ICDT 2018

The Thirteenth International Conference on Digital Telecommunications

ISBN: 978-1-61208-623-1

April 22 - 26, 2018

Athens, Greece

ICDT 2018 Editors

Stan McClellan, Texas State University - San Marcos, USA
George Koutitas, Texas State University - San Marcos, USA

ICDT 2018

Forward

The Thirteenth International Conference on Digital Telecommunications (ICDT 2018), held between April 22, 2018 and April 26, 2018 in Athens, Greece, continued a series of special events focusing on telecommunications aspects in multimedia environments. The scope of the conference was to focus on the lower layers of systems interaction and identify the technical challenges and the most recent achievements.

High quality software is not an accident; it is constructed via a systematic plan that demands familiarity with analytical techniques, architectural design methodologies, implementation polices, and testing techniques. Software architecture plays an important role in the development of today's complex software systems. Furthermore, our ability to model and reason about the architectural properties of a system built from existing components is of great concern to modern system developers.

Performance, scalability and suitability to specific domains raise the challenging efforts for gathering special requirements, capture temporal constraints, and implement service-oriented requirements. The complexity of the systems requires an early stage adoption of advanced paradigms for adaptive and self-adaptive features.

Online monitoring applications, in which continuous queries operate in near real-time over rapid and unbounded "streams" of data such as telephone call records, sensor readings, web usage logs, network packet traces, are fundamentally different from traditional data management. The difference is induced by the fact that in applications such as network monitoring, telecommunications data management, manufacturing, sensor networks, and others, data takes the form of continuous data streams rather than finite stored data sets. As a result, clients require long-running continuous queries as opposed to one-time queries. These requirements lead to reconsider data management and processing of complex and numerous continuous queries over data streams, as current database systems and data processing methods are not suitable. Event stream processing is a new paradigm of computing that supports the processing of multiple streams of event data with the goal of identifying the meaningful events within those streams.

The conference had the following tracks:

- Next generation wireless systems and services
- 5G Testbeds, Applications, Standards and New Business Models

We take here the opportunity to warmly thank all the members of the ICDT 2018 technical program committee, as well as all the reviewers. The creation of such a high quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated their time and effort to contribute to ICDT 2018. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

We also gratefully thank the members of the ICDT 2018 organizing committee for their help in handling the logistics and for their work that made this professional meeting a success.

We hope that ICDT 2018 was a successful international forum for the exchange of ideas and results between academia and industry and to promote further progress in the field of digital communications. We also hope that Athens, Greece, provided a pleasant environment during the conference and everyone saved some time to enjoy the historic charm of the city.

ICDT 2018 Chairs

ICDT Steering Committee
Constantin Paleologu, University Politehnica of Bucharest, Romania
Jaime Lloret Mauri, Polytechnic University of Valencia, Spain
Ioannis Moscholios, University of Peloponnese - Tripolis, Greece
Sathiamoorthy Manoharan, University of Auckland, New Zealand
Bernd E. Wolfinger, University of Hamburg, Germany
Stan McClellan, Texas State University - San Marcos, USA

ICDT Industry/Research Advisory Committee
Tomohiko Taniguchi, Fujitsu Laboratories Limited, Japan
Scott Trent, IBM Research – Tokyo, Japan

ICDT 2018
Committee

ICDT Steering Committee

Constantin Paleologu, University Politehnica of Bucharest, Romania
Jaime Lloret Mauri, Polytechnic University of Valencia, Spain
Ioannis Moscholios, University of Peloponnese - Tripolis, Greece
Sathiamoorthy Manoharan, University of Auckland, New Zealand
Bernd E. Wolfinger, University of Hamburg, Germany
Stan McClellan, Texas State University - San Marcos, USA

ICDT Industry/Research Advisory Committee

Tomohiko Taniguchi, Fujitsu Laboratories Limited, Japan
Scott Trent, IBM Research – Tokyo, Japan

