and show the proposed system provides significant performance improvement.

II. SYSTEM DESCRIPTION

Fig. 1 shows a block diagram of the transmitter of the proposed system. There are N simultaneous users and each have two encoders that are assigned different sequences. We use OOC, whose code length and number of weights are F and k, respectively. Here, the two sequences of user i is denoted by $C_{i,1}$ and $C_{i,2}$, respectively. An information bit is coded by both encoder 1 with $C_{i,1}$ and encoder 2 with $C_{i,2}$. The output of the encoder 1 and encoder 2 are modulated by optical power P and 2P, respectively.

 ISBN: 978-1-61208-174-8

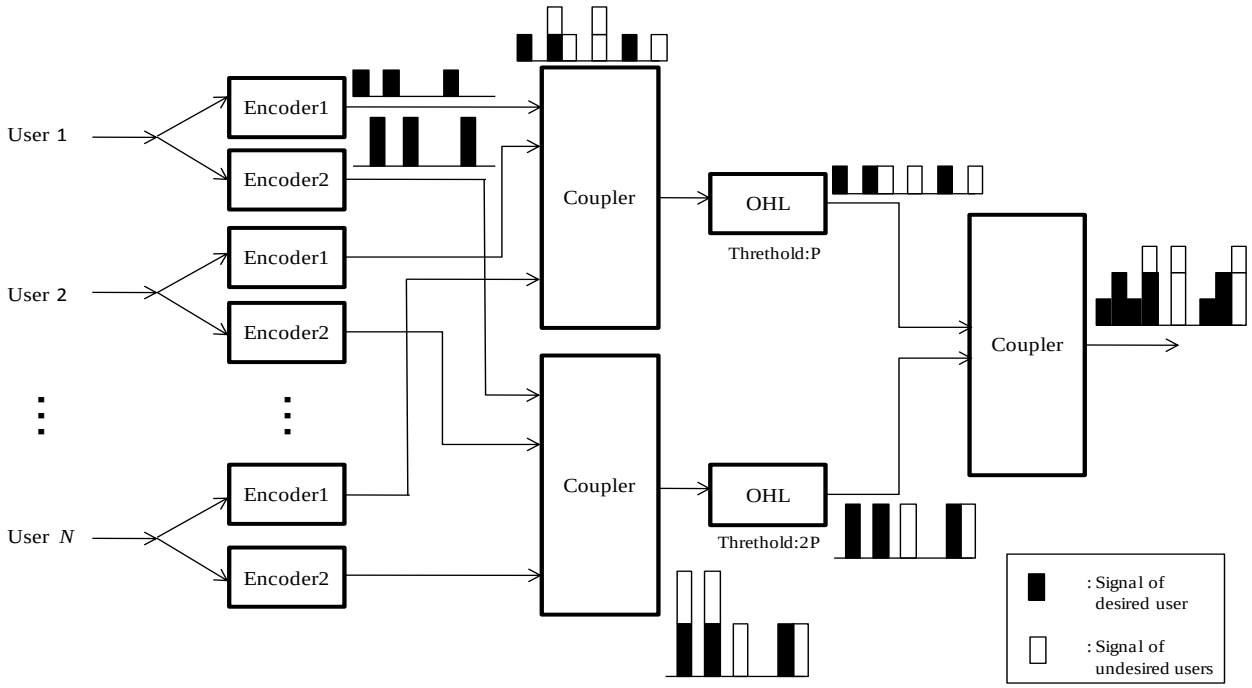


Figure 1. Block diagram of the transmitter of the proposed system.

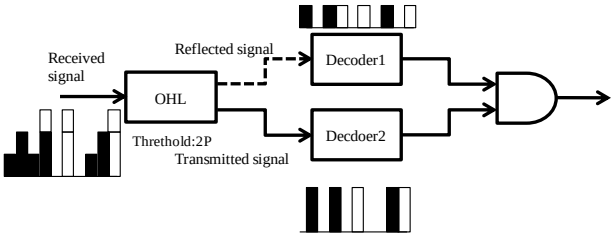


Figure 2. Block diagram of the receiver of the proposed system.

TABLE 1. Input and output characteristics of the OHL

Received signal	Transmitted signal	Reflected signal
0	0	0
P	0	P
2P	2P	0
3P	2P	P

Signals of the same power from all transmitters are once coupled at the coupler, and are fed into OHLs with threshold P and 2P only to transmit the minimum optical signal to reduce interference. The output signal power I_{tr} of the OHL for input x is given by [13]

$$I_{tr}(x) = \begin{cases} th, & x \geq th \\ 0, & 0 \leq x < th \end{cases}, \quad (1)$$

where th is the output threshold of the OHL, and given by P and 2P, respectively. The outputs of two OHLs are coupled by the coupler and the output of the coupler is sent to optical fiber networks. The output of the coupler consists of signals of power P, 2P and 3P, which are single sum of power P and 2P.

Fig. 2 shows a block diagram of the receiver of the proposed system. The received signal is fed into OHL with threshold 2P. The output of OHL consists of reflected signal and transmitted signal. TABLE 1 shows intensity of transmitted signal and reflected signal for the received signal. Thus, encoded signals by the encoder 1 and 2 appear at the output as the transmitted signal and the reflected signal at the output of OHL, respectively. Decoder 1 with the sequence $C_{i,1}$ decodes the sequence coded by encoder 1 modulated with P, and also, decoder 2 with the signature $C_{i,2}$ decodes the sequence coded by encoder 2 modulated with 2P. Namely, two signals modulated by the different power do not interfere with each other. At the each decoder, when the correlation with the corresponding sequence is larger than the threshold k, which is the weight of the sequence, the output of decoders is "1". In the case of both outputs of two decoders are "1", received bit is determined as "1", otherwise as "0".

III. CODE CONSTRUCTION

In this section, we explain construction method of the sequence. $C_{i,1}$ is OOC with $\lambda_a = \lambda_c = 1$ generated by Greedy algorithm [4]. $C_{i,2}$ is given by cyclic shift of $C_{i,1}$ by Si. Then, the element of the sequence is $C_{i,2}(n) = C_{i,1}(n \oplus S_i)$, $(0 < n < F\text{-}1)$, where $C_{i,1}(n)$ denotes nth bit of the sequence of 0 or 1 and "$\oplus$" denotes modulo-F addition. $C_{i,1}$ and $C_{i,2}$ are used in the same users and transmitted synchronously, so we can use cyclic shifted

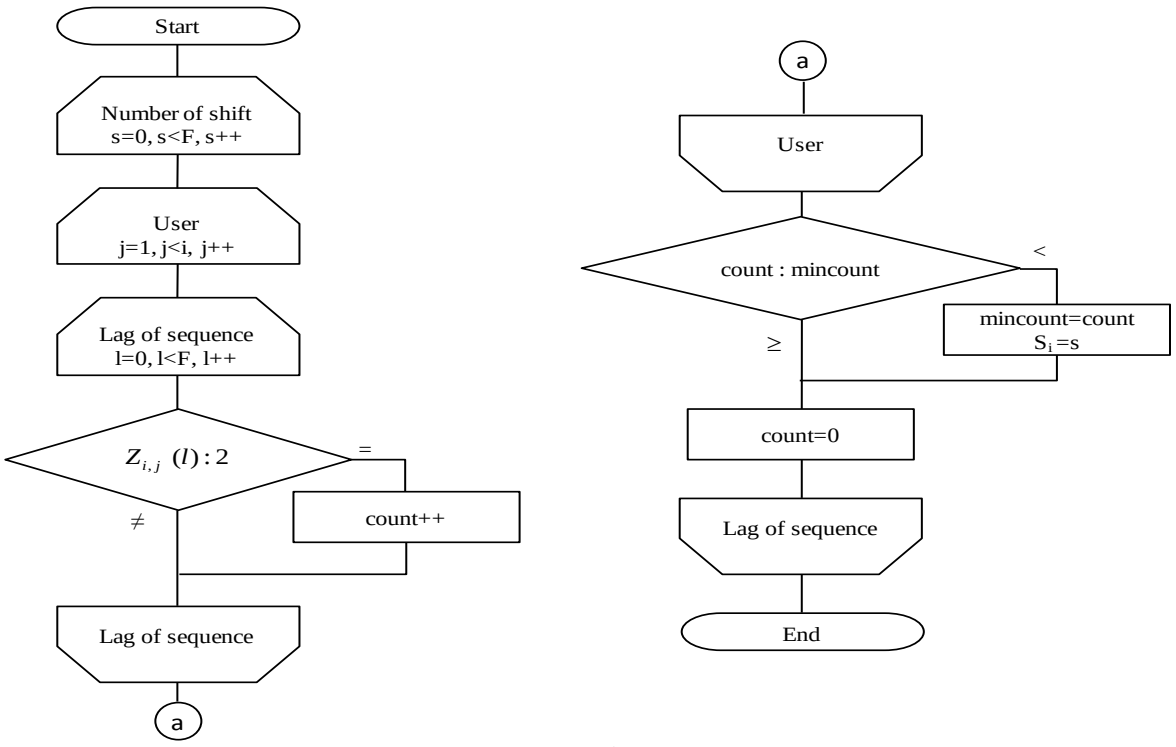


Figure 3. The algorithm of determining *Si*.

TABLE 2. Number of sequence

N, *k*, *F*	20, 3, 127	30, 4, 464	50, 4, 777	50, 5, 1514
Number of sequecne	10	11	22	21

TABLE 3. Event probability of frequency of cross correlation

N, *k*, *F*	20, 3, 127	30, 4, 464	50, 4, 777	50, 5, 1514
$Z_{i,j}(l) = 0$	20741	187969	912688	1793494
$Z_{i,j}(l) = 1$	3389	13871	39137	61156
$Z_{i,j}(l) = 2$	31	49	63	94

sequence code of $C_{i,1}$ as $C_{i,2}$. Thus, although the proposed system requires two sequence codes for each user, we can assign sequence codes without decreasing the number of users. Since there are no interferences between signals of different power, maximum value of cross correlation between sequences of same power is 1. However, we should consider the case that $C_{i,1}$ and $C_{i,2}$ are interfered simultaneously from one user. Thus, we define cross correlation between sequences of user $C_{i,1}$ and $C_{i,2}$ to determine S_i. Considering that $C_{i,1}$ and $C_{i,2}$ are used synchronously, the cross correlation can be expressed as

$$Z_{i,j}(l) = \sum_{n=0}^{F-1} \{C_{i,1}(n) \cdot C_{j,1}(n \oplus l) + C_{i,2}(n) \cdot C_{j,2}(n \oplus l)\} \quad for \quad 0 \le l < F. \tag{2}$$

Note that the maximum value of $Z_{i,j}(l)$ is two, when the correlations of both code coincidently take value one, and in this case the influence of MAI increases. For reducing interferences of undesired users, S_i is given so as to satisfy $Z_{i,j}(l) < 1$ for all l. We show an algorithm of determining S_i in Fig. 3.

We determine S_i using the algorithm shown in Fig. 3, then the number of sequence satisfying $Z_{i,j}(l) \le 1$ is shown in TABLE 2. To accommodate more users, we can also use the sequences with S_i so that the total number of l being $Z_{i,j}(l) = 2$ is as small as possible. TABLE 3 shows frequency of $Z_{i,j}(l) =$ 0, 1 and 2 for all combinations of l and pairs of i and j. It is found that the influence of $Z_{i,j}(l) = 2$ is small, because the event probability of $Z_{i,j}(l) = 2$ in the all combinations is very few. In OOC with F = 127, k = 3, N = 20, the event probability of $Z_{i,j}(l) = 2$ in the all combinations is only 1.3×10^{-3}.

IV. PERFORMANCE ANALYSIS

In this section, we derive the BER of the proposed system. To investigate the basic performance of the proposed system, we do not consider any noise in this analysis and we only consider MAI to corrupt signal. We assume that only sequences satisfying $Z_{i,j}(l) \le 1$ are used. A bit error occurs when desired user sends "0" and both of correlation values of two decoders exceed a decision threshold simultaneously. The probability q that a pulse from an undesired user overlaps with a pulse of the desired user is given by

 ISBN: 978-1-61208-174-8

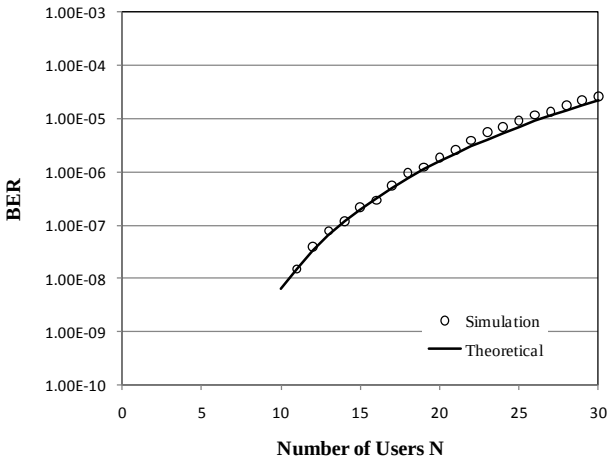


Figure 4. BER versus the number of users N for theoretical and simulation (F=464, k=4).

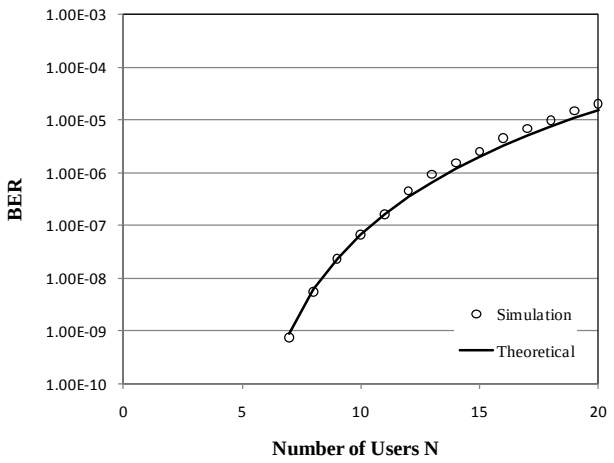


Figure 5. BER versus the number of users N for theoretical and simulation (F=127, k=3).

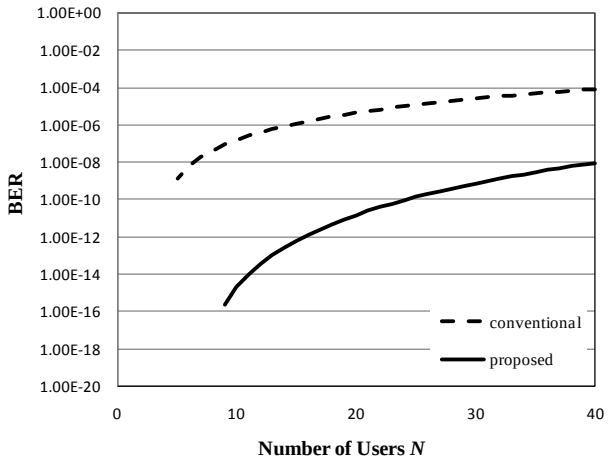


Figure 6. BER versus the number of users N for conventional optical CDMA and our proposed system (F=620, k=4).

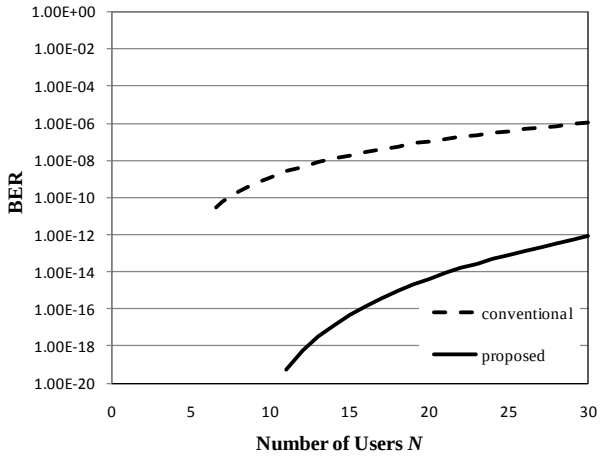


Figure 7. BER versus the number of users N for conventional optical CDMA and our proposed system (F=890, k=5).

$$q = 1 - \frac{k}{2F}. \quad (3)$$

The probability that at least one undesired user overlaps the pulse is given by $1-q^{N-1}$. Since total number of chips is 2*k*, the BER, *Pe*, can be expressed as

$$Pe \le \frac{1}{2}\prod_{m=0}^{2k-1}(1-q^{N-1-m}). \quad (4)$$

V. NUMERICAL RESULTS

Figs. 4 and 5 show the BER given by (4), and simulation. We use OOC with F = 464, k = 4 in Fig. 4, and OOC with F = 127, k = 3 in Fig. 5. In both figures, when N is smaller than number of sequence in TABLE 2, simulation value is almost equal with theoretical value. The BER is slightly degraded when N is large, because (4) is on the assumption that all sequence is satisfy $Z_{i,j}(l) \le 1$. Although combinations of sequences to be $Z_{i,j}(l) = 2$ slightly exist, the probability of $Z_{i,j}(l) = 2$ is very small as in TABLE 3, and the degradation can be ignored.

Fig. 6 shows the BER of the proposed system given by (4) versus the number of users N for F = 620, k = 4 together with conventional system. As the conventional system, we assume the optical CDMA system using OOC having the same parameter and placing OHL in front of receivers [5]. In order to keep BER less than 10^{-8}, the conventional system can accommodate only 5 users, but our proposed system can accommodate 40 users.

Fig. 7 shows BER versus the number of users for proposed system for F = 890, k=5 together with conventional system. In case of N=30, BER of conventional system is 10^{-6}, while BER of the proposed system achieves 10^{-12}.

VI. CONCLUSION

In this paper, we have proposed a new optical CDMA using dual encoding with different optical powers in order

to improve system performance. Each user has two signature sequences, and an information bit is modulated by these sequences with different power levels. By using the proposed transmitter and receiver, there are no interferences between signals of different power. We show construction method of sequence so that the maximum value of cross correlation is small. We also derive BER theoretically. As a result, it is shown that our proposed systems significantly improve the performance of optical CDMA systems.

REFERENCES

[1] J. A. Salehi, "Code division multiple-access techniques in optical fiber networks—part I: fundamental principles," *IEEE Trans. Commun.*, vol. 37, pp. 824–833, Aug. 1989.

[2] C. Argon and R. Ergül, "Optical CDMA via shortened orthogonal codes based on extended sets," *Opt. Commun.*, vol. 116, no. 4–6, pp. 326–330, May 1995.

[3] F. R. A. Chung, J. A. Salehi, and V. K.Wei, "Optical orthogonal codes: Design, analysis, and application," *IEEE Trans. Inf. Theory*, vol. 35, no. 3, pp. 595–604, May 1989.

[4] T. M. S. Khattab and H. M. Alnuweiri, "Optical orthogonal code construction using rejected delays reuse for increasing," *J. Lightw. Technol.* vol. 24, no. 9, pp.3280–3287, Sep. 2006

[5] J. A. Salehi and C. A. Brackett, "Code division multiple-access techniques in optical fiber networks—part II: system performance analysis," *IEEE Trans. Commun.*, vol. 37, pp. 834–842, Aug. 1989.

[6] M. Azizoglu, J. A. Salehi, and Y. Li, "Optical CDMA via temporal codes,"*IEEE Trans. Commun.*, vol. 40, no. 7, pp. 1162–1170, Jul. 1992.

[7] T. Ohtsuki, "Performance analysis of direct-detection optical asynchronous CDMA systems with double optical hard limiters," *J. Lightwave Technol.*, vol. 15, no. 3, pp. 452–457, Mar. 1997.

[8] J.-J. Chen and G.-C. Yang, "CDMA fiber-optic systems with optical hard limiters," *J. Lightwave Technol.*, vol. 19, no. 7, pp. 950–958, Jul. 2001.

[9] J. Y. Kim and H. V. Poor, "Turbo-coded packet transmission for an optical CDMA network", *J. Lightwave Technol.*, vol. 18, pp. 1905–1916, Dec. 2000.

[10] B. M. Ghaffari and J. A. Salehi, "Multiclass, multistage, and multilevel fiber-optic CDMA signaling techniques based on advanced binary optical logic gate elements," *IEEE Trans. Commun.*, vol. 57, no. 5, pp. 1424–1432, May 2009.

[11] H. Yashima and T. Kobayashi, "Optical CDMA with time hopping and power control for multirate networks," *J. Lightwave Technol.*, vol. 21, pp. 695–702, Mar. 2003.

[12] E. Inaty, L. A. Rusch, and P. Fortier, "Multirate optical fast frequency hopping CDMA system using power control," in *Proc. IEEE Global Telecommunications Conf.*, vol. 2, pp. 1221–1227, Nov.2000.

[13] J.-J. Chen and G.-C. Yang, "CDMA fiber-optic systems with optical hard limiters," *J. Lightwave Technol.*, vol. 19, no. 7, pp. 950–958, Jul. 2001.

 ISBN: 978-1-61208-174-8

A Comparison Study on Data Vortex Packet Switched Networks with Redundant Buffers and with Inter-cylinder Paths

Qimin Yang
Harvey Mudd College, Engineering Department
Claremont, California, USA
E-mail: Qimin_yang@hmc.edu

***Abstract*-Optical switching fabric networks become essential systems in high capacity communication and computing systems. This paper focuses on Data Vortex network architecture with two alternative implementations for improved performance. Either a buffer is added within the routing node or inter cylinder paths are provided for enhanced routing performance. Since the extra hardware required for both implementations are the same, the network with better routing performance provides a better solution. The comparison study has demonstrated that networks with inter cylinder paths provide significantly lower latency and better throughput, therefore this approach provides more effective sharing of the routing resources within the network compared with the node buffering implementation. The difference in performance is also shown to be more significant under higher load conditions and for larger networks.**

Keywords-data vortex network; packet switched network; optical; network; buffering.

I. INTRODUCTION

Switching fabric networks are important subsystems in high capacity communication networks and computing systems. Typical space switch uses rich connectivity to handle dynamic traffic coming from a large number of I/O ports while maintaining a high data throughput and small latencies. In high end multi-processor computing applications, the number of I/O ports or processors can reach ~1000 and each could run at tens of Gbit/s data rate, and at the same time low latency (tens or hundreds of μs) must be maintained through such networks. Multistage self-routing network architectures often provide better system scalability with distributed routing nodes incorporating relatively simple routing logics which lead to cost-effective implementation and shorter processing delay. In order to provide higher data throughput, such networks can be implemented using optical fibre or optical switching technology.

Many recent researches have focused on developing optical switching fabric networks and network testbeds [1][2][3][4]. While it is relatively easy to achieve higher transmission bandwidth with Wavelength Division Multiplexing (WDM) within a single fibre, the routing logics as well as the handlings of traffic contention are hard to manage within the optical domain. In particular, Data Vortex packet switched network architecture is developed for the ease of photonics implementation, and such network is highly scalable to support a large number of I/O ports where each runs at high data rate and the network maintains a small routing latency [4][5]. The combination of its high spatial connectivity and an electronic traffic control mechanism among routing nodes lead to bufferless operation and a much simpler routing logic within the nodes. Even though it uses deflection based routing, the special connectivity avoids large deflection penalty and overall probability of deflection; therefore, it is advantageous compared with other commonly used interconnection architectures.

Previous researches have shown that with sufficient network redundancy, Data Vortex network scales to support a large number of I/O ports while achieving high throughput and low latency performance. On the other hand, at extremely high load conditions, and less redundant network conditions, the throughput tends to be limited by traffic backpressure in the deflection based routing. There have been several approaches suggested to enhance the routing performance of the Data Vortex networks, especially for these less ideal operating conditions [6][7][8][9]. In general, these performance enhancement methods require additional routing paths or routing resources, thus detailed cost and performance analysis must be carried out in comparison with the original networks. There is no comparison between different enhancement method, so in this paper, we emphasize such comparison of two methods using buffering and using extra inter-cylinder paths. These two methods are of particular interests because of they share the same cost with reasonable hardware increase in comparison to the original network and their easy implementation. The performance will be compared to each other as well as to the original Data Vortex networks.

The paper is organized as follows: in Section II, the original Data Vortex network architecture will be explained in details. In Section III, two previously proposed enhancement methods, the nodal buffering method as well as inter-cylinder path method are illustrated and compared in details. The routing performance comparison will be provided in Section IV for various network conditions, and the conclusion is given in Section V.

 ISBN: 978-1-61208-174-8

II. DATA VORTEX ARCHITECTURE

The Data Vortex architecture arranges its routing nodes in three dimensional multiple stage configuration as shown in Fig. 1. While the cylindrical levels (c=0 at the outermost cylinder to $c = \log_2 H$ at the innermost cylinder) provide the multiple levels in the routing stages, the angular dimension with repeated connection patterns provides multiple open paths to the destination therefore results in a much smaller latency penalty as deflection occurs. Inter-cylinder paths are not shown for a better view, and they are simply parallel links that maintain the height position of the packets when they propagate from outer to inner cylinders.

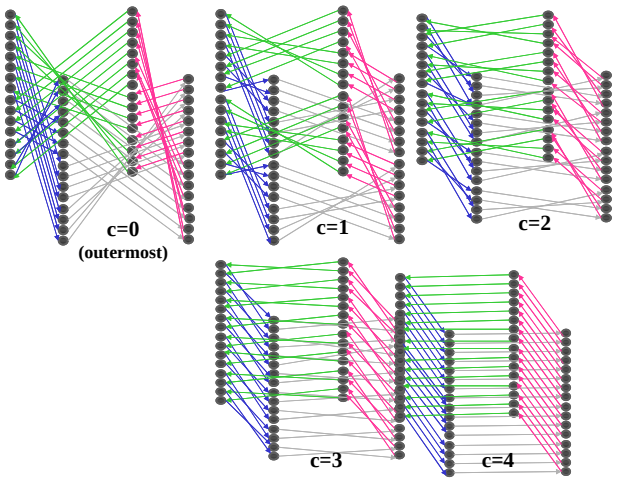


Figure 1. Data Vortex network with Angle=4, Height=16 and Cylinder=5 and its layout of routing node at different cylinders

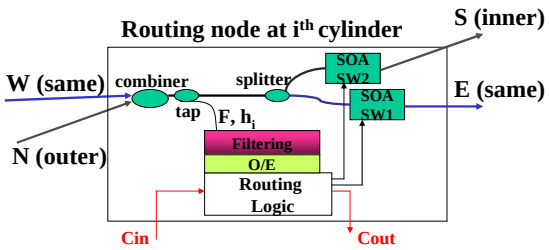


Figure 2. Routing node implementation

Packet's destination height is encoded in binary, and in the physical layer each of these binary bits is modulated onto a distinct wavelength, so that simple passive wavelength filtering can be used to extract and decode the single header bit h_i at the ith cylinder level. This is shown within the node structure in Fig. 2. Only a small amount of optical power will be tapped and used for header decoding purpose so that packets stay in optical domain as they travel through the network. Each node accepts two input paths *West* (*W*) for same cylinder input or *North* (*N*) for outer cylinder input or new injection for the outermost cylinder. Only a single input can be present at the same time, and it is routed to *East* (*E*) to the same cylinder or *South* (*S*) path to the inner cylinder by turning on the proper Semiconductor Optical Amplifier (SOA) switch (SW). Each provides power amplification to balance the power loss at the node due to tap and 3-dB power splitter between *E* and *S* paths.

Data Vortex network combines a traffic control mechanism with deflection routing. Control signals stay in the electronic domain for a simple implementation. As seen in the routing node in Fig. 2, a control signal C_{in} dictates whether *South* path to the inner cylinder is indeed "open" or "blocking". The routing node also generates a proper C_{out} to inform the current cylinder's traffic condition for its outer cylinder node. These distributed control signals allow the neighbouring nodes to coordinate properly for the single packet processing condition for each routing node within the network. Every time a packet is to stay at the current cylinder or to the *East* path, it creates a "blocking" control C_{out} for its outer cylinder contender. In the case the outside traffic receives a "blocking" control, the packet that is intended for *South* path will be deflected by staying on its current outer cylinder and wait for the next open path in two hops. The single packet routing arrangement eliminates optical buffers within the node as the network serves as a virtual buffer when the packet travels on the cylinders.

The last cylinder is typically added for exit buffering purpose so packets are looping around in the last cylinder without changing height positions. As a result, the total number of cylinders is given by $C = \log_2 H + 1$. Note that inter-cylinder paths and intra-cylinder paths are slightly different to allow for the establishment of the control signal. The inner cylinder nodes always make the routing decision slightly earlier than their outer neighbour to inform the traffic condition, so by making the inter-cylinder travel slightly shorter, packets can arrive the same cylinder node at the same time frame regardless of their origins. Detailed traffic control and routing performance have been reported in earlier studies [4-5], and it is shown that Data Vortex network's overall routing performance is very reasonable even as the network scales up to thousands of I/O ports. In addition, many physical layer limitations have also been studied and addressed in these studies.

III. MODIFIED DATA VORTEX IMPLEMENTATION

As Data Vortex networks run at high load conditions or less redundant configurations, i.e. more input angles are attached to I/O ports for incoming traffic, the traffic backpressure could build up between the cylinders, so it takes longer to go through the network and the overall throughput also drops significantly. Due to the physical degradation of the optical signal through each node, reduction of the latency is highly desired as well as maintaining the high data throughput. There have been several approaches suggested to enhance the routing performance of the Data Vortex networks with additional hardware. The detailed analysis of cost and

 ISBN: 978-1-61208-174-8

performance comparison to the original network has been reported earlier in these studies. This paper emphasizes comparison of the two methods using buffering and using extra inter-cylinder paths respectively. Because the increase in hardware in the two methods is reasonably low and the costs are close to each other, a comparison of the two methods under the same network operation conditions will be of great interests.

A. *Buffering*

The original Data Vortex network is attractive for its bufferless operation. However, for enhanced performance, separate buffers can be added within the routing nodes with slightly more complicated routing logic. This allows for less deflection when the packets wait in the buffer instead of circulating around the cylinders. As shown in Fig. 3, an additional switch (SW3) is used to provide the third routing path to the buffer unit. However, to inform the presence of the traffic within the buffer path so that other traffic is not allowed to enter the node during the same time slot, the buffer unit must have at least two slot delays. Even though previous studies also show that two simultaneous packets routing scheme are possible and it provides much better performances, the required hardware is significantly more [6]. So this study only focuses on the buffer implementation that maintains a single packet routing principle through a two hop delay buffer unit.

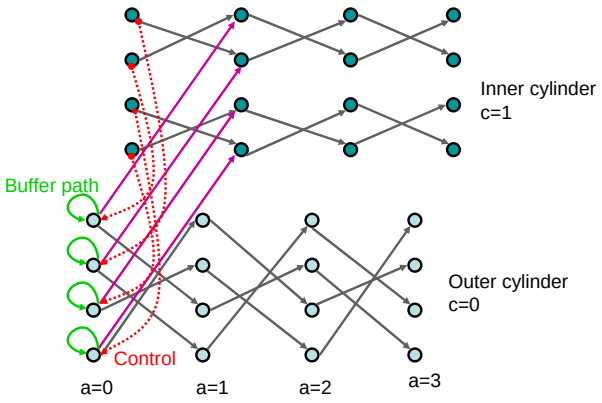


Figure 3. Data Vortex network with buffers within node shown at a=0.

The detailed node implementation is shown in Fig. 4. This implementation requires the network to have roughly 50% more hardware in number of switches and in routing paths compared to that in the original network. There is slight modification of routing logic within the node. If a packet is not able to reach S output, it will be sent to the buffer unit and be routed within the same node in two time slots. If the buffer packet is entering the node, it will not accept W or N input from the other nodes to maintain the single packet routing rule. Overall the priority is given to the packet within the buffer, and if there is no buffer traffic, then the same cylinder traffic gets the priority over the outer cylinder traffic as that in the original Data Vortex network. The additional control signal has to inform both the same cylinder neighbour and the outer cylinder neighbour to avoid contention.

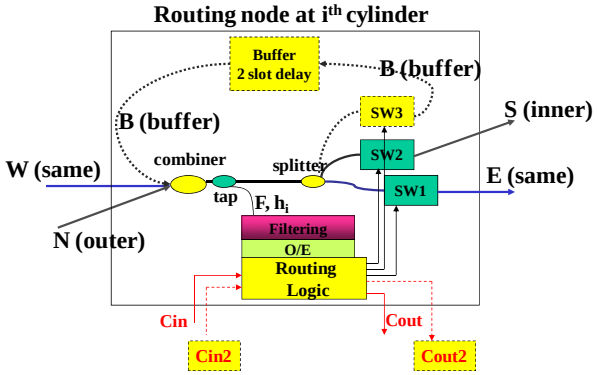


Figure 4. Routing node with buffer implementation: a 2-slot delay for buffer path is necessary to setup the control signal on time and additional controls C_{out2} are used to inform the state of buffer

B. *Inter cylinder paths*

In addition to buffering, there are also proposals for additional routing paths between the cylinders because these paths are critical to channel the traffic through the cylinders as fast as possible [7-8]. Lack of such routing resource would result in deflection thus the building up of the traffic backpressure. Here we allow the packet to be routed to a secondary inter-cylinder path S_2 output if there is no other traffic (from regular West and North path) entering that same node. We will only focus on this inter cylinder paths implementation, which is the same as that reported in [7] because a separate study has shown very similar results for implementations in [7] and [8] under various traffic and network conditions. An additional injection path is provided at each of the injection ports so that packets are less likely to be blocked by the traffic that is already circulating around the outermost cylinder. The setup of extra links and controls are shown in Fig. 5, and a detailed node implementation is shown in Fig. 6. The single packet routing rule is maintained for simplicity and an additional switch (SOA-SW_3) is added to provide the third routing path as shown in the routing node. In this case, an additional control is also necessary to inform the same cylinder traffic so that the traffic that goes to the regular S_1 output obtains the higher priority over the traffic that requires the S_2 output path. The height choice for the secondary inter-cylinder path must maintain the same binary bits for all the previous cylinders as those in the primary inter-cylinder path's height. As an example, for routing at c^{th} cylinder, the secondary height or the height of its S_2 path node can simply inverts the $(c+1)^{th}$ header bit of the current height.

 ISBN: 978-1-61208-174-8

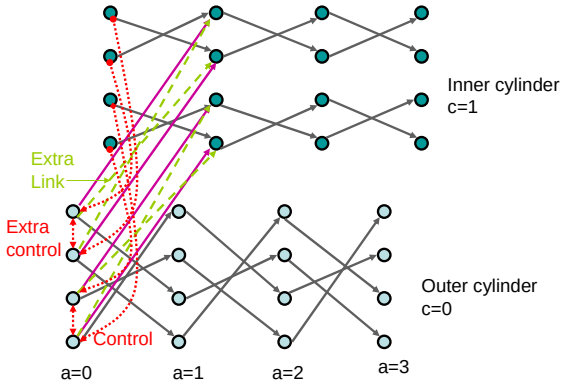


Figure 5. Additional inter cylinder path in Data Vortex network with required extra control

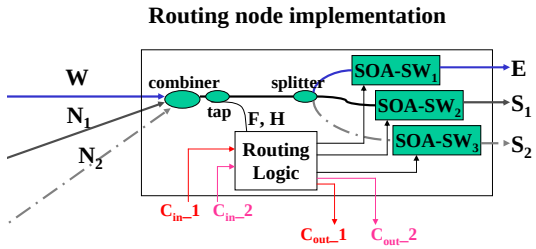


Figure 6. Modified routing node

The inter cylinder paths implementation requires about 50% more hardware in the number of switches and number of routing paths; therefore, it has comparable cost to the buffering implementation.

III. PERFORMANCE EVALUATION

In order to compare the effect of node buffering and the extra inter-cylindrical path for routing, a simulation in C/C++ is written to study the routing performance such as latency and data throughput. The networks under comparison are of the same size and same load conditions. Only random and uniform traffic pattern is studied for the purpose. Latency is measured as the average latency of packets that reach the destination for a long period of simulation time after the initial injection transient period. The network throughput is measured as the successful injection rate at the input port as previously reported. Once the packet reaches the correct target height, it exits the network immediately, therefore no angular resolution is considered in the simulation study. The networks that incorporate the two enhancement methods are compared where A=5, C=9 and H=256 as an example. Fig. 7 and Fig. 8 shows the results of the delay and throughput performance where two redundant conditions are considered. Because both methods are for performance enhancement purpose when the Data Vortex network is heavily loaded or under less redundant operation, we choose A_{in}=3 and A_{in}=5 for the study. Keep in mind, for the buffer implementation, each buffer stay requires a two packet slots delay even though the number of node hop is one.

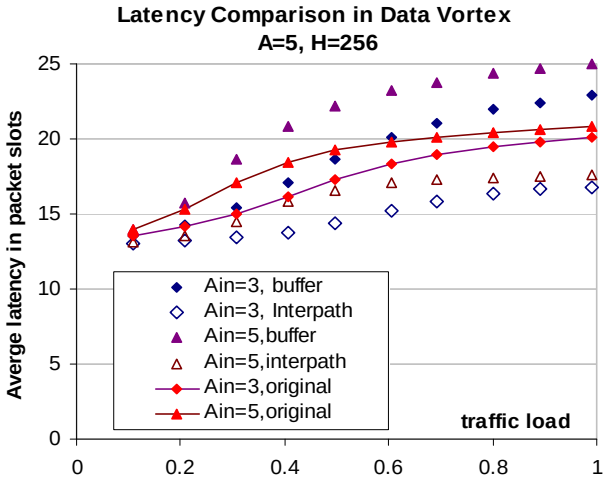


Figure 7. Latency comparison under various traffic load and redundant conditions

For comparison purpose, the original network performances are shown with the solid lines. From these results, we can see that the inter-cylinder paths provide a smaller latency in general compared to that with an additional buffer within the routing node. In fact, the latency is worse for the case of node buffering compared to the original network especially at higher load conditions and less redundant network conditions. This is mainly because of the two hop delay requirement on the buffer path for timing issue, which does not provide efficient reduction of latency even though the deflection events are reduced by keeping the packet at the open path to inner cylinder. The traffic backpressure remains significant because as the buffer packet re-enters the node for routing, there is no acceptance of additional traffic from neighbouring nodes. On the other hand, the inter-cylinder paths provide a better shared configuration of the redundant resource because when such resource is available, the additional routing paths always push more traffic through towards the inner cylinders. As a result, the traffic backpressure has been effectively reduced.

A similar performance edge in inter-cylinder path implementation is also reflected in the data throughput comparison as shown in Fig. 8. In this rather busy network conditions, the buffer implementation has little improvement compared with the original networks, while the inter-cylinder path approach provides much more visible improvement. The results follows very similar trend for the two different redundant conditions. In reference [6], more detailed cost performance study is provided on this buffer implementation in comparison to the original network and a two input buffer scheme which uses much more hardware. Similar conclusion is provided that the overall the

 ISBN: 978-1-61208-174-8

improvement in throughput and latency in this buffer scheme is rather limited and this implementation is only attractive for certain network conditions. In our comparison for more heavily loaded network conditions, the results have proved that the buffered implementation could even degrade the overall network performance once the system reaches saturation in load. On the other hand, the inter-path approach maintains the performance enhancement in both throughput and latency, and it provides a much more attractive implementation for the same amount of hardware cost. Such performance enhancement also scales to very demanding network conditions.