ICDT 2018 Technical Program Committee

Arsalan Ahmad, National University of Sciences and Technology (NUST), Islamabad, Pakistan
Akbar Sheikh Akbari, Leeds Beckett University, UK
Atallah Mahmoud AL-Shatnawi, Al al-Byte University, Jordan
Maria Teresa Andrade, FEUP / INESC Porto, Portugal
Mario Arrigoni Neri, Università di Bergamo, Italy
Ilija Basicevic, University of Novi Sad, Serbia
Andrzej Borys, Gdynia Maritime University, Poland
Fernando Cerdan, Universidad Politecnica de Cartagena, Spain
Paskorn Champrasert, Chiang Mai University, Thailand
Sruti Das Choudhury, University of Nebraska-Lincoln, USA
Eleonora D'Andrea, University of Pisa, Italy
Wanyang Dai, Nanjing University, China
Abhishek Das, All India Council for Technical Education, Kolkata, India
Tan Do-Duy, Universitat Autònoma de Barcelona, Spain
Alexander Dudin, Belarusian State University, Belarus
Jingyuan Fan, University of New York at Buffalo, USA
Qiang Fan, New Jersey Institute of Technology, USA
Mahmood Fathy, Iran University of Science and Technology, Iran
Mário F. S. Ferreira, University of Aveiro, Portugal
Rita Francese, Università di Salerno, Italy
Francois Gagnon, Cégep Sainte-Foy, Canada
Katja Gilly, Universidad Miguel Hernández, Spain
Félix J. García Clemente, Universidad de Murcia, Spain
Andre Leon S. Gradvohl, University of Campinas, Brazil
Carlos Guerrero, University of Balearic Islands, Spain
Onur Günlü, TU Munich, Germany
Ahmad Yusairi Bani Hashim, Universiti Teknikal Malaysia Melaka, Malaysia

Daniela Hossu, University Politehnica of Bucharest, Romania
Yasin Kabalci, Nigde University, Turkey
Dattatraya Vishnu Kodavade, D.K.T.E. Society's Textile & Engineering Institute Ichalkaranji – Rajwada, India
Wen-Hsing Lai, National Kaohsiung First University of Science and Technology, Taiwan
Jan Lansky, University of Finance and Administration, Czech Republic
Isaac Lera, Universitat de les Illes Balears, Spain
Xiuhua Li, University of British Columbia, Vancouver, Canada
Jaime Lloret Mauri, Polytechnic University of Valencia, Spain
Malamati Louta, University of Western Macedonia, Greece
Stephane Maag, Telecom SudParis, France
Sathiamoorthy Manoharan, University of Auckland, New Zealand
Stan McClellan, Texas State University, USA
Manar Mohaisen, Korea University of Technology and Education, Republic of Korea
Ioannis Moscholios, University of Peloponnese - Tripolis, Greece
Masayuki Murata, Osaka University, Japan
Dmitry Namiot, Lomonosov Moscow State University, Russia
Patrik Österberg, Mid Sweden University, Sweden
Constantin Paleologu, University Politehnica of Bucharest, Romania
Euthimios (Thimios) Panagos, Vencore Labs, USA
Liyun Pang, Huawei German Research Center, Germany
Nada Philip, Kingston University, UK
Juha Röning, University of Oulu, Finland
Prasan Kumar Sahoo, Chang Gung University | Chang Gung Memorial Hospital, Taiwan
Abdel-Badeeh M. Salem, Ain Shams University, Cairo, Egypt
Konstantin Samouylov, RUDN University, Moscow, Russia
Panagiotis Sarigiannidis, University of Western Macedonia, Greece
Sayed Chhattan Shah, Hankuk University of Foreign Studies, South Korea
Tarun Kumar Sharma, Amity University Rajasthan, India
Shazmin Aniza Abdul Shukor, Universiti Malaysia Perlis, Malaysia
Sabrina Sicari, University of Insubria, Italy
Edvin Škaljo, BH Telecom, Bosnia and Herzegovina
María Estrella Sousa Vieira, University of Vigo, Spain
Cristian Stanciu, University Politehnica of Bucharest, Romania
Dimitrios Stratogiannis, National Technical University of Athens, Greece
Tatsuya Suda, University Netgroup Inc, Fallbrook, CA, USA
David Suendermann-Oeft, Dialog, Multimodal, and Speech research center (DIAMONDS) | Educational Testing Service, USA
Mahbubur R. Syed, Minnesota State University, USA
Salvatore Talarico, Huawei Technologies, Santa Clara, USA
Tomohiko Taniguchi, Fujitsu Labs, Japan
Yoshiaki Taniguchi, Kindai University, Japan
Tony Thomas, Indian Institute of Information Technology and Management-Kerala (IIITM-K), Thiruvananthapuram, India