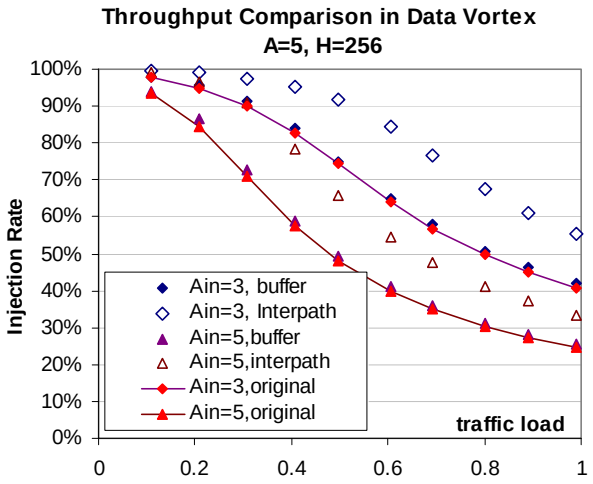


Figure 8. Throughput comparison under various load and redundant conditions

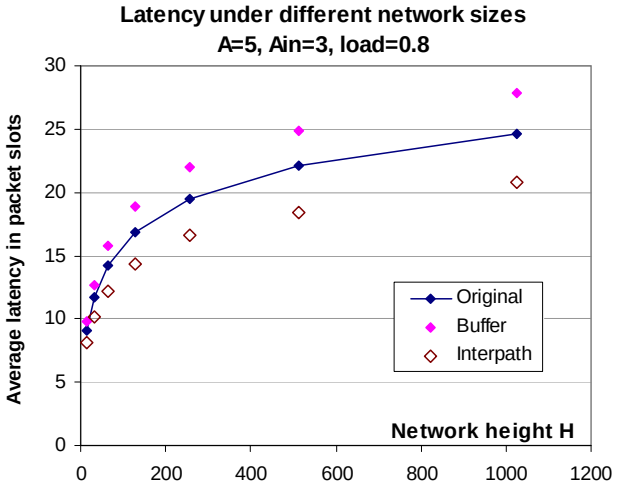


Figure 9. Latency performance comparison at different network sizes

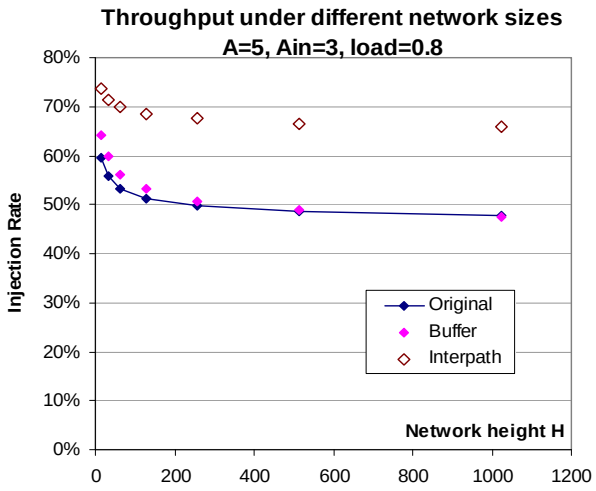


Figure 10. Throughput performance comparison at different network sizes

In order to study the scalability of such performance comparison, networks of different heights are also compared in the study. In Fig. 9 and Fig. 10, networks with A=5 and injection angles of A_{in}=3 with buffer and with inter-cylinder paths are compared and the original Data Vortex network performances are also shown as references. All cases shown are with a medium to high traffic load of 0.8. It is seen that for all network sizes, the inter-path cylinder approach provides better performance over the buffer implementation, and there is especially significant difference for larger networks.

IV. CONCLUSIONS AND FUTURE WORKS

This study focuses on two different modification schemes for Data Vortex networks improvement. With similar hardware cost and complexity, the inter-cylinder paths provide better configuration of shared redundant routing resource. Such arrangement effectively reduces the traffic backpressure present in the original network at high load network conditions, and it provides much better performance in latency and data throughput compared with the modified network with buffering implementation. Future developments in switching device integration are important and relevant for this investigation, and allow us to further quantify the benefits of different modification schemes. For future development in novel enhancement methods, researchers should consider not only the hardware cost but also the routing performance in both delay and throughput especially for less ideal network operation conditions so that a fair and effective evaluation of the proposal can be achieved.

 ISBN: 978-1-61208-174-8

REFERENCES

[1] Keren Bergman, *Optical Fiber Telecommunications, B: Systems and Networks (Editor Ivan P. Kaminow, Tingye Li, Alan E. Willner), Chapter 19*, "Optical interconnection networks in advanced computing systems", Academic Press.

[2] Ronald Luijten, Cyriel Minkenberg, Roe Hemenway, Michael Sauer, and Richard Grzybowski, "Viable opto-electronic HPC interconnect fabric", *Proceedings of the 2005 ACM/IEEE SuperComputing*, Seattle, pp. 18-18, November 2005.

[3] Roberto Gaudino, Guido A. Gavilanes Castilo, Fabio Neri, and Jorge M. Finochietto, "Can Simple Optical Switching Fabrics Scale to Terabit per Second Switch Capacities?", *Journal of Optical Communication Networks*, Vol.1, No.3, pp. B56-B68, August 2009.

[4] Odile Liboiron-Ladouceur, Assaf Shcham, Benjamin A. Small, Benjamin G.Lee, Howard Wang, Caroline P. Lai, Aleksandr Biberman, and Keren Bergman, " The Data Vortex Optical Packet Switched Interconnection Network", *Journal of Lightwave Technology*, Vol. 26, No. 13, pp. 1777-1789, July 2008.

[5] Cory Hawkins, Benjamin A. Small, D.Scott Wills, and Keren Bergman, "The Data Vortex, an All Optical Path Multicomputer Interconnection Network", *IEEE Transactions on Parallel and Distributed Systems*, Vol. 18, Issue 3, pp. 409-420, March 2007.

[6] Assaf Shacham and Keren Bergman, "On contention resolution in the data vortex optical interconnection networks", *Journal of Optical Networking*, Vol.6, pp. 777-788, 2007.

[7] Qimin Yang, "Enhanced control and routing paths in data vortex interconnection networks", *Journal of Optical Networking*, Vol. 6, No.12, pp. 1314-1322, December 2007.

[8] Neha Sharma, D. Chadha, and Vinod Chandra, "The augmented data vortex switch fabric: An all-optical packet switched interconnection network with enhanced fault tolerance", *Optical Switching and Networking*, Elsevier, Vol. 4**,** pp. 92-105, 2007.

[9] Qimin Yang, "Performance Evaluation of *k*-ary Data Vortex Networks with Bufferless and Buffered Routing Nodes", *Asia Photonics and Communication Conference (ACP) 2009*, pp. 1-2, Shanghai, November 2009.

 ISBN: 978-1-61208-174-8

Multicasting over OBS WDM Networks

Pınar Kırcı
Computer Engineering Department
Istanbul University
Istanbul,Turkey
pkirci@istanbul.edu.tr

A. Halim Zaim
Computer Engineering Department
Ticaret University
Istanbul, Turkey
azaim@iticu.edu.tr

***Abstract*—In this study, we present a new protocol structure for multicasting on OBS networks. We investigate the performance on Just Enough Time (JET) and Just In Time (JIT) between multicast and unicast traffic. We examine behaviors of data and control planes of the network. We also study how to add and remove a client or a node from a constructed multicast tree. The results show that our proposed multicast protocol structure produces low multicast traffic join request drop rates.**

Keywords-OBS; multicast; JIT; JET; WDM network.

I. INTRODUCTION

The number of Internet users continuously increases. Internet traffic growth rates exponentially ascend; therefore, there is a great bandwidth demand emerges in backbone networks. The main reason of the growing number of Internet users is the recent advances in networking. These advances are for instance, video conferencing, distributed games, HDTV, interactive distance learning and many more multimedia real time applications. These are multi destination communication-based, most popular network applications. The bandwidth need of these multicast applications may be met by optical networks and dividing fiber into numerous channels with WDM technology. In multicasting, the messages and data are transmitted from one or more sources to a set of destinations in a multicast group of a WDM network over a multicast tree. To construct a tree, a route should be decided from source to all of the related destinations, a wavelength should be assigned to the links and also QoS should be provided. In WDM networks, for multicast trees, delay and cost are the most important QoS factors for network efficiency. In optical networks, Optical Burst Switching (OBS) is a good solution for data transmission since OBS combines the advantages of Optical Packet Switching (OPS) and Optical Circuit Switching (OCS). The first generation of optical networks was based on OCS. The most important detail of the switching in OCS is the construction of a lightpath between the source and destination before data transmission. By the way, OPS is presented as a good alternative for OCS since OPS is more efficient for dynamic and bursty traffic transmissions. However, today OBS seems as the best solution for data transmission since OBS evolves the performance of WDM on optical networks with bursty traffic [1], [2].

In this paper, we study a multicasting protocol for OBS networks. In the network, communication of clients and the source is performed by join-request messages. We concentrate on how to avoid gathering of traffic crowd in definite parts of the network by using thresholds at intermediate switches with minimum join-request message loss rates. Most of the multicasting protocols depend on a source initiated structure. This type of structures produces excessive message round trip times because of using acknowledgement messages. In our study, a leaf initiated structure is considered. In the protocol, a join-request message notifies the source about the request, composes the path and informs the switches without producing extreme message round-trip times. In this protocol, the main aim is to achieve the routing and bandwidth allocation of the network by using less number of messages and message round-trip times.

The rest of the paper is organized as follows. In Section II, a brief overview of all optical networks, OBS and multicasting is presented. In Section III, the proposed protocol structure is introduced and the numerical results are given. Finally, the paper is concluded in Section IV.

II. MULTICASTING AND OBS

All optical networks with WDM transmissions include many optical cross connects (OXCs). These OXCs connect client networks over lightpaths or light trees. An optical signal arrives at an OXC over an input fiber wavelength. Then it is switched to the same wavelength over an output fiber. However, the arriving signal can be switched to a different wavelength over the output fiber by the help of the converters. Optical switching has many advantages like protocol transparency and less power consumption rates [4].

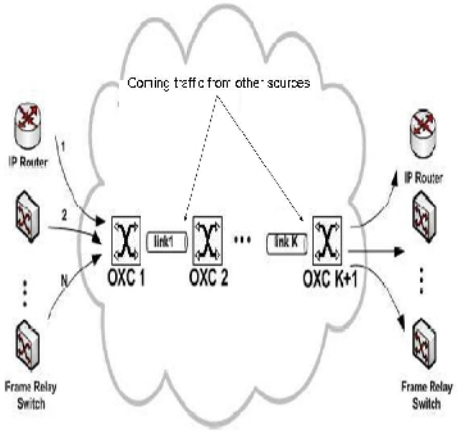


Figure 1. OBS network structure [7].

 ISBN: 978-1-61208-174-8

OBS is presented as a solution for great bandwidth demand on Internet traffic. The network structure with OXCs is presented in Fig. 1. The data packets are assembled into bursts at the ingress node and transmitted by fiber optics on the network in optical form. In the network, to arrange the reservations for the coming burst, a control packet is sent before the burst. Furthermore, in OBS networks there is a definite time period between the control packet and the burst which is called as offset. Separated data and control channels are the most remarkable specialty of OBS networks. Data is transmitted in optical form with optical switching but control packets are transmitted in electronic form with electronic switching in OBS networks [3], [6].

Other than OBS, another solution for increasing bandwidth demand is multicasting. There are many multicast applications in the literature according to the multipoint destinations. Document distribution and on demand video distributions are one to many type applications, which are sent from only one source to many destinations. Many to one applications are sent from more than one sources to a destination like group polling and resource discovery. In many to many, there are more than one sources and many destinations as in multimedia conferencing and distributed simulations [5].

III. THE MULTICAST OPTICAL NETWORK DESIGN

We consider an optical WDM network with 14 node NSFNET topology. The constructed network model is composed of a source, an ingress switch, intermediate switches, edge switches and clients. The nodes of the networks are connected to each other by fiber links that carry two wavelengths with 10 Gb/s capacity. In the network, all of the nodes are multicast capable. In our study there is a source and multiple destinations as clients. These clients access the backbone network by edge switches. The multicast sessions have different amount of multicast traffic and different number of destinations. As distinct from the source initiated multicast tree structures, we consider to construct leaf initiated multicast trees for each of the multicast sessions.

The multicast session begins with the video context announcement with broadcasting to the clients then the join requests are considered. If a client is interested in one of the video contexts' video then it sends a join request to the connected edge switch. The edge switch sends this request to the closest intermediate switch. The intermediate switch controls both its timetable and threshold. If the threshold value of the intermediate switch is not exceeding the predetermined value, then the timetable is checked if the intermediate switch will be available during the video transmission time. If both of the conditions are ensured, join request is accepted and transmitted to the next node but if one of the conditions are not ensured, the join request is rejected. Then the edge node sends a re_join_request message to another closest intermediate switch. If none of the intermediate switches accepts the join request, in that case the join request is dropped. The threshold control takes place at the intermediate switches, which are directly connected with edge switches for preventing potential traffic collisions. Besides, the clients whose join requests are accepted have to send their keep alive messages on specific time intervals to the source to notify that they are still alive and when the video transmission starts, the bursts are sent from source to the related clients.

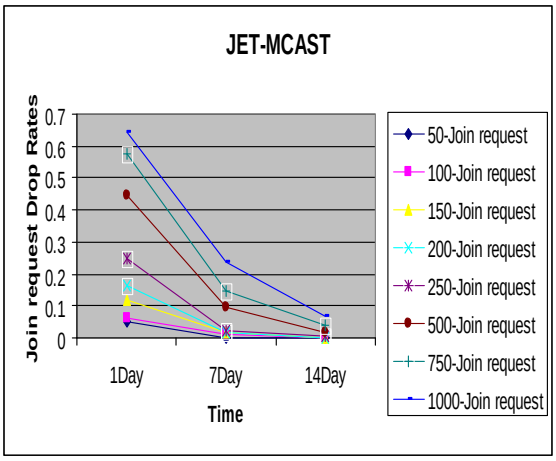


Figure 2. The multicasting on JET reservation protocol.

The main function of our protocol is multicasting the bursts; for comparison, we also designed a unicast data traffic. Fig. 2 and Fig. 3 present the results of our simulation.

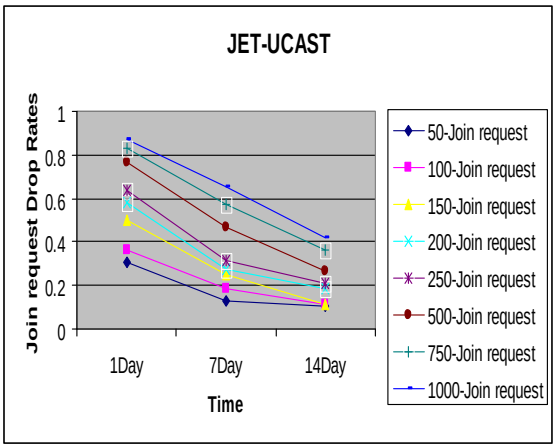


Figure 3. The unicasting on JET reservation protocol.

 ISBN: 978-1-61208-174-8

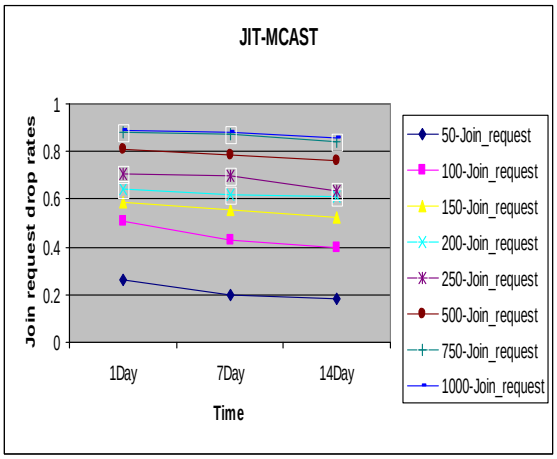


Figure 4. The multicasting on JIT reservation protocol

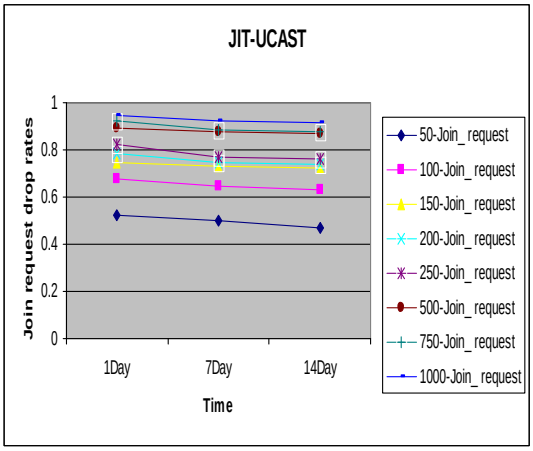


Figure 5. The unicasting on JIT reservation protocol.

The join request drop rates in JET with multicast gives better results. The drop rates change according to the time periods and they decrease as the time periods increase.

Fig. 4 and Fig. 5 show the join request drop rates in JIT reservation protocol. As the time period increases from one day to fourteen days according to the increasing number of join requests, the join request drop rates change with respect to the variable traffic.

IV. CONCLUSION

In this paper, a promising protocol structure for multicasting is presented. We considered the drop rates of join requests according to definite time periods. We investigated the protocol performance on JET and JIT with multicasting and unicasting. Our performance results show that multicasting provides better drop rates than unicasting in both of the reservation protocols. Furthermore, when we compare the simulation results, the protocol structure gives best values on JET with the least drop rates.

REFERENCES

[1] J.P. Jue, W.H. Yang, Y.C. Kım and Q. Zhang, “Optical packet and burst switched networks:a review,” IET Communications, vol. 3, Issue 3, 334-352, 2009.

[2] L. Xu, H.G. Perros and G. Rouskas, “Techniques for optical packet switching and optical burst switching,” IEEE Communications Magazine, 0163-6804-01, 2001.

[3] J. Ramamirtham and J. Turner, “Time sliced optical burst switching,” IEEE Infocom 2003, 0-7803-7753-2.

[4] Y. Xin and G.N. Rouskas, “Multicast routing under optical layer constraints,” IEEE Infocom, 0-7803-8356-7.

[5] A.E. Kamal, “Algorithms for multicast traffic grooming in WDM mesh networks,” IEEE Communications Magazine, 0163-6804—06.

[6] R.C. Rajesh and V.M. Vokkarane, “Dynamic load balanced manycasting over optical burst switched (OBS) networks,” 2009, IEEE 978-1-55752-865-0.

[7] T. Battestilli and H. Perros, “A Performance study of an optical burst switched networks with dynamic simultaneous link possesion,” Computer Networks, vol. 50, Issue 2, pp. 219-236, Optical Networks.

 ISBN: 978-1-61208-174-8

Performance Study of Interconnected Metro Ring Networks

Van T. Nguyen, Tülin Atmaca, Glenda Gonzalez,
Lab. CNRS/Samovar
Institut Telecom/Telecom SudParis
Evry – France
e-mail: nvt2302@gmail.com,{tulin.atmaca,
Glenda.Gonzalez}@it-sudparis.eu

Joel Rodrigues
Instituto de Telecomunicações
University of Beira Interior
Covilhã - Portugal
e-mail: joelr@ieee.org

Abstract— **Metropolitan ring networks are usually used to connect the high speed backbone networks with the high speed access networks. Until now, metropolitan networks and access networks are gained much attention of researchers just as in separate direction. In this work, we study an interconnected Multi-Ring Network (MRN) architecture in which a Metropolitan Access (MA) Ring is interconnected by a Metropolitan Core (MC) Ring via a Hub Node who is in charge of the synchronization between them. The synchronization in this architecture is the major problem. To solve this problem, we propose a new mechanism called Common-Used Timer Mechanism (CUTM) inspired from CoS-Upgrade Mechanism (CUM) to create well filled optical packets in the hub. CUTM is developed and also integrated as a module to the software Network Simulator 2 (NS2), to simulate the behavior of the MRN considered. We compare the performance of this mechanism with the opportunistic one. The results have shown that, compared to existing solutions, the CUTM enhances the network throughput, optimizes the use of resources, and also offers a solution to the synchronization problem.**

Keywords-Interconnected Ring Networks; Synchronization; Performance; Simulation.

I. INTRODUCTION

Optical technology is being developed more and more in all levels of networks. It led the innovation of broadband networks. Passive Optical Networks (PON) have attracted much attention of researchers because it is an excellent solution for low cost broadband services. The next generation of Metropolitan Area Network (MAN) requires flexible, scalable and manageable architectures to provide different type of services to their customers at the access or backbone networks. With passive devices on the transmission line signal, it is easy to build and maintain the PON. So PON becomes the first choice for metropolitan area network.

Metropolitan ring networks are generally used to connect the high speed backbone networks with the high speed access networks. The metro rings can be interconnected transparently through a single access node (Hub node) or multiple access nodes. Current metro networks are typically SONET/SDH-over-WDM rings which carry the huge amount of bursty data traffic. The metro core and regional networks are normally both 2-fiber rings. A fiber failure in a metro access ring does not affect the traffic in the core and other access rings. The network thus becomes more reliable. Dual Bus Optical Ring Network (DBORN) has been proposed as one of the first passive architecture, known for the metropolitan networks. However, new transparent optical network providing packet-level granularity architectures have been proposed and studied called ECOFRAME. This architecture is studied and a prototype is developed as for next generation of MAN in the ANR/ECOFRAME (France) project. Its important characteristic is that it can be used as MA and/or MC Network. Until now, metropolitan networks and access networks are gained much attention of researchers in separate direction. Recently, the end-to-end metropolitan performance of a multi-ring architecture (in which MANs are interconnected by a metropolitan core network) has been investigated [1]. We consider a multi-ring architecture, in which MANs are interconnected by MC Network. The interconnection of MC and MA networks is made via Hub node that is in charge of the synchronization of the two ring networks. Other functions to be operated by the hub are similar to those of access node.

Some works have presented new architectures to integrate in a transparent way metro-access and metro-core ring networks [2]. Other works [3] have studied the design and the development of new devices to interconnect Metro Access and Metro Core Ring networks. However, the synchronization problem between the networks has been neglected and a major research opportunity exists in this sense. Several mechanisms to create optical packets that improve the performance of the multi-ring network have been proposed in the literature. In this paper, we present a new mechanism CUTM to create optical packets well filled and we compare the results obtained with the well known "opportunistic" mechanism in terms of waiting time, end to end delay, filling ratio, and jitter.

The rest of this paper is organized as follows. In Section II, selected Metro Access and Metro Core architectures have been summarized. In Section III the studied architecture is presented. In Section IV, our proposed mechanism CUTM is introduced. In Section V, our simulation scenario is described and in Section VI the simulation results are presented. Finally, we conclude our work.

 ISBN: 978-1-61208-174-8

II. METRO ACCESS AND METRO CORE ARCHITECTURES

According to the physical distribution of the network components, the bus, star and ring topologies can be implemented. Ring topologies have been widely adopted and studied for MAN because it is easy to construct and maintain with low cost, and bidirectional rings inherently provide fast restoration. Statistical multiplexing of data traffic flowing from different nodes over the shared medium provides efficient utilization of optical fibers. Some optical MANs in ring topologies are: Resilient Packet Ring (RPR), DBORN, ECOFRAME. DBORN [4][5] uses a double bus WDM rings topology with spectral separation and it functions in the asynchronous mode. This topology consists of two unidirectional buses: upstream and downstream. In the upstream bus, access nodes share a common transmission medium for carrying their traffic to a centralized node (hub) while the downstream bus carries traffic from Hub node to all access nodes. For the cost-effective solution, each ring node possesses passive components; lead in to the fact that they can not drop any transit packets in upstream line. Although of its simplicity, this architecture has several drawbacks as positional priority, fairness issues and bandwidth fragmentation. In our work we use synchronous DBORN version (slotted ring, fixed size packet) as a MA ring. Compared to the previously mentioned architecture, ECOFRAME [6] pays special attention to the deployment of optical technologies "low cost" to ensure good network performance while remaining competitive with electronic technology in terms of cost, service transparency and modularity. ECOFRAME ring uses fixed optical packet size, separate data and control channels and it can be used as MA and MC ring. In our work, ECOFRAME is used as a MC network.

III. STUDIED ARCHITECTURE

In this section, we present an architecture, which is composed of two segments: Access Network and Core Network (Fig. 1). For the access network synchronous DBORN architecture is considered and for core network ECOFRAME architecture. The interconnection is made via a hub node.

We distinguish two traffic flows: 1) the traffic flowing from the access network to the core network through the hub, and 2) the traffic flow circulating in the core network. In an access node of MA, the electronic packets are encapsulated in optical packets and transported through the hub. In the hub O/E/O converter is used to build new optical packets fill well coming from different nodes and going to same destination. These packets are stored in the queue in the hub. Hub architecture is presented in Fig; 2. It is composed of two parts: electronic part and optical part. In the electronic part, the packets are converted and stored in the buffer before processing. In optical part, it is used FDL.

One of the roles of Hub is to create new optical packets well filled. The creation of new optical packets can be made using three mechanisms: 1) mutual combination (electronic packets coming from different access nodes can be combined

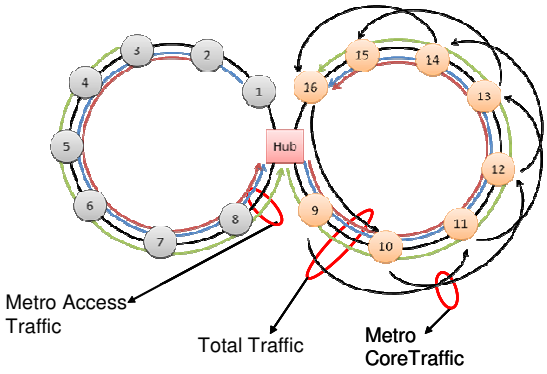


Figure 1. Networks Interconnection

together), 2) local combination (combined with local electronic packets of the hub) and 3) total combination (two combinations mentioned), totally according to class of service. According to the access control mechanism used in the Hub, the optical packets coming from MA can be placed directly in the optical buffers (to be ready to be routed through the Hub) or they are converted in electronic packets by O/E converter and wait in the electronic buffer corresponding to their CoS until timer is expired and new timer is reset. New optical packets are created using a packet creation mechanism and are sent to the core network.

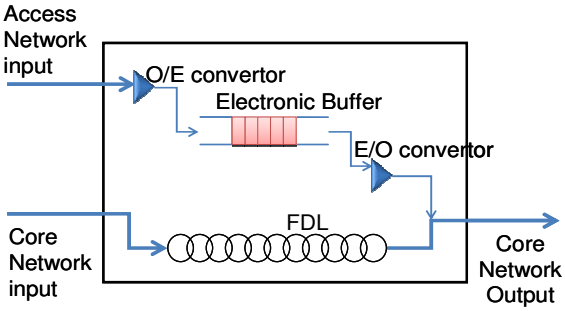


Figure 2. Architecture of Hub

We present a new packet creation mechanism in the Section IV. At the Hub node, the packets in transit in the core network have higher priority than traffic of access network; therefore, the E/O converter is performed if there is no packet in the optical FDL. The associated times to the creation process are specified in Fig. 3. One of the problems of interconnection between the rings is the synchronization of timer between them. Each ring is already synchronized but each one has different size of slot time and optical packet. Therefore, it is needed to synchronize data inputs and outputs at the Hub.

 ISBN: 978-1-61208-174-8

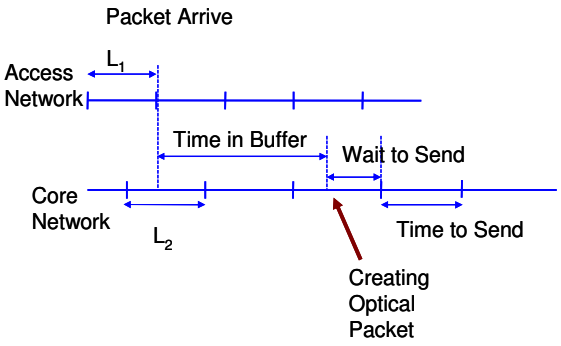


Figure 3. Times for the Optical Packet Creation Process

The synchronization problem is solved using electronic buffers in the hub. Packet creating process introduces the delay which helps to synchronize the two rings. Fig. 3 shows the transmission time slot of two rings with different sizes. L1 is the transmission time of a packet in the optical metro access and L2 corresponds to the transmission time of a packet in the core network. The correlation of the variables L1 and L2, and synchronization lag Δt affect the network performance.

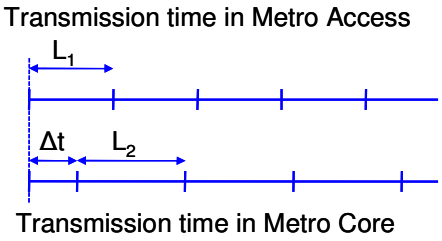


Figure 4. Transmission time in MA and MC

IV. CUTM MECHANISM

Some mechanisms have been proposed in literature to decide the time to create optical packets. A well known scheme is the opportunistic mechanism [6], this mechanism is simple: if a slot in transit is free, the optical packet is built and transmitted on the ring. The purpose of this mechanism is to reduce the load of the hub and use fewer resources. However, other mechanisms have been developed to optimize the optical packet filling. We proposed a mechanism [1], called CoS-Upgrade mechanism (CUM), to synchronize the traffics in the access nodes. This mechanism can be used not only for the access nodes (in the access network and core network) but also for the Hub to solve the problem of creating fixed size optical packet, so it uses timers in deciding when the optical packet is constructed. CUM mechanism has many advantages but also some limitations: 1) it uses several timers and buffers, 2) the hub will be over loaded when all timers are running, and 3) when the order of packets is changed building the packet at the receiver side is complicated. To improve the limitations of CUM, we propose Common-Used Timer Mechanism (CUTM), which uses a single timer for all classes of service. The principal of CUTM is shown in Fig. 5. CUTM principal corresponds to the creation process descript before in this Section, according to the Hub function. To use CUTM, we need a single buffer to hold the optical packets.

V. STUDIED SCENARIOS

The traffic flow in network is shown in Fig. 1. All the access nodes in the first ring will send the data to the node 16 (the last node). In the second ring network, there are 2 types of traffic flow: one coming from the access network and one is the local traffic (core network). So in each link connect 2 core nodes, there are 8 traffic flow from access network and 2 local traffics from other core nodes. The traffic in second network is symmetric. We consider 8 classes of service for electronic packets and 4 CoS for optical packets with different traffic sources models and packet sizes (Table I).

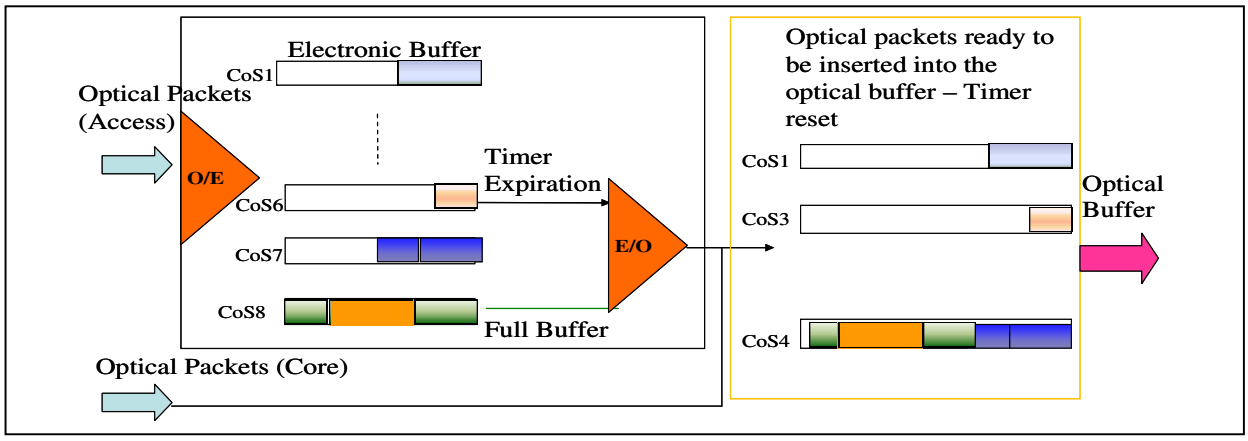


Figure 5. CUTM Principal

 ISBN: 978-1-61208-174-8

TABLE I. CLASSES OF SERVICE

	CoS 1 – CoS 2 Premium		CoS 3 – CoS 4 Silver		CoS 5 – CoS 6 Bronze		CoS 7 – CoS 8 Best Effort	
% CoS	10.4%	10.4%	13.2%	13.2%	13.2%	13.2%	13.2%	13.2%
Electronic Packet Size (Octet)	810	810	50 500 1500	50 500 1500	50 500 1500	50 500 1500	50 500 1500	50 500 1500
Source	CBR	CBR	MMPP	MMPP	MMPP	MMPP	MMPP	MMPP
Optical buffer size	1600 KOctets		4000 KOctets		4000 KOctets		8000 KOctets	

We simulate considered interconnected architecture with 3 scenarios (Table II). In the first scenario load of 2 rings are different and in the second scenario transmission rates, packets sizes and loads in two rings are differents for each ring. The scenario three uses same parameters of second one but only packet sizes are differents.

TABLE II. SIMULATION SCENARIOS

	Scenario 1		Scenario 2		Scenario 3	
	Metro Access	Metro core	Metro Access	Metro core	Metro Access	Metro core
Bit rate	10Gb/s	10Gb/s	10Gb/s	40Gb/s	10Gb/s	40Gb/s
Optical packet size	10µs – 12500 octets	10µs - 12500 octets	10µs – 12500 octets	5µs – 25000 octets	10µs – 12500 octets	10µs – 50000 octets
Load	35% - 3.5Gb	50% - 5Gb	60% - 6Gb	70% - 28Gb	60% - 6Gb	70% - 28Gb
Node traffic	437.5Mb/s	2.5Gb/s	750Mb/s	14Gb/s	750Mb/s	14Gb/s

Qos requirements are specified in Table III according to the MEF recommendations.

TABLE III. QOS REQUIREMENTS

Class of service	Characteristic of service	Service Performance		
		Loss rate	Delay	Jitter
Premium	Telephone or real-time video application	< 0.001%	<5ms	< 1ms
Silver	Applications require less loss and delay	< 0.01%	<5ms	N/S
Bronze	Applications require guaranteed bandwidth	< 0.1%	<15ms	N/S
Standard	Best effort services	< 0.5%	<30ms	N/S

VI. NUMERICAL RESULTS

In this work, several performance criteria for the given architecture are evaluated by simulation using NS2 tool and the results are presented in terms of the waiting time in the hub, end to end delay, throughput, jitter and filling ratio, loss rate at node 16. Firstly, we fix the value of Δt = 1µs and study the interaction of L1 and L2 depending on the bandwidth and packet size in each network. CUTM uses a timer equal to 100µs. The results in Fig. 6 show the jitter for the 3 considered scenarios at node 16, both mechanisms ensure the jitter condition for data flow specified on Table III.

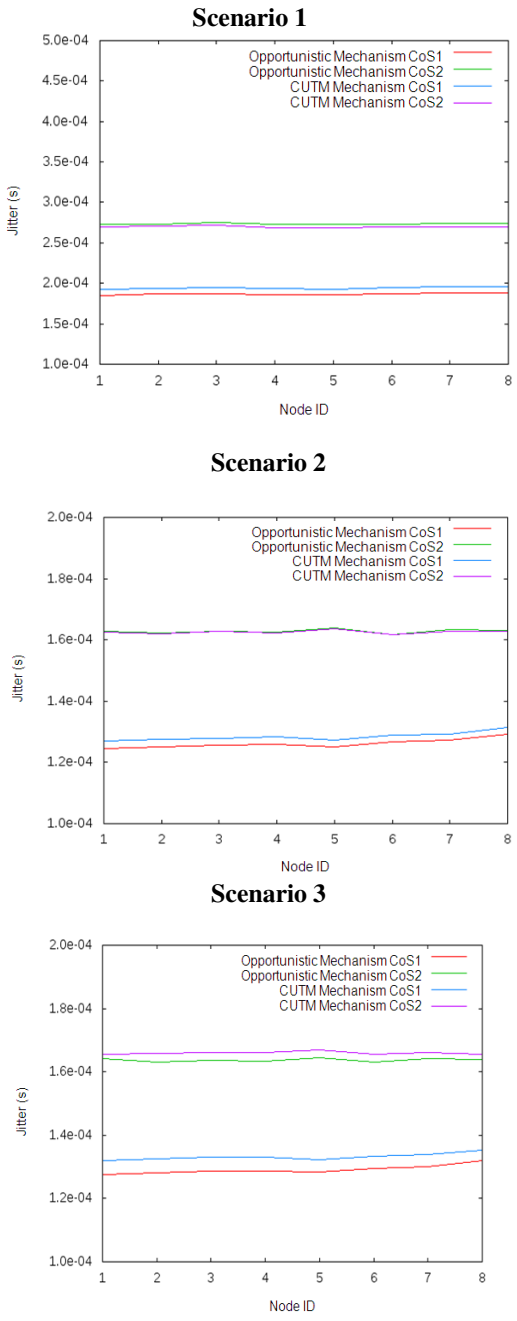


Figure 6. Jitter at Node 16

Fig. 7 shows the waiting time in hub, the average waiting time of packets in the electronic buffers with opportunistic mechanism is smaller than that of the CUTM. Based on these results we can say that CUTM is independent of L1&L2 correlation but depends on the capacity of the MC. By using opportunistic mechanism, the performance of hub does not depend on the capacity of MC; but it is sensitive to the correlation of L1 and L2.

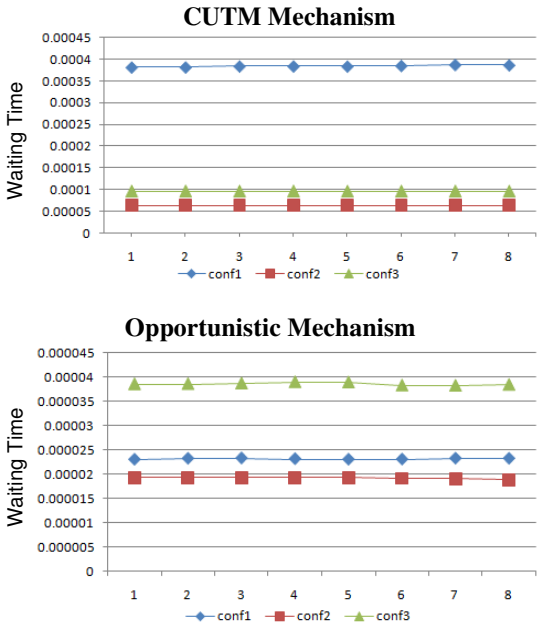


Figure 7. Waiting time in hub $\Delta t = 1\mu s$ vs CoS

Fig. 8 shows the End to End delay for both mechanisms considered, the results are better with opportunistic mechanism, however it is important to remark that CUTM uses the timer 100µs.

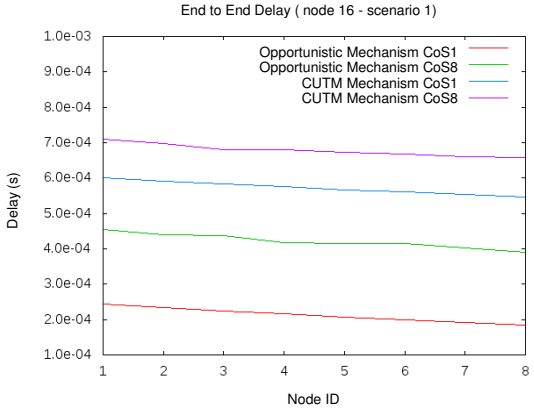


Figure 8. End to End delay (Node 16 – scenario 1)

Fig. 9 shows the throughput obtained for scenarios 2 and 3, here the opportunistic mechanism uses the network resources less effectively than CUTM.

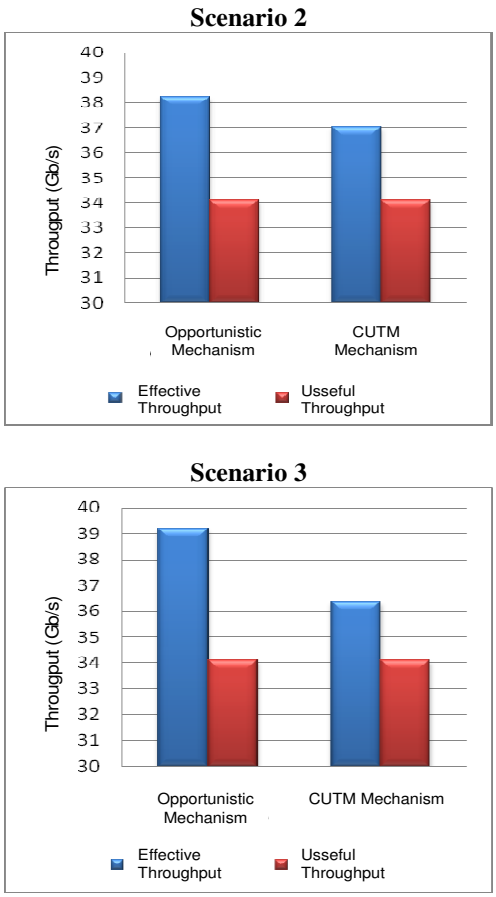


Figure 9. Throughput

Optical packet filling ratio is presented in Fig. 10; it shows that CUTM has a better result that opportunistic mechanism.

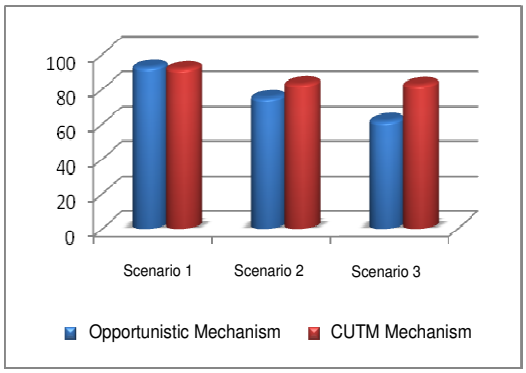


Figure 10. Filling ratio

 ISBN: 978-1-61208-174-8

The scenario 3 results show that the opportunistic mechanism uses more metro core bandwidth than CUTM, also that there is not loss at the hub and the nodes in the MC. To evaluate the loss rate we change the scenario 3 parameters. We increase the load of MA from 60% to 70%. With this change, we have the 3-1 scenario with MA load =70% ~ 8Gb/s, MC load = 70% ~ 28Gb/s, it means a total load = 35Gb/s ~ 87.5%@40Gb/s. The results in Fig. 11 show that for nodes 9 to 16 there is the loss of electronic packets. These nodes lack the bandwidth to send local traffic. The loss rate is zero at node 16 because node 16 is the destination of data flows from the Metro Access. So the loss rate at node 9 is higher than at node 10.

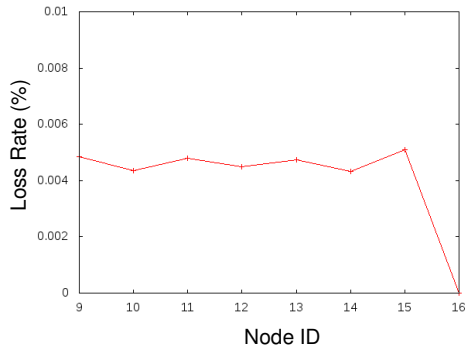


Figure 11. Loss rate for scenario 3-1.

The impact of Δt is analyzed in varying it from 1µs to 21µs (20µs = 2 x L2) on the performance of network and on hub. The Fig. 12 shows that the value of Δt does not significant impact on the network performance. The results are the same with the opportunistic mechanism.

Our results show that the opportunistic mechanism is better than CUTM. However, the filling ratio of CUTM is better than the opportunistic mechanism. Consequently CUTM mechanism saves more bandwidth than the opportunistic mechanism and provides good packet filling ratio.

VII. CONCLUSION AND FUTURE WORK

We have studied and analyzed the performance of interconnected MAN rings (MA and MC) and especially the synchronization problem between them. Performance comparison of two mechanisms: Opportunistic and CUTM has been presented. CUTM offers a solution to solve the problem of synchronization and provides good network utilization. CUTM is independent of the correlation between L1&L2, but depends on the core network capacity. Performance of opportunistic mechanism does not depend on core network capacity but it is sensitive to the correlation of L1 and L2. There is not a real impact of Δt on the network performance, variation in waiting time at hub is very small. We wish to study the impact Δt on the performance with other traffic models.

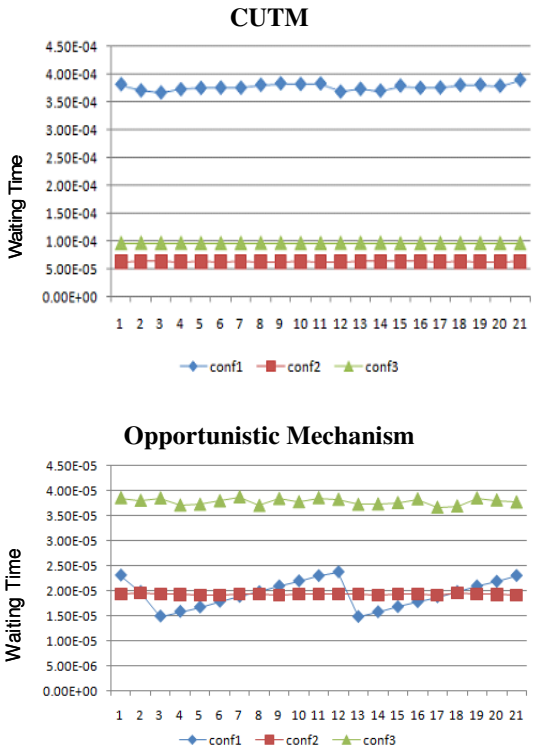


Figure 12. Waiting time in hub Δt = 1µs to 21 µs

REFERENCES

[1] T. Atmaca and T. D. Nguyen, "End-to-End Performance Evaluation of Interconnected Optical Multi-ring Metropolitan Networks", vol. 327, Springer 2010, ISBN: 978-3-642-15475-1, pp. 206-216.

[2] T. Orphanoudakis, H. Leligou, E. Kosmatos, and A. Stavdas, "Future Internet Infrastructure Based on The Transparent Integration of Access And Core Optical Transport Network". Journal of Optical Communications and Networking, vol. 1, Issue 2, july 2009, pp. A205-A218, doi:10.1364/JOCN.1.00A205

[3] R. Bonk, P. Vorreau, D. Hillerkuss, W. Freude, G. Zarris, D. Simeonidou, F. Parmigiani, P. Petropoulos, R. Weerasuriya, S. Ibrahim, A. D. Ellis, D. Klonidis, I. Tomkos, and J. Leuthold, "An All-Optical Grooming Switch for Interconnecting Access and Metro Ring Networks [Invited]". Journal of Optical Communications and Networking, Vol. 3, Issue 3, 2011, pp. 206-214, doi:10.1364/JOCN.3.000206.

[4] N. Le Sauze, E. Dotaro, and A. Dupas, "DBORN: A Shared WDM Ethernet Bus Architecture for Optical Packet Metropolitan Network", Photonic in Switching, July 2002.

[5] N. Le Sauze, E. Dotaro, and L. Ciavaglia, "DBORN (Dual Bus Optical Ring Network) An Optical Metropolitan Ethernet Solution", Research Report – Alcatel, 2004.

[6] T. Atmaca, T. Eido, T. Nguyen, P. Gravey, A. Gravey, M. Morvan, J. Roberts, S. Oueslati, T. Ronald, D. Barth, and D. Chiaroni, "Définition du Plan de Transport (MAC, Protocoles)", livrable D2.1, French ANR Project/ECOFRAME (Eléments de convergence pour les futures réseaux d'accès et métropolitain à haut débit), Conventions n°2006 TCOM 002, Project Report, January 2008.

 ISBN: 978-1-61208-174-8

A MAC Layer Covert Channel in 802.11 Networks

Ricardo Goncalves
Department of Electrical and Computer Engineering, Naval Postgraduate School
Monterey, California
santana.goncalves@marinha.pt

Murali Tummala
Department of Electrical and Computer Engineering, Naval Postgraduate School
Monterey, California
mtummala@nps.edu

John C. McEachen
Department of Electrical and Computer Engineering, Naval Postgraduate School
Monterey, California
mceachen@nps.edu

***Abstract*—Covert channels in modern communication networks are a source of security concerns. Such channels can be used to conduct hidden communications, facilitate command and control of botnets or inject malicious contents into unsuspected end user devices or network nodes. The vast majority of the documented covert channels make use of the upper layers of the OSI model. In this work, we present a proof of concept on a new covert channel in IEEE 802.11 networks, making use of the Protocol Version field in the MAC header. This is achieved by forging modified CTS and ACK frames. Forward error correction mechanisms and interleaving were implemented to increase the proposed channel's robustness to error. A laboratory implementation of the proposed channel and the results of tests conducted on the proposed channel, including measurements of channel errors and available data rate for transmission, are presented. The results validate the viability of the proposed covert channel and demonstrate that robustness of the channel to frame errors can be improved by using well known forward error correction and interleaving techniques.**

Keywords - IEEE802.11 MAC frame; frame forging; covert channel; protocol version

I. INTRODUCTION

As wireless networks become more ubiquitous, so do our dependencies on them. According to an industry report, in 2012 over a billion devices will be shipped with technology based on this standard onboard and the number is projected to be over two billion in 2014 [1]. Mobility and ease of access of wireless networks are very attractive characteristics to the end users, but along with them come additional security concerns [2].

In order to protect wireless networks from being exploited, we need to constantly evaluate their vulnerabilities and devise techniques to mitigate them. Finding possible covert channels presents an ongoing challenge, and the potential uses for such channels range from well-intentioned authentication mechanisms [3] to malware propagation [4], exfiltration [5] or command and control of botnets [6].

Many covert channels have been documented over the years and reflect the technological stage of the networks at which they were documented. The idea of network covert channels was documented 25 years ago by Girling [7], although the concept of a system-based covert channel was initially presented by Lampson in 1973 [8]. The vast majority of academic research has focused on documenting covert channels in layer 3 (network layer) or above (transport, session, presentation and application layers) of the OSI model [9]. These types of covert channels based on higher layer protocols span a wider variety of networks, since they are not limited by the physical or medium access mechanisms. The two most explored protocols above layer 2 (data link layer) are IP and TCP [10]. Even higher layer protocols, such as ICMP, HTTP or DNS, have several documented covert channels [10].

Recently, researchers began investigating wireless networks, specifically identifying covert channels in the MAC layer [11,12,13]. Frame forging plays a key role in this type of covert channel. Creating fake frames with modified header bits is a recurring theme to implement such channels. MAC header fields such as the sequence number [12], initialization vector [12] or destination address [13], have been used to hide the covert information.

Frikha, et al. [12] proposed two different implementations of a covert channel, both using fields in the 802.11 MAC header. The first one uses the 8 most significant bits of the sequence control field; the second implementation applies to networks that use Wired Equivalent Privacy (WEP) where the initialization vector subfield is used to carry the covert message. Another covert channel, as proposed by Butti [13], uses part of the destination address field of ACK frames to hide the payload. Each of these approaches relies on the forging of frames by manipulating the contents of the MAC header in order to hide the covert information.

In this paper, a covert channel that will use the MAC header of control frames is proposed to hide the covert information. This will be achieved by forging frames that use the protocol version bits in a way that was not intended by the designers of the IEEE 802.11 standard. Specifically, the protocol version field and selected control bits in the MAC header field are used to accomplish this. Our work also addresses the error robustness and throughput of the channel, supported by experimental results.

The rest of the paper is organized as follows. Section II presents an overview of the IEEE802.11 MAC frame fields and an analysis of network frame traffic. The proposed covert channel is described in Section III. Section IV presents the results of experiments.

 ISBN: 978-1-61208-174-8

II. IEEE802.11 NETWORKS AND FRAME TRAFFIC

IEEE 802.11 based wireless nodes share a common medium for communication. The fundamental building block of the 802.11 architecture is called the Basic Service Set (BSS). One BSS may be connected to other BSSs via a Distribution System (DS). Within this framework, stations can connect in ad-hoc mode or infrastructure mode. The simpler case is ad-hoc mode, where two stations can connect directly, point to point, without a DS and an Access Point (AP). If we have the stations connecting via an AP and making use of a DS, then we say they are setup in infrastructure mode.

A. 802.11 MAC frame format

A generic MAC format for an 802.11 MAC frame can be seen in Figure 1. The frame consists of the MAC header, the frame body and the Frame Check Sequence (FCS).

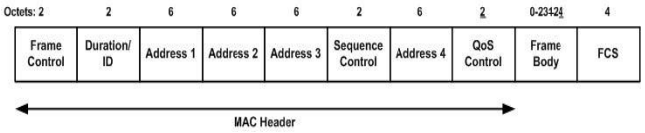


Figure 1. MAC frame format (from [14]).

The first field in the MAC header is the Frame Control (FC), consisting of two octets, and its contents are shown in Figure 2, with the protocol version field highlighted. This field consists of two bits that represent the version number of the 802.11 protocol being used. As of this writing, PV is expected to be set to zero [14]. This value may change in the future if a newer version of the standard is released.

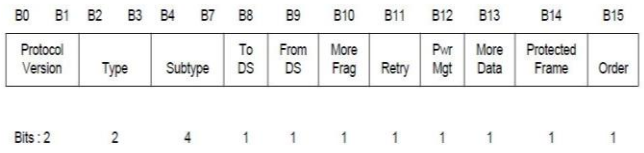


Figure 2. Frame control field (from [14]).

In the proposed covert channel, we utilize the remaining three possible combinations of the PV field to hide the covert information.

B. Frame Types of Interest

Four different types of frames exist in the 802.11 protocol: management, data, reserved and control frames.

Control type frames facilitate the exchange of data frames between stations. Within the existing control subtypes, we are interested in the smaller sized frames, the Acknowledgement (ACK) and the Clear To Send (CTS). These frames also tend to be present in large volume.

The IEEE 802.11 MAC layer makes use of the CSMA/CA scheme, in order to minimize the number of collisions and subsequent frame loss. To address the hidden node problem, a RTS/CTS handshake mechanism is used. The CTS is a 14-byte long frame whereas the RTS is 20 bytes long.

The ACK frame is generated when a station correctly receives a packet, and it is intended to signal the source station that the reception was successful. For this reason, this type of frame also tends to be very common in an operational wireless network. The length of this frame is the same as the CTS, 14 bytes.

Both frames share the same format and they only differ in one bit in the subtype field within the frame control. The ACK frame has the subtype value set to 1101; the CTS sets it to 1100.

C. Network Analysis

A heavily used 802.11 network on campus is monitored to collect frame traffic on multiple channels. From the MAC frame traffic collected, channel 1 is found to be the one with most traffic volume and number of users. We collected over 22 million packets to analyze the following frame basic characteristics.

Ideally, we want a frame that is short in length, common in occurrence, and still valid if some bits are changed. Additionally, its presence in bursts shouldn't be a rare event. These features are desirable for achieving a reasonable throughput while providing covertness.

The results of our analysis are shown in Figure 3 as a pie chart, which represents the frequency of occurrence of different types of frames. The data frames are dominant, followed by CTS, ACK and beacons. The "others" refers to the sum of all other frames that represent less than 1% individually. From this plot we can clearly see that two types of control frames matching our needs stand out, the ACK and the CTS.

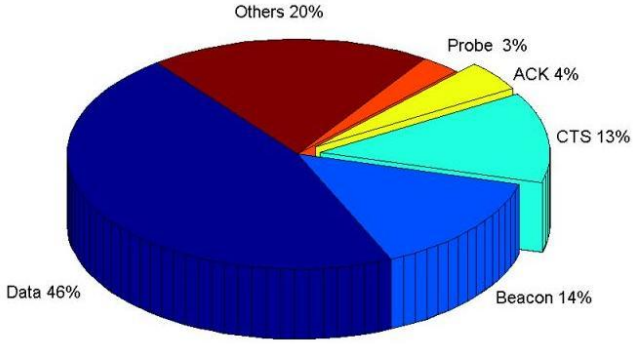


Figure 3. Frequency of occurrence of the monitoried frame types.

D. Choosing the Frame Type

In the process of choosing a frame for the covert channel, several frames were considered, such as RTS and ACK. These frames could serve as well as the CTS, but they were found to be less frequent than CTS. Also, among these three frames, RTS is the longest one with 20 bytes, and the CTS and ACK have only 14 bytes. For this reason we narrowed the options to ACK and CTS.

From monitoring of frame traffic on the campus wireless network and empirical analysis, we found that the CTSs occur with a frequency two times higher than that of the ACKs. The monitoring was conducted in different traffic scenarios, ranging from low traffic periods to high levels of

 ISBN: 978-1-61208-174-8

utilization of the network. We chose to use CTS for building the proposed covert channel as the CTS traffic volume is large and is of same frame size as ACK. By choosing CTS, we can minimize the chance of causing a traffic anomaly based on the type and frequency of packets flowing through the network.

Since CTS and ACK have a similar frame structure, it is easier to switch from one to the other, according to our objectives. The main concept of the proposed covert channel applies equally to both frames. It is even possible to have one end of the channel transmitting ACK frames, and the other transmitting CTS frames, without any loss or degradation of performance. Alternating frame types, such as transmitting a forged ACK followed by a forged CTS is also viable. Many other variations are also feasible.

The fact that both CTS and ACK frames do not contain a source address also contributes to a higher level of stealthiness, since it is not possible to immediately identify the source of the transmission.

III. PROPOSED COVERT CHANNEL

This section describes the proposed covert channel and the use of forward error correction and bit interleaving mechanisms to improve its performance.

A. MAC Header Manipulation

In the proposed covert channel we use two bits in the protocol version field of the MAC header of an 802.11 CTS packet to carry hidden information. The proposed covert channel uses the protocol version bits in a variety of ways to signal the beginning and end of the transmission as well as to carry the information, one bit at a time. A graphical representation of the bits being used is shown in Figure 4.

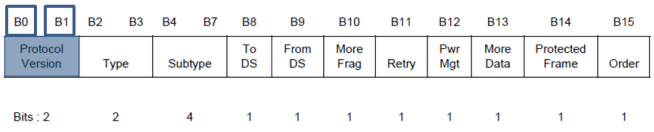


Figure 4. Manipulated bits in Frame control field (blue squares).

In order to facilitate communication in the proposed covert channel, we divided the transmission into three segments: start message delimiter, message, and end message delimiter. The start and end delimiters are realized by transmitting a sequence of five frames with 01 in the protocol version field. The message bits are transmitted using combinations of 10 as binary "0" and 11 as binary "1" in the protocol version field. The message is organized into 8-bit ASCII characters.

B. Forward Error Correction

Since we are operating in a shared media, collisions will eventually occur. This will be interpreted as an error, since a frame carrying covert payload will be lost. To mitigate the effect of frame losses, and thus reduce the number of errors in the covert channel, the use Forward Error Correction (FEC) was considered.

There are several options for implementing FEC: block codes such as Hamming and Reed-Solomon, convolutional codes, turbo codes, or low density parity check codes. In this work, however, a convolutional code was used for error correction.

A convolutional coder takes an m - bit message and encodes it into an n - bit symbol. The ratio m/n is known as the code rate. In our case a code rate of $2/3$ was used, meaning the encoded message will be one and a half times as long as the original message. This will increase the time needed to transmit the same message as before, since a higher number of bits is being sent.

Another important parameter in convolutional coding is the constraint length. This parameter, *k*, represents the number of bits in the encoder memory that affect the generation of the n output bits [15]. A constraint length of 4 is used in our experiments.

Forward error correction is typically applied to a transmission of a stream of bits sent and received sequentially. In our case, however, the bits are embedded into independent frames, which are prone to loss. As a result, when a frame is lost, the receiver has no indication that a bit was missing. Consequently, we now need to know exactly which frames were lost in order to apply the FEC correctly.

One option is to use the eight flag bits in the frame control field of the MAC header to index a longer sequence number, which makes determining the location of lost frames an easier task. These flag bits will not carry any covert information but serve only the error correction function. However, it is important to state that applying this use of the flag bits will increase the probability of detection of the covert channel, since unexpected flag attributions will be present. In this situation, we move from a minimum deviation of two bits (as in Figure 4) to a maximum of 10 bits (as in Figure 5). This presents a tradeoff between detectability and error performance, and the user must exercise the option to choose one over the other as dictated by the application. In order not to use the flag bits one could use the type and subtype fields of the MAC header. The IEEE802.11 standard defines some bit combinations of the subtype field as "Reserved". Exploring these combinations could be an option, although we did not test it.

Figure 5 is a representation of how we accommodated the information and sequence bits within the MAC header.

The blue squares represent our covert channel bits. These bits are used in the same way as before: the first bit (B0) signals the presence of the channel and the second is payload (B1). The red circles refer to the sequence bits, which are placed in the flag bits of the frame control field.

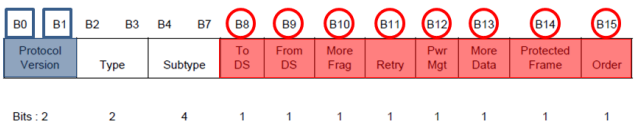


Figure 5. Representation of the frame structure using the flag bits for sequencing (red circles).

 ISBN: 978-1-61208-174-8

Given that we have eight flags, this gives us a total of 256 possible sequence numbers. This alone provides a reasonable amount of protection against a long burst of frame losses, when compared to the previous approach.

C. Forward Error Correction and Interleaving

We now consider sending more than one bit of information per forged frame.

Since each frame now carries more than one information bits, the loss of one or more frames has a bigger impact on the number of errors in the channel. In order to mitigate this effect, we interleave the bit string resulting from the convolutional coder. This consisted of breaking the coded message in blocks of 8 bits, building a matrix with each block in a different row. By reading the matrix out by column, from top to bottom, we generate a new string of bits, effectively interleaving all the 8 bit blocks. The number of rows depends on the length of the message we are transmitting.

Figure 6 is a schematic representation of this idea. At the output of the convolutional coder we interleave the bits in groups of 8 bits. This will result in a new string of zeros and ones, which goes into the covert channel processing block. Here the string is separated in groups of n bits, and each group will become the payload of the forged frames.

Notice that only information bits are encoded and interleaved; in this implementation the convolutional coder is applied after we have the complete message we want to transmit.

Figure 6. FEC and interleaving block diagram.

One possible implementation is to use six bits for payload. The frame is forged as follows: six information bits are placed in the selected flag bits, three other bits are used for sequence numbers, and the first PV bit is set to one, indicating the use of the covert channel. Figure 7 illustrates the proposed structure. The blue squares indicate payload bits, and the red circles are sequence numbers. The green diamond (B0) indicates the presence of the covert channel. Bits B1, B8 and B9 form the sequence number yielding a sequence length of 8. Bits B10-B15 form the payload of six bits to carry the message.

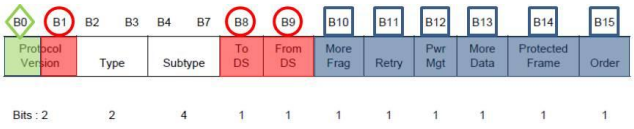


Figure 7. Representation of the frame structure using three bits for sequencing (red squares) and six bits for payload (blue squares).

IV. EXPERIMENTS AND RESULTS

In order to implement the proposed covert channel, we developed the necessary code to forge, transmit, and receive frames. Python was the chosen programming language, due to its simplicity, available libraries and extension modules that facilitated our task. Regarding the OS, a Linux environment was elected, for being more flexible, open source and GNU licensed.

The code is divided into three threads running simultaneously. One thread runs as the receiver, another one as the transmitter, and the third one as a control mechanism in order to handle possible discrepancies in the identification of the beginning and end of the covert communication. Other version 1 frames (with bad checksums) where found circulating in the network, and become noise to our version 1 frames forming the start and end delimiters. Thread3 is responsible for filtering out these unwanted frames.

A. Test bed

Frame traffic was recorded over operational wireless networks, during week days, in order to capture the real-world scenarios.

Three different scenarios were considered and tested. All scenarios consisted of transmitting similar messages during approximately the same time of day. The difference between the scenarios is the way the data was transmitted since we varied the number of payload bits and applied different error mitigation mechanisms.

It is important to notice that stations A and B were operating in the ad-hoc mode, outside the infrastructure wireless network being monitored. The stations transmit without any coordination from the access point. This likely causes collisions, and thus frame losses, which are interpreted as errors for analysis purposes.

A standard sentence was used for all scenarios, with a total of 1408 bytes. In the first scenario, the messages were sent without any error control. The second scenario introduced the use of FEC, and the third used a combination of FEC and interleaving, in order to improve the error robustness of transmitted message. In the following analysis, in order to have a performance benchmark, we used the first scenario as the baseline for comparison with the FEC scenarios,.

B. Results

1) Scenario 1

In Figure 8(a) we can see the profile of the traffic collected for a period of about ten hours on channel 1. Figure 8(b) displays the percentage of errors detected upon reception of the test sentence.

Summarizing this analysis, we observed an average error of approximately 3% for the sentence over a total of 30 sets of transmissions. No error correction or sequencing is at work in this scenario.

 ISBN: 978-1-61208-174-8

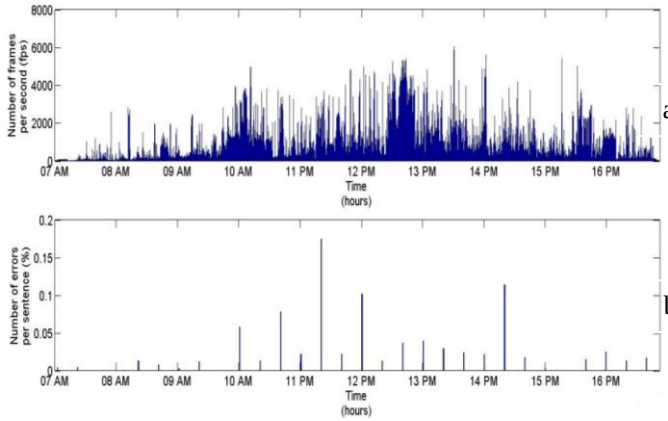


Figure 8. Network traffic profile and percentage of errors for sentence and sequence receptions in channel 1.

2) Scenario 2

The percentage of errors as a function of 15 repeated transmissions of the sentence, in channel 1, over a period of 4 hours, is shown in Figure 9. The length of the transmitted sentence is now 2,112 bits long because we applied a $^{2}/_{3}$ rate encoder on a 1,408-bit string. The red stems (x) represent the number of errors detected in the received sentence, and the blue stems (o) the number of errors in the received sentence with FEC. In most cases the number of errors drops to zero or is significantly reduced.

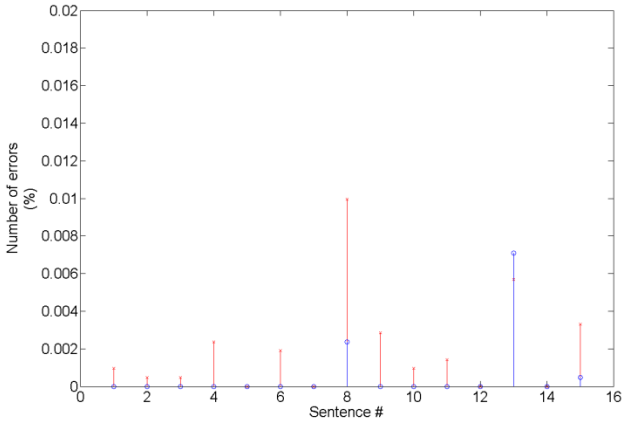


Figure 9. Percentage of erros before (red cross) and after FEC (blue circle) per received sentence, using flag bits for sequencing.

This is consistent with our expectations. We have one outlier in that for the 13th repetition of the sentence we got a higher number of errors with FEC.

We recorded a total of 67 errors in this experiment (without FEC), which translates into an average of 4.5 errors per sentence, or an average error percentage of 0.21%. After the execution of FEC, the total number of errors dropped to 21, resulting in an average of 1.4 errors per sentence or an overall average of 0.09%, relative to the 1,408 bits of the original message. However, this gain was the direct result of having to transmit more bits to send the same message, when compared to the first scenario with no FEC, thus reducing the data rate.

3) Scenario 3

The percentage of errors per sentence repetition can be seen in Figure 10. From this figure we can notice an outlier at repetition 12, actually gaining errors after the FEC. This was an isolated event and it was excluded from this analysis. The result is an average number of 1.53 errors per repetition or 0.07% of the total amount of bits sent per sentence. Following the sequence number tracking, de-interleaving and correcting the bit sequence, the total number of errors is reduced to zero. These are significant results; however, the sample space is small, and we cannot conclude that this level of robustness will be achieved in every reception.

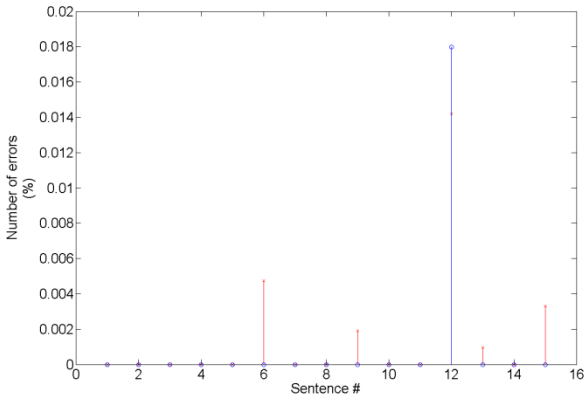


Figure 10. Percentage of erros before (red cross) and after FEC (blue circle) per received sentence with interleaving.

C. Throughput Analysis

In order to evaluate the throughput offered in each scenario, the rate at which the frames were transmitted was measured. Being a proof of concept, code efficiency was not a major concern, and the results are presented for analysis purpose only, meaning significant improvements may be easily achieved. This was done using Airopeek and by averaging the rate of the forged frames on a per second (fps) basis. Depending on the network usage at the time, the frame rate varies significantly. Another factor responsible for this variation is the continuous adjustment of the maximum data rate of the network as dictated by the channel conditions. For IEEE 802.11b networks the maximum network data rate possible values are 1, 2, 5.5, and 11 Mbps [14].

To obtain a benchmark for performance comparison, we first determine the maximum data rate possible for the covert channel under optimal conditions. The following conditions are assumed: (i) The channel is ideal with no errors; (ii) there is only one station with frames to transmit; and (iii) we use a data rate of 2 Mbps, the highest possible for 802.11b control frames (basic rate set) [14].

 ISBN: 978-1-61208-174-8

The medium access scheme has to obey some predetermined timing constraints, set by the standard. Figure 11 is a graphical representation of the timing requirements for transmitting a frame.

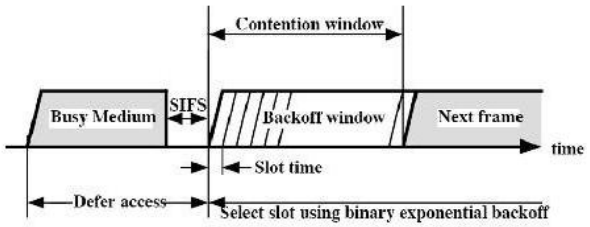


Figure 11. Timing constraints in an 802.11 frame transmission [After 16].

Applying the work of Xiao and Rosdhal [17] to the proposed covert channel, the minimum amount of time necessary to transmit a forged CTS is $t_{min} = 376$ µs, corresponding to a maximum of 2659 forged frames per second. At one bit per frame the maximum bit rate is 2659 bps; at six bits per frame we get 15.954 kbps. The measured throughput values, however, will be significantly smaller.

When we transmit one bit of information in each forged frame, we have an overhead of the start and end delimiters for a total of 10 signaling frames. The measured average frame rate was 61 frames per second. Since each frame represents a bit, and considering our message payload of 1408 bits, we transmit a total of 1418 bits. At 61 fps this corresponds to a total transmission time of 23.25 sec, and a useful bit rate or throughput of 60.5 bits per second (bps).

On the other hand, when we transmitted 6 bits per forged frame and introduced the use of interleaving, the measured average transmission rate was 32 fps. By transmitting a total of 2122 bits, we obtained a total transmission time of 11 seconds. The resulting throughput value is 127.4 bps, considerable improvement over the previous case.

V. CONCLUSIONS

This work presented, implemented and tested a previously undocumented covert channel in an IEEE802.11 network. We used the protocol version field in the MAC header to hide and transfer the covert information. Robustness to errors in the covert channel is improved by the use of forward error correction and bit interleaving. The proposed covert channel was implemented by developing the necessary code in Python. A GUI chat console is used for message transmission. The test bed used for experiments operated in a Linux environment. Preliminary results indicate significant improvement in the error performance of the channel. The achieved throughput of the covert channel is measured and the maximum channel data rate is also determined. The case of 6-bit payload along with convolutional coding and interleaving yielded the highest measured throughput.

REFERENCES

[1] D. McGrath, "WLAN chip set shipments projected to double," in EE Times, 2/17/2011. (accessed March 17, 2011) http://www.eetimes.com/electronics-news/4213260/WLAN-chip-set-shipments-projected-to-double

[2] Y. Xiao, C. Bandela, and Y. Pan, "Vulnerabilities and security enhancements for the IEEE 802.11 WLANs," in Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM) 2005, pp. 1655-1659, 2005.

[3] T.E. Calhoun, R. Newman, and R. Beyah, "Authentication in 802.11 LANs Using a Covert Side Channel," in Communications, 2009. ICC '09., IEEE International Conference, pp. 1-6, 14-18 June 2009.

[4] E. Couture, "Covert Channels," SANS Institute InfoSec Reading Room (accessed January 17, 2011). http://www.sans.org/reading_room/whitepapers/detection/covert-channels_33413

[5] A. Giani, V. H. Berk, and G. V. Cybenko, "Data Exfiltration and Covert Channels," Process Query Systems, Thayer School of Engineering at Dartmouth (accessed February 02, 2011). http://www.pqsnet.net/~vince/papers/SPIE06_exfil.ps.gz

[6] D.T. Ha, G. Yan, S. Eidenbenz, and H.Q. Ngo, "On the effectiveness of structural detection and defense against P2P-based botnets," in Dependable Systems & Networks, 2009. DSN '09. IEEE/IFIP International Conference, pp. 297-306, June 29 2009-July 2 2009.

[7] C.G. Girling, "Covert Channels in LAN's," in Software Engineering, IEEE Transactions, vol. SE-13, no. 2, pp. 292-296, Feb. 1987.

[8] B. Lampson, "A note on the confinement problem," in Communications of the ACM, vol. 16, pp. 613-615, October 1973.

[9] H. Zimmermann, OSI Reference Model, IEEE Transactions on Communications, Vol. COMM-28(4), April 1980.

[10] M. Smeets and M. Koot, "Research report: covert channels," Master's thesis, University of Amsterdam, February 2006.

[11] T. Calhoun, X. Cao, Y. Li, and R. Beyah, "An 802.11 MAC layer covert channel," in Wireless Communications and Mobile Computing, Wiley InterScience (accessed January 2011). http://onlinelibrary.wiley.com/doi/10.1002/wcm.969/pdf

[12] L. Frikha, Z. Trabelsi, and W. El-Hajj, "Implementation of a Covert Channel in the 802.11 Header," in Wireless Communications and Mobile Computing Conference, 2008. IWCMC '08., pp. 594-599, 6-8 August 2008.

[13] L. Butti, Raw Covert (accessed September 2010) http://rfakeap.tuxfamily.org/#Raw_Covert

[14] Institute of Electrical and Electronics Engineers, 802.11, Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications (accessed January 17, 2011). http://ieeexplore.ieee.org

[15] S. Lin and D. J. Costello., Error Control Coding: Fundamentals and Applications, Pearson Prentice Hall, New Jersey, 1983.

[16] W. Stallings, Wireless Communications and Networks, Second edition, Pearson Prentice Hall, New Jersey, 2005.

[17] Y. Xiao and J. Rosdahl, "Throughput and delay limits of IEEE 802.11," in Communications Letters, IEEE, vol.6, no.8, pp. 355- 357, Aug 2002.

 ISBN: 978-1-61208-174-8

Design Time Reliability Predictions for Supporting Runtime Security Measuring and Adaptation

Antti Evesti, Eila Ovaska
VTT Technical Research Centre of Finland
Oulu, Finland
{antti.evesti, eila.ovaska}@vtt.fi

***Abstract*—The reliability of a quality-critical software component affects the security level that is achieved. There is currently no runtime security management approach that uses design time information. This paper presents an approach to exploiting design time reliability predictions in runtime security management. The Reliability and Availability Prediction (RAP) method is used to predict reliability at software design time. The predicted reliability values are stored in ontology form to support runtime use. The use case example illustrates the presented approach. The presented approach makes it possible to use design time reliability predictions at runtime for security measuring and adaptation. Hence, the reliability of security mechanisms is taken into account when security adaptation is triggered.**

Keywords - information security; quality; evaluation; metric; architecture

I. INTRODUCTION

A variety of quality prediction and testing techniques are used at software design time. The results of these predictions are used to enhance architecture designs, select better component alternatives, and reveal implementation errors. The use of these prediction results ends when satisfactory quality is achieved for a component or system and the product is delivered. However, these prediction results could also be used in runtime situations. This is reasonable, especially in reliability and security management. Reliability is an important factor in achieving a required security level, as can clearly be seen from the security decomposition presented in [1]. Weak reliability of a security-related software component ruins the offered security. Hence, the reliability information of component is valuable for security-related decision-making. This paper therefore presents an approach to bring the design-time reliability prediction results for runtime security measuring and adaptation purposes. To achieve this, ISMO (Information Security Measuring Ontology) [2] is extended in a way that allows prediction results to be stored at design time.

In the literature, different security adaptation approaches exist. The adaptive SSL presented in [3] sets parameters for the SSL session based on the environment information. An Extensible Security Adaptation Framework [4] adds a middleware layer for security mechanisms. The application sets the required security policy and, based on the policy, the middleware layer selects security mechanisms. Context-sensitive Adaptive Authentication [5] uses time and location information to calculate a confidence level for the authentication. In some situations, a low confidence level is sufficient while others require adaptation of the authentication method used. Our earlier work presents an approach that uses ontologies and risk-based measures for security adaptation [6]. These adaptation approaches are intended to work at runtime by observing the system's resources and environment. Based on the observations, different security mechanisms or parameters are set. To our knowledge, none of the existing approaches uses design-time information for adaptation purposes.

Figure 1. presents the broader context of the contribution of this paper. In the first phase, the Reliability and Availability Prediction (RAP) method [7] is used to predict future reliability from software designs. The prediction results are stored in ontology form in order to ensure exploitation at runtime. In this paper we will focus on this first phase. In the second phase, application security is measured at runtime. Reliability predictions are used as input information for security measuring. The third phase is security adaptation, which is triggered by the measuring phase. The adaptation also uses reliability predictions to select the most suitable security mechanism for different situations. After the adaptation, the execution returns to the measuring phase.

Figure 1. Broader scope

The contribution of this paper makes it possible to use design-time reliability predictions for runtime security measuring and adaptation. Hence, a wider information set is available for triggering and making a decision on the adaptation. In other words, information for runtime use can be collected in different phases of the application lifecycle. Thus, the adaptation is not only based on the measurements made just before adaptation but also on knowledge of the whole life cycle of the component.

The paper is organised as follows. After the introductory section, background information is presented. Next, Section 3 is divided into three parts describing the design steps towards applications with security adaptation, design time

 ISBN: 978-1-61208-174-8

reliability predictions, and a way to transform prediction results into the ontology form. Section 4 illustrates the presented approach by means of a case example. A conclusion and future work ideas close the paper.

II. BACKGROUND

Reussner et al. define reliability as the probability of failure-free operation of a software system for a specified period of time in a specified environment [8]. ISO/IEC defines security as follows: The capability of the software product to protect information and data so that unauthorized persons or systems cannot read or modify them and authorized persons or systems are not denied access to them [9].

The RAP method evaluates the reliability of a designed software system and its components already at architecture design time [7, 10]. The RAP method reveals design flaws and critical components from the reliability viewpoint. The evaluation is based on architectural models, which means that the first evaluation results are available before any implementation effort is required. Hence, modifications can be performed easily. The RAP method uses state-based models, i.e., Markov models, to predict the reliability of components. The path-based models are used to predict the reliability of a single execution path and the whole software system. The RAP method produces the following reliability values, known as probability of failure (pof) values: 1) independent pof values for software components, 2) pof values for execution paths, 3) the component's pof value in each execution path, 4) the components' system-dependent pof values, and 5) the pof value for the whole software system. The RAP method supports the feedback loop from software testing [11]. The prediction results can therefore be replaced with more accurate values when measured reliability values are available from the software testing. Tool support for the RAP method, called the RAP tool, is also available. The RAP tool reads architectural models from UML diagrams, i.e., state, component, and sequence diagrams. In addition, the RAP tool uses usage profiles that describe system usage, i.e., how many times each execution path is called. The usage profiles make it possible to perform own predictions for different user groups, e.g., professional and normal users. In this work, the results from the RAP tool will be made available for runtime use.