Copyright Information

Table of Contents

Data-centric Operations in Oil & Gas Industry by the Use of 5G Mobile Networks and Industrial Internet of Things (IIoT)

Spandonidis C. Christos
Prisma Electronics
Athens, Greece
email: c.spandonidis@prismael.com

Giordamlis Christos
Prisma Electronics
Athens, Greece
email: christos@prismael.com

Abstract - **For many years, the Oil & Gas Industry has been collecting huge amounts of data, turning thus slowly and gradually to “Data-centric Operations”. Unfortunately, this data collection has typically happened via many independent pieces of equipment and systems – each with its own data and interfaces. The promising features of the forthcoming 5th generation (5G) mobile networks have enhanced the Industrial Internet of Things providing technology improving safety and optimizing performance. In this work a dedicated platform aiming at the continuous collection of critical information from multiple nearshore assets and transmission of the data through a 5G network is proposed. Important details and technology challenges that drive to a competitive proposal are discussed.**

Keywords - Offshore monitoring; Industrial Internet of Things; 5th Generation (5G).

I. Introduction

The ability to utilize data to obtain knowledge, predictions and insights gives today the tools for continuous process improvements and optimal performance throughout the lifetime of assets. According to World Economic Forum [1], the Oil & Gas industry is seeking to leverage new developments in digital technologies to unlock a value at $1.6 trillion by 2025. Major Energy companies like Aramco, BP and Statoil are leading the way in the “digitalization” of the oil field and recently presented their plans to invest in new digital technologies through 2020, in order to improve safety, security, and efficiency of their operations (see for example [2]). Based on the existing data flows from vessels, we are exploiting a unique business window opportunity for a platform architecture that allows the reliable and uninterrupted reception and management of data from a wide range of sensors within a near-shore offshore structure network.

The rest of the paper is structured as follows. In Section II, experience gained by shipping industry is provided in a critical way. Characteristic examples of data that lead to effective decision making and cost reduction are discussed. In Section III, the main challenges when it comes to Oil & Gas Industry are briefly described, while in Section IV a platform based on the development of the existing LAROS-based platform [3] that allows reliable and uninterrupted reception and management of data from a wide range of sensors within a wide range of offshore structures is proposed.

II. Experience from Shipping

LAROS is a dedicated platform aiming at the continuous collection of critical information from the ship's inputs, the transmission of the data through a wireless network, centralization and homogenization of information in central computing, and analysis of measurements to support the decision-making mechanism of shipping companies. The system is using the vessel’s communication systems (satellite) to transmit all collected & synchronized data to the Headquarters, in a very efficient manner in terms of cost, speed, and security. Transferred data are further processed using LAROS Data Analysis System.

A. System description

In more detail, the Data Collection process can be described as follows:

LAROS Smart Collectors are connected using the appropriate interface to analog or digital signals coming from different sensors and instruments of the vessel. Smart Collectors analyze the signals and calculate the required parameters. The sampling rate, as well as rate of the parameters calculations can be set from 100msec up to 30 minutes. Smart Collectors setup a wireless secure network inside the vessel to transmit the processed data to the Gateway with a user-defined sampling rate and ability to maintain and customize them remotely. The wireless protocol is based on IEEE 802.15.4 MESH [4] with additional layers and data format to cover the requirements of the vessel environment and increase the network Quality of Service. Through the Gateway, all the measured and processed parameters are stored in LAROS Server (onboard). All data are stored in LAROS server’s database for a long period (up to 1 year depending on the number of sensors and on sampling rate). In addition, there are options to forward the data to any third party systems on board avoiding costly cabling or other infrastructure implications. LAROS On board Server periodically produces binary files and compresses them in order to reduce the size of the data to be sent via normal satellite broadband. The compressed files are transmitted through File Transfer Protocol (FTP) to the data center that will be selected by the operator. In the data center, there is a service that decompresses the incoming files and stores the new measurements in the main data base. In case the system is connected to a weather site, the weather data are stored in the main data base in the same format.