Evesti et al. present the ISMO ontology in [2]. The ISMO composes security ontology and general software measuring terminology. The ISMO thus offers a generic and extendable way to present security measures. These measures are connected to security threats and/or supporting mechanisms, depending on the measure. Measures are divided into base measures, derived measures, and analysis models. The base measure is the simplest measure and is used for more complex measures, i.e., derived measures and analysis models. The ISMO is instantiated as an example using authentication measures, especially Authentication Identity Structure (AIS) measures [1] for password-based authentication. The ISMO thus contains measures for password age and type, i.e., length and the number of different symbols. The software application uses different measures from those of the ISMO to measure its security level at runtime. In this work, the ISMO is extended to contain design time reliability predictions.

Savola et al. present Basic Measurable Components (BMCs) for security attributes (e.g., authentication, confidentiality, etc.) in [1]. BMCs are derived by means of the decomposition approach. The idea of BMCs is to divide security attributes into smaller pieces that can be measured. For example, authentication is divided into five BMCs in [1] as follows: Authentication Identity Uniqueness (AIU), Authentication Identity Structure (AIS), Authentication Identity Integrity (AII), Authentication Mechanism Reliability (AMR), and Authentication Mechanism Integrity (AMI).

III. RELIABILITY PREDICTIONS FOR SUPPORTING SECURITY MEASURING AND ADAPTATION

This section is dived into three subsections. Firstly, high-level design steps for the application with security adaptation features are described. Secondly, a design time reliability prediction is presented. Finally, a way to store the prediction results in ISMO in a way that supports runtime measuring is described.

A. Designing an Application with Security Adaptation Features

This subsection lists design steps that a software architect has to take when designing an application with security adaptation features. Figure 2. illustrates these design phases. The last three phases of the process are iterative. This is not depicted in the figure, however, for reasons of clarity.

1) Required security attributes

In the first phase, the software architect has a set of required security attributes for the application, for instance, S1 for communication confidentiality, S2 for user authentication, and S3 for data integrity requirements. S refers to a security requirement in general.

2) Adaptable security attributes

The software architect has to design adaptation features separately for each security attribute. From the above-listed required security attributes, the architect has to select which ones to implement in an adaptable manner, i.e., variation will take place at runtime [12]. In Figure 2. user authentication S2 is selected for the adaptable security attribute. Other security attributes are thought of as static security requirements from the runtime viewpoint. In other words, the possible variation in these attributes is taken into account at design time.

3) Mechanisms for adaptable security attributes

The adaptable security requirement has to be met by security mechanisms that can be changed or that have parameters that can be modified at runtime. For example, in the adaptable user authentication case, the architect designs two alternative user authentication mechanisms for the application, e.g., password-based and voice-based authentications. Another alternative is to design one security mechanism and set different parameters for it at runtime.

4) Measurements for triggering adaptation

 ISBN: 978-1-61208-174-8

Software measures are designed for the application in parallel with the mechanism design phase. In particular, this means base measures that require measuring probes inside the application. Adaptation at runtime will be triggered based on derived measures and analysis models, which both depend on base measures. In other words, base measures are used to compose derived measures and analysis models. Hence, the software architect has to implement these base measures for the application.

5) Architecture design

The architect designs the architecture for the system. From the security adaptation viewpoint, it is important that variation points are designed with care. For adaptable user authentication, this means that an authentication feature can be called without knowing the currently used authentication mechanism.

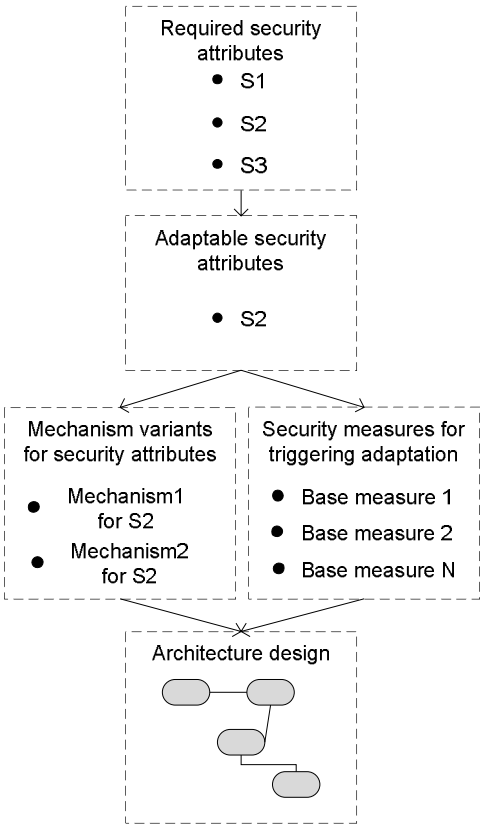


Figure 2. Steps towards adaptable application

B. Design Time Predictions

This subsection describes how the software architect predicts the reliability of components from the architectural designs. The architect uses the RAP method to perform these predictions. Based on the steps listed in the previous subsection, the architect has design documents for the application. Firstly, the component diagram describes the structure of the application. Secondly, the internal behaviour of components is described by means of state diagrams. Finally, sequence diagrams describe the mutual behaviour of components, i.e., how the component calls other components.

The RAP method contains state-based and path-based reliability prediction methods. For runtime security measuring and adaptation purposes, the RAP method is used to predict the probability of failure (pof) values for security mechanism components, i.e., mechanisms designed in phase 3 of the previous subsection.

The state-based prediction method calculates reliability for one independent software component by means of state diagrams. In state diagrams, pof values are given for each state to describe the failure probability in that particular state. Moreover, transition probabilities between states are described. Based on this information, the RAP tool automatically adds a separated failure state and calculates the component's pof value using a state transition matrix *p* and a probability vector *p(n)* as follows:

$$p = \begin{bmatrix} p_{SS} & p_{SA} & p_{SB} & p_{SF} \\ p_{AS} & p_{AA} & p_{AB} & p_{AF} \\ p_{BS} & p_{BA} & p_{BB} & p_{BF} \\ p_{FS} & p_{FA} & p_{FB} & p_{FF} \end{bmatrix} \quad (1)$$

$$p(n+1) = p(n)*p \quad (2)$$

In transition matrix p, p_{SA} presents the probability of transit from the start state S to state A. Similarly, p_{AF} presents a probability of transit from state A to the failure state F. In the beginning, the probability vector takes the form p(0) = [1, 0, 0, 0], which means that the probability of being in the start state is 1 at time moment 0.

The state-based prediction produces independent pof values for the components. These values are further used to calculate the component's pof values in different execution paths. Execution paths are presented by means of sequence diagrams in architectural models. The following equation is used to calculate a component's pof value in a particular execution path:

$$p_{ij} = 1 - (1 - p_i)^{Nij} \quad (3)$$

The previously calculated independent pof of the component is substituted in p_i, and Nij represents the number of execution times of the component in that execution path. Execution paths describe how the particular component is called in different execution paths.

As mentioned in Section 2, the RAP tool is also able to calculate pof values for each execution path, the component belonging to the particular software system, and for the whole software system. The equations for these calculations are presented in [11]. However, our interest is in bringing the previously presented component-related pof values for runtime use.

C. Storing Prediction Results in a Runtime-Applicable Way

After the RAP predictions, the software architect has the components' independent pof values and the components' pof values for the execution paths. Initially, the RAP tool was only intended for use at design time. Thus, the RAP tool

 ISBN: 978-1-61208-174-8

stores these reliability values in the component diagram by means of a UML profile. Hence, the values are available during the implementation and testing phases. Figure 3. shows the security mechanism part of the component diagram after the RAP predictions. Now, the mechanism alternatives designed earlier contain the predicted pof values. This is not practical for runtime purposes however. Reading the pof values from the UML profile requires a connection to a UML tool, which cannot be offered at runtime.

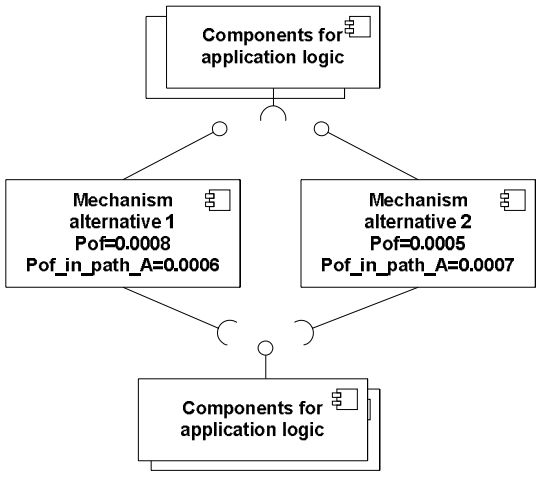


Figure 3. Component diagram after reliability predictions

As mentioned in Section 2, the ISMO supports runtime security measurements. The ISMO is therefore extended to store the components' reliability values. The following information needs to be stored in the ISMO:

1. Software component name
2. Software component version number
3. Which security mechanism the component implements
4. Reference to a place where the pof values are stored
5. Information on the execution path used to calculate path-specific pof values

The component name is intended to separate different alternatives of the mechanisms and is the name taken directly from the component diagram. It is natural to create an instance in the ISMO with a component name. This is because each software component is an individual element.

The version number separates different implementations of the same component. For instance, a new component version that contains bug fixes has a better pof value than the old version. This information therefore has to be separated in the ISMO. The version number is combined with the component name, i.e., an instance name in the ISMO. This naming convention also ensures that the ISMO does not contain instances with the same name.

Information on the security mechanism that the component implements is required because components use different security mechanisms to meet the required security, i.e., the mechanism alternatives in Figure 3. use different security mechanisms. For example, two components can use different authentication mechanisms to achieve user authentication. Countermeasures are described as concepts, i.e., classes, in the ISMO. Thus, it is reasonable to create the instance from the software component under the right countermeasure concept. Figure 4. presents instances created from software components from different countermeasure concepts.

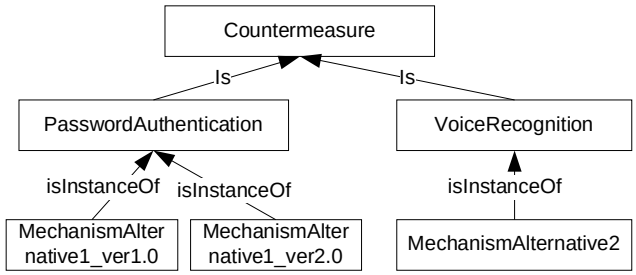


Figure 4. Component instances in the ISMO

The most important information is the components' pof values, and the above-described additions to the ISMO make it possible to put this information into the ISMO. Figure 5. shows a way of presenting pof values in the ISMO. A new base measure called *pof* is added to the ISMO. This is able to offer the component's pof values for the runtime measuring. In the ISMO, each measure is defined for *attribute*, i.e., path-specific pof and independent pof. The attribute relates to *MeasurableConcept*, i.e., Authentication Mechanism Reliability (AMR). Previously, the ISMO contained only measures related to the Authentication Identity Structure (AIS). Both attributes are connected to the countermeasure instance, i.e., MechanismAlternative_ver1.0 in this case, with the *hasMeasurableAttribute* property. Other mechanism instances also contain these attributes. However, for reasons of clarity, these are not presented in the figure.

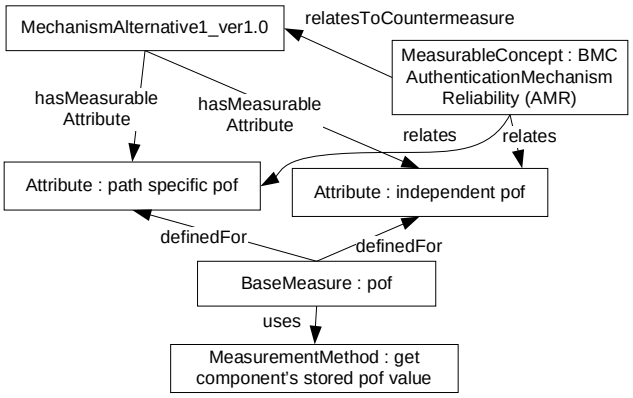


Figure 5. Update for the ISMO

Both attributes use the same pof base measure. The purpose of this base measure is to use *MeasurementMethod*, which retrieves the components' pof values. The measurement method is a concrete measuring probe that is able to retrieve pof values. Hence, it has to know the format

 ISBN: 978-1-61208-174-8

that is used to store pof values. The following structure is used: componentName, componentPof, the component's pof in execution path 1, the component's pof in execution path 2, etc. This structure therefore offers information on the execution path used to calculate path-specific pof values. It is possible to store pof values in a separated file or structure inside the application code. The separated file offers more flexibility, however, i.e., pof values can be updated without knowing the program code. The architect decides where the pof values are stored and creates an appropriate measurement method.

The reason why pof values are not stored directly in the attributes is twofold. Firstly, the measurement part of the ISMO – inherited from the Software Measurement Ontology (SMO) [13] – defines that attributes only define things that can be measured. Secondly, storing pof values outside the ISMO makes the ontology and pof values manageable. An application with security adaptation can contain several security mechanism components and each component can belong to several execution paths. Storing all these values into the ISMO will increase its size and complicate the updating of pof values.

IV. USE CASE EXAMPLE

This section gives a use case example of the presented approach. The purpose of the example is to show how the reliability of the security mechanism component is predicted. The results are stored in a runtime-applicable way in the ISMO.

The software architect designs a software application with security adaptation features. Communication confidentiality and user authentication are required securities for the application, c.f. Figure 2. From these security requirements, it is decided to implement user authentication in an adaptable manner. Hence, the architect designs alternative mechanisms for achieving user authentication, for example, password-based and fingerprint authentication. At the same time, base measures for measuring the user authentication are designed for the application. One of these base measures is pof. The value of the pof base measure is retrieved using a measurement method. It is notable, that the base measures and related measurement method implementations are reusable. Hence, the same base measure is also applicable to other security mechanisms.

After these design steps, there will be a component diagram, state diagrams of components, and sequence diagrams. Both authentication mechanisms are implemented as one independent software component called *passwordAuthentication* and *fingerprintAuthentication*.

Figure 6. presents a state diagram for the password authentication component. In this case, each transition probability is 1, i.e., only one leaving transition from each state. The architect sets the pof values for each state heuristically, and these pof values then affect the transition probabilities. In other words, the state's pof value reduces the occurrence probability of the right state transition respectively. Based on values from Figure 6. the RAP tool automatically adds the failure state and builds the transition matrix p as described in Section 3. From the transition matrix, the RAP tool calculates the pof value for the *passwordAuthentication* component. In this case, the pof value for the *passwordAuthentication* component is 0.000482. Similarly, pof values are given for states in the *fingerpringAuthentication* component, and the pof value of the component is calculated.

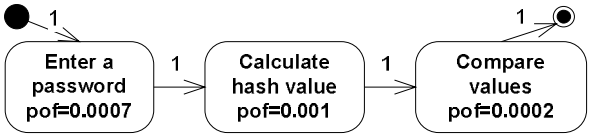


Figure 6. State diagram for the passwordAuthentication component

To exemplify path-specific pof values, the sequence diagram in Figure 7. is used. The RAP tool uses this sequence diagram, previously calculated pof value, and equation 3 to calculate the path-specific pof value. Hence, a pof value of 0.000482 is attained for the *passwordAuthentication* component in this specific execution path. In this case, the independent and path-specific pof values are the same because the *passwordAuthentication* component is only called once in this sequence diagram, c.f. equation 3.

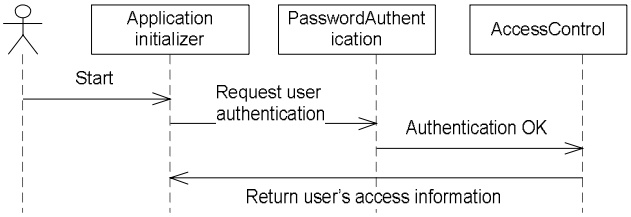


Figure 7. Sample execution path for password authentication

The architect stores this information in the ISMO in the form defined in the previous section and illustrated in Figure 8. In the figure, grey is used to describe information added in this case example. The component name is now *passwordAuthentication* and the version number is 1.0. Hence, the instance named *passwordAuthentication_ver1.0* is created under the password authentication concept in the ISMO. Similarly, the instance for the *fingerprintAuthentication* component is created. Both of these instances contain previously mentioned attributes. Attributes for the *fingerprintAuthentication* are not presented in the figure, however, for reasons of clarity. Calculated pof values are stored in the specific file called pofs. This file is presented in dark grey in Figure 8. because it is a separate part from the ISMO. *MeasurementMethod* contains a link to that file and is able to read pof values from the file. In this case, the file contains pof values for the *passwordAuthentication* and *fingerpringAuthentication* components.

 ISBN: 978-1-61208-174-8

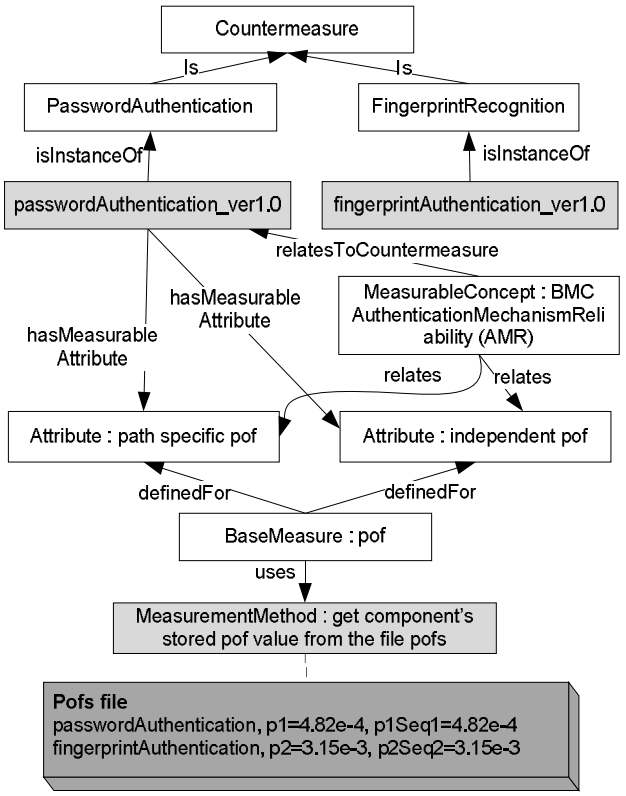


Figure 8. The content of ISMO after design time predictions

V. CONCLUSION AND FUTURE WORK

There is a clear connection between software reliability and security. An unreliable software component that performs security-related actions can ruin the security of the whole application. In this work, an approach was introduced to bring the results from design-time reliability predictions for runtime security measuring and adaptation purposes. Hence, the reliability of the security mechanisms can be taken into account when security adaptation is triggered. The work presented steps on how to produce an application with security adaptation features. Thereafter, reliability was predicted from design documents. Finally, these prediction results were stored in the ISMO, which makes it possible to use the prediction results at runtime. Storing the components' pof values in the ISMO required some extensions to the ontology. Firstly, the way to present individual security mechanism components in the ISMO was added. Secondly, the attributes for pof values were added and finally, a new base measure for pof values was introduced in the ISMO.

To our knowledge, there is no security measuring and adaptation approach that also uses design time information. Thus, the introduced approach is the first step towards enabling the use of the design time reliability predictions for runtime security measuring and adaptation. Reliability values are stored in a way that supports fast and easy updating. This is important when bug fixes for the security components are made. Furthermore, the real use of a component may produce different reliability to that initially predicted and it is then important to update the pof values. The presented approach is not restricted to one particular security mechanism or attribute. Hence, the software architect can make the decision of which attributes will be implemented in an adaptable manner on a case-by-case basis.

In the future, it is important to develop security measures that use the components' pof values in runtime security measuring. Current pof values of components can be used to compare different security components. Moreover, combining the reliability information and security level supported by the component offers valuable information for adaptation purposes. This means that the ISMO will be enhanced by new analysis models. The RAP tool also needs new features for storing information automatically to the ISMO.

ACKNOWLEDGMENT

This work is being carried out in the ARTEMIS SOFIA project funded by Tekes, VTT, and the European Commission.

REFERENCES

[1] R. Savola and H. Abie. "Development of measurable security for a distributed messaging system", *International Journal on Advances in Security, 2(4)*, pp. 358-380, 2010.

[2] A. Evesti, R. Savola, E. Ovaska, and J. Kuusijärvi, "The Design, Instantiation, and Usage of Information Security Measuring Ontology", *MOPAS'2011*, pp. 1-9, 17th April, 2011. 2011.

[3] C. J. Lamprecht and A. P. A. van Moorsel, "Runtime Security Adaptation Using Adaptive SSL", *Dependable Computing, 2008. PRDC '08. 14th IEEE Pacific Rim International Symposium*, pp. 305-312, 2008.

[4] A. Klenk, H. Niedermayer, M. Masekowsky, and G. Carle, "An architecture for autonomic security adaptation", *Ann Telecommun, 61(9-10)*, pp. 1066-1082. 2006.

[5] R. Hulsebosch, M. Bargh, G. Lenzini, P. Ebben, and S. Iacob. "Context sensitive adaptive authentication", *Smart Sensing and Context*, pp. 93-109, 2007.

[6] A. Evesti and E. Ovaska, "Ontology-Based Security Adaptation at Run-Time", *4th IEEE International Conference on Self-Adaptive and Self-Organizing Systems (SASO)*, pp. 204-212, 2010.

[7] A. Immonen, "A method for predicting reliability and availability at the architecture level", in *Software Product Lines* T. Käkölä and J. Dueñas, Eds., 2006.

[8] R. H. Reussner, H. W. Schmidt, and I. H. Poernomo, "Reliability prediction for component-based software architectures", *J. Syst. Software, 66(3)*, pp. 241-252. 2003.

[9] ISO/IEC 9126-1:2001. *Software Engineering – Product Quality – Part 1: Quality Model.* 2001.

[10] E. Ovaska, A. Evesti, K. Henttonen, M. Palviainen, and P. Aho, "Knowledge based quality-driven architecture design and evaluation", *Information and Software Technology, 52(6)*, pp. 577-601. 2010.

[11] M. Palviainen, A. Evesti, and E. Ovaska, "The reliability estimation, prediction and measuring of component-based software", *J. Syst. Software, 84(6)*, pp. 1054-1070. 2011.

[12] E. Niemela, A. Evesti, and P. Savolainen, "Modeling quality attribute variability", *ENASE – Proc. Int. Conf. Eval. Novel Approaches Software Eng.*, pp. 169-176, 2008.

[13] F. García, M. F. Bertoa, C. Calero, A. Vallecillo, F. Ruíz, M. Piattini, and M. Genero, "Towards a consistent terminology for software measurement", *Information and Software Technology, 48(8)*, pp. 631-644. 2006.

 ISBN: 978-1-61208-174-8

Incident Detection for Cloud Environments

Frank Doelitzscher, Christoph Reich, Martin Knahl
Cloud Research Lab
Furtwangen University
Furtwangen, Germany
{*Frank.Doelitzscher, Christoph.Reich, Martin.Knahl*}
@hs-furtwangen.de

Nathan Clarke
Centre for Security, Communications and Network Research
University of Plymouth
Plymouth PL4 8AA, United Kingdom
N.Clarke@plymouth.ac.uk

***Abstract*—Security and privacy concerns hinder a broad adoption of cloud computing in industry. In this paper we identify cloud specific security risks and introduce the cloud incident detection system Security Audit as a Service (SAaaS). SAaaS is built on autonomous distributed agents feeding a complex event processing engine, informing about a cloud's security state. In addition to technical monitoring factors like number of open network connections business process flows can be modelled to detect customer overlapping security incidents. In case of identified attacks actions can be defined to protect the cloud service assets. As contribution of this paper we provide a high-level design of the SAaaS architecture and a first prototype of a virtual machine agent. We show how an incident detection system for a cloud environment should be designed to address cloud specific security problems.**

***Keywords*-cloud computing; security; autonomous agents.**

I. INTRODUCTION

Enterprise analysts and research identified cloud specific security problems as the major research area in cloud computing [1][2][3][4]. Since security is still a competitive challenge for classic IT environments it is even more for cloud environments due to its characteristics like shared resources, multitenancy, access from everywhere, on-demand availability and 3rd party hosting. Although existing recommendations (ITIL), standards (ISO 27001:2005 and laws (e.g., Germanys Federal Data Protection Act) provide well-established security and privacy rulesets for data center providers, research [5][1] is showing they are not sufficient for cloud environments. In classic IT infrastructures security audits and penetration tests are used to document a datacenter's compliance to security best practices or laws. But, the major shortcoming of a traditional security audit is that it only provides a snapshot of an environments' security state at a given time (time of the audit was performed). This is adequate since classic IT infrastructures don't change that frequently. But because of the mentioned cloud characteristics above it is not sufficient for auditing a cloud environment. A cloud audit needs to consider the point of time when the infrastructure changes and the ability to decide if this change is considered as normal. Knowledge of underlying business processes is needed, for example that a new Virtual Machine (VM) gets created after a user's scalability threshold for its Webshop has been exceeded.

Therefore, we introduce an incident detection system for cloud environments based on autonomous agents, which collect data directly at the source, analyse and aggregate information and distribute it considering the underlying business process. To achieve this data interpretation gets supported by a Security Service Level Agreements (SSLA) policy modelling engine that allows to define monitoring events which consider business process flows. The usage of autonomous agents enables a behaviour anomaly detection of cloud components while maintaining the cloud specific flexibility. Our system respects the following cloud specific attributes:

- high number of distributed systems
- Frequently changing infrastructure due to the scalability advantages
- Multitenancy of users who are "owning" participating systems with administrator rights.

In the remainder of this paper, we first describe related work (Section II). Section III introduces the Security Audit as a Service (SAaaS) architecture which targets to solve the mentioned problems above. Why the paradigm of autonomous agents is valuable for incident detection in cloud environments is discussed in Section IV and a first SAaaS agent prototype gets presented. Subsequently (Section V), we discuss cloud specific security issues, which are addressed by the presented SAaaS architecture. Section VI concludes the paper and informs about future work.

II. RELATED WORK

This section covers related research work. First, we show current literature identifying cloud security issues. Following, we are discussing other cloud security research projects in contrast to SAaaS and the usage of autonomous agents for systems security .

The most comprehensive survey about current literature addressing cloud security issues is given by Vaquero et al. in [3]. It categorises the most widely accepted cloud security issues into three different domains of the Infrastructure as a Service (IaaS) model: machine virtualization, network

 ISBN: 978-1-61208-174-8

virtualization and physical domain. It also proposes prevention frameworks on several architectural levels to address the identified issues. While Chen et al. state in [4] that many IaaS-related cloud security problems are problems of traditionally computing solved by presented technology frameworks it also demands an architecture enabling "mutual trust" for cloud user and cloud provider. Both papers confirm and complete the cloud specific security issues identified by our research.

Raj et al. [6] introduce a virtualization service implemented as Xen VM extensions, which provides Role Based Access Control (RBAC) based on a trust value of a VM. This trust is based upon a VMs attributes like number of open network connections. Access to different cloud services like file access is given on a VMs' trust value. The presented implementation methods are following the same idea as the SAaaS architecture: trust generation via behavioural monitoring to build a "normal" cloud usage profile. The implementation presented is mainly based on Xen tools. Since SAaaS is build upon the CloudIA infrastructure which uses KVM corresponding tools need to be identified/implemented.

Zamboni et al. present in [7] how traditional Intrusion Detection Systems (IDS) can be enhanced by using autonomous agents. They confirm the advantages of using autonomous agents in regards to scalability and system overlapping security event detection. In contrast to our SAaaS architecture their research is focusing on the detection of intrusions into a relatively closed environment whereas our work applies an open (cloud) environment where incidents like abuse of resources needs to be detected. Mo et al. introduce in [8] an IDS based on distributed agents using the mobile technology. They show how mobile agents can support anomaly detection thereby overcoming the flaws of traditional intrusion detection in accuracy and performance. The paradigm of cooperating distributed autonomous agents and its corresponding advantages for IDS' is shown by Sengupta et al. in [9]. The presented advantages apply for our SAaaS agents as well.

III. Security Audit as a Service architecture

While distributed monitoring sensors are a well known procedure in intrusion detection systems (IDS) for traditional IT systems they do not cover the security needs of cloud environments. They are not flexible enough to monitor such a complex environment in a user manageable fashion. Mostly because existing architectures are built around a single monolithic entity which is not scalable enough to do data collection and processing in an efficient and meaningful way [10]. To mitigate this, we propose an autonomous agent-based intrusion detection system for cloud computing: Security Audit as a Service (SAaaS). The SAaaS architecture aims to support the following scenarios.

A. SAaaS Target Scenarios

A) Monitoring and audit of cloud instances User VMs run in a cloud infrastructure are equipped with an SAaaS agent. The user defines Security Service Level Agreements defining which behaviour of this VM in considered "normal", which VM components are to be monitored and how to alert in case of system deviation from the defined manner. The status gets conditioned in a user friendly format in a webportal - the SAaaS security dashboard. This continuous monitoring creates transparency about the security status of a user's cloud VMs hence increasing the user's trust into the cloud environment.

B.) Cloud infrastructure monitoring and audit The security status of the entire cloud environment, especially the cloud management system, access to customer data and data paths are monitored. This way customer-spanning monitoring is used by the cloud provider as well as a 3rd party, like a security service provider to ensure the overall cloud security status. Standardised interfaces enable security audits of a cloud infrastructure which can lead to a cloud security certification.

B. Typical SAaaS Use Case

To fullfill the presented scenarios we are proposing to use an autonomous agent system to monitor cloud environments. Before explaining the advantages of autonomous agents in detail we briefly want to explain the whole SAaaS event processing sequence. To support this consider the following example. Given a typical web application system consisting of a webserver, a load balancer and a database backend deployed at three VMs in a cloud. All VMs are equipped with SAaaS agents. The user's administrator installs the three VMs with the necessary software, e.g., Apache webserver, Tomcat load balancer, MySQL database. After the functional configuration the monitoring configuration gets designed in form of **S**ecurity **S**ervice **L**evel **A**greements (SSLAs). This can be technical rules like allowed user logins, allowed network protocols and connections between VMs, or that the webserver configuration is finished and an alarm should be raised if changes to its config files are detected. Furthermore SSLAs allow to design rules considering the system's business flow. For example: if a request (using the allowed protocols) to the load balancer or database VM without a preceding service request to the web application is detected this is rated as an abnormal behaviour which does not occur in a valid business process flow. Therefore, a monitoring event should be generated. If an event gets generated it first will be preprocessed by the SAaaS agent which is responsible for the monitoring target. This is important to reduce the overall messages sent to the cloud event processing system especially in large cloud computing environments. The SAaaS agent filters out possible VM dependent events like a started web application session from *IP 1.2.3.4*. A

 ISBN: 978-1-61208-174-8

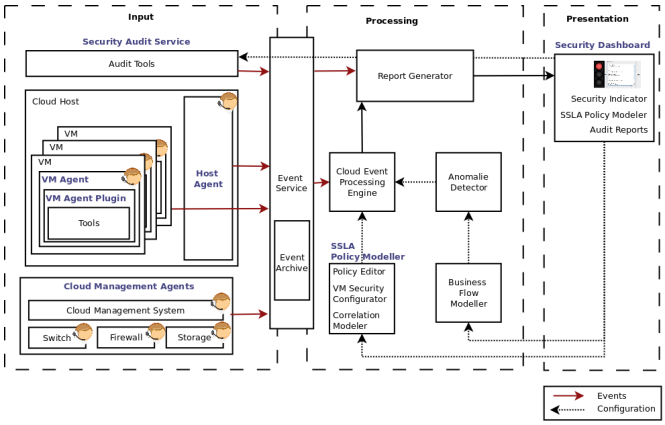


Figure 1. SAaaS event processing sequence

more abstracted event gets send to the cloud event processing system (a complex event processing (CEP) engine) to detect (possible) user overlapping security incidences. This could be a message containing the number of not completed web shop transactions by IP *IP 1.2.3.4* to pre-detect a Denial of Service attack. If the CEP engine detects an abnormal behaviour actions can be executed like warning the cloud provider's Computer Emergency Response Team, adjusting firewall settings or informing the cloud customer's admin.

Figure 1 gives a high level overview how events are generated, preprocessed, combined and forwarded within the SAaaS architecture. It can be divided into three logical layers: input, processing and output.

Input: The SAaaS architecture gets its monitoring information from distributed agents which are positioned at key points of the cloud's infrastructure to detect abnormal activities in a cloud environment. Possible key points are: running VMs of cloud users, the VM hosting systems, data storage, network transition points like virtual switches, hardware switches, firewalls, and especially the cloud management system. A `VM agent` integrates several monitor and policy enforcing tools. Therefore, it loads necessary VM agent plugins to interact with stand-alone `tools` like process monitor, intrusion detection system or anti virus scanner. It gets installed on a VM likewise on a cloud host. A logging component is recording the chronological sequence of occurrences building `audit trails`.

Processing: Each SAaaS agent receives security policies from the SSLA policy modeller component. Through security policies each agent gets a rule set (its intelligence) specifying actions in case of a specific occurrence (e.g., modification of a freezed config file). Thus every occurrence gets first preprocessed by an agent which reduces communication between VM agent and Cloud Management Agent. Self learning algorithms will be evaluated to improve an agents' intelligence. The `Security Service Level Agreements policy modeller` consists of a policy editor, a VM security configurator and a semantic correlation modeller to enable cloud user to design Security Service Level Agreements and security policies. An example for a SSLA rule could be: "In case of a successfully detected rootkit attack on a VM running on the same cloud as a users VM, the user VM gets moved to a different host to minish the risk of further damage." whereas a security policy could state: "In case a modification attempt of a file within /etc/php5/ gets detected, deny it and send an email to the cloud administrator." Security policies get send from the Security Audit Service to the corresponding agents. Using the monitoring information of the distributes agents in combination with the SSLAs a `cloud behaviour model` is build up for every cloud user. SSLAs are also used as input for the Cloud Management Agent to detect user overlapping audit events. Forwarded higher level events are processed by a `complex event processing (CEP) engine`. It is also fed with the modelled business flows from the `Business Flow Modeller` to aggregate information and detect behaviour anomalies. Countermeasures can then be applied to early detect and prohibit security or privacy breaches. The `Report Generator` conditions events, corresponding security status as well as audit report results in a human friendly presentation.

Presentation: As a single interaction point to cloud users the `Security Dashboard` provides usage profiles, trends, anomalies and cloud instances' security status (e.g., patch level). Information are organised in different granular hierarchies depending on the information detail necessary. At the highest level a simple three colour indicator informs about a users cloud services overall status.

Communication between the distributed agents and the Security Dashboard is handled by an `Event Service`. Events will use a standardised message format which is not defined yet. Our first prototype implements the Intrusion Detection Message Exchange Format (IDMEF). Events are also stored in an `Event Archive`.

IV. Anomaly Detection Using Autonomous Agents

In this section, we are showing the advantages of using distributed autonomous agents for incident detection in a cloud environment. Therefore, we first give a definition what can be considered as an autonomous agent.

A. Agent Definition

An agent can be defined as [11]:
"... a software entity which functions continuously and autonomously in a particular environment ... able to carry out activities in a flexible and intelligent manner that is responsive to changes in the environment ... Ideally, an agent that functions continuously ... would be able to learn from its experience. In addition, we expect an agent that inhabits

 ISBN: 978-1-61208-174-8

an environment with other agents and processes to be able to communicate and cooperate with them ..."

Since the agents in the SAaaS architecture are running independently, not necessarily connected to a certain central instance, are self-defending and self-acting, we term them *autonomous*. Agents can receive data from other instances, e.g., the policy module and send information to other instances like other agents or SAaaS' event processing system. The "central" event processing system gets itself implemented as an agent which can be scaled and distributed over multiple VMs.

B. How agents can improve incident detection

Incident detection in cloud environments is a non trivial task due to its characteristics as discussed in Section I. Therefore, it is important to have a high number of sensors capturing simple events. Preprocessed and combined complex events can be generated reducing the possibility of "event storms". Combined with knowledge about business process flows it will be possible to detect security incidents like discussed in Section V, while keeping the network load low.

The usage of autonomous agents delivers this possibility because agents are *independent units* that can be added, removed or reconfigurated during runtime without altering other components. Thus, the amount of monitoring entities (e.g., network connections of a VM, running processes, storage access, etc.) of a cloud instance can be changed without restarting the incident detection system. Simultaneously using agents can *save computing resources* since the underlying business process flow can be taken into account. Imagine a business process of a web application `P1` where user Bob adds a new user to a user database by filling out a web form. By pressing the "Save" button a legal request R gets executed as part of business process `P1`. An agent `A` monitoring database access can get moved at the beginning of `R` to the request-executing VM `V1`, monitoring the data access during process time and gets deleted from `V1` after `P` is finished. Furthermore agents can be updated to new versions (depending their interface remains unchanged) without restarting the whole incident detection system or other SAaaS agents running at a VM.

While single agents can monitor simple events (e.g., user login on VM) and share them with other agents *complex events can be detected*. Given the scenario of a successful unauthorised login of an attacker at a virtual machine `VM2`, misusing a webserver's directory to deposit malicious content for instance a trojan. Agent `A1` monitors the user login, agent `A2` detects the change of a directory content and agent `A3` detects a download of a not known file (the trojan). Instead of sending those three simple messages to a central event processing unit a VM agent can collect them conditioning one higher level event message that `VM2` was hijacked. This can result in a predefined action by the Cloud Management Agent, e.g., moving a hijacked VM into a quarantine environment, alerting the user and simultaneously starting a fresh instance of `VM2` based on its VM image.

By ordering agents in a hierarchical structure and preprocessing of detected events reduces network load originated from the incidents detection system. Furthermore this makes the system more scalable by reducing data sent to upper system layers. This is introduced and used in [12]. Combining events from system deployed agents (e.g., VM agent, host agent) and infrastructure monitoring agents (network agent, firewall agent) incident detection is not limited to either host or network based sensors which is especially important for the characteristics of cloud environments.

Furthermore using autonomous agents has advantages in case of a system failure. Agents can monitor the existence of co-located agents. If an agent stops for whatever reasons this stays not undetected. Concepts of asymmetric cryptography or Trusted Platform Module (TPM) technology can be used to guarantee the integrity of a (re-)started agent. If an agent stops the damage is restricted to this single agent or a small subset of connected agents which are requiring information from this agent.

C. SAaaS Agent prototype

For the SAaaS architecture we evaluated existing agent frameworks with the following requirements:

- Agents can be deployed, moved, updated during runtime
- Agent performance
- Open Source software platform
- Documentation & community support

Since our cloud environment at HFU's Cloud Research Lab CloudIA [13] is build around the cloud management system Open Nebula another requirement was the agent programming language: Java. As a result we choose the Java Agent DEvelopment Platform (JADE), which enables the implementation of multi-agent systems and complies to FIPA (IEEE Computer Society standards organisation for agent-based technology and its interoperability with other technologies) specifications. Furthermore it already provides a user interface which alleviates agents creation, deployment and testing.

Figure 2 illustrates a basic agent architecture we already assumed in the SAaaS Use Case presentation in Section III-B. It shows three SAaaS VM agents. Agents life in an agent platform which provides them with basic services such as message delivery. A platform is composed of one or more Containers. Containers can be executed on different hosts thus achieving a distributed platform. Each container can contain zero or more agents [14]. To provide monitoring functionality a VM agent interacts through agent plugins with stand-alone tools like process monitor, intrusion detection system or anti virus scanner, as depicted in Figure

 ISBN: 978-1-61208-174-8

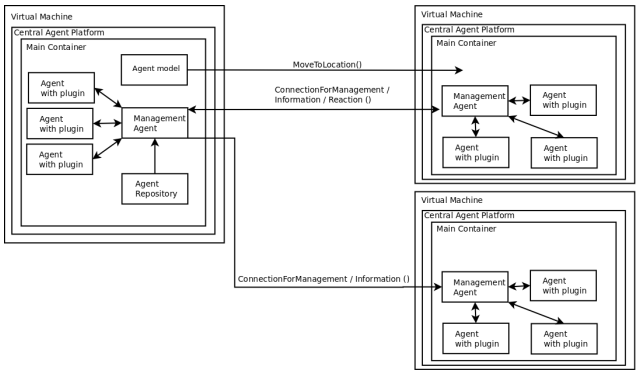


Figure 2. Basic SAaaS agent design

2. To harness the potential of cloud computing an agent can be deployed to a VM on-demand according to the SSLA policies a user defines. Different agents based on modelled business processes are stored within an agent repository. To be able to move a JADE agent to a running cloud instance the Inter Platform Mobility Service (IPMS) by Cucurull et al. [15] was integrated. This supports the presented advantage of deploying agents on-demand if a designed business process flow was started (as described in Section IV-B). Though this implementation is up to future work.

As a first prototype, a two layered agent platform was developed, consisting of a `VM agent` running inside a VM, a `Cloud Management Agent` running as a service at a dedicated VM feeding information to a `Security Dashboard`. Since all cloud VMs in CloudIA are Linux based, only Open Source Linux tools were considered during our research. Two notification mechanisms were implemented: a) the tool sends agent compatible events directly to the agent plugin; b) the tool writes events in a proprietary format into a logfile which gets parsed by an agent plugin. As for mechanism a) the filesystem changes monitoring tool *inotify* was used, whereas for mechanism b) *fail2ban* [15], an intrusion prevention framework was chosen.

For demo purposes a simple web frontend was written which offers to launch several attack scenarios on a VM agent equipped VM in CloudIA. Before/After tests were performed to validate, that an attack was detected and (depending on the plugin's configuration) prohibited. A prototype version of the Security Dashboard, showing a signal light indicator informed about occurring events. When started it shows a green light. After launching an attack, the Security Dashboard indicator light changes its colour to yellow or red as defined in a severity matrix given the type of detected attack.

D. Agent Performance Test

It is essential for the SAaaS architecture that the agents are very efficient not causing a high offset of resource consumption. JADE agent performance is very low as demonstrated by E.Cortese et al. in [16]. They show that CPU overhead is very low. Average round trip time of a message between to agents (request message, answer message) with a message content of seven characters takes only 13,4 ms. Jurasovic et al. [17] show that even with increasing message size the round trip time does not increase significantly. Also the message overhead by the agent communication does not increase significantly with increasing message size. Details about the used test lab are given in the mentioned literature.

In our first prototype, we wanted to see how fast an agent can be deployed to a new platform. All tests were done at the university's research cloud infrastructure CloudIA. Hardware of machines hosting the VMs was: 8x CPU: Intel(R) Xeon(R) CPU E5504 @ 2.00GHz 64-bit architecture, 12 GB of memory and 1 Gigabit Ethernet. Each VM was assigned with 512 MB RAM, 274 MB Swap, 1 CPU and 4GB HDD local storage. Over all test runs we confirmed that the average time for an agent move is below 1,5 seconds. This proves the applicability of the JADE agent platform to support the presented SAaaS use case.

V. Discussion - Cloud Specific Security Issues addressed by SAaaS

The German Federal Office for Information Security publishes the IT baseline protection catalogues enabling enterprises to achieve an appropriate security level for all types of information. The catalogues were extended by a special module covering virtualization in 2010. In a comprehensive study on all IT baseline protection catalogues as well as current scientific literature available [1][18][2][3][4], we made a comparison between classic IT-Housing, IT-Outsourcing and cloud computing. The following cloud specific security issues were identified as solvable by the SAaaS system:

Abuse of cloud resources Cloud computing advantages are also used by hackers, enabling them to have a big amount of computing power for a relatively decent price, startable in no time. Cloud infrastructure gets used to crack WPA, and PGP keys as well as to host malware, trojans, software exploits used by phishing attacks or to build botnets like the Zeus botnet. The problem of malicious insiders also exists in classical IT-Outsourcing but gets amplified in cloud computing through the lack of transparency into provider process and procedure. This issue affects authorisation, integrity, non-repudiation and privacy. Strong monitoring of user activities on all cloud infrastructure components is necessary to increase transparency. The presented SAaaS scenario A) Monitoring and audit of cloud instances addresses this problem.

Missing security monitoring in cloud infrastructure Security incidents in cloud environments occur and (normally) get fixed by the cloud provider. But to our best knowledge no cloud provider so far provides a system

which informs user promptly if the cloud infrastructure gets attacked, enabling them to evaluate the risk of keeping their cloud services productive during the attack. Thereby the customer must not necessarily be a victim of the attack, but still might be informed to decide about the continuity of his running cloud service. Furthermore no cloud provider so far shares information about possible security issues caused by software running directly on cloud host machines. In an event of a possible 0-day exploit in software running on cloud hosts (e.g., hypervisor, OS kernel) cloud customer blindly depend on a working patch management of the cloud provider. The presented SAaaS scenario B) Cloud infrastructure monitoring and audit addresses this problem.

Defective isolation of shared resources In cloud computing isolation in depth is not easily achievable due to usage of rather complex virtualization technology like VMware, Xen or KVM. Persistent storage is shared between customers as well. Cloud provider advertise implemented reliability measures to pretend data loss like replicating data up to six times. In contrast customer have no possibility to prove if all these copies get securely erased in case they quit with the provider and this storage gets newly assigned to a different customer. While the presented SAaaS architecture does not directly increase isolation in depth it adds to the detection of security breaches helping to contain its damage by the presented actions.

VI. Conclusion and Future Work

In this paper, we introduced the Security Audit as a Service architecture to mitigate the shortcoming traditional audit systems suffer to audit cloud computing environments. We showed the advantages of using autonomous agents as a source for sensor information. We explained how incident detection in clouds can be done by adding business process information to technical monitored events to perform anomaly detection in clouds.

As for future work, we identified the following tasks: a) comprehensive research in anomaly detection algorithms, b) comprehensive research in complex event processing, c) development of the SSLA policy modeller, d) development of SAaaS agents.

Acknowledgment

This research is supported by the German Federal Ministry of Education and Research (BMBF) through the research grant number 01BY1116.

References

[1] Cloud Security Alliance, "Security Guidance for Critical Areas of Focus in Cloud Computing v2.1," 12 2009.

[2] European Network and Information Security Agency, "Cloud Computing Security Risk Assessment," Tech. Rep., 11 2009.

[3] L. Vaquero, L. Rodero-Merino, and D. Morn, "Locking the sky: a survey on iaas cloud security," *Computing*, vol. 91, pp. 93–118.

[4] Y. Chen, V. Paxson, and R. H. Katz, "What's New About Cloud Computing Security?" EECS Department, University of California, Berkeley, Tech. Rep. UCB/EECS-2010-5, 01 2010.

[5] F. Doelitzscher, C. Reich, and A. Sulistio, "Designing cloud services adhering to government privacy laws," in *Proceedings of 10th IEEE International Conference on Computer and Information Technology (CIT 2010)*, 2010, pp. 930–935.

[6] H. Raj and K. Schwan, "Extending virtualization services with trust guarantees via behavioral monitoring," in *Proceedings of the 1st EuroSys Workshop on Virtualization Technology for Dependable Systems*, ser. VDTS '09. New York, NY, USA: ACM, 2009, pp. 24–29.

[7] J. Balasubramaniyan, J. Garcia-Fernandez, D. Isacoff, E. Spafford, and D. Zamboni, "An architecture for intrusion detection using autonomous agents," in *Computer Security Applications Conference, 1998, Proceedings., 14th Annual*, dec 1998, pp. 13 –24.

[8] Y. Mo, Y. Ma, and L. Xu, "Design and implementation of intrusion detection based on mobile agents," in *IT in Medicine and Education, 2008. ITME 2008. IEEE International Symposium on*, dec. 2008, pp. 278 –281.

[9] J. Sen, I. Sengupta, and P. Chowdhury, "An architecture of a distributed intrusion detection system using cooperating agents," in *Computing Informatics, 2006. ICOCI '06. International Conference on*, june 2006, pp. 1 –6.

[10] E. H. Spafford and D. Zamboni, "Intrusion detection using autonomous agents," *Computer Networks*, vol. 34, no. 4, pp. 547 – 570, 2000, recent Advances in Intrusion Detection Systems.

[11] J. M. Bradshaw, *An introduction to software agents*. Cambridge, MA, USA: MIT Press, 1997, pp. 3–46.

[12] S. Staniford-chen, S. Cheung, R. Crawford, M. Dilger, J. Frank, J. Hoagl, K. Levitt, C. Wee, R. Yip, and D. Zerkle, "Grids - a graph based intrusion detection system for large networks," in *In Proceedings of the 19th National Information Systems Security Conference*, 1996, pp. 361–370.

[13] A. Sulistio, C. Reich, and F. Doelitzscher, "Cloud Infrastructure & Applications - CloudIA," in *Proceedings of the 1st International Conference on Cloud Computing (CloudCom'09)*, Beijing, China, 2009.

[14] D. Grimshaw, "JADE Administration Tutorial," http://jade.tilab.com/doc/tutorials/JADEAdmin, 06.09.2011.

[15] J. Cucurull, R. Mart, G. Navarro-Arribas, S. Robles, B. Overeinder, and J. Borrell, "Agent mobility architecture based on ieee-fipa standards," *Computer Communications*, vol. 32, no. 4, pp. 712 – 729, 2009.

[16] E. Cortese, F. Quarta, G. Vitaglione, T. I. Lab, C. Direzionale, J. Message, and T. System, "Scalability and performance of jade message transport system," 2002.

[17] K. Jurasovic, G. Jezic, and M. Kusek, "A performance analysis of multi-agent systems." *ITSSA*, vol. 1, no. 4, pp. 335–342, 2006, http://dblp.uni-trier.de/db/ journals/itssa/itssa1.html#JurasovicJK06, 06.09.2011.

[18] Cloud Security Alliance, "Top Threats to Cloud Computing V1.0," 2010, https://cloudsecurityalliance.org/topthreats.html, 06.09.2011.

 ISBN: 978-1-61208-174-8

Multipath Routing for Survivability of Complex Networks Under Cascading Failures

Preetha Thulasiraman
Department of Electrical and Computer Engineering
Naval Postgraduate School
Monterey, CA, USA
pthulas1@nps.edu

***Abstract*—Complex infrastructure networks have been characterized as being scale-free and therefore maintain a heterogeneous node distribution. While scale free networks (SFN) have been investigated using vulnerability assessments, particularly that of cascading node failures, existing research has not dealt with the aftermath of these failures. This paper addresses the problem of discovering end to end paths in a SFN in the presence of cascading failures such that survivability is achieved for each source-destination pair. We first develop a model to capture cascading failures in SFNs while redistributing traffic load to neighboring nodes. Given the traffic distribution after the cascade of failures, we develop a routing algorithm such that backup connections are constructed for each source-destination pair. We formulate the routing algorithm by exploiting the multipath topology of SFNs and the different priorities of the traffic flows. We compare our routing approach in a SFN with that of a random network in which node distributions are homogeneous. We show that our routing algorithm performs well under intentional node attacks and efficiently considers the classification of the traffic when constructing alternate routing paths.**

Keywords – **scale free networks; multipath routing; cascading failures; load redistribution**

I. INTRODUCTION

Complex networks such as the Internet, electrical power grid and telecommunication and transportation systems are an essential part of the global society. These infrastructure networks are not random but rather known to be scale free, with some nodes having a tremendous number of connections, whereas others have only a few connections [1]. This highly heterogeneous node distribution has led researchers to prove that the degrees of the nodes in scale free networks (SFN) follow a power law distribution: the probability that any node is connected to k other nodes, $P(k)$, is proportional to $\frac{1}{k^n}$, where n is a parameter whose value is typically in the range $2 < n < 3$ [1]. Due to its topology, SFNs are robust against accidental failures (which tend to affect low degree nodes), but are vulnerable to coordinated attacks that target highly connected nodes in order to inflict maximum damage by disabling numerous connections.

The fragile properties of SFNs become more evident when the intrinsic dynamics of the network flows are taken into account. Specifically, due to the existence of many simultaneous traffic flows, the removal of a single, highly or moderately connected node can cause large scale cascading failures. This domino effect results in the interruption of traffic flow, service, and distribution of network resources. Thus, the vulnerability and reliability of SFNs in the face of attacks must be investigated.

The notion of survivability is an essential aspect of reliable communications. Survivability consists of the ability of the network to continue to deliver and preserve essential services in the presence of failures. These failures can occur due to natural faults and other unintentional errors or due to malicious adversaries. From the viewpoint of network resilience and survivability, a key question is whether a SFN can, in the face of dependent and correlated node failures, retain its functionality in terms of maintaining some sense of global communication. In this regard, traffic redistribution and robustness of routing policies for SFNs is a central problem which is gaining increased attention with a growing awareness to safeguard critical infrastructure networks.

A. Related Work and Motivations

Over the years, researchers have investigated the cascade based attack vulnerability of either specific infrastructure SFNs, such as the power grid [2], [3], or that of general SFNs with heterogeneous traffic load distributions [4], [5], [6]. In these works, different cascading failure models are analyzed to determine the best manner in which traffic load should be redistributed to maintain service. With advances in cyber based communication systems and their logical coupling to infrastructure networks [7], it is imperative that the vulnerabilities and consequences of node failures are studied from the perspective of network survivability [8], [9]. Survivability of networks depends on three key capabilities: resistance, recognition, and recovery [10]. While resistance repels failures from happening, recognition and recovery deal with and evaluate the failures to provide network restoration protocols. Thus far, the research on providing survivable network solutions to infrastructure networks has been tailored to focus on failure modeling and vulnerability assessments rather than network management [11], [12]. It is important not only to understand how to recognize faults and vulnerabilities but also how to recover from them.

Multipath routing has long been recognized as an effective strategy to increase reliability. To improve the transmission reliability, the multiple paths can be selected to be node

 ISBN: 978-1-61208-174-8

disjoint. Disjoint multipath routing provides better robustness and a greater degree of fault tolerance than compared to the generic multipath routing scheme. Due to these advantages, disjoint multipath routing has been researched in order to enhance network survivability [13], [14].

SFNs are inherently highly connected, thus there always exists two or more paths between each source-destination pair. When a node fails in a SFN, potentially causing a cascade of node failures, the traffic flows that use the failed node should be maintained and the services they provide must be sustained. The aim of this paper is to ensure end to end survivability by bypassing failed nodes using efficient, robust multipath routing in the presence of cascading failures, while redistributing the corresponding traffic loads accordingly.

B. Contributions and Organization

The contributions of this paper are two-fold. First, we develop a local traffic redistribution model for a failed node by redistributing its load uniformly among its neighbors, taking into consideration that this redistribution can possibly overload the neighboring nodes, causing a series of cascading failures. Second, given the redistribution of the load, we establish survivable shortest disjoint multipath routes that bypass the failed node(s). The shortest disjoint paths are determined by the priority of the traffic flows; some flows, due to its service requirements, require backup paths that are more reliable than others (i.e., to ensure service availability). Therefore, the backup path for each traffic flow should be determined using local topological and connection information. In other words, the shortest paths for increasing traffic flow priority are those between a source and destination that cumulatively traverse the least number of highly connected nodes.

The rest of this paper is organized as follows: Section II discusses the system model. Section III discusses the load redistribution model based on cascading failures and Section IV develops the disjoint multipath route selection procedure based on traffic priority. We show our performance analysis using simulations in Section V and conclude the paper in Section VI.

II. System Model

The topology used in this paper is that of a SFN. The Barabasi-Albert (BA) model is used to generate SFNs with a power law degree distribution. The BA model is a well known algorithm for generating random SFNs using a preferential attachment mechanism [15]. Without loss of generality, for the purposes of this work, we construct the underlying network structure using the BA network model.

We consider a SFN consisting of N nodes ($n = 1, ..., N$) and A directed arcs ($a = 1, ..., A$). We assume that there are K ($k = 1, ..., K$) different traffic flows that are routed through the SFN, where a traffic flow is defined as a set of demands from a source to destination. Each traffic flow has a level of service that has to be maintained, therefore a certain amount of capacity is required along each arc of the route taken by a traffic flow, k. Within these K traffic flows, there are M classes of priority numbered from 0 to $M-1$, where Class 0 represents the highest class and $M-1$ represents the lowest class. Because highly connected nodes in a SFN are more vulnerable to outside attack, it is critical that high priority traffic flows route along paths that contain the least number of highly connected nodes to ensure end to end route survivability.

A manner in which node connectivity is measured is by the betweenness centrality (BC) parameter of SFNs. The BC is a measure of the number of shortest paths that go through a node n [1]. Nodes that occur on many shortest paths have higher betweenness than those that do not and are therefore more vulnerable to a coordinated attack. The BC of a node is denoted as

$$BC(n) = \sum \frac{\delta_n(p,q)}{\delta(p,q)} \tag{1}$$

where $\delta_n(p,q)$ is the number of shortest paths between nodes p and q and $\delta_n(p,q)$ is the number of shortest paths between p and q that run through node n. The BC parameter provides information about the physical connectivity of each node for the purposes of routing.

For the purpose of modeling network node failures, the actual traffic load of a node (the amount of traffic that each node processes) must be considered. The traffic load of a node is directly related to its BC; the higher the BC, the higher the traffic load that the node has to support. In this paper we assume that highly connected nodes are more susceptible to attacks than those that are not highly connected. Therefore, our proposed cascading failure model and routing algorithm are developed under the scenario that a highly connected node has failed and has caused a series of cascading failures.

III. Cascading Failure Model: Local Redistribution of Failure Load

Each source-destination pair in a SFN has an active path. This is the path on which a traffic flow is typically routed. Active paths often run through highly connected nodes and are thus exposed to attacks. In order to find active paths on a shortest path basis, a cost is defined, κ_a, of an arc a as

$$\kappa_a^m = \frac{m}{M-1} d_a + \frac{(M-1)-m}{M-1} BC(n) \tag{2}$$

where m is the current class of traffic flow, $m = 0, 1, 2, ..., M-1$, d_a is the length of arc a, and $BC(n)$ is the betweenness centrality parameter.

The active paths that are determined with the above cost are used as the default routing connection. However, when a node fails, the path that uses this node and its load has to be redistributed. The redistribution of the load may cause further node failures due to overload. Fig. 1 illustrates an example of a failed node's traffic being redistributed to its neighbors. Note that SFNs are always at least 2-connected, meaning that each node will have at least two disjoint paths to every other node in the network [16]. Not all the connections for each node to show 2-connectivity are shown in Fig. 1. The network of Fig. 1 is simply for illustration of the load redistribution concept.

Within a SFN, we assume that every node has a minimum load value, L_{min} and a maximum value, L_{max}. All nodes have

 ISBN: 978-1-61208-174-8

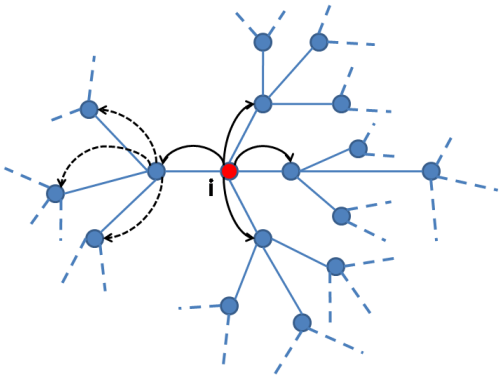


Fig. 1. Illustrates the load redistribution triggered by the failure of node i due to a coordinated attack. Node i is removed and its load is redistributed to the neighboring nodes

the same limit of operation, L_{fail}, beyond which they fail. To start the cascade, an initial disturbance causes the failure of a node. The algorithm for simulating the cascading failures proceeds in successive stages as shown in Fig. 2.

Cascading Failure Model

Step1: At stage $i = 0$, all N nodes are initially working under independent uniformly random initial loads $L_1, L_2, ..., L_N \ \forall \ [L_{min}, L_{max}]$, with $L_{max} < L_{fail}$.

Step2: An initial disturbance, causes a node to fail. This initial disturbance can either be a direct attack on the node itself or an overload.

Step3: The nodes' loads are incremented taking into account the neighboring topology of the failed node. Given that a node n has failed, $L*_n > L_{fail}$, its load $L*_n$ is spread uniformly among its neighbors. Each neighbor receives $\frac{L*_n}{d_n}$ portion of the load where d_n is the degree of the failed node. That is the total load of the failed node n is divided by the number of nodes to which node n is connected to in order to determine the amount of load each neighbor is incremented with.

Step4: If the neighborhood of the failed node is empty (i.e., if there are no functioning nodes connected to it), then the failure propagation comes to an end.

Fig. 2. Steps to model cascading failures given an initial disturbance

IV. Disjoint Multipath Survivable Routing under Cascading Failures

Once load is redistributed, the new topology configuration has to be considered when shortest paths are determined for each source-destination pair. When a node fails, the traffic flows that traverse that node need to be routed on alternate shortest paths that are disjoint from the original active paths. These backup paths allow for redistribution of end to end routing between nodes. The backup paths are determined based on the priority of the traffic flow. In this paper, our objective is to protect infrastructure networks against coordinated attacks on highly connected nodes. Therefore, high priority traffic must traverse the least number of highly connected nodes. This can deliver backup paths that are longer in length than other possible paths. Low priority traffic, if link capacities allow, can use backup paths with shortest hops as long as they are not taking away resources for high priority traffic. We formulate the discovery of backup paths as an integer linear program (ILP). We assume that a series of cascading failures does not partition the network, meaning that there will always exist at least one path between each pair of nodes in the network. The nomenclature used in the ILP formulation is shown in Table I.

TABLE I
Nomenclature Used in ILP

p_k - source node of a traffic flow k
q_k - destination node of a traffic flow k
λ_a - number of available channels on arc a
$\alpha^{\lambda}_{k,a}$ - takes value of 1 if channel λ of an arc a is used by an active path of traffic flow k; 0 otherwise
$\beta^{\lambda}_{k,a}$ - takes values of 1 is channel λ of an arc a is used by a backup path of traffic flow k; 0 otherwise
κ^{m}_{a} - cost of arc a (shown in Eq. 2)
$s_{k,a}$ - cost of an arc a calculated for traffic flow k on the backup path
C_a - capacity of an arc a
x - vector of all components of flows (variables)

Before developing the ILP, two boundary cases are worth mentioning. For Class 0, the highest priority class of traffic flow, the cost of an arc is calculated only on the basis of the $BC(n)$. This can be seen from Eq. 2. This results in finding backup paths that omit highly connected nodes. This causes the backup connections of Class 0 traffic to have a low probability of breaking. However, the backup paths may not be the shortest ones. For Class $M-1$, the lowest priority traffic flow, the backup connections do not have to be guaranteed service continuity. For these flows, the cost of each arc is determined solely by the length of the arc, d_a. For all other classes of traffic flows, the cost of the arcs are determined using Eq. 2.

The ILP shown below finds backup paths while minimizing the linear cost of the paths.

Objective Function

$$\varphi(x) = \text{minimize} \quad \sum_{k=1}^{K} \sum_{a=1}^{A} \sum_{\lambda=1}^{\lambda_a} (\kappa^{m}_{a} \cdot \alpha^{\lambda}_{k,a} + s_{k,a} \cdot \beta^{\lambda}_{k,a}) \quad (3)$$

subject to the following constraints

a) Capacity constraints on the number of available channels on an arc a

$$\sum_{\lambda=1}^{\lambda_a} \sum_{k=1}^{K} (\alpha^{\lambda}_{k,a} + \beta^{\lambda}_{k,a}) \leq C_a, \forall a \in A \quad (4)$$

b) Flow balance constraints for each channel λ and for each demand k

For a source node of an active path

$$\sum_{a=(q_k,j),j\neq q_k} \alpha_{k,a}^{\lambda} - \sum_{a=(i,p_k),i\neq p_k} \alpha_{k,a}^{\lambda} = 1, \quad \forall i,j \in N, \forall k \in K, \forall a \in A, \forall \lambda \in \lambda_a \quad (5)$$

For a destination node of an active path

$$\sum_{a=(q_k,j),j\neq q_k} \alpha_{k,a}^{\lambda} - \sum_{a=(i,p_k),i\neq p_k} \alpha_{k,a}^{\lambda} = -1, \quad \forall i,j \in N, \forall k \in K, \forall a \in A, \forall \lambda \in \lambda_a \quad (6)$$

For intermediate nodes of an active path

$$\sum_{a=(i,j),i,j\neq q_k,i,j\neq p_k} \alpha_{k,a}^{\lambda} - \sum_{a=(i,j),i,j\neq q_k,i,j\neq p_k} \alpha_{k,a}^{\lambda} = 0, \quad \forall i,j \in N, \forall k \in K, \forall a \in A, \forall \lambda \in \lambda_a \quad (7)$$

For a source node of a backup path

$$\sum_{a=(q_k,j),j\neq q_k} \beta_{k,a}^{\lambda} - \sum_{a=(i,p_k),i\neq p_k} \beta_{k,a}^{\lambda} = 1, \quad \forall i,j \in N, \forall k \in K, \forall a \in A, \forall \lambda \in \lambda_a \quad (8)$$

For a destination node of a backup path

$$\sum_{a=(q_k,j),j\neq q_k} \beta_{k,a}^{\lambda} - \sum_{a=(i,p_k),i\neq p_k} \beta_{k,a}^{\lambda} = -1, \quad \forall i,j \in N, \forall k \in K, \forall a \in A, \forall \lambda \in \lambda_a \quad (9)$$

For intermediate nodes of a backup path

$$\sum_{a=(i,j),i,j\neq q_k,i,j\neq p_k} \beta_{k,a}^{\lambda} - \sum_{a=(i,j),i,j\neq q_k,i,j\neq p_k} \beta_{k,a}^{\lambda} = 0, \quad \forall i,j \in N, \forall k \in K, \forall a \in A, \forall \lambda \in \lambda_a \quad (10)$$

c) Constraints to ensure node disjointness of active and backup paths.

$$\sum_{\lambda=1}^{\lambda_a} \sum_{a=(i,j),j\neq i,i\neq p_k} (\alpha_{k,a}^{\lambda} + \beta_{k,a}^{\lambda}) \leq 1, \quad \forall i,j \in N, \forall a \in A, \forall k \in K \quad (11)$$

$$\sum_{\lambda=1}^{\lambda_a} \sum_{a=(i,j),j\neq i,j\neq p_k} (\alpha_{k,a}^{\lambda} + \beta_{k,a}^{\lambda}) \leq 1, \quad \forall i,j \in N, \forall a \in A, \forall k \in K \quad (12)$$

The constraint given in Eq. 4, assures that the total number of channels, reserved for survivable connections on an arc a, will not exceed the capacity of this arc. For each channel and each demand, flow balance for the active paths is assured by Eqs. 5-7. Eq. 7 simply states that the intermediate nodes do not store traffic. Eqs. 8-10 describe the flow balance constraints for backup paths. Eqs. 11 and 12 reflect the requirement that the active and backup paths be node disjoint.

V. Performance Evaluation

In this section, we evaluate the performance of our cascading failure model and end to end survivable routing algorithm via simulations. We consider a SFN generated by the BA model [15] and compare it to a random graph generated by the algorithm given in [17]. For fairness, the number of nodes and number of links in both are randomly set to be 1470 and 3131, respectively. The number of nodes and links chosen generate networks that are at least two connected to ensure disjoint paths can be obtained for each node in the face of a cascading failure. The BA model follows the power law degree distribution, while the degree distribution of a random graph is Poisson. Unlike a SFN, the random graph is a homogeneous network, in which there is no node with an enormous number of connections. In each network, we randomly generate 4000 traffic flows (i.e., K=4000). The source and destination nodes for each flow are chosen randomly. Once a source-destination pair is chosen, a shortest path between them is determined using the cost metric given in Eq. 2. The capacity of the links in the network are determined by the traffic loads. Intuitively, links from highly connected nodes need larger capacity since more traffic loads go through them. Thus, the capacity of an arc (i,j) is given as

$$C_{ij} \propto BC(i) + BC(j) \quad (13)$$

where C_{ij} is the capacity of the arc, which is proportional to the sum of the betweenness of node i and node j. Comparing the definition of betweenness with the routing rule of the traffic flows, it can be concluded that the betweenness characterizes the average traffic load of a node [18]. In addition, each directed arc in the networks have 8 channels (i.e., $\lambda_a = 8$) available to them and are of equal length (i.e., d_a = 175km).

A. Simulation Results and Discussion: Cascading Failure Model

Given a network, to start a cascade, an initial disturbance is imposed on a node in the form of an extra load, D, which results in the failure of that node due to overload. This failure occurrence leads to the redistribution of the load to neighboring nodes, which may cause further failures. As the nodes become progressively more loaded, the cascade continues. The cascade propagation algorithm is embedded in a Monte Carlo simulation framework implemented in Matlab version 7.11.0. The damage caused by the cascades for any initial load, $[L_{min}, L_{max}]$, is quantified in terms of the number of nodes that have failed on average. This is referred to as the cascade size, S. It is assumed that each node operates in such a manner that the initial node loads are normalized between the range $L_{min} = 0$ to $L_{max} = L_{fail} = 1$. Large load values

represent highly loaded nodes where each node is on average operating close to its limit capacity, $L_{fail} = 1$. The range of load conditions is normalized from 0 to 1 so that the model for cascading failures is not limited to the propagation of failures in specific applications. As the simulation is repeated for different ranges of initial load, $[L_{min}, L_{max}]$, with $L_{max} = 1$ and $L_{min} \in [0, 1]$ the pair (L, S) is recorded.

Fig. 3 portrays the effect of propagation of failure. The analysis is performed for values of D that span the entire feasibility range $D \in [0, 1]$. Eight different initial disturbance values are used $D \in [0.8, 0.6, 0.4, 0.2, 0.1, 0.01, 0.001, 0.0001]$. The results reflect the simulation of the generated SFN.

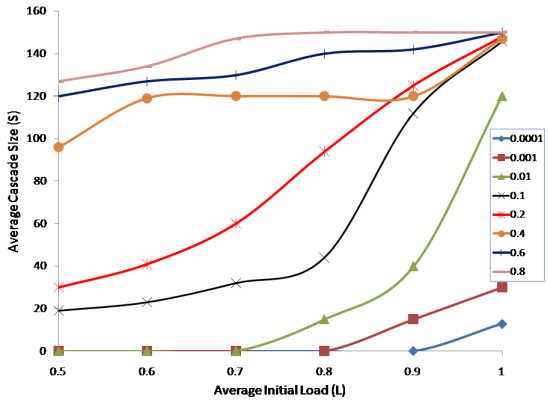


Fig. 3. Illustrates the average cascade size, S, versus the average initial load, L, for eight different values of the initial disturbance D. Each point is averaged for the same range of initial load $[L_{min}, L_{max}]$

From Fig. 3, it can be seen that a low D value causes almost no cascading failures, thus as D increases, the number of failures also increases. This intuitively makes sense since the value of D determines the strength of the disturbance.

B. Simulation Results and Discussion: Disjoint Multipath Survivable Routing Algorithm

To evaluate the performance of our disjoint multipath routing algorithm, we adopt the following performance metrics:

- Restoration time: restoration time is defined as the amount of time needed by the algorithm to construct an alternate path after the failure of a node.
- Bandwidth utilization ratio: the utilization ratio of the bandwidth is the total bandwidth used by the backup path to the total bandwidth provided (capacity) for different classes of traffic. This metric describes how well the backup paths use the bandwidth for different classes of traffic.

The routing algorithm was implemented using Matlab 7.11.0 and IBM's ILOG CPLEX optimizer. In these simulations we do not consider any route signalling mechanisms. We first look at the average restoration time of broken connections due to node failure as a function of the class of service. The results are shown in Fig. 4. We assume that there are 5 traffic classes, with Class 0 being the highest and Class 5 the lowest. The results shown were averaged over 10 simulation trials in which each trial has a different node failing, thereby causing a different cascading failure sequence and load distribution. It can be seen that the proposed multipath routing algorithm leads to a significant reduction in the restoration time for high traffic classes versus lower priority traffic. Thus, our routing algorithm efficiently takes into consideration the priority of the connection when constructing a backup path between a source-destination pair. Given the limited published research in routing for networks with cascading failures, the restoration time of our approach can not be compared at this time with existing fast recovery techniques.

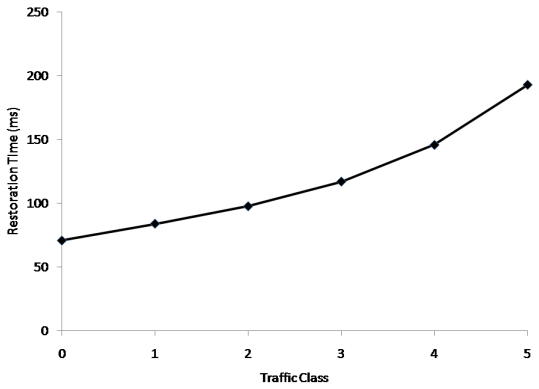


Fig. 4. Illustrates the average restoration time in milliseconds for 6 classes of traffic

We next look at the bandwidth utilization ratio for different classes of traffic versus the total capacity available. Fig. 5 shows the performance of the SFN network generated by the BA algorithm for a random failure and intentional failure (i.e., highly connected node removed) compared to the utilization for the original intact network for Class 0 traffic. The random attack curve in Fig. 5 overlaps with the original one, whereas the intentional attack curve is approximately 14% lower in terms of bandwidth utilization. The utilization ratio of the bandwidth decreases as the total capacity rises, which means that a higher percentage of bandwidth is wasted. The results obtained for Class 5 traffic are shown in Fig. 6. It can be seen that the bandwidth utilization for Class 5 traffic is higher than the Class 0 traffic results. This difference in utilization ratio results from the backup paths being longer for lower priority connections and therefore using more bandwidth. The results of both Figs. 5 and 6 indicate that the BA generated SFN is robust under random attack but fragile under intentional attack.

By contrast, Fig. 7 shows the results for a randomly generated graph. It can be seen that the random graph is robust to both random and intentional attacks; both curves perform similarly to the original curve. There is only a slight decline in utilization ratio when the network is intentionally attacked. Because the random graph is homogeneous, the traffic is well distributed among all the nodes. Therefore, the attack on one node (no matter randomly or intentionally) has little effect on the traffic performance of the whole network. Due to space limitations, only the results for Class 0 traffic are shown for the random graph. Similar results were obtained for lower traffic

 ISBN: 978-1-61208-174-8

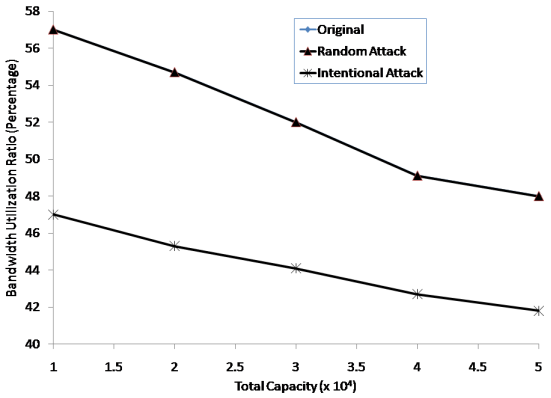


Fig. 5. Utilization ratio of the bandwidth in a BA generated SFN for Class 0 traffic (highest priority)

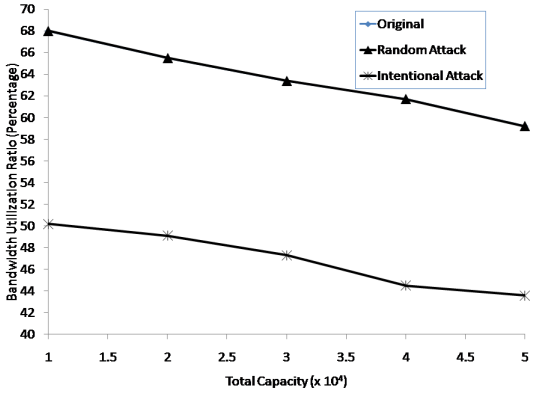


Fig. 6. Utilization ratio of the bandwidth in a BA generated SFN for Class 5 traffic (lowest priority)

classes.

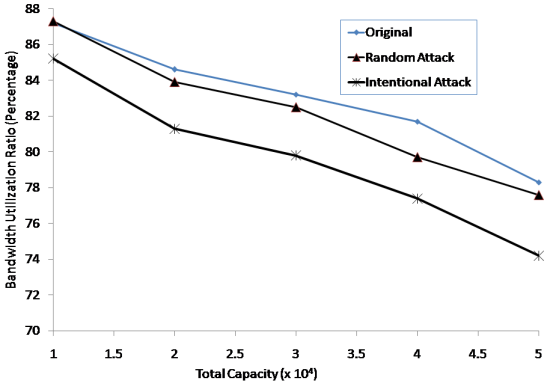


Fig. 7. Utilization ratio of the bandwidth in a randomly generated network for Class 0 traffic (highest priority)

VI. Conclusion

In this paper we have developed an end to end disjoint multipath survivable routing algorithm for SFNs in the presence of cascading node failures. We show that our algorithm effectively constructs alternate paths in a SFN considering the priority of the different traffic classes. We also show that our routing algorithm fares well when an intentional attack occurs. In our future work, we will look at improving the cascading failure model by redistributing load onto neighboring nodes based on the capacity of the nodes rather than using a uniform distribution. We will also introduce resource allocation mechanisms into the cascading failure routing scheme.

Acknowledgement

This work was funded by the Research Initiation Program (RIP) at the Naval Postgraduate School, Monterey, CA, USA.

References

[1] A. Laszlo and E. Bonabeau, "Scale free networks," *Scientific American*, pp. 50–59, May 2003.

[2] I. Dobson, B.A. Carrerras, V.E. Lynch, and D.E. Newman, "Complex systems analysis of series of blackouts: Cascading failure, critical points and self-organization," *Chaos, American Institute of Physics*, vol. 17, no. 2, pp. 1–13, June 2007.

[3] J.-W. Wang and L.-L. Rong, "Cascade-based attack vulnerability on the US power grid," *Safety Science (Elsevier)*, vol. 47, no. 10, pp. 1332–1336, December 2009.

[4] H.J. Sun, H. Zhao, and J.J. Wu, "A robust matching model of capacity to defense cascading failure on complex networks," *Physica A (Elsevier)*, vol. 387, no. 25, pp. 6431–6435, November 2008.

[5] X. Wang, S. Guan, and C.H. Lai, "Protecting infrastructure networks from cost-based attacks," *New Journal of Physics*, vol. 11, no. 3, pp. 1–9, March 2009.

[6] J.-W. Wang and L.-L. Rong, "A model for cascading failures in scale-free networks with a breakdown probability," *Physica A (Elsevier)*, vol. 388, no. 7, pp. 1289–1298, April 2009.

[7] J. Kopylec, A. D'Amico, and J. Goodall, *Visualizing Cascading Failures in Critical Cyber Infrastructures*, Springer, 2007.

[8] R. Zimmerman, "Decision-making and the vulnerability of interdependent critical infrastructure," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, 2004, pp. 4059–4063.

[9] R. Zimmerman and C. Restrepo, "The next step: Quantifying infrastructure interdependencies to improve security," *International Journal of Critical Infrastructures*, vol. 2, no. 23, pp. 215–230, February 2006.

[10] "Survivable mobile wireless networks: Issues, challenges, and research directions," in *Proceedings of ACM Workshop on Wireless Security*, 2002, pp. 31–40.

[11] Y.Y. Haimes, B.M. Horowitz, J.H. Lambert, J.R. Santos, C. Lian, and K.G. Crowther, "Inoperability inputoutput model for interdependent infrastructure sectors," *Journal of Infrastructure Systems*, vol. 11, pp. 67–709, June 2005.

[12] Y. Xia and D.J. Hill, "Attack vulnerability of complex communication networks," *IEEE Transactions on Circuits and Systmes-II: Express Briefs*, vol. 55, no. 1, pp. 65–69, January 2008.

[13] X. Huang and Y. Fang, "Multiconstrained qos multipath routing in wireless sensor networks," *Wireless Networks*, vol. 14, no. 4, pp. 465–478, 2008.

[14] P. Thulasiraman, J. Chen, and X. Shen, "Multipath routing and max-min fair qos provisioning under interference constraints in wireless multihop networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 22, no. 5, pp. 716–728, 2011.

[15] A.-L. Barabasi and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, October 1999.

[16] P. Thulasiraman, S. Ramasubramanian, and M. Krunz, "Disjoint multipath routing to two distinct drains in a multi-drain sensor network," in *Proceedings of IEEE International Conference on Computer Communications (INFOCOM)*, 2007, pp. 643–651.