B. Maintaining the Integrity of the Specifications

Table 1 summarizes the main functionality modules, the needed signals and the collection points onboard:

TABLE I. INDICATIVE FUNCTIONAL MODULES - SHIPPING

Module	***Needed signals***	***Connection points***
Proppeler – Hull Performance	Vessel Speed, Shaft Revolutions per Minute (RPM), Shaft Power.	Speedlog, Torque-meter- RPM Indicator.
Engine Performance	Fuel Oil Consumption (FOC), Power (Specific Fuel Oil Consumption - SFOC), Diesel Generator (DG) Output	Flowmeters [Fuel Oil (FO) flow], FO temp, FO density, DG Power Analyzer
FO Consumption	FOC, Vessel Speed through water, Shaft RPM, Boiler Status	Flowmeters (FO flow), FO temp, FO density, Boiler status indicator
On-line bunkering	Tank level, FO temperature	Cargo Control Console, Engine Control Room (ECR)/ Cargo Control Room (CCR) Indicators.
Maintenance managementt	Pressures, Temperatures, Alarms from critical systems	Alarm Monitor System (AMS), ECR Indicators
Power management	DG Output, Reefers Power Consumption	DG/Reefer Power Analyzers
Environmental conditions	Wind speed & direction, Water depth, Ambient temperature & Pressure	Anemometer, Echo-Sounder, weather station
Operational profile	Ground Speed, Drafts, Trim, Rudder angle	GPS, strain gage, Inclinometer

Next, we present two characteristic examples of performance increase. In the first example, real time data were used to identify a problem in condition of crucial system, while the second is an example of using historical data to find a root cause on increased operational expenditure.

Example 1: The operator was unaware about any issues with the Diesel Generator. As shown in Figure 1, the visualization of data pointed increasing deviation in Lube Oil (LO) inlet pressure which normally is an indication of the main axis of Diesel Generator cracking. The crew inspected the axis and fixed the issue in the next stop. As a result, the operator saved thousands of Euros, and his liability in the case this issue caused a serious accident.

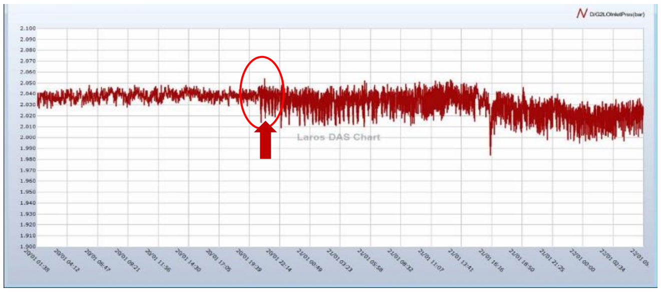


Figure 1. Observed increasing deviation in LO inlet pressure.

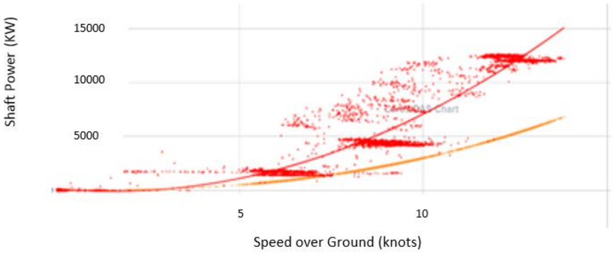


Figure 2. Shaft power vs vessel speed. With red dots are indicated the actual measurements. Sea trial baseline is indicated with yellow curve.

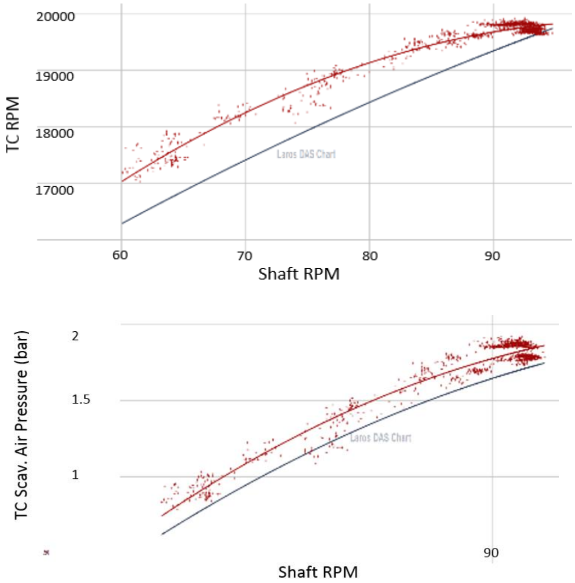


Figure 3. Turbo Charger (TC) Rounds per Minute (RPM) vs Shaft RPM (up) and TC Scav. Air Pressure vs Shaft RPM (down).