[17] P. Erdos and A. Renyi, "On the evolution of randome graphs," in *Mathematical Institute of Hungarian Academy of Sciences*, 1960, pp. 17–60.

[18] K.-I Goh, B. Kahng, and D. Kim, "Universal behavior of load distribution in scale free networks," *Physical Review Letters*, vol. 87, no. 27, pp. 278–281, December 2001.

 ISBN: 978-1-61208-174-8

Association Control for Throughput Maximization and Energy Efficiency for Wireless LANs

Oyunchimeg Shagdar, Suhua Tang, Akio Hasegawa, Tatsuo Shibata, and Sadao Obana
ATR Adaptive Communications Research Laboratories,
Hikaridai, Keihanna Science City, Kyoto, Japan
{oyunaa, shtang, ahase, shibata, obana}@atr.jp

***Abstract*—Because the access points (APs) and the stations (STAs) of a community network are deployed at the users' desired places, the APs and STAs tend to concentrate in certain areas. A concentration of STAs often results in the AP(s) and STAs in that particular area suffering from severe congestion. On the other hand, a concentration of APs causes energy wastage. A proper association control can effectively alleviate congestion and improve network throughput. In this paper, we analytically formulate the throughput maximization problem and show that the existing association control schemes do not necessarily maximize throughput. Furthermore, while load balancing tends to use all the existing APs, sufficient performance can be achieved by utilizing fewer APs, especially in AP-concentrated areas. This enables lower energy consumption by putting unused APs in power-saving mode. To this end, we propose an association control scheme that aims at maximizing network throughout and reducing energy consumption. Using both computer simulations and testbed experiments, we confirm that the proposed scheme provides excellent performance and that it is feasible using the off-the-shelf WLAN devices.**

Keywords- association control, throughput maximization, congestion alleviation, energy efficiency

I. INTRODUCTION

Due to the increasing popularity of WLAN technology, users (i.e., STAs) are often in the vicinity of one or more APs deployed at offices, school campuses, hotspot areas (e.g., cafes, train stations, airports), and individuals' homes. Such a widespread deployment of WLAN triggered a launch of community networks, including FON [1], which are built exploiting user-owned APs. As community networks enable users to enjoy ubiquitous Internet access, it has the potential to play an important role in the future networking paradigm.

The fundamental difference of community networks from e.g., enterprise wireless access networks is that the APs of a community network are deployed at the users' (i.e., the owners of the APs) desired places and they are generally not movable in a systematic manner. APs are often concentrated in areas such as a residential street, and their distribution is highly non-uniform. Figure 1 shows a FON map for an approximately 600mx600m area in Tokyo (near Akihabara station) where 14 APs are installed in total. However, 8 APs are concentrated in a small area, approximately 1/8 of the overall area, in the upper left part of the map. In fact, the majority of these APs are deployed in a condominium building, which has residential homes, a café, conference spaces, and a fitness centre. The remaining 6 APs are installed in the rest of the overall area (approximately 7/8 of the overall area). STAs (i.e., users) are, on the other hand, expected to be concentrated in public places, such as cafes and train stations. It is clear that STAs and APs are not necessarily concentrated in a same area. A concentration of STAs often results in the AP(s) and STAs in that particular area suffering from severe congestion [2]- [6]. On the other hand, a concentration of APs causes energy wastage [7].

Congestion can be effectively alleviated by proper association control. Indeed a large number of association control schemes are proposed for WLANs, mainly aiming at load balancing among cells under different definitions of load (e.g., load is defined as the number of nodes in [2], [3], channel condition in [4], and traffic rate in [5], [6]).

In this paper, we analytically formulate the throughput maximization problem, and show that the existing schemes do not necessarily operate towards throughput maximization. Furthermore, while load balancing tends to use all the existing APs, the same or even improved network performance can be achieved by utilizing fewer APs especially in AP-concentrated areas. This provides a positive impact on energy efficiency since the unused APs can now be in power-saving mode. To this end, we propose an association control scheme that is to maximize network throughput and reduce energy consumption. We investigate the efficiency and feasibility of the scheme by both computer simulations and testbed experiments.

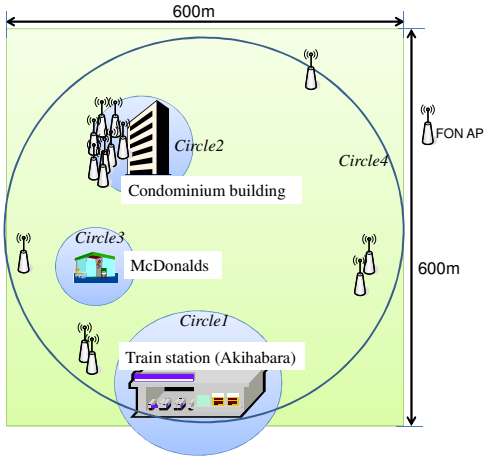


Figure 1. A map showing locations of FON APs in 600mx600m area in Tokyo (the information is taken from maps.fon.com on July 20, 2010).

 ISBN: 978-1-61208-174-8

II. PROBLEM FORMULATION AND RELATED WORK

The aggregate throughput of a DCF (Distributed Coordination Function) system is expressed as follows [8].

$$s = \frac{P_s P_{tr} E[P]}{(1-P_{tr})\sigma + P_{tr} T} \quad (1)$$

Here, E[P] is the average data size, P_{tr} is the probability that there is at least one station transmits in the sensing range, P_s is the probability of a successful transmission. σ is the slot time and T is the average time required for transmission of a data (includes DIFS and etc.). The numerator of (1) corresponds to the successfully transmitted payload length and the denominator is the total time. The former and the latter components of the denominator are the time that channel is idle and busy (either due to successful or unsuccessful transmissions), respectively.

Data transmission is successful if the frame is not collided and the frame does not contain errors (due to poor link quality). Thus letting P_{sc} and P_{se} represent the priorities of the former and the latter events, respectively, $P_s = P_{sc} \times P_{se}$. P_{sc} (in what follows we call P_{sc} as success probability) is expressed as

$$P_{sc} = \frac{n\tau(1-\tau)^{n-1}}{1-(1-\tau)^n} \quad , \quad (2)$$

where n is the number of nodes. τ is the channel access probability, which is determined by the contention window size (CW) and the probability that a node has a pending packet in its transmission queue [9]. Letting $r \equiv P_{se} \times E[P]/T$, we rewrite (1) as

$$s = P_{sc} \times r \times \frac{P_{tr} T}{(1-P_{tr})\sigma + P_{tr} T} \quad . \quad (3)$$

E[P]/T, the average frame transmission rate, is variable if rate adaptation is deployed at MAC, and it is fixed otherwise. If rate adaptation is deployed, the better the link quality (stronger RSSI), the higher is the selected transmission rate. Furthermore, if the rate adaption operates such that PER is minimized (i.e., P_{se} is maximized), the second component of (3), r, depends mainly on the selected transmission rate, i.e., $r \approx E[P]/T$. The last component of (3) is the ratio of time the channel is determined to be busy to the total time. Letting ATR (air-time ratio) represent the last component, the throughput of a DCF system is a multiplication of P_{sc}, r, and ATR:

$$s = P_{sc}(n) \times r(RSSI) \times ATR \, . \quad (4)$$

Finally, the throughput maximization problem for a network that consists of multiple overlapping WLANs is the maximization of

$$S = \sum_{i=1}^{N} s_i \quad , \quad (5)$$

where s_i is the throughput of cell i. It should be noted that overlapping cells, which operate under a same frequency channel, share the same P_{sc} and ATR.

In the traditional AP selection policy, a STA associates with the AP corresponding to the strongest RSSI. Thus such a scheme takes account of only the second component of (4), r. However, increasing r alone does not necessarily increase the total throughput, especially when STAs' distribution is highly non-uniform. In such a scenario, it is possible that most of the STAs select a same AP (because they are closer to that AP). As it can be seen in (2), P_{sc} decreases sharply with the increase of n (because the numerator of approaches zero and denominator approaches 1). Thus, in such a case, the throughput of the traditional scheme is poor because of a small P_{sc} for the crowded cell(s) and likely a small ATR for the scarce cell(s). This suggests distributing STAs to the existing cells. Indeed several schemes are proposed to distribute the number of STAs among cells, and some of them take account of the link quality (RSSI in [2], and PER in [3]). A drawback of these schemes is that they do not consider ATR, the channel availability.

Since DCF requires some idle slots for e.g., backoff procedures, ATR is upper-limited by a value smaller than 1 (let ATR_{max} represent the saturation value). Furthermore, it has been recognized that ATR has an optimal value, ATR_{th} ($ATR_{th} < ATR_{max}$), where the channel utilization is maximized [10]. This means that P_{sc} is maximized when ATR is equal to or smaller than ATR_{th}.

Reference [4] proposed to balance effective channel busy-time (i.e., time the channel is busy for successful transmissions) among cells. Because it does not discriminate the time corresponding unsuccessful transmissions and the idle time, [4] may suffer from such estimation errors. Moreover, because [4] ignores offered traffic volume, it might take a long time until load is balanced.

ATR is also the ratio of the amount of bandwidth consumed for transmissions to the total bandwidth. Thus, an increase of the offered traffic volume (injected traffic) increases ATR until it reaches its saturation value (ATR_{max}). Further increase of the offered traffic, however, results in congestion (i.e., buffer overflow) that significantly hampers communication performance. Since a cell with a small ATR can accommodate additional traffic, the overall throughput can be improved by moving STAs from a congested cell to such a non-congested cell. References [5], [6] proposed schemes that balance traffic volume among cells. A drawback of the schemes is that they do not consider the fact that the overlapping cells operating under the same frequency channel share the same channel resource. Moreover, [4]-[6] do not consider the link quality (the second component of (2)), and thus they may force STAs to use links with poor quality, degrading the users' throughput.

To the best of our knowledge, none of existing schemes takes account of the overall of (4), thus they do not necessarily maximize throughput. To this end, we propose an association control scheme that takes account of the overall of (4). The direct target of the proposed scheme is maximizing the sum of the multiplication of the second and third components of (4), r and ATR, by taking account of RSSI, ATR, and the offered traffic volume. The success probability, Psc, is indirectly maximized by maintaining ATR smaller than ATR_{th}. An important difference of the proposed scheme from the previous schemes is that because both the offered traffic and channel availability are estimated for each cell, the proposed scheme does not aim at load balancing. This enables the scheme to utilize fewer APs, providing a positive impact on energy efficiency.

 ISBN: 978-1-61208-174-8

III. PROPOSED SCHEME

A user-owned AP is integrated into a community network by a common equipment/software program provided by the entity (e.g., FON [1]). Besides integrating users' APs, the entity can play an important role for e.g., network management for better communication quality. Hence, community networks are centrally controllable and we expect that the members (i.e., the users) are cooperative to such a control. To this end, while a distributed mechanism can also be designed, we propose a centralized association control scheme due to ease of management and better performance [4]-[5], [7]. The proposed scheme aims at maximizing network throughput by a congestion alleviation mechanism and reducing energy consumption by a cell aggregation mechanism. For a large community network, the network can be divided into sub-networks, which are independently and separately controlled.

Figure 2 shows the network architecture. Besides APs and STAs, the network has an information server (server) and control manager (manager). The server and manager can be physically separated or coexist in a same machine. APs and STAs periodically inform the server of information on link quality and so on. Periodically referring to the information, the manager triggers STAs' handover for improved network throughput and/or energy efficiency.

A. Estimation of Channel Availability and Offered Traffic

STAs and APs measure RSSI, ATR, and the offered traffic volume, and inform the server of the information. The manager uses the information to estimate channel availability and traffic condition for each cell, and changes STAs' associations.

- Frame Transmission Rate

By periodically performing channel scanning, STAs learn the existence of the neighboring APs and the corresponding $RSSI_{AP,STA}$. Such background scanning is supported by the off-the-shelf wireless LAN cards [11], and some efforts have been made for fast channel scanning [12]. Transmission rate for frame payload field is estimated from $RSSI_{AP,STA}$ and finally the frame transmission rate, $r_{AP,STA}$ ($\approx E[P]/T$), for each pair of STA and AP is calculated.

- ATR

For ease of implementation and without much loss of generality, ATR can be defined as the ratio of the channel busy time to the total time (the numerator does not contain inter-frame spaces (DIFS, SIFS)). In the proposed scheme, each AP measures ATR on its operating channel. Such a measurement can easily be made using the existing WLAN cards [4],[7],[13]. We empirically found that the appropriate values for ATR_{max} and ATR_{th} are 0.6 and 0.58, respectively.

- Potential Throughput

The manager estimates the maximum achievable throughput for each pair of STA and its neighboring AP (which is not the STA's currently associated AP). Let $PT_{STA,AP}$ (potential throughput) represent the maximum achievable throughput for such a STA and an AP. As (4) shows, the throughput is a function of P_{sc}, the estimated transmission rate ($r_{AP,STA}$), and ATR_{AP}.

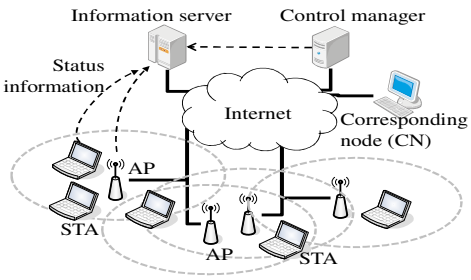


Figure 2. Network architecture.

As discussed in the previous section, however, P_{sc} can be maximized by ensuring ATR below ATR_{th}. This enables PT be estimated from only $r_{AP,STA}$ and ATR_{AP}. Thus the manager calculates PT for each STA and its neighboring AP as

$$PT_{AP,STA} = \begin{cases} 0, & ATR_{AP} \geq ATR_{th} \\ (ATR_{th} - ATR_{AP}) \times r_{AP,STA}, & \text{otherwise} \end{cases} . \quad (6)$$

The upper equation is to not move the STA to the AP because ATR_{AP} exceeds ATR_{th}. Otherwise, the AP is a candidate destination AP for the STA and the maximum achievable throughput at the candidate cell is expressed as the lower equation. In our previous work [13], we confirmed that such an estimation of PT can be achieved with a high accuracy using the existing WLAN cards.

- Offered Traffic

A STA can be moved to a neighboring AP, if it does not cause congestion at the neighboring cell. The condition can be checked by comparing the offered traffic volume for the STA and $PT_{STA,AP}$ (to be discussed later). Letting $EnqueueRate_{A,B}$ be the rate at which packets destined to node B are inserted into the IP queue at node A, the offered traffic volume for a STA is defined as

$$OfferedRate_{STA} \equiv EnqueueRate_{STA,AP} + EnqueueRate_{AP,STA} . \quad (7)$$

If the WLAN is the bottleneck link of the end-to-end route, the OfferedRate is approximately equal to the traffic generation rate.

B. Congestion Alleviation

A cell is considered to be congested if

$$ATR > ATR_{th} . \quad (8)$$

If the condition (8) is met for a cell, the manager checks if the aggregate offered rate exceeds the aggregate packet throughput for that cell, i.e.,

$$\alpha \sum_{STAs} OfferedRate_{STA} > \sum_{STAs} PacketThroughput_{STA} . \quad (9)$$

Here $\alpha(<1)$ is a system parameter to absorb rate fluctuation. $PacketThroughput_{STA}$ is the sum of the rates at which packets are successfully transmitted on the uplink and downlink for the STA. The condition (9) indicates that one or more nodes in the cell are suffering from buffer overflow. It is possible that a cell satisfies (8) but not (9), in a case, when the cell does not have much traffic but the channel is congested due to the overlapping cells.

A cell that satisfies both (8) and (9) becomes a target cell of congestion control. Letting APt be the AP of the target

 ISBN: 978-1-61208-174-8

cell, the manager selects STAs to move from APt to its neighboring cells. The association control is made based on the following policies:

1. Moving STAs are selected such that the number of handovers is minimized.
2. A STA should be moved only if it will not cause congestion at the destination cell.
3. Among the candidate destination APs, the STA should be moved to the AP corresponding to the strongest RSSI.

The more the handovers, the larger is the control overhead. Thus policy 1 is to minimize the number of moving STAs. To support this objective, we define "load" of a STA, as $Load_{STA} \equiv OfferedRate_{STA}/TxRate_{STA,APt}$. Here $TxRate_{STA,APt}$ is the rate used for transmissions of frame payload fields between the STA and APt. Obviously, the larger the offered traffic and/or the lower the transmission rate, the heavier loaded is the STA for APt. Due to policy 1, heavier loaded STAs are preferred to be moved from APt over lighter loaded STAs. For policy 2, a STA is moved to a neighbouring AP, APd, only if the condition (10) is met.

$$OfferedRate_{STA} < PT_{STA,\ APd} \quad (10)$$

In other words, the STA is moved to APd only if the destination cell can accommodate the offered traffic volume for the STA. Finally, among the candidate destination APs (i.e., the APs that satisfy (10) for the STA), the AP corresponding to the strongest RSSI is selected as the destination AP for the STA.

When a STA, STA_m, is selected to be moved from APt, the manager updates the aggregate offered rate (see (9)) for the target cell by decrementing it by $OfferedRate_{STAm}$. Moreover ATR for the destination cell and its overlapping cells (which operate using the same channel) is incremented by $OfferedRate_{STAm}/r_{APd,STAm}$. After updating the values, the manager checks if the target cell still satisfies (9). If it does, the manager selects the next moving STA.

- Discussion on TCP traffic

TCP reacts to congestion and controls its rate based on AIMD algorithm. However such a rate change occurs in the order of RTT (milliseconds) which is much shorter than the control period at the manager (in order of seconds). Therefore we expect that the proposed scheme does not react to the AIMD-based rate fluctuation, but only the gradual change of the average rate. Hence, the proposed scheme and TCP can stably coexist. Moreover because TCP adjusts its traffic rate, $OfferedRate_{STA}$ for STA might be largely changed due to the STA's movement. However, it should be reminded that a STA is moved to a neighbouring cell only if the destination cell can accommodate the current OfferedRate for the STA (see (10)). Thus we expect that TCP throughput will be increased or maintained after the STA's movement. Furthermore, since some amount of channel resource is released at the previous cell of the moving STA, STAs in that cell can now increase their rate.

C. Cell Aggregation

Since both of the offered traffic volume and the channel availability are known for each cell, load balancing among cells is not necessary. Indeed, if all the associated STAs of an AP can be moved to its neighboring cells without hampering the network throughput, such STA movements should be encouraged for energy efficiency, since the AP can now be in power-saving mode. Our scheme can provide such an association control based on the following policies:

1. The target AP is an AP that is associated with preferably a few STAs, which can all be moved to the neighboring cells.
2. To suppress channel interference, the target AP should have overlapping cells that operate using the same frequency channel.
3. Policy 2 described in the previous subsection.
4. Policy 3 described in the previous subsection.

The manager triggers a handover only if destination APs are found for all the STAs of the target AP. A detailed protocol design for actually putting APs in power-saving mode is left for our future work. The main concern of such a protocol design is to ensure newly arriving STAs are covered by the network. For such a control, Wake-on-WLAN technology [14] can be used.

D. Changing STA's Association

To change a STA's association, the manager sends a control frame to the STA, indicating the destination AP and the corresponding channel information. Such a network directed association control can be supported by the upcoming IEEE 802.11v [15], which enables APs to explicitly request STAs to re-associate with an alternate AP.

IV. PERFORMANCE EVALUATION

A. Simulation Evaluations

Using Scenargie network simulator [16], we investigate the proposed scheme with and without cell aggregation capability. The performance of the proposed scheme is compared against:

- Strongest RSSI: The traditional AP selection scheme.
- LB(NumofSTAs): A load balancing scheme [2], which takes account of the number of STAs and RSSI.
- LB(Traffic): A load balancing scheme [5], where load of a cell is defined by traffic rate.

In each scheme, STAs initially associate with APs based on the strongest RSSI policy. In the proposed scheme, STAs inform the server of the measured information using a 160 bytes packet. The manager checks the collected information in every 1s. 20-bytes of packets are used for handover requests and replies between the manager and moving STAs. α (see (9)) is set to 0.98.

Performances of the schemes are investigated for the real-world scenario depicted in Figure 1, where 14 APs are non-uniformly distributed in a 600mx600m area. 8 APs are concentrated in the small area (around the condominium), the other 6 APs are installed in the remaining area as shown in the figure. The network operates using IEEE 802.11a [17], where 3 orthogonal frequency channels are available. Frequency channel allocation (to the APs) is made based on a simple graph coloring technique.

 ISBN: 978-1-61208-174-8

STAs (users) can be anywhere, but it is natural to expect that users are especially attracted to the public places specifically, the train station, condominium building (for the café), and McDonalds in the target map. Thus, in our simulations, 40 STAs are distributed in the circle-shaped areas centered at the 1) train station, 2) condominium building, 3) McDonalds, and 4) the center of the map, with a radius of 100m, 10m, 10m and 300m, respectively (see Figure 1). The first three areas are set to create users' concentration in the public places, while the last area is for uniform distribution of users in the overall area. Table I shows the simulation scenarios, which have different number of STAs in each circle-shaped area. The smaller the scenario number, the more uniform is the STAs' distribution.

Figures 3 and 4 show the results of the simulations targeted at TCP and UDP traffic, respectively. In TCP simulations, STAs upload 5MB of file using FTP/TCP-SACK. In UDP simulations, STAs have uplink and downlink CBR traffic generated at a random rate in the range of [0Mbps, 1.2Mbps].

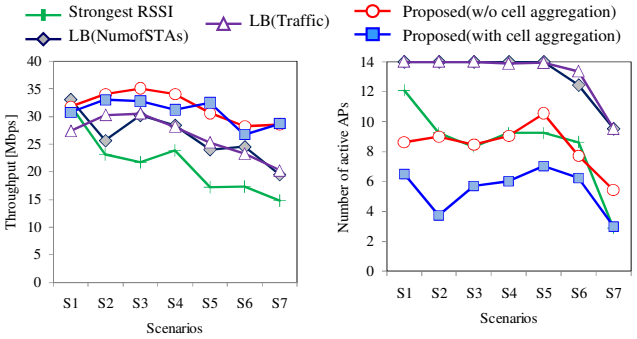
As the figures a) show, for both the TCP and UDP traffic, the performance of Strongest RSSI scheme is inferior to the remaining schemes and the proposed scheme shows the best performance. The proposed scheme with the cell aggregation mechanism achieves around the same throughput performance as the scheme without cell aggregation especially for UDP traffic. The load balancing schemes have lower throughput than that of the proposed scheme, because they do not take account of the overall of (4).

The figures a) show that, for any scheme, the throughput tends to be smaller when STAs' distribution is less uniform (e.g., the throughput of S7 is smaller than that of S5). As discussed in Section II, the reason is clear for the strongest RSSI scheme (many STAs select the same AP). The reason for the remaining schemes is as follows. When STAs are highly concentrated around a particular AP, the schemes have to move some of the STAs to farther APs. This reduces the frame transmission rate, r, for the moving STAs, resulting in lower throughput compared to that of a scenario where STAs' distribution is more uniform. Nevertheless, compared with the strongest RSSI scheme, the overall throughput is improved due to an increase of the 1st and 3rd components of (4). Finally, S1 (where STAs' distribution is uniform), however, does not show the largest throughput due to the non-uniform AP distribution.

Figures 3b) and 4b) compare the number of active APs, i.e., the number of APs that actually serve for the users. As the figures show, the load balancing schemes use all the existing APs (except in scenario S7, where STAs are concentrated only at the train station). Strongest RSSI scheme, on the other hand, does not use many APs due to its simple AP selection policy. The proposed scheme without cell aggregation utilizes around the same number of APs as that of Strongest RSSI scheme. Finally, the scheme with cell aggregation utilizes the smallest number of APs. This is especially attractive because, compared to the existing schemes, the proposed scheme largely improves throughput while utilizing fewer APs. Our future work includes a study on how much energy reduction can be achieved by such an association control.

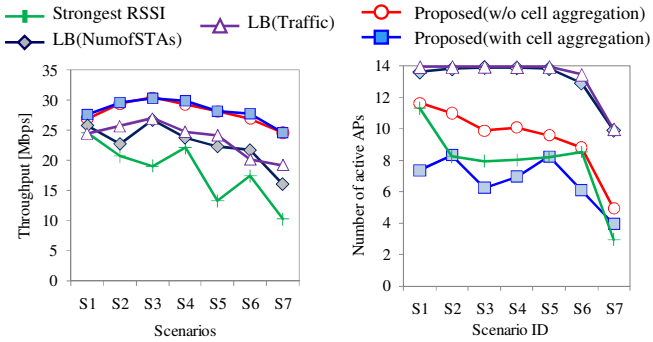
TABLE I. SIMULATION SCENARIOS (USER DISTRIBUTION)

	Circle1 (station)	Circle2 (condominium)	Circle3 (McDonalds)	Circle4 (overall area)
S1	0	0	0	40
S2	10	10	10	10
S3	10	20	0	10
S4	10	20	10	0
S5	20	10	0	10
S6	30	0	0	10
S7	40	0	0	0



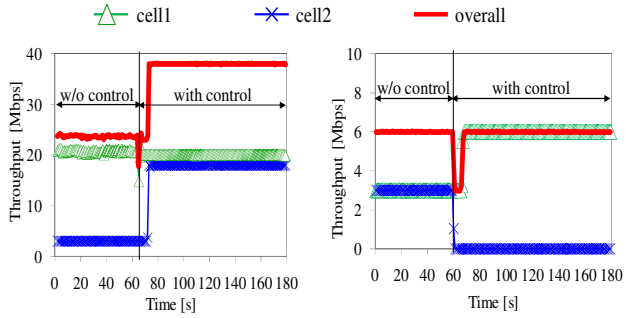
a) Aggregate throughput b) The number of active of APs

Figure 3. Performance comparison for TCP traffic.



a) Aggregate throughput b) The number of active of APs

Figure 4. Performance comparison for UDP traffic.



a) Congestion alleviation b) Cell aggregation

Figure 5. Empirical results.

 ISBN: 978-1-61208-174-8

B. Testbed Evaluations

We implemented the proposed scheme in a wireless testbed and evaluated its throughput performance. Computers with Cent OS 5.5 (kernel 2.6.25-17) are used for STAs, APs, and the manager (note that the manager and server coexist in a same machine). The APs, the manager, and a source computer (acts as a corresponding node (CN)) are connected to a 100Mbps Ethernet. The APs and STAs are equipped with 802.11a WLAN card made by NEC (Aterm WL54AG). We modified the Atheros device driver to enable measurements of ATR and PacketThroughput (see (9)). The packet transmission rate from the kernel to the device driver is monitored to measure EnqueueRate (see (7)). TCP is used for information collection at the server and for handover requests and replies. The testbed evaluations target scenarios that consist of 2 APs, AP1 and AP2, which use different frequency channels and are each initially associated with 3 STAs. An experiment lasts about 200 s. Two scenarios (scenario1 and 2) are set to evaluate the congestion alleviation and cell aggregation mechanisms. In scenario1, CN transmits 15, 10, and 10 Mbps CBR traffic to the STAs, which are initially associated with AP1, and 1Mbps traffic to each STA initially associated with AP2. For scenario 2, CN transmits 1Mbps CBR traffic to each STA. To see the impacts of the proposed mechanisms, the manager is activated at app. 60 s for both the scenarios.

Figure 5 shows the time series plots of the throughput for the scenarios. In Figure 5a), when the manager is not activated, the aggregate throughput is 22Mbps and packet loss ratio is app. 40%. Upon the activation of the manager, the STA with 15 Mbps traffic is moved to cell2. This association control maximizes the throughput (aggregate throughput is 38Mbps) and eliminates packet loss ratio. For scenario 2 (see Figure 5b)), as both the cells are lightly loaded, congestion is not an issue, thus the system shows 6Mbps throughput and 0% of packet loss ratio even before the activation of the manager. Upon the activation of the manager, the cell aggregation is performed and all the STAs of cell2 are moved to cell1.

Unfortunately, as it can be seen in the figure, it took app. 6 seconds to complete the handover (without depending on the number of moving STAs). The 6 seconds are used for 1) MAC layer disassociation/association, 2) IP route advertisement, 3) IP duplicate address detection, 4) MIP binding update, and 5) MIP binding acknowledgement. Among the operations, there was a software bug corresponding to 2) and we confirmed that by fixing this bug, handover time can be reduced down to 3 seconds. We are now working on this issue.

V. CONCLUSION

In this paper, we analytically formulated the throughput maximization problem for WLANs and showed that the existing association control schemes do not necessarily maximize network throughput. Furthermore, since most of the existing schemes aim at load balancing among cells, they tend to use all the existing APs. However, sufficient network performance can be achieved by utilizing fewer APs, providing positive impact on energy efficiency. To this end, we proposed an association control scheme to maximize throughput and reduce energy consumption. The simulation results showed that compared to the existing schemes, the proposed scheme can provide much larger throughput while utilizing fewer APs regardless of traffic type and node density. The testbed experiment shows that proposed scheme is feasible using the off-the-shelf wireless LAN cards. Our future work includes a study on energy efficiency induced by the cell aggregation mechanism.

ACKNOWLEDGMENT

This research was performed under research contract of "Research and Development for Reliability Improvement by The Dynamic Utilization of Heterogeneous Radio Systems", for the Ministry of Internal Affairs and Communications, Japan.

REFERENCES

[1] FON, www.fon.com

[2] S. Sheu and C. Wu, “Dynamic Load Balance Algorithm (DLBA) for IEEE 802.11 Wireless LAN,” Tamkang Journal of Science and Engineering, vol. 2, no. 1, pp.45-52, 1999.

[3] Y. Fukuda and Y. Oie, “Decentralized Access Point Selection Architecture for Wireless LANs,” IEICE Trans. Commun. vol. E90-B, no. 9, pp. 675-684, 2007.

[4] F. Guo and T-C. Chiueh, “Scalable and Robust WLAN Connectivity Using Access Point Array,” IEEE DSN’05, pp. 288-297, 2005.

[5] I. Jabri, N. Krommenacker, T. Divoux, and A. Soudani,”IEEE 802.11 Load Balancing: An Approach for QoS Enhancement,” Springer, Int J Wireless Inf Networks, vol 15, pp.16-30, 2008.

[6] H. Velayos, V. Aleo, and G. Karlsson, “Load Balancing in Overlapping Wireless LAN Cells,” IEEE ICC 2004, Vol. 7, pp. 3833-3836, 2004.

[7] A. J. Jardosh, K. Papagiannaki, E. M. Belding, K. C. Almeroth, G. Iannaccone, and B. Vinnakota, “Green WLAN: On-demand WLAN Infrastructures,” Springer, Mobile Networks and Applications, Vol. 14, No. 6, pp. 798-814, Dec. 2008.

[8] G. Bianchi, “Performance Analyis of the IEEE 802.11 Distributed Coordination Function,” IEEE Journal on Selected Areas in Comm., vol. 18, no. 3, pp. 535-547, March 2000.

[9] F. Daneshgaran, M. Laddomada, F. Mesiti, and M. Mondin, “Unsaturated Throughput Analysis of IEEE 802.11 in Presence of Non Ideal Transmission Channel and Capture Effects,” IEEE Trans. Wireless Comm. vol. 7, issue: 4, pp. 1276 – 1286, April 2008.

[10] H. Zhai, X. Chen, and Y. Fang, “How well can the IEEE 802.11 Wireless LAN suport Quality of Service,” IEEE Trans. Wireless Comm. Vol. 4., no. 6, pp. 3084-3094, Nov. 2005

[11] http://madwifi.org/

[12] I. Ramani and S. Savage, “SyncScan: practical fast handoff for 802.11 infrastructure networks,” INFOCOM 2005, Vol.1, pp. 675-684, 2005.

[13] S. Tang, N Taniguchi, O. Shagdar, M. Tamai, H. Yomo, A. Hasegawa, and et al., “Potential Throughput Based Access Point Selection,” IEEE APCC 2010.

[14] IEEE 802.11v: Wireless Network Management, http:://grouper.ieee.org/groups/802/11/Reports/tgv_update.htm.

[15] N. Mishra, K. Chebroulu, B. Raman, and A. Pathak, “Wake-on-WLAN,” ACM, WWW, pp. 761-769, 2006.

[16] Scenargie, http://www.spacetime-eng.com

[17] IEEE Computer society LAN MAN Standards Committee, Wireless LAN Medium Access Protocol (MAC) and Physical Layer (PHY) Specification, IEEE Std 802.11-1997, IEEE, 1997.

 ISBN: 978-1-61208-174-8

Multipath Routing Management using Neural Networks-Based Traffic Prediction

Melinda Barabas, Georgeta Boanea, Virgil Dobrota
Communications Department
Technical University of Cluj-Napoca
28 Memorandumului Street, 400114 Cluj-Napoca, Romania
Email: {Melinda.Barabas, Georgeta.Boanea, Virgil.Dobrota}@com.utcluj.ro

***Abstract*—Embedding forecasting algorithms into routing management systems can play an important role in guaranteeing QoS in IP networks. In this paper, we propose an intelligent routing framework, consisting of a situation aware multipath routing algorithm and a routing management system involving neural networks-based predictors with multi-task learning. The solution is characterized by QoS-awareness, load balancing and self-management. The main goal is to offer a proof-of-concept by practical implementation of predictive QoS-aware multipath routing in a real test environment. The proposed solution is compared with the OSPF and ECMP routing protocols in case of congested network links. The experimental results show that traffic prediction enables proactive routing management and improves the global performance of the network through congestion control and avoidance.**

Keywords–cross-layer QoS; multipath routing; neural networks; self-management; traffic prediction

I. INTRODUCTION

Ensuring Quality of Service (QoS) over the Internet is difficult, especially in the case of real-time multimedia applications where the retransmission of packets is not a viable option. The occurrence of congestion can severely degrade the quality of transmissions due to packet losses, increased delay and jitter [1]. Embedding forecasting algorithms into routing management systems can play an important role in guaranteeing QoS in IP networks. Traffic prediction enables proactive network management which improves the global performance of the network through congestion control and prevention.

Initially, it was believed that adaptive routing protocols, such as OSPF (Open Shortest Path First) [2], can react to congestion. Unfortunately, congested links often remain undetected because of the way OSPF assesses link connectivity. If a link flaps constantly due to congestion, but at least 1 out of every 4 Hello messages is received, OSPF does not detect the problem. If the congestion is severe and no Hello messages are received from a neighbor, it is automatically considered *down* because OSPF makes no distinction between hardware failures and congestion. Thus, the involved router will not be further used and all the traffic will be rerouted to a different link which in turn can also become congested. The solution adopted by OSPF does not resolve the underlying problem, that of transmitting too much traffic on a single link.

In the present Internet, congestion control mechanisms rely on queue management algorithms (dropping packets randomly or based on their priority) or TCP (Transmission Control Protocol) congestion avoidance (reducing the sending rate). From the end-user perspective, these solutions are not optimal because they mean lost packets or a reduced bitrate, both affecting the quality of transmission, especially the QoE (Quality of Experience) of multimedia content.

The main motivation for this paper is to resolve the above mentioned limitations of legacy routing protocols and congestion control mechanisms by applying the multipath routing paradigm. We focus on the problems caused by congested network links and our goal is to improve the overall network performance by load balancing and prediction of network traffic. We envision an intelligent routing framework, consisting of the SAMP (Situation Aware Multipath) routing algorithm [3] and a routing management system.

The routing strategy presented herein is characterized by self-management and QoS-awareness, achieved via monitoring link resources through cross-layering techniques. QoS-aware routing means that not only shortest paths, but traffic-aware shortest paths are computed for optimal network performance. The traffic predictors integrated into the routing management system enable proactive decision-making, as opposed to reacting to past events. Employing a prediction-based approach helps to match network resources to the traffic demand [4]. Thanks to the early warning, a prediction-based approach will be faster, in terms of congestion detection and elimination, than reactive methods which detect congestion only after it significantly influenced the operation of the network, as demonstrated in [5].

The rest of this paper is organized as follows. Section II briefly presents previous work regarding prediction used in combination with routing systems. In Section III, neural network traffic predictors with multi-task learning approaches are described. Section IV presents the multipath routing framework. In Section V, the practical testbed is described, followed by the experimental results in Section VI. Section VII concludes the paper and discusses future work.

II. RELATED WORK

In the literature, several works address the topic of network parameter prediction techniques integrated into single-path [6],

 ISBN: 978-1-61208-174-8

[7], [8] or multipath routing solutions [5], [9], [10].

The authors of [6] propose the PBS (Prediction Based Routing) heuristic mechanism that predicts the availability of links/routes and selects routes without taking into account network state information. In [7], a neural networks-based queuing delay prediction mechanism is integrated with a MANET proactive routing protocol OLSR (Optimized Link State Routing), increasing the packet delivery ratio and reducing the end-to-end delay. Masip-Bruin *et al.* [8] designed a routing technique based on CBR (Constraint-Based Routing) that combines the strength of prediction with an innovative link-state cost. CBR is applied in circuit-switched networks and it reduces the impact of routing inaccuracy on the blocking probability.

A data forwarding algorithm over multipath is described in [5]. The proposed solution is based on linear prediction and particle swarm optimization and it improves the QoS of real-time applications. Li *et al.* [9] proposed a Multipath Routing Algorithm based on Traffic Prediction (MRATP) to be used in Wireless Mesh Networks (WMN) in order to guarantee end-to-end QoS. A method for multipath selection based on prediction in wireless networks is introduced in [10] where neural networks are used to infer the types of the links and the paths are chosen based on predicted incremental throughput.

In the literature, a predictive approach is taken into consideration either for single-path routing approaches or for multipath routing over wireless networks. Based on this observation, we chose to integrate a network parameter prediction algorithm into a multipath routing solution over wired networks. In this way, the routing metrics will depend on predicted traffic conditions. Thereby, we intend to identify congestion in the network faster than through simple monitoring. This is achieved by predicting the available transfer rate on unidirectional network links, as opposed to other solutions which predict: *a)* the rate of packet losses [5], *b)* the delay in routing queues [7], *c)* the type of wireless links and the incremental throughput [10] or *d)* the bitrate of video flows [11], etc. Reaction to congestion is manifested by rerouting traffic, unlike alternatives such as: *a)* reduced video bitrate [1], *b)* advanced allocation of transfer rate for future transmissions [4], [8], [11], *c)* controlled dropping of packets [12], etc.

III. Neural Networks-based Prediction

The prediction of network traffic is possible because it presents a strong correlation between chronologically ordered values. The most widely used traffic forecasting methods involve Neural Networks (NN) [13], [14], [15], etc. NNs are employed for modeling and predicting traffic because of their strong self-learning and self-adaptive capabilities through which they are able to learn complex patterns. NNs are characterized by nonlinear mapping and generalization ability, robustness, fault tolerance, parallel processing, etc.

A NN consists of several layers of interconnected nodes (neurons): *a)* an input layer, *b)* one or more hidden layers and *c)* an output layer. The most popular NN architecture is feed-forward in which the information travels through the network in the forward direction: from the input layer towards the output layer. The NN model represents a nonparametric and adaptive modeling approach, the architecture and the parameters being determined solely by the observed data.

Using a NN as a predictor involves two phases: *a)* the training phase and *b)* the prediction phase. In the training phase, the training set is presented at the input layer and the parameters of the NN are dynamically adjusted to achieve the desired output. The prediction phase represents the testing of the NN. A new input (not included in the training set) is presented to the NN and the output, which represents the predicted value, is calculated.

Usually, NN predictors have a single output node and they focus on a single main task, i.e., predicting x_{t+1} based on $\{x_1, x_2, \ldots, x_t\}$. Thereby, the predictor neglects information hidden in other tasks (e.g. the relationship between the historical data and x_{t+2}, although both tasks belong to the same dataset). The Multi-Task Learning (MTL) paradigm is introduced to improve the generalization performance of NNs. A main task is trained simultaneously with extra tasks, sharing the hidden layer of the NN, as shown in Figure 1. By learning multiple tasks simultaneously, the NN can achieve better prediction accuracy. For time series forecasting through the MTL concept, usually, two extra tasks are chosen, namely the prediction of x_t and x_{t+2}, which are closely related to the main task x_{t+1}, as in [16] and [17].

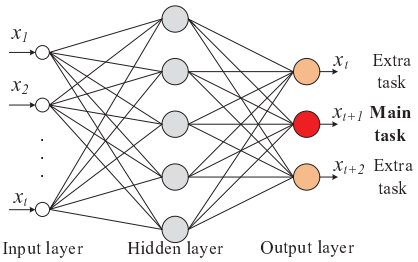


Fig. 1. NN predictor with multi-task learning

In our experiments, we selected a NN with only one hidden layer because more layers would make the network more time- and resource-consuming, but added complexity would not be justified by the improvement of the prediction accuracy.

IV. Multipath Routing Framework

The main idea of the proposed multipath routing framework is to separate the monitoring from the routing process itself: the link monitoring and the communication between neighboring nodes is realized by the routing management system, while the multipath routing algorithm deals with the routing decisions and the packet forwarding. Thus, the information regarding the state of the network becomes reusable.

A. Multipath Routing Algorithm

The multipath routing algorithm used in this paper is called SAMP (Situation Aware Multipath). Practical implementation of SAMP and simulation results are presented in detail in our previous work [18]. To ensure efficient and high quality

 ISBN: 978-1-61208-174-8

transmissions, SAMP relies on the information provided by the routing management system regarding the network status. The key features of this solution are load balancing and congestion avoidance by fast rerouting.

1) Load balancing

To overcome the problem of inefficient link resource utilization, load balancing is employed. In order to divide the traffic among multiple routes, a split granularity at flow level is used, avoiding the problem of out-of-order packet arrivals. A flow is identified by the triplet: source IP address, destination IP address and destination port. Because SAMP takes into account the physical state of the network, the flows will be routed along paths that ensure the application's requirements in terms of transfer rate and delay.

With the purpose of providing scalability and decreasing the complexity of the solution, the network is divided into multiple routing domains, each consisting of two types of routers:

- *AR (Adaptive Router)*: located inside the domain and performing situation aware routing (reacting in case of congestion);
- *AMR (Adaptive Multipath Router)*: located at the edge of the domain. Besides the situation aware routing features, it also achieves load balancing for the traffic coming from outside the domain.

The traffic is divided into elastic end inelastic flows [20]. The elastic flows are handled by the main routing table because they are not sensitive to delay- and throughput variations. The inelastic flows (e.g. video, VoIP, etc.) are identified and transmitted on multiple paths. This forwarding method is carried out using the VRF (Virtual Routing Forwarding) concept: depending on the path that is allocated to a flow, the corresponding routing table is used.

The AMR dictates how a flow is routed in a domain. This is possible because all the nodes have a global view of the network, possessing the necessary information concerning the behavior of all other routers in any situation. The proposed solution does not impose any restrictions regarding the number of multipath domains/nodes, but the complexity increases along with the number of domains.

2) Congestion avoidance by fast rerouting

In case of congestion on one of the links, flows transmitted along the affected link are gradually rerouted, one by one, until the congestion disappears. The new selected path for a flow will be the one that offers the highest available transfer rate and has the lowest delay. Only multimedia flows are rerouted, the rest of the traffic being considered background traffic. Because all paths in the network are precomputed, the algorithm does not dependent on the number of congested/failed links.

B. Routing Management System

The Routing Management System (RMS) is a highly distributed self-managing system, which is capable to dynamically adapt to external events, minimizing the need for human intervention. It consists of *Local Management Entities* (LME) located on every node of the network (Figure 2). LMEs located on different nodes communicate through XML messages, discovering the network topology and exchanging network status information that are stored in local databases.

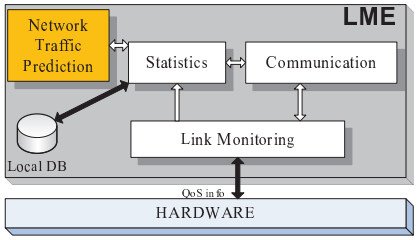


Fig. 2. Local Management Entity

LME performs real-time monitoring of the inbound links, measuring the Available Transfer Rate (ATR) and the One Way Delay (OWD) at the data link layer, as well as the missing packets at the application layer. Dropped packets can be considered an early indicator for congested links and overloaded routers. These are monitored only for multimedia streams, whose quality is the most likely to be influenced by congestion. Thereby, the system employs a cross-layer optimization strategy by making decisions at the network layer based on information derived from lower- and upper layers.

A previous version of the proposed routing management system is described in [21]. There are three main differences from the previous implementation. First of all, the RMS is a highly distributed system, as opposed to the previous solution where the congestion detection mechanism and the network status updates followed a centralized approach. Another difference lies in the monitoring of dropped packets. This enables the identification of the most affected multimedia streams which will have priority in the rerouting process. The third and most significant difference is the integration of a network traffic prediction module into LMEs. This represents a key component for the adaptive congestion control scheme. It forecasts the values of the ATR for inbound links for the next time interval (1 second). LME can detect congestion on its monitored links and it broadcasts this information through the network. It indicates to the routing algorithm when to update the routing tables. The system being highly distributed, routing decisions are not taken synchronously on every node.

V. Performance Evaluation

The practical testbed illustrated in Figure 3 is used to evaluate the performance of the proposed solution. This network offers sufficient paths between the source and destination nodes in order to employ multipath routing, but it is simple enough to allow practical implementation. The testbed consists of: *a)* six routers (`R1`, `R2`, `R3`, `R4`, `R5`, and `R6`), *b)* a source node (`S`) and *c)* two destination nodes (`D1` and `D2`). All nodes in the network are Linux-based computers with Fedora operating system. On each machine, several software applications written in C++ are running:

- multipath routing application (SAMP);
- Local Management Entity (LME);

 ISBN: 978-1-61208-174-8

- NNs based traffic predictors.

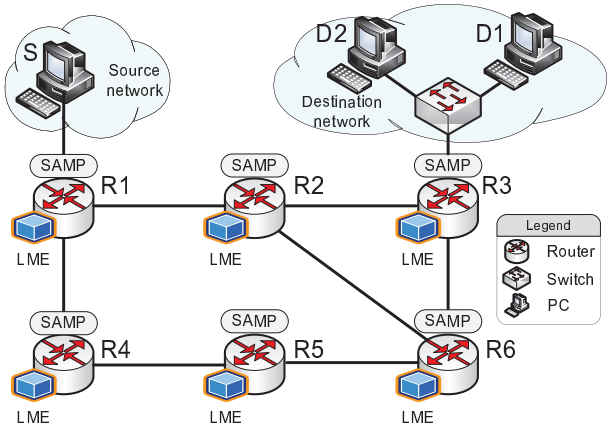


Fig. 3. Practical testbed

Providing good video quality is a major problem in congested networks since video traffic is both massive and intolerant to packet loss or latency. We demonstrate the improvements brought by the described predictive QoS-aware multipath routing framework by sending video streams from the source node S to the destinations D1 and D2 when two of the network links are affected by congestion. During the experiments, three different MPEG-4 video flows, each having an average bitrate of 1 Mbps, are sent by a VLC client over RTP/UDP: *1) Stream #1* from S to D1, *2) Stream #2* from S to D1, and *3) Stream #3* from S to D2.

The test scenario has a duration of 5 minutes. We generate background traffic in the network using the iperf network testing tool. Congestion is introduced on links R2–R3 and R2–R6 after 1 minute and after 2 minutes, respectively. As a result, packet losses can appear because the ATR drops below the required rate to transmit the streams and the OWD increases.

Experiments are performed in order to compare the following intra-domain routing approaches: *Case 1)* **OSPF** (Open Shortest Path First) – the most widely used routing protocol in large networks; *Case 2)* **ECMP** (Equal-Cost Multi-Path) [22] – the only multipath solution supported by current IP routers, and *Case 3)* **SAMP using NNs-based traffic prediction**.

The performance of the different routing solutions is measured in terms of their ability to reduce the negative effects of congested links. We take into consideration the following objective Video Quality (VQ) metrics of the received streams:

- *Number of lost packets*: determined by examining the sequence number in the RTP (Real-time Transport Protocol) header;
- *Magnitude of loss*: the number of packets that are dropped at each loss event, i.e., how many packets are missing between two consecutive received packets (a magnitude of 0 means the packet arrived successfully);
- *Discontinuity counter*: the frequency of detected discontinuities (i.e. packet drops);
- *Success Ratio* (SR): the number of packets received successfully divided by the total number of packets sent.

VI. Experimental Results

Case 1 (OSPF)

In order to evaluate the OSPF protocol on Linux-based machines, the Quagga Routing Software Suite [23] is used which is an advanced routing software package that provides a suite of TCP/IP based routing protocols.

For the tested network topology, OSPF determines the same path between the source and the destination nodes for all three streams, namely R1–R2–R3. After 1 minute, we introduce congestion between R2 and R3, but OSPF does not modify the routing tables because it does not take into consideration the physical state of the links. As an effect, we observe packet losses at the destination nodes and a very poor quality of experience. At 2 minutes from the beginning of the experiment, congestion is introduced on link R2–R6, but this is also not detected by the OSPF protocol.

The VQ parameters of interest for the received video streams in case of OSPF routing are shown in Table I. As we can observe, each of the streams is characterized by significant losses. A total number of 32485 packets are missing at the destinations D1 and D2 out of 64977 packets sent by the source node S, i.e., 49.99% of the transmitted packets were dropped due to congestion.

Figure 4 presents the magnitude of loss events for each video stream. In the first minute of the experiment, the magnitude of loss is 0, indicating that all packets are received at the destination nodes. It can be observed that losses appear constantly after the link R2–R3 gets congested.

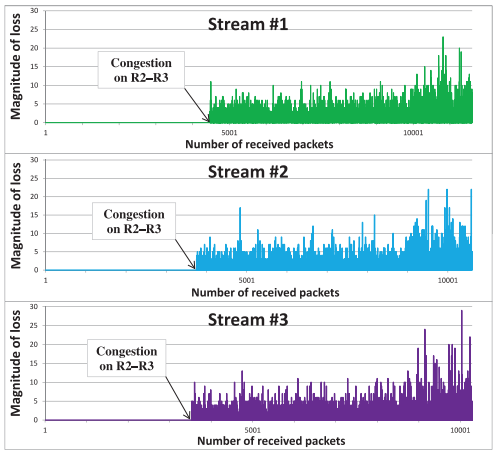


Fig. 4. Case 1 (OSPF) – Magnitude of loss

In Figure 5, the Success Ratio (SR) of the different transmissions is shown over the experiment duration. The SR corresponding to all streams starts to fall steadily after congestion is introduced on link R2–R3, reaching a minimum of 51.41% for Stream #1, 49.62% for Stream #2 and 48.89% for Stream #3 at the end of the experiment. The global final SR is 50.01%.

Case 2 (ECMP)

In our experiments, the ECMP routing approach is used in conjunction with the OSPF routing protocol in Quagga.

 ISBN: 978-1-61208-174-8

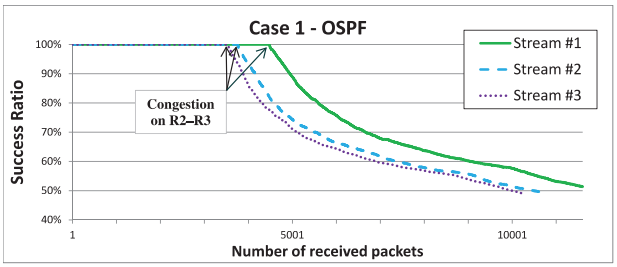


Fig. 5. Case 1 (OSPF) – Success ratio over experiment duration

This represents the only available multipath solution for Linux which allows load balancing by per-flow routing. In Linux, a flow is defined by the *source IP address* and the *destination IP address*. This means that, in our experiment, ECMP will identify only two flows which will be routed on different paths: *1)* Stream #1 and Stream #2 between S and D1, and *2)* Stream #3 sent from S to D2.

A major limitation of ECMP is that it only uses paths having equal costs. Initially, in our network topology the costs of all links were by default 10. Because there exist no multiple paths between the source and destination networks with the same cost, to be able to use ECMP, the cost of link R2–R3 is set to 20. Thereby, ECMP identifies two paths with the same cost (40): *1)* R1–R2–R3: used for the first flow (Stream #1 and #2) and *2)* R1–R2–R6–R3: used for the second flow (Stream #3).

The parameters of the received video streams in case of ECMP routing are shown in Table I. As we can observe, the percentage of lost packets for Streams #1 and #2 is more pronounced, than for Stream #3. This can be explained by the fact that they are routed on different paths: the first two are affected by congestion for a period of 4 minutes, while the third only experiences congestion in the last 3 minutes. Out of the total number of 64977 packets sent by S, only 44860 reached the destination nodes, i.e., 30.96% were dropped.

Figure 6 illustrates the magnitude of loss for the video streams. In the case of Stream #1 and #2, losses appear constantly after the first minute, while packets of Stream #3 are dropped only after 2 minutes, leading to a lower frequency and average magnitude of loss events.

Figure 7 shows the SR of the different transmissions over the experiment duration. The success ratio corresponding to Stream #1 and Stream #2 starts to fall steadily after congestion is introduced on link R2–R3, reaching at the end of the experiment a minimum value of 68.86% and 65.26%, respectively. The SR corresponding to Stream #3 decreases after link R2–R6 is also congested, its final value being 73.08%. The global final SR is 69.04%.

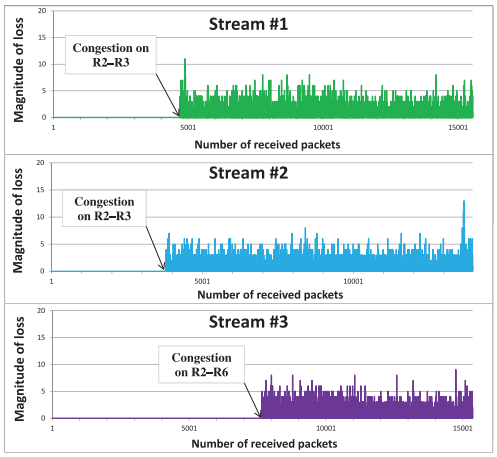


Fig. 6. Case 2 (ECMP) – Magnitude of loss

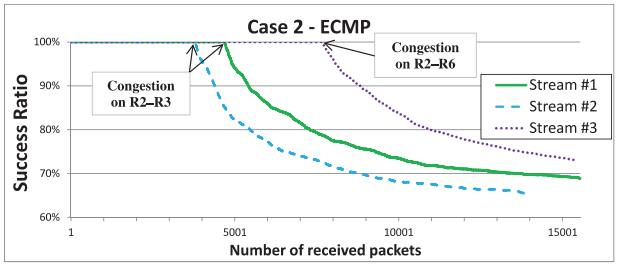


Fig. 7. Case 2 (ECMP) – Success ratio over experiment duration

Case 3 (SAMP with prediction)

The NN predictors integrated into the LMEs are implemented using *Flood*, an open source NNs C++ library [24]. A different NN is utilized to predict the ATR on every inbound link monitored by a LME. In order to reduce the overall complexity, the NNs have a small topology: 4–5–3, i.e., 4 input nodes, 5 hidden neurons and 3 output neurons. As a training algorithm, the Quasi-Newton method is used. The training lasted for 100 epochs and the learning rate was set to 0.01. The NNs are trained *offline* (i.e. before starting the

TABLE I
PARAMETERS OF THE RECEIVED VIDEO STREAMS

	Case 1 – OSPF			Case 2 – ECMP			Case 3 – SAMP with prediction		
	Stream #1	**Stream #2**	**Stream#3**	**Stream #1**	**Stream #2**	**Stream#3**	**Stream #1**	**Stream #2**	**Stream#3**
Sent packets	22567	21376	21034	22567	21376	21034	22567	21376	21034
Received packets	11602	10606	10284	15539	13950	15371	22356	21238	21034
Lost packets	10965	10770	10750	7028	7426	5663	211	138	0
% of lost packets	48.59%	50.38%	51.11%	31.14%	34.74%	26.92%	**0.94%**	**0.65%**	**0%**
Avg. magnitude of loss	0.945	1.015	1.045	0.452	0.532	0.369	0.009	0.006	0
Max. magnitude of loss	23	22	29	11	13	9	7	6	0
Discontinuity counter	4395	4386	4290	4283	4453	3355	139	81	0

experiments) with a special dataset of length 200, to detect congestions.

Until there is no congestion in the network, as illustrated in Figure 8, the proposed multipath solution sends each stream on a different path: *1)* Stream #1 on R1–R2–R3; *2)* Stream #2 on R1–R2–R6–R3, and *3)* Stream #3 on R1–R4–R5–R6–R3.

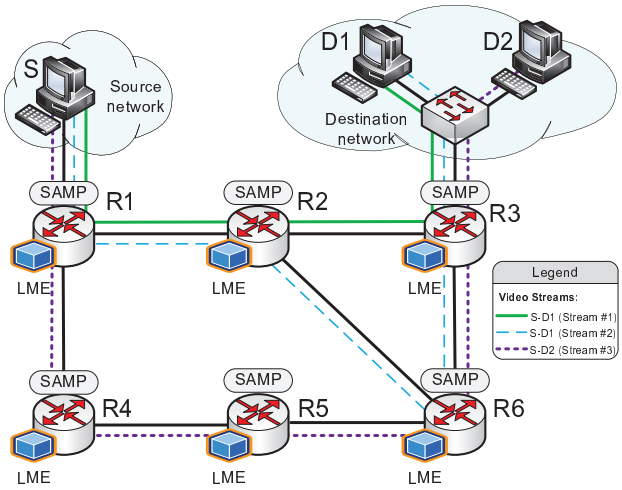


Fig. 8. Case 3 (SAMP) – No congestion

After 1 minute, we introduce congestion on link R2–R3 by starting several UDP streams between the two nodes. Thereby, the ATR on the affected link will decrease. By examining values of the ATR, the local management entity located on R3 will predict the appearance of congestion 1 second before it would be detected through simple monitoring. LME will trigger an alarm, indicating to the routing algorithm to recalculate the routes. The new best path followed by the affected Stream #1 is: R1–R2–R6–R3, as shown in Figure 9. The selection is based on the current state of the network links, choosing the path with the highest ATR and the lowest OWD.

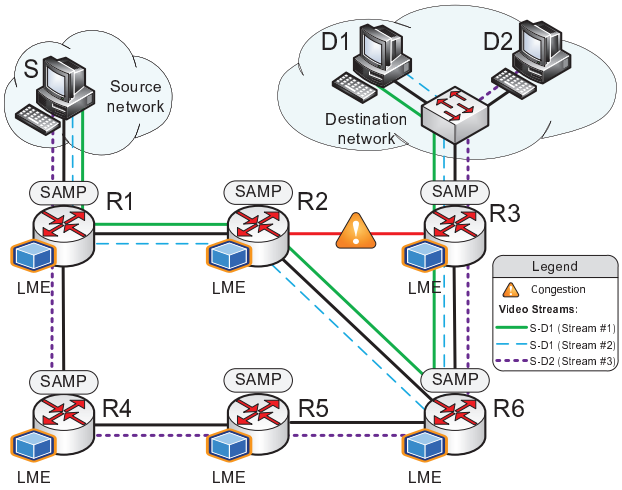


Fig. 9. Case 3 (SAMP) – Congestion on link R2--R3

The quality of the streaming is affected just for a very short period of time, mainly as a result of router reconfigurations. These are not carried out synchronously due to the distributed nature of the routing management system. At this moment, only 121 packets corresponding to Stream #1 are lost. Note that packets are also considered lost when they arrive out-of-order, because rearranging them at the destination is not feasible in case of video transmissions.

After an additional minute, congestion is introduced between nodes R2 and R6. LME on R6 detects lost packets and predicts the congestion. As a result, the multipath routing application will reroute the affected streams to the path used by Stream #3, namely R1–R4–R5–R6–R3, as presented in Figure 10. During this situation, 90 packets corresponding to Stream #1 and 138 packets from Stream #2 will be considered lost at the destination. The percentage of lost packets at the end of the experiments is 0.54% of the total number of packets sent.

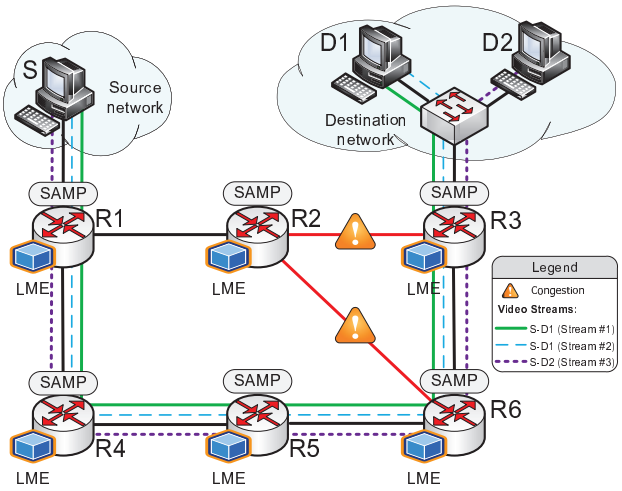


Fig. 10. Case 3 (SAMP) – Congestion on link R2--R3 and R2--R6

The VQ metrics of the received video streams when using the proposed predictive multipath routing framework are shown in Table I. Figure 11 illustrates the magnitude of loss events for the received videos. In case of Stream #1 there are two short time-intervals and for Stream #2 there is one short period in which losses occur. These correspond to the appearance of congestion and the rerouting.

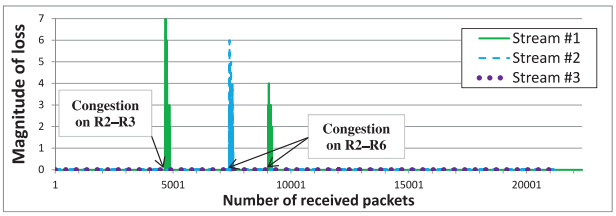


Fig. 11. Case 3 (SAMP) – Magnitude of loss

In Figure 12, the SR over the experiment duration can be observed. The SR for Stream #1 presents two local minimums: *1)* 97.56% when R2–R3 is congested and *2)* 97.76% when R2–R6 is congested; but after that, it recovers, increasing to the final value of 99.06%. For Stream #2 the SR drops for

 ISBN: 978-1-61208-174-8

a short time to 98.21% when R2–R6 is congested, but it increases by the end of the experiment to 99.35%. The SR for Stream #3 has a constant value of 100% because it is not affected by congestion. The value of the global final success ratio is 99.46%.

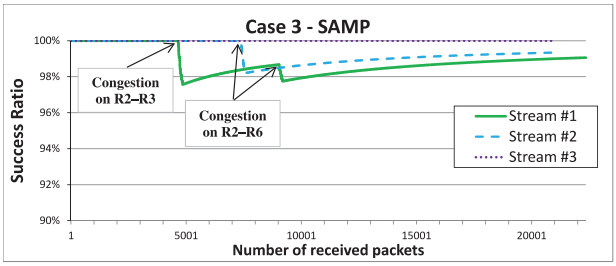


Fig. 12. Case 3 (SAMP) – Success ratio over experiment duration

Over the duration of the experiment, the prediction accuracy of the ATR is very high, in terms of NMSE (Normalized Mean Square Error) and E (Efficiency coefficient). Note that for a perfect prediction: NMSE $= 0$ and $E = 100\%$. In case of link R2–R3, we obtain NMSE $= 0.00091$ and $E = 99.91\%$, while for link R2–R6 we get NMSE $= 0.0011$ and $E = 99.89\%$.

In order to evaluate the beneficial effect of traffic forecasting, we performed the same test without predictors. In this case, the congestions is detected at a later moment, leading to a larger percentage of lost packets: 3.51%, 1.69%, and 0% for Stream #1, #2, and #3, respectively. In conclusion, if no prediction is used, the total loss (1.78%) is more than three times higher then in the case of embedding NNs-based predictors into the multipath routing framework (Table II).

TABLE II
PERCENTAGE OF LOST PACKETS

	OSPF	ECMP	SAMP no prediction	SAMP with prediction
Lost packets [%]	49.99	30.96	1.78	**0.54**

VII. CONCLUSION AND FUTURE WORK

This paper presented a multipath routing framework able to improve the global performance of the network, in case of congestion, by applying a predictive congestion control scheme. The goal was to offer a proof-of-concept by practical implementation in a real test environment. In our test scenario, the total lost percentage was: *1)* 49.99% with OSPF, *2)* 30.96% when employing ECMP, and *3)* **0.54%** when implementing our proposed solution. This approach significantly improves the link utilization and reduces the loss rate. We cannot demonstrate it at the moment, but we foresee that similar results would be obtained in a larger network topology. As future work, we intend to verify the results through simulation.

REFERENCES

[1] Y. Li, Z. Li, M. Chiang, and A. R. Calderbank, "Content-Aware Distortion-Fair Video Streaming in Congested Networks", IEEE Trans. on Multimedia, Vol. 11, No. 6, pp. 1182–1191, 2009.

[2] J. Moy, RFC 2328 "OSPF Version 2", Internet Engineering Task Force, 1998.

[3] G. Boanea, M. Barabas, A. B. Rus, and V. Dobrota, "Design Principles and Practical Implementation of a Situation Aware Multipath Routing Algorithm", Intern. Conf. on Software, Telecommunications and Computer Networks 2010, pp. 321–325. Split–Bol, Croatia, 2010.

[4] Z. Fan, "Bandwidth Allocation for MPEG-4 Traffic in IEEE 802.11e Wireless Networks Based on Traffic Prediction", Future Generation Communication and Networking, pp. 191–196. Jeju-Island, Korea, 2007.

[5] L. Cai, J. Wang, C. Wang, and L. Han, "A Novel Forwarding Algorithm over Multipath Network", Intern. Conf. on Computer Design and Applications, pp. V5-353–V5-357. Qinhuangdao, China, 2010.

[6] E. Marin-Tordera, X. Masip-Bruin, S. Sanchez-Lopez, J. Domingo-Pascual, and A. Orda, "The Prediction Approach in QoS Routing", IEEE Intern. Conf. on Communications, pp. 1020–1025. Istanbul, Turkey, 2006.

[7] Z. Guo, S. Sheikh, C. Al-Najjar, H. Kim, and B. Malakooti, "Mobile ad hoc network proactive routing with delay prediction using neural network", Wireless Networks, Vol. 16, No. 6, pp. 247–262, 2009.

[8] X. Masip-Bruin, E. Marin-Tordera, M. Yannuzzi, R. Serral-Gracia, and S. Sanchez-Lopez, "Reducing the Effects of Routing Inaccuracy by Means of Prediction and an Innovative Link-State Cost", IEEE Communications Letters, Vol. 14, No. 5, pp. 492–494, 2010.

[9] Z. Li, R. Wang, and J. Bi, "A Multipath Routing Algorithm Based on Traffic Prediction in Wireless Mesh Networks", Intern. Conf. on Natural Computation, pp. 115–119. Tianjin, China, 2009.

[10] Suyang Ju and J.B. Evans, "Intelligent Multi-Path Selection Based on Parameters Prediction", ICC Workshops, pp. 529–534. Beijing, China, 2008.

[11] Y. Liang, "Real-Time VBR Video Traffic Prediction for Dynamic Bandwidth Allocation", IEEE Trans. on Systems, Man, and Cybernetics, Part C: Applications and Reviews, Vol. 34, No. 1, pp. 32–47, 2004.

[12] Y.-R. Chuang, C.-S. Hsu, and J.-W. Chen, "Implementation of a Smart Traffic Prediction and Flow Control Mechanism for Video Streaming", Intern. Conf. on Intelligent Information Hiding and Multimedia Signal Processing, pp. 240–243. Darmstadt, Germany, 2010.

[13] P. Cortez, M. Rio, M. Rocha, and P. Sousa, "Internet Traffic Forecasting using Neural Networks", Intern. Joint Conf. on Neural Networks, pp. 2635–2642. Vancouver, Canada, 2006.

[14] D.-C. Park, "Prediction of MPEG Traffic Data Using a Bilinear Recurrent Neural Network with Adaptive Training", Intern. Conf. on Computer Engineering and Technology, pp. 53–57. Singapore, 2009.

[15] V. B. Dharmadhikari and J. D. Gavade, "An NN Approach for MPEG Video Traffic Prediction", Intern. Conf. on Software Technology and Engineering, pp. V1-57–V1-61. San Juan, USA, 2010.

[16] S. Sun, "Traffic Flow Forecasting Based on Multitask Ensemble Learning", ACM/SIGEVO Summit on Genetic and Evolutionary Computation, pp. 961–964. Shanghai, China, 2009.

[17] J. Rodrigues, A. Nogueira, and P. Salvador, "Improving the Traffic Prediction Capability of Neural Networks Using Sliding Window and Multi-task Learning Mechanisms", Intern. Conf. on Evolving Internet, pp. 1–8. Valencia, Spain, 2010.

[18] G. Boanea, M. Barabas, A. B. Rus, V. Dobrota, and J. Domingo-Pascual, "Performance Evaluation of a Situation Aware Multipath Routing Solution", RoEduNet Intern. Conf. Iasi, Romania, 2011.

[19] S. Kandula, D. Katabi, S. Sinha, and A. Berger, "Dynamic load balancing without packet reordering", ACM SIGCOMM Computer Communication Review, Vol. 37, No. 2, pp. 51–62. 2007.

[20] R. Li, L. Ying, A. Eryilmaz, and N. B. Shroff, "A Unified Approach to Optimizing Performance in Networks Serving Heterogeneous Flows", IEEE/ACM Trans. on Networking, Vol. 19, No. 1, pp. 223–236. 2011.

[21] M. Barabas, G. Boanea, A. B. Rus, and V. Dobrota, "Routing Management Based on Statistical Cross-Layer QoS Information Regarding Link Status", Intern. Conf. on Knowledge in Telecommunication Technologies and Optics, pp. 8-13. Szczyrk, Poland, 2011.

[22] D. Thaler and C. Hopps, RFC 2991 "Multipath Issues in Unicast and Multicast Next-Hop Selection", Internet Engineering Task Force, 2000.

[23] Quagga Routing Software Suite, http://www.quagga.net/ (Last accessed: 20.09.2011)

[24] R. Lopez, (2010). Flood: An Open Source Neural Networks C++ Library (Version 3) [software]. Retrieved from www.cimne.com/flood (Last accessed: 20.09.2011)

 ISBN: 978-1-61208-174-8

Blocking Performance of Multi-Rate OCDMA Passive Optical Networks

John S. Vardakas*, Ioannis D. Moscholios†, Michael D. Logothetis‡ and Vassilios G. Stylianakis§
*WCL, Dept. of Electrical & Computer Engineering, University of Patras, 265 04, Patras, Greece
Email: jvardakas@wcl.ee.upatras.gr
†Dept. of Telecommunications Science and Technology, University of Peloponnese, 221 00, Tripolis, Greece
Email: idm@uop.gr
‡WCL, Dept. of Electrical & Computer Engineering, University of Patras, 265 04, Patras, Greece
Email: m-logo@wcl.ee.upatras.gr
§WCL, Dept. of Electrical & Computer Engineering, University of Patras, 265 04, Patras, Greece
Email: stylian@wcl.ee.upatras.gr

***Abstract*—Optical Code Division Multiple Access (OCDMA) is one promising candidate for the provision of moderate security communications with large dedicated bandwidth to each end user. We present a new teletraffic model for the call-level performance of an OCDMA Passive Optical Network (PON) configuration that accommodates multiple service-classes. Parameters related to the local blocking, Multiple Access Interference (MAI) and user activity are incorporated to our analysis, which is based on a two-dimensional Markov chain. In this paper we focus on the derivation of recursive formulas for the efficient calculation of blocking probabilities. To evaluate the proposed model, the analytical results are compared with simulation results to reveal that the model' s accuracy is quite satisfactory.**

***Keywords-PON; OCDMA; Multiple Access Interference; Loss; Blocking Probability; Recursive Formula; Markov chain*.**

I. INTRODUCTION

The popularity of broadband applications, including bandwidth-hungry multimedia services and Internet Protocol Television (IPTV), has promoted the fiber-to-the-end-user as a feasible access networking technology. Among different optical access solutions, Passive Optical Networks (PONs) have received a tremendous attention from both industrial and academic communities, mainly due to the low operational cost, the enormous bandwidth offering and the absence of active components between the central office and the customers premises [1].

Over the years, several standards for PONs have evolved, in the form of the G.983 ITU-T recommendations series, which include Asynchronous Transfer Mode PONs (ATM-PONs), Ethernet PONs (EPONs) and Broadband PONs (BPONs). These architectures are based on a Time Division Multiple Access (TDMA) scheme and they typically use a 1550 nm wavelength for the downstream traffic and a 1310 nm for the upstream traffic [2]. While these TDMA-PONs employ two wavelengths for the upstream and downstream direction, respectively, the Wavelength Division Multiplexing (WDM)-PON utilizes multiple wavelengths, so that two wavelengths are allocated to each user for down/upstream transmissions. A different approach for the provision of multiple access in PONs is the Optical Code Division Multiple Access (OCDMA). In contrast to the other multiple access schemes OCDMA can multiplex a number of channels on the same wavelength and on the same time-slot [3]. In addition, OCDMA offers full asynchronous transmission, soft capacity on demand, low latency access, simple network control and better security against unauthorized access [4].

In OCDMA, each communication channel is distinguished by a specific optical code. The encoding procedure involves the multiplication of each data bit by a code sequence either in the time domain [5], in the wavelength domain [6], or in a combination of both [7]. The decoder receives the sum of all encoded signals (from different transmitters) and recovers the data from a specific encoder, by using the same optical code. All the remaining signals appear as noise to the specific receiver; this noise is known as Multiple Access Interference (MAI) and is the key degrading factor of the networks performance.

To perform service differentiation in OCDMA networks, different solutions have been investigated. A simple approach is based on the utilization of multi-length codes [8]; however, under multi-length coding, short-length codes introduce significant interference over long-length codes, while high error probability emerges for high rate users. Optical fast-frequency hopping has been also proposed for multi-rate OCDMA networks [9]. This technique is based on multiple wavelengths; therefore it requires multi-wavelength transmitters with high sensitivity on power control. Another way to provide service differentiation is the assignment of several codes to each service-class. This procedure is known as the parallel mapping technique [10]. In this case the number of codes is proportional to the data rate of the assigned service-class.

The research activity on OCDMA networks is mainly focused on the performance of several OCDMA components; only a few analytical models have been presented in the literature involving the computation of blocking probabilities in OCDMA networks. Goldberg and Prucnal [11] provide

 ISBN: 978-1-61208-174-8

analytical models for the determination of blocking probabilities and for the teletraffic capacity in OCDMA networks. A similar study is performed in [12]. In both cases, a single service-class is considered, while the different sources of noise that are present in an OCDMA network are not taken into account. In [13] we provided a call-level analysis of hybrid WDM-OCDMA PONs, which is based on a similar teletraffic model for the call-level performance modeling of Wireless OCDMA (W-OCDMA) networks [14].

In [15] we proposed a call-level performance analysis of OCDMA PONs, where multiple service-classes of infinite traffic source population are accommodated. The shared medium is modeled by a two dimensional Markov chain. Based on it, we provided an approximate recursive formula for the calculation of blocking probabilities in the PON. The analysis takes into account the user activity, by incorporating different service times for active and passive (silent) periods. In this paper, we provide the proof of the recursive formula that describes the distribution of the occupied bandwidth in the PON that we presented in [15]. We present the necessary assumptions that have to be taken into account in order to derive the approximate recursive formula. The capacity of the PON is defined by the total interference caused by both active and passive users. An arriving call may be blocked if the resulting interference exceeds a predefined maximum threshold. This case defines the Hard Blocking Probability (HBP). A call may also be blocked in any other system state, due to the existence of different forms of additive noise (thermal noise, shot noise, beat noise). The latter case is expressed by the Local Blocking Probability (LBP). The accuracy of the proposed algorithm is evaluated through simulation and is found to be quite satisfactory.

This paper is organized as follows. Section II includes the system model. In Section III, we derive recursive formulas for the blocking probabilities calculation. Section IV is the evaluation section. We conclude in Section V.

II. System Model

We consider the OCDMA PON of Fig. 1. A number of Optical Network Units (ONUs) located at the users' premises are connected to a centralized Optical Line Terminal (OLT) through a Passive Optical Splitter/Combiner (PO-SC). The PO-SC is responsible for the broadcasting of traffic from the OLT to the ONUs (downstream direction) and for the grouping of data from the ONUs and the transmission of the collected data to the OLT through one fiber (upstream direction). We study the upstream direction; however the following analysis can be applied to the downstream direction. Users that are connected to an ONU switch between active and passive (silent) periods. The PON supports *K* service-classes. The service-differentiation in this OCDMA system is realized using the parallel mapping technique. When a single codeword is applied to an active call, the interference that this call causes to the receiver is denoted by I_{unit}.

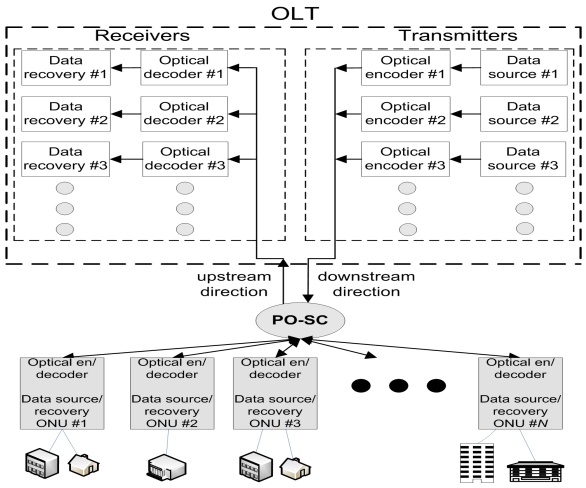


Figure 1. A basic configuration of an OCDMA PON

Since different service-classes require different data-rates, a number of codewords is assigned to each service-class; therefore the interference I_k that an active service-class *k* call causes to the receiver is proportional to I_{unit}. We define the bandwidth requirement b_k, of service-class *k*, as the number of codewords assigned to the specific service-class, which is equal to the ratio of I_k to I_{unit}, therefore:

$$b_k = \frac{I_k}{I_{unit}} \tag{1}$$

Calls that are accepted for service start an active period and may constantly remain in the active state for the entire duration of the call, or alternate between active and passive states. Throughout an active state, the traffic source sends bursts, while during a passive state no transmission of data occurs. When a call is transferred from the active state to the passive state, the bandwidth (which is expressed by the interference that this call produces to the receiver) held by the call in the active state is released and this bandwidth becomes available to new arriving calls. When a call attempts to become active again, it re-produces the same amount of interference (as in the previous active state); if the total interference at the receiver does not exceed a maximum value, a new active period begins; if not, burst blocking occurs and the call remains in the passive state. At the end of the active period the total interference at the receiver is reduced by I_k and the call either jumps to the passive state with probability v_k, or departs from the system with probability $1 - v_k$. At the end of the passive period the call returns to an active state only if the call will not be blocked due to the presence of the additive noise. Furthermore, calls that belong to service-class *k* arrive to an ONU according to the Poisson process; the total arrival rate from all ONUs is denoted λ_k. The service time of service-class *k* calls in state *i*, (*i*=1 indicates the active state, *i*=2 the passive state) is exponentially distributed with mean μ_{ik}^{-1}.

 ISBN: 978-1-61208-174-8

According to the principle of the CDMA technology, a call should be blocked if it increases the noise of all in-service calls above a predefined level, given that a call is noise for all other calls. This noise is known as MAI. We distinguish the MAI from other forms of noise, the shot noise, the thermal noise and the fiber-link noise. The thermal noise is generally modeled as Gauss distribution (0, *th*), and the fiber-link noise is modeled as Gauss distribution (0, σ_{fb}) [16]. The shot noise is modeled as a Poisson process where its expectation and variance are both denoted by *p* [16]. According to the central limit theorem, we can assume that the additive shot noise is modeled as Gauss distribution (μ_N, σ_N), considering that the number of users in the PON is relatively large. Therefore, the interference I_N caused by the the three types of noise is modeled as a Gaussian distribution with mean $\mu_N = p$ and variance $\sigma_N = \sqrt{\sigma_{th}^2 + \sigma_{fb}^2 + p^2}$.

The Call Admission Control (CAC) in the OCDMA PON under consideration is performed by measuring the total interference at the receiver. When a new call arrives (which automatically enters an active state), the CAC checks if two conditions are valid. The first condition (condition A) estimates the total received interference and if it exceeds a maximum value I_{max}, the call is blocked and lost. Therefore, condition A is expressed by the following relation:

$$\sum_{k=1}^{K}(n_k^1 I_k) + I_k + I_N > I_{\max} \Leftrightarrow \frac{I_N}{I_{\max}} > 1 - \sum_{k=1}^{K}(n_k^1 \frac{I_k}{I_{\max}}) - \frac{I_k}{I_{\max}} \quad (2)$$

where n_k^1 represent the number of the service-class *k* calls in the active system. The same condition is used at the receiver, when a passive call jumps to an active state. Based on (2), we define the LBP $lb_k(n_k^1)$ that a service-class *k* call is blocked due to the presence of the additive noise, when the number of active calls is n_k^1:

$$lb_k(n_k^1) = P\left(\frac{I_N}{I_{\max}} > 1 - \sum_{k=1}^{K}\left(n_k^1 \frac{I_k}{I_{\max}}\right) - \frac{I_k}{I_{\max}}\right) \quad (3)$$

or

$$1 - lb_k(n_k^1) = P\left(\frac{I_N}{I_{\max}} \leq 1 - \sum_{k=1}^{K}\left(n_k^1 \frac{I_k}{I_{\max}}\right) - \frac{I_k}{I_{\max}}\right) \quad (4)$$

Since the total additive noise I_N follows a Gaussian distribution (μ_N, σ_N), the variable I_N/I_{max}, which is used for the LBP calculation also follows a Gaussian distribution (μ_N/I_{max}, σ_N/I_{max}) Therefore the right-hand side of (4), which is the Cumulative Distribution Function (CDF) of I_N/I_{max}, is denoted by $F_n(x) = P(I_N/I_{\max} \leq x)$ and is given by:

$$F_n(x) = \frac{1}{2}(1 + erf(\frac{x - (\mu_N/I_{max})}{(\sigma_N/I_{max})\sqrt{2}})) \quad (5)$$

where $erf(\bullet)$ is the well-known error function. Using (2) and (4) we can calculate the LBP, $lb_k(n_k^1)$ by means of the substitution $x = 1 - \sum_{k=1}^{K}(n_k^1 \frac{I_k}{I_{\max}}) - \frac{I_k}{I_{\max}}$:

$$lb_k(x) = \begin{cases} 1 - F_n(x), & x \geq 0 \\ 1, & x < 0 \end{cases} \quad (6)$$

III. The Distribution Of The Number Of Active And Passive Calls

The following analysis is inspired by the multi-rate ON-OFF model for the call-level performance of a single link, presented in [17], [18] , which considers discrete state space. The discretization of the total interference I_{max} is performed by using the interference caused by a single-codeword call:

$$C_1 = \left|\frac{I_{\max}}{I_{unit}}\right| \quad (7)$$

When a call is at the passive state, it is assumed that it produces a fictitious interference of a fictitious system, with a discrete capacity C_2: This *passive system* is used to prevent new calls to enter the system when a large number of calls are at the passive state. In order to employ the analysis presented in [17], we use the following notations:

1) the total interference of the in-service active calls represents the occupied bandwidth of the active system, and is denoted by j_1.
2) the total (fictitious) interference of the in-service passive calls represents the occupied bandwidth of the passive system, and is denoted by j_2.

Based on the analysis presented in the previous section, a new call will be accepted for service if the call's interference (which is expressed by the bandwidth requirement c_k together with the interference caused by all the in-service active calls, (which is expressed by the parameter j_1), will not exceed I_{max}. Moreover, in order to avert the acceptance of new calls when a large number of calls are at the passive state, the interference of the new call together with the total interference of all in-service active calls and the fictitious interference of all in-service passive calls should not exceed the fictitious capacity (which is expressed by the discrete value C_2). Based on this analysis, a new service-class *k* call will be accepted for service in the system, if it satisfies both the following constraints:

$$j_1 + b_k \leq C_1 \text{ and } j_1 + j_2 + b_k \leq C_2 \quad (8)$$

If we denote by Ω the set of the permissible states, then the distribution $\vec{j} = (j_1, j_2)$, denoted as $q(\vec{j})$ can be calculated by the proposed two-dimensional approximate recursive formula:

$$\sum_{i=1}^{2}\sum_{k=1}^{K} b_{i,k,s} p_{ik}\left(\vec{j}\right) q(\vec{j} - B_{i,k}) = j_s q(\vec{j}) \quad (9)$$

where

$$\vec{j} \in \Omega \Leftrightarrow \left\{\left(j_1 \leq C_1 \cap \left(\sum_{s=1}^{2} j_s \leq C_2\right)\right)\right\} \quad (10)$$

 ISBN: 978-1-61208-174-8

The parameter *s* refers to the systems (*s*=1 indicates the active system, *s*=2 the passive system), while *i* refers to the states (*i*=1 specifies the active state, *i*=2 specifies the passive state). Also,

$$b_{i,k,s} = \begin{cases} b_k, & \text{if } s = i \\ 0, & \text{if } s \neq i \end{cases} \quad (11)$$

and $B_{i,k} = (b_{i,k,1}, b_{i,k,2})$ is the *i,k* row of the ($2K \times 2$) matrix *B*, with elements $b_{i,k,s}$. Also, $p_{ik}(\vec{j})$ is the utilization of the *i*-th system by service-class *k*:

$$p_{i,k}\left(\vec{j}\right) = \begin{cases} \frac{\lambda_k[1-lb(j_1-b_k)]}{(1-v_k)\mu_{1k}} & \text{for } i = 1 \\ \frac{\lambda_k \sigma_k}{(1-v_k)\mu_{2k}} & \text{for } i = 2 \end{cases} \quad (12)$$

Moreover, j_s is the occupied capacity of the system:

$$j_s = \sum_{i=1}^{2} \sum_{k=1}^{K} n_k^i b_{i,k,s} \quad (13)$$

Proof: In order to derive the recursive formula of (9) we introduce the following notation:

$$\begin{array}{l} \vec{n} = (n^1, n^2),\ n_k^i = (n_1^i, n_2^i, ..., n_K^i), \\ n_{k+}^i = (n_1^i, ..., n_k^i + 1, ..., n_K^i), \\ n_{k-}^i = (n_1^i, ..., n_k^i - 1, ..., n_K^i), \\ \vec{n}_{k+}^1 = (n_{k+}^1, n^2),\ \vec{n}_{k+}^2 = (n^1, n_{k+}^2), \\ \vec{n}_{k-}^1 = (n_{k-}^1, n^2),\ \vec{n}_{k-}^2 = (n^1, n_{k-}^2) \end{array} \quad (14)$$

Having determined the steady state of the system $\vec{n} = (n^1, n^2)$, we proceed to the depiction of the transitions from and to state $\vec{n}$, as it is shown in Fig. 2. The horizontal axis of the state transition diagram of Fig. 2 reflects the arrivals on new calls and the termination of calls. More specifically, when the system is at state (A) it will jump to state (B) with a rate λ_k, when a new service-class *k* call arrives at the system. This rate is multiplied by the probability $1 - lb(n_k^1 - 1)$ that this call will not be blocked due to the presence of the additive noise. Similarly, we define the rate from state (C) to state (A). From state (B) the system will jump to state (A) $\mu_{1,k}\left(n_k^1+1\right)(1-v_k)$ times per unit time, since one of the n_k^1+1 active calls of service-class *k* (in state (B)) will depart from the system with probability $(1 - v_k)$. The transition from state (A) to state (C) is defined in a similar way.

The vertical axis of the state transition diagram of Fig. 2 defines the transition from the active state to the passive state and vice versa. In particular, when the system is at state (A) it will jump to state (D) $\mu_{2,k} n_k^2 \left[1 - lb_k\left(n_k^1\right)\right]$ times per unit time. In this case a transition from the passive state to the active state occurs; this transition will be blocked only due to the presence of the additive noise, which is expressed by the LBP. The reverse transition (from state (D) to state (A)) occurs when one of $n_k^1 + 1$ active calls jumps to the passive state with probability v_k. Similarly, we can define the transitions between states (E) and (A).

Let $P(\vec{n})$ be the probability of the steady state of the state transition diagram. Assuming that local balance exists between two subsequent states, we derive the local balance equations:

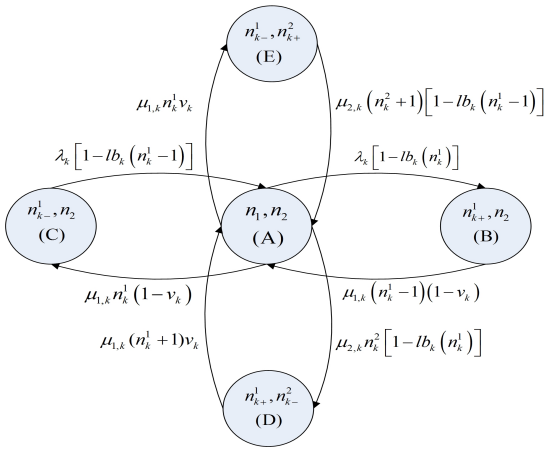


Figure 2. State transition diagram of the OCDMA system with active and passive users.

$$\begin{array}{l} P(\vec{n})\mu_{ik} n_k^1 v_k = P(\vec{n}_{k-+})\mu_{2k}(n_k^2+1)[1-lb(n_k^1-1)] \\ P(\vec{n})\lambda_k[1-lb(n_k^1)] = P(\vec{n}_{k+}^1)\mu_{1k}(n_k^1+1)(1-v_k) \\ P(\vec{n})\mu_{2k} n_k^2[1-lb(n_k^1)] = P(\vec{n}_{k+-})\mu_{1k}(n_k^1+1)v_k \\ P(\vec{n})\mu_{1k} n_k^1(1-v_k) = P(\vec{n}_{k-}^1)\lambda_k[1-lb(n_k^1-1)] \end{array} \quad (15)$$

We assume that the system of (15) has a Product Form Solution (PFS):

$$P(\vec{n}) = \frac{1}{G} \prod_{i=1}^{2} \prod_{k=1}^{K} \frac{p_{ik}^{n_k^1}(n_k)}{n_k^1!} \quad (16)$$

where *G* is a normalization constant and $p_{ik}(n_k)$ is given by

$$p_{i,k}\left(n_k\right) = \begin{cases} \frac{\lambda_k\left[1-lb_k\left(n_k^1-1\right)\right]}{(1-v_k)\mu_{1k}} & \text{for } i = 1 \\ \frac{\lambda_k \sigma_k}{(1-v_k)\mu_{2k}} & \text{for } i = 2 \end{cases} \quad (17)$$

In order for (16) to satisfy all equations of the system of (15) using (17), we assume that $1 - lb(n_k^1) \approx 1 - lb(n_k^1 - 1)$, i.e. the acceptance of one additional call in active state does not affect the LBP. This is the first assumption that we take into account in order to derive (9). By using (16), the probability $P(\vec{n}_{k-}^i)$ can be expressed by:

$$n_k^i P(\vec{n}) = p_{ik}(n_{k-}) P(\vec{n}_{k-}^i) \quad (18)$$

The probability of $q(\vec{j})$ is given by:

$$q(\vec{j}) = P(\vec{j} = \vec{n} \cdot B) = \sum_{\vec{n} \in \Omega_{\vec{j}}} P(\vec{n}) \quad (19)$$

where $\Omega_{\vec{j}} = \left\{\vec{n} \in \Omega_{\vec{j}} : \vec{n}B = \vec{j},\ \mathrm{n}_k^i \geq 0,\ k = 1, ..., K\right\}$. By multiplying both sides of (19) with $b_{i,k,s}$, and summing over *k*=1,...,*K* and *i*=1,2, we have:

$$P(\vec{n}) \sum_{i=1}^{2} \sum_{k=1}^{K} b_{i,k,s} n_k^i = \sum_{i=1}^{2} \sum_{k=1}^{K} b_{i,k,s} p_{ik}(n_{k-}) P(\vec{n}_{k-}^i) \quad (20)$$

By using (13) and summing both sides of (20) over the set of all states of $\Omega_{\vec{j}}$, we have:

$$j_s \sum_{\vec{n}\in\Omega_{\vec{j}}} P(\vec{n}) = \sum_{i=1}^{2}\sum_{k=1}^{K} b_{i,k,s} p_{ik} \sum_{\vec{n}\in\Omega_{\vec{j}}} p_{ik}(n^i_{k-}) P(\vec{n}^i_{k-}) \quad (21)$$

The second assumption that we consider is that:

$$\sum_{\vec{n}\in\Omega_{\vec{j}}} p_{ik}(n^i_k) P(\vec{n}^i_{k-}) \approx p_{ik}(\hat{n}^i_k) \sum_{\vec{n}\in\Omega_{\vec{j}}} P(\vec{n}^i_{k-}) \quad (22)$$

Based on the fact that $(\vec{n}B = \vec{j}) \Rightarrow (n^i_{k-}B = j - \mathrm{B}_{i,k})$ (17) is equal to (12) and (19) can be rewritten as:

$$q(\vec{j} - B_{i,k}) = \sum_{\vec{n}\in\Omega_{\vec{j}}} P(n^i_{k-}) \quad (23)$$

By using the assumption of (22) and substitute (19) and (23) to (22), we derive the recursive formula of (9). The LBP is a function of the total interference of the in-service active calls j_1, i.e. $lb(n^1_k) = lb(j_1)$, since

$$x = 1 - \sum_{k=1}^{K}(n^1_k \frac{I_k}{I_{\max}}) - \frac{I_k}{I_{\max}} = 1 - \sum_{k=1}^{K}(n^1_k \frac{b_k}{C_1}) - \frac{b_k}{C_1} = 1 - \frac{j_1}{C_1} - \frac{b_k}{C_1} \quad (24)$$

CBP is calculated by combining LBP and HBP as follows:

$$Pb_k = \sum_{\vec{j}\in\Omega-\Omega_h} lb_k(j_1) q(\vec{j}) + \sum_{\vec{j}\in\Omega_h} G^{-1} q(\vec{j}) \quad (25)$$

where $\Omega_h = \left\{\vec{j} \mid [(b_{i,k,1} + j_1) > C_1] \cup [(b_{i,k,2} + j_1 + j_2) > C_2]\right\}$. The first summation of the right part of (25) refers to the probability that a new call could be blocked at any system state due to the presence of the additive noise. The second summation signifies the HBP, which is derived by summing the probabilities of all the blocking states that are defined by (8). Note that the bounds of the first summation in (25) are accidentally different than those of the corresponding equation in [16] due to a misprint in (10) of [16].

IV. Evaluation

In this section we evaluate the proposed anaytical model through simulation. To this end we simulate the OCDMA PON of Fig.1 by using the Simscript II.5 [19] simulation tool. The simulation results are mean values from 6 runs with confidence interval of 95%. The resulting reliability ranges of the simulation measurements are small and therefore we present only mean results. We consider that the OCDMA PON supports $K = 2$ service-classes. The traffic parameters $(I_k, \mu_{1k}, \mu_{2k}, \sigma_k)$ of each service-class are (5, 1, 1.2, 0.9) and (1, 0.7, 1, 1), where I_k is expressed in μW and μ_{ik} in sec^{-1}. The threshold of the total interference at the receiver is 60 μW, while the fictitious interference that describes the passive system is 70 μW. The total additive noise follows a Gaussian distribution (1, 0.1) μW. The interference that a single-codeworded call is assumed to be 0.05 μW. In Fig. 3 we comparatively present analytical and simulation CBP results vs. the arrival rate, which is assumed that is the same for both service-classes. The results of Fig. 3 show that the accuracy of the proposed analysis is quite satisfactory. Small declinations between analytical and simulation results are due to the assumptions that were taken into account in order to prove (9).

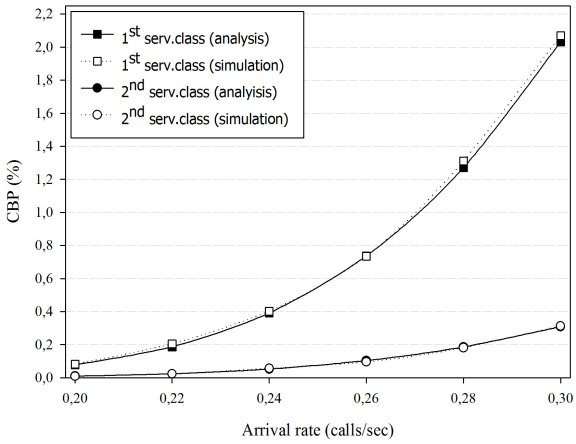


Figure 3. Analytical and simulation CBP results vs. offered traffic load of each service-class.

In order to demonstrate the effect of the additive noise to the CBP, in Fig. 4 we present analytical CBP results of the first service-class versus the arrival rate, for different noise distributions. As the results of Fig. 4 reveal, the increment of the additive noise results in the increase of the CBP.

We also study the effect of the fictitious capacity of the passive system to the CBP. To this end, we present analytical CBP results of both service-classes vs. different values of the fictitious capacity. the arrival rate of both service-classes is 0.3 calls/sec. The increment of C^* results in lower CBP, since the passive system can accommodate more calls, and therefore less interference is present in the active system. However, this increment results in a higher probability that a call is blocked at its transition from passive to active state.

V. Conclusion

We propose a new multi-rate loss model for the calculation of blocking probabilities in an OCDMA-PON. Our analysis takes into account the user activity and different service times for active and passive periods. We provide and prove an approximate recurrent formula for the efficient calculation of the CBP, which is a function of the LBP, and of the HBP. The accuracy of the proposed analysis is quite satisfactory, as it was verified by simulations.As a future work we will incorporate a finite population of traffic sources in the CBP calculation, while we will study the case where the receiver has an interference cancellation capability.

 ISBN: 978-1-61208-174-8

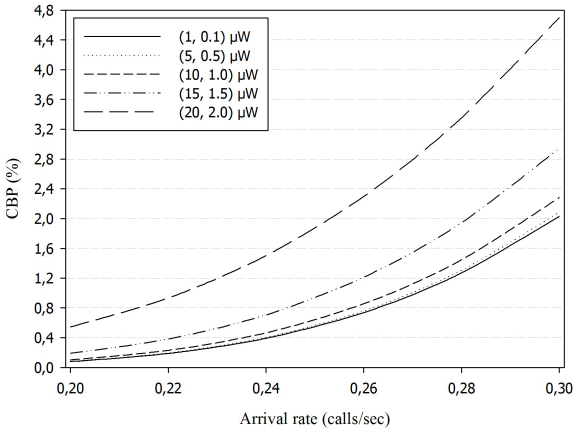


Figure 4. Analytical CBP results of the first service-class for different additive noise distributions.

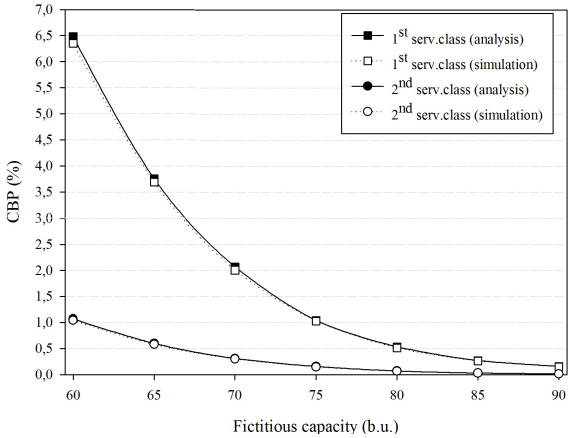


Figure 5. Analytical and simulation CBP results vs. the fictitious capacity.

ACKNOWLEDGMENT

Work supported by the Research Program Caratheodory (Research Committee, University of Patras, Greece).

REFERENCES

[1] F. Effenberger, D. Cleary, O. Haran, G. Kramer, R. Li, M. Oron, T. Pfeiffer, "An Introduction to PON Technologies", *IEEE Communications Magazine*, March 2007, pp. S17-S25.

[2] T. Koonen, "Fiber to the Home/Fiber to the Premises: What, Where, and When?",*Proceedings of the IEEE*, Vol. 94, No. 5, May 2006, pp. 911-934.

[3] K. Fouli and M. Maier, "OCDMA and Optical Coding- Principles, Applications, and Challenges", *IEEE Communications Magazine*, August 2007, pp. 27-34.

[4] P. R. Prucnal, *Optical Code Division Multiple Access: Fundamentals and Applications*, New York: Taylor & Francis, 2006.

[5] M. Azizoglu, J.A. Salehi, Y. Li, "Optical CDMA via temporal codes", *IEEE Transactions on Communications*, Vol.40, No. 7, July 1992, pp. 1162-1170.

[6] D. Zaccarin, M. Kavehrad, "An optical CDMA system based on spectral encoding of LED", *IEEE Photonics Technology Letters*, Vol. 5, No. 4, 1993, pp. 479-482.

[7] M. Morelle, C. Goursaud, A. J.-Vergonjanne, C. A.-Berthelemot, J.-P. Cances, J.-M. Dumas, and P. Guignard, "2-Dimensional optical CDMA system performance with parallel interference cancellation", *Microprocessors and Microsystems*, Vol. 31, 2007, pp. 215-221.

[8] W.C. Kwong, G.C. Yang,"Multiple-Length extended Carrier-Hopping Prime Codes for Optical CDMA systems supporting Multirate Multimedia services", *IEEE/OSA Journal of Lightwave Technology*, Vol. 23, No. 11, November 2005, pp. 3653-3662.

[9] E. Intay, H. M. H. Shalaby, P. Fortier, and L. A. Rusch, "Multirate optical fast frequency-hopping CDMA system using power control", *IEEE/OSA Journal of Lightwave Technology*, Vol. 20, No. 2, February 2002, pp. 166-177.

[10] A.R. Forouzan, N-K. Masoumeh, N. Rezaee, "Frame Time-Hopping patterns in Multirate Optical CDMA Networks using Conventional and Multicode schemes", *IEEE Transactions on Communications*, Vol. 53, No. 5, May, 2005, pp. 863-875.

[11] S. Goldberg and P. R. Prucnal, "On the Teletraffic Capacity of Optical CDMA", *IEEE Transactions on Communications*, Vol. 55, No. 7, July 2007, pp. 1334-1343.

[12] M. Gharaei, C. Lepers, O. Affes, and P. Gallion, "Teletraffic Capacity Performance of WDM/DS-OCDMA Passive Optical Network", *NEW2AN/ruSMART* 2009, LNCS 5764, Springer-Verlag Berlin Heidelberg, 2009, pp. 132-142.

[13] John. S. Vardakas, Vasillios G. Vassilakis and Michael D. Logothetis, "Call-Level Analysis of Hybrid OCDMA-WDM Passive Optical Networks", *in Proc. of the 10th ICTON* 2008, Athens, Greece, 22-26 July 2008.

[14] V. G. Vassilakis, G. A. Kallos, I.D. Moscholios, and M. D. Logothetis, "The Wireless Engset Multi-Rate Loss Model for the Call-level Analysis of W-CDMA Networks", *in Proc. of the 8th IEEE PIMRC* 2007, Athens 2007.

[15] John S. Vardakas, Ioannis L. Anagnostopoulos, Ioannis D. Moscholios, Michael D. Logothetis, Vassilios G. Stylianakis, "A Multi-Rate Loss Model fol OCDMA PONs", *in Proc. of the 13th ICTON* 2011, Stockholm, Sweden, 26-30 June 2011.

[16] W. Ma, C. Zuo, and J. Lin, "Performance Analysis on Phase-Encoded OCDMA Communication System", *IEEE/OSA Journal of Lightwave Technology*, Vol. 20, No. 5, May 2002.

[17] M. Mehmet Ali, "Call-burst blocking and call admission control in a broadband network with bursty sources", *Performance Evaluation*, 38, 1999, pp. 1-19.

[18] I.Moscholios, M. Logothetis and G. Kokkinakis, "Call-burst blocking of ON-OFF traffic sources with retrials under the complete sharing policy", *Performance Evaluation*, 59/4, 2005, pp. 279-312.

[19] Simscript II.5, http://www.simscript.com/

 ISBN: 978-1-61208-174-8

A Multimedia Capture System for Wildlife Studies

Kim Arild Steen
Department of Engineering,
Aarhus University
Faculty of Science and Technology
Dalgas Avenue 2, Aarhus, Denmark
KimA.Steen@agrsci.dk

Henrik Karstoft
Department of Engineering,
Aarhus University
Faculty of Science and Technology
Aarhus, Denmark
hka@iha.dk

Ole Green
Department of Engineering,
Aarhus University
Faculty of Science and Technology
Aarhus, Denmark
Ole.Green@agrsci.dk

***Abstract*—This paper presents a system for video and audio recording of wildlife geese in their natural environment. The system enables remote controlled recording, and is designed for an outdoor environment. The recordings lasted 1 month, where 4-5 hours of geese video and audio were succesfully captured. Data recorded using the system is a part of ongoing research to design a method for automatic recognition of animal behaviour and species based on audio and video recordings. Automatic recognition could potentially lead to systems capable of reducing habituation.**

***Keywords*-video; audio; recording; wildlife surveillance; remote**

I. INTRODUCTION

In modern society, we often experience unwanted encounters between groups of animals and human activities, such as in agricultural fields or at airports. This can be a costly affair and often inflicts damages both to the animals as well as humans. In the case of agricultural fields, visual and acoustic stimuli may be used as mechanisms for scaring away unwanted animals. However, these methods often have limited success rates, as the animals habituate to the stimulus [2].

Recently, computer technology has been applied for characterising animal behaviour using computer vision for tracking animal trajectories [6][8] and audio processing for recognition of animal vocalizations [1][5][7]. These approaches may lead to systems capable of recognizing specific species and behaviours, and scare off the animals before they inflict damage or get hurt. In [2] different approaches to scaring off animals is reviewed, such as guard animals, gas exploders and distress calls, with the latter showing good results.

In the process of linking wildlife animals vocalizations with specific behaviour, a system for recording video and audio data in a wildlife setting is presented. Wildlife surveillance systems have been previously described [3][4]; however, these systems are designed for specific scenarios. Likewise, the present system is specific to the context of video and audio recording of wildlife birds foraging in agricultural fields. The system described in [3] support both audio and video recording in a harsh environment (humid environment), however the data recorded is used for manual inspection and not research regarding automatic recognition.

The system has been used for video and audio recording of wildlife geese foraging in agricultural fields. The main purpose of this system is to record and store images and audio of geese as they land, eat and flee. The data provided by the system will be used in further research regarding automatic recognition of geese behaviour. Geese were chosen in this study, as they inflict much damage in agriculture, and they are very vocal.

The structure of this paper is as follows: In Section II the requirements for the system is presented followed by a description of the system in Section III. The implemented infrastructure is presented in Section IV and preliminary results, using the system, are presented in Section V. The discussion in Section VI follows up on the results and experience gained while using the system. The paper ends with a conclusion in Section VII.

II. SYSTEM REQUIREMENTS

Agricultural fields are wide open spaces, implying windy conditions during wildlife recordings. Wind reduction is therefore necessary to preserve the quality of the audio recordings, and can be accomplished for instance through use of a casing. Furthermore, the remote location of agricultural fields reduces access to power grids. Consequently the consideration of a power source and power consumption is important, as the system requires a standalone power source.

Barnacle geese are highly mobile with flight speed up to $20\frac{m}{s}$[9], and the video recording equipment needs to provide adequate frame rates to capture their movements. It is not desirable to reduce the image quality or add computations by adding compression, as this could degrade performance of later image processing and potentially cause fluctating frame rates, as compression time could be affected by information in the images.

As it is impossible to pinpoint in advance where the geese will land and eat, video and audio recordings need to be inspected during the study. Consequently remote access is an important system requirement, to avoid frightening the geese during inspection.

 ISBN: 978-1-61208-174-8

A further system requirement is a minimum uptime long enough to capture video and audio as the geese return to the location, to avoid interference caused by installation of the system. An uptime of 36 hours has been chosen, as the geese are likely to return to the same location because of the availability of food, however it is not certain that they will return the same day. Therefore, a harddrive with a large enough capacitiy, must be chosen to ensure no loss of data.

To summarize the most important requirements for a multimedia capture device for wildlife studies:

- Reduction of wind noise in the audio recording equipment
- Standalone power source (limitations to power consumption)
- Adequate frame rates ($20-30$ frames per second (fps)), due to the mobile animals
- Remote access, to monitor the recording without scaring off the animals
- Minimum uptime of 36 hours
- High harddrive capacity ($>$ 1 TB), due to the long uptime and no compression

III. System description

The requirements, specified above, led to the system setup, described in this section and illustrated in Figure 1.

The power source needs to be standalone, and two soultions were considered: car batteries and solar panels. The power produced by a solar panel is dependent on the weather, and, as sunlight is not garanteed on the west coast of Denmark, risks downtime. Another risk of solar panels is to scare off or interfere with the animals' behaviour, because of their shiny surface. Car batteries, were therefore chosen as the power source, as they are reliable, however they eventually run out and need to be replaced and charged. To aviod unnecassary power consumption, the system is set to stand-by during the night and automatically restarted the next morning.

The lifetime of average car batteries are highly reduced when they are drained, which would be the case in the system setup. Deep cycling batteries are therefore preferable, as they are designed to cope with this kind of treatment.

The system is a work in progress, and it was chosen to use DC/AC converters, as a part of the power source, for more flexibility. This ensures easy expansion if other equipment were to be used at a later stage, however it also introduces a loss in efficiency. The chosen converter has an efficiency of 90%.

The overall power consumption of the system is 60 W, and with a 90% efficiency. The minimum uptime must be 36 hours, which requires batteries of approximately 200 Ah (two 95 Ah were chosen), however this is derived from the worst case power consumption scenario and without the planned stand-by hours.

System components	
Component	**Details**
Battery	12 V Deep Cycling
DC/AC converter	Sine wave converter
uEye Camera UI-1245LE-C	Lens: 6 mm 640 x 480
Harddrive	3 TB
3G connection	5 MB
Sennheiser MKE 400 Microphone	Shotgun
Asus Eee Laptop	1.6 Ghz 1 Gb memory

Table I
TABLE OF SYSTEM COMPONENTS USED IN THE SETUP

To preserve quality in the audio recording a directional shotgun microphone, with wind reduction filter was chosen. For the connection, a 10 m long multiple shielded audio extension cable is being used, which enables different placement of the microphone.

The high frame rates are provided by the chosen camera, which enables $20 - 30$ fps depending on the resolution of the image. The camera uses a global shutter, which reduces blurring caused by movements. It is powered via the USB connection, which is also used for data transfer. The recorded images are not compressed, which requires a high capacity harddrive. The SSD technology would be preferable, because of the low power consumption, however due to dollar/GB, this was not chosen for the system.

For remote connection, a 3G connection was chosen. Due to lack of coverage, this solution can potentially lead to loss of connection, however the location chosen for the recording had good 3G coverage. A lack of coverage would not be vital for the recordings, however remote access would be affected. A list of the specific items used in this setup can be seen in table III.

IV. Infrastructure and data description

The main purpose of the system is to record and store large amounts of data. In Figure 2 an overview of the data flow and connections are shown. With a frame rate of 20 fps, an image is captured from the camera and stored on the external harddrive. Meanwhile an audio file is saved on the harddrive every 5 minutes. This is accomplished by a loop-recording software, which increments filenames and records while storing the files. The audio recordings were done with a sample rate of $44.1kHz$ and 16 bit resolution, which is the default settings of the loop-recording software.

The images captured from the camera, are stored as the raw bayer pattern. This reduced the file size (from 900 kb to

 ISBN: 978-1-61208-174-8

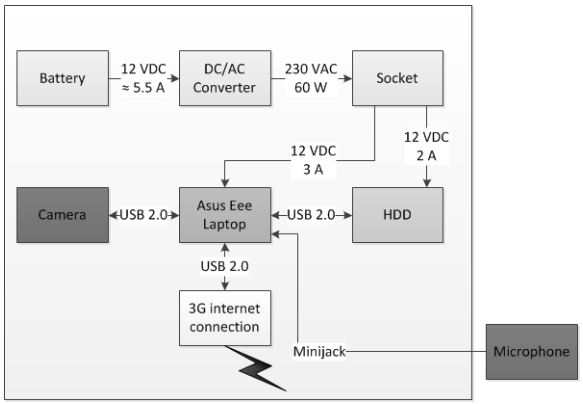


Figure 1. Block diagram of the system

300 kb) and the CPU load as image encoding is not being done. The demosaic and encoding of the captured images is done offline in the analysis fase of the research.

The USB 2.0 protocol used for data transfer offers a theoretical maximum rate at 60 MB/s. The image capture requires 6 MB/s, which lies within the specifications. The audio file is saved every 5 minutes, and does not affect the ongoing audio recording. This means that a transfer rate of approximately $\frac{50}{(60 \cdot 5)} \approx 0.2$ MB/s would be sufficient for storing the audio.

The 3G internet connection is used for remote access and uploading files to an FTP-server. The purpose of the file transfer is to monitor the video recording, and as the images are not encoded it is not possible to view the images on the surveillance system laptop. The newest captured image is being uploaded every hour, and accessed from another laptop in the laboratory.

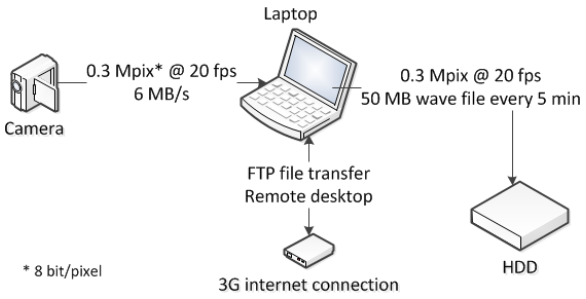


Figure 2. Overview of the infrastructure of the system setup, including description of data

The dataflow and software considerations are summarized here:

- An image is captured every 1/20 second and saved on the external hdd
- Every five minutes an audio file (.wav file) is saved on the external hdd using loop recorder (see www.looprecorder.de)
- Every hour a batch script uploads the newest image to an ftp server
- At sunset, the system is set to stand-by, and at sunrise the system wakes up again (3G connection is automically started to enable remote access)

With a frame rate of 20 fps and chosen audio encoding (.wav files), the system records a data rate of approximately 22 GB/hour.

V. Results

The described system was used for recording wild life geese over a period of one month, where the only downtime was due to replacement of batteries. Over 4.5 hours of geese audio and video, capturing landing, eating and fleeing, were succesfully recorded. During the recordings, the weather conditions were diverse, including storms and sunny weather, and the system and the recordings were not affected.

The power supply used for the setup was car batteries, and with two 95 Ah batteries, the system was able to run for approximately three and a half days. This was accomplished by putting the system to stand-by every night.

During the recording, the average CPU load was $45 - 50\%$, with a peak load of 70%. The memory load was constantly on 40%. If processing of the signals were to be implemented, the system has to be upgraded, however this was not the scope of the system at hand.

In Figure 3 an example of the recorded audio is shown. The geese vocalizations are clearly visible in the shown waveform, as they appear with high amplitude in the recordings. Visual inspection of the recorded video also verify this, as the geese were standing close to the microphone. Recordings of the three desired behaviours were present in the recorded data (both video and audio).

The waveform shown in Figure 3 was recorded on a windless day, however some noise is present in the recording. This is due to the amplification of the audio signal in the built-in sound card in the laptop.

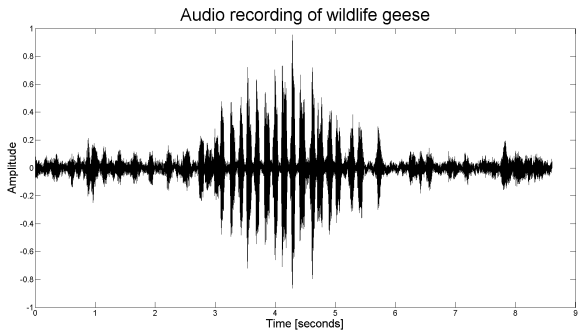


Figure 3. A sample waveform of the recorded audio while geese were foraging. The vocalizations are easily detected, as they appear with high amplitude

VI. Discussion

As mentioned in the results section, the chosen laptop cannot be used if processing of the recorded audio or video were to be done, which will be a part of further development of the system. A laptop with more memory and a faster CPU is desirable. Other limitations with the chosen laptop, are the frame rate and resolution of the images. The camera supports both higher frame rates and higher resolution, and a faster laptop could increase both without over-burdening the CPU.

The overall power consumption was reduced by setting the system to stand-by at nighttime. Another approach could be to trigger the recording, and only record when animals are present. This was not chosen as a loss of usefull data, due to potential trigger errors, could delay the further research. However a suggested modification could be to use a computer vision approach to trigger the recording.

The microphone used in the experiment, was a directional shotgun microphone. Another approach could be parabolic microphones, which are directional and amplify the sound before the digitizing of the audio. This was not used in this recording due to the physical size of the microphone.

Some noise is present in the recordings, even on windless days. This is due to the low quality sound card in the laptop, and an external sound card or a microphone amplifier could provide a better quality recording, however the vocalizations were clear in the recordings. A spectrogram analysis of the recorded vocalizations have shown that all information in the geese vocalizations is within 10 kHz, which means that sample rate could be reduced, which is preferable if more processing were to be performed in the system.

The remote access allowed for adjustments of audio and video recording, however aperture could not be adjusted as this must be done manually. Adjustments were not made during the recordings, as it was hard to verify the image quality on the remote access, which was mainly used to verify the presence of geese.

The recorded data contained examples of the three desired behaviours: land, eat and take off. Automatic behaviour recognition research, based on these recordings, could lead to a system capable of reducing habituation, as scaring mechanism could be targeted towards specific behaviours or species.

Even though the system was used for a specific scenario, it is applicable to other wildlife studies where audio and video material is essential. It is designed to cope with different weather conditions and the remote access makes it possible to verify recordings.

VII. Conclusion

Based on the described system setup, it was possible to record geese in order to analyze the link between their vocalizations and behaviour. The geese quickly grew accustomed to the setup, and only two days after the installation of the system, the geese landed and foraged.

Data provided by the described system is a part of ongoing research to automatically recognize animal behaviour based on audio and video recordings. The results of this research are to be tested using a modfication of the descibed system, where both audio and video processing will be a part of the system.

Acknowledgment

We would like to thank Morten Laursen for contributing to the overall design, and I would also like to thank Esben Rasmussen and Claus Andersen for support in developing the imaging capture software.

References

[1] Seppo Fagerlund. Bird Species Recognition Using Support Vector Machines. *EURASIP Journal on Advances in Signal Processing*, 2007:1–9, 2007.

[2] Jason M Gilsdorf, Scott E Hygnstrom, and Kurt C Vercauteren. Use of frightening devices in wildlife damage management. *Integrated Pest Management Reviews*, 7(1):29–45, 2002.

[3] Roman Gula, Jörn Theuerkauf, Sophie Rouys, and Andrew Legault. An audio / video surveillance system for wildlife. *European Journal of Wildlife Research*, 56(5):803–807, 2010.

[4] Andrea M. Kleist, Richard A. Lancia, and Phillip D. Doerr. Using Video Surveillance to Estimate Wildlife Use of a Highway Underpass. *Journal of Wildlife Management*, 71(8):2792–2800, 2007.

[5] C Lee, C Chou, C Han, and R Huang. Automatic recognition of animal vocalizations using averaged MFCC and linear discriminant analysis. *Pattern Recognition Letters*, 27(2):93–101, January 2006.

[6] Maja Matetić, Slobodan Ribarić, and Ivo Ipšić. Qualitative Modelling and Analysis of Animal Behaviour. *Applied Intelligence*, 21(1):25–44, July 2004.

[7] Vlad M Trifa, Alexander N G Kirschel, Charles E Taylor, and Edgar E Vallejo. Automated species recognition of antbirds in a Mexican rainforest using hidden Markov models. *The Journal of the Acoustical Society of America*, 123(4):2424–31, April 2008.

[8] D. Tweed and A. Calway. Tracking multiple animals in wildlife footage. In *Proceedings of the 16th International Conference on Pattern Recognition*, pages 24–27. IEEE Comput. Soc, 2002.

[9] S Ward, C M Bishop, a J Woakes, and P J Butler. Heart rate and the rate of oxygen consumption of flying and walking barnacle geese (Branta leucopsis) and bar-headed geese (Anser indicus). *The Journal of experimental biology*, 205(Pt 21):3347–56, November 2002.

 ISBN: 978-1-61208-174-8