Example 2: Often dry-docking is a scheduled event and it is a costly one. However, a persistent problem may be much more costly if it is not addressed as soon as possible. Operator had the insights of very accurate data, proving that loss was greater than gain to keep operating. By using historical data, it was able to find out the propulsion performance and power analysis on the main engine's performance, influence of wind, waves, swell, current, shallow water, trim, use of rudder, possible drift, resulting in 73% deviation from Sea Trials/Model test report (Figure 2). Further analysis on engine performance (Figure 3) identified a well operating engine with no deviation, which was indicative of a clear hull fouling problem. This results in estimated 65% excess FOC.

III. OIL & GAS CHALLENGES

For many years, the Oil & Gas Industry has been collecting huge amounts of data (e.g., one rig can generate 1 terabyte of data per day), turning thus slowly and gradually to "Data-centric Operations". Wang et al. [5] provide a comprehensive review of the recent developments in field monitoring for offshore structures. In their work, the authors present a detailed list with typical monitoring projects of offshore platforms for the last 3 decades. Table II summarizes the main monitoring scopes and related sensing technologies. Unfortunately, this data collection has typically happened via many independent pieces of equipment and systems – each with its own data and interfaces. Integrated Marine Monitoring Systems (IMMS) have been developed to overcome such problems. A review of IMMS systems that enable the synchronization of collected data is presented by Yuan [6], while Wu [7] described the aspects of software and hardware to ensure the long term operation of such systems. Von Aschwege [8] presented the principles of design as well operational and technical considerations an Independent Remote Monitoring System (en entirely independent back up system used to transit critical data during hurricane conditions) should follow.

Despite the significant number of offshore platform monitoring related projects conducted all over the world, they are almost entirely limited to single offshore platforms. Data exchange between long-distance remote located offshore structures (e.g. between offshore platform and hurricane prediction stations, pipeline /reservoir monitoring systems, ocean ecological environment monitoring) or even between platform and support vessels or on-shore low-coverage areas, is a difficult to solve the problem.

Through our experience implementing LAROS on ships, we have highlighted several issues in the transmission of information, when it comes to multi structure network architectures, mainly due to (a) adverse conditions within specific offshore environments, (b) long distances between the nodes and (c) dependence of the system on the main power supply network (d) latency/bandwidth limitation on communication network, (e) cybersecurity issues. The promising features of the forthcoming 5th generation (5G) mobile networks assisting the aim of integrating multiple offshore structures into a wider ecosystem for the exchange of *large dimensional structured information* leading to an efficient and comprehensive operation of the offshore assets through Information Intelligence.

IV. INTEGRATED 5G BASED PROPOSAL

The proposed platform is based on the development of the existing LAROS-based platform to allow the reliable and uninterrupted reception and management of data from a wide range of sensors within a wide range of offshore structures (e.g. platform, Support Vessel, floating buoy) with emphasis on the requirements described by Wang et al. [1]:

(a) Need for smart monitoring instruments with less power requirement, higher accuracy, ability for synchronization, and applicable in a variety of sensor types (Table II) and ideally on fiber-optic sensors.

(b) Need for efficient wireless communication network that allows multi-measurement acquisition and real time data exchange between different offshore structures and between offshore structures and third parties (e.g., port authorities).

TABLE II. INDICATIVE MONITORING SCOPES – OIL & GAS

Scope	***Needed signals***	***Sensing technologies***
Metocean	Wind	Anemometer
	Sea Waves	Remote wave buoys X-band Radar
	Current	Acoustic Doppler Current Profiler (ADCP)
	Internal waves	SAR
	Ice	Moored Upward Looking Sonar (ULS)
	Environmental conditions	Humidity, pressure, Ambient Temperature
	Tide	Pressure & Water Density sensor
Structural Motions	Positioning	Differential Global Positioning Systems (DGPS), Inertial Navigation System (INS)
Structural operational status	Platform Hull	Altering Current Field Method (ACFM), Field Signature Method (FSM), underwater robot probe, Remoted Operated Vehicle (ROV)
	Riser	Tension riser Monitoring system
	Mooring line	Load cells, inclinometers
	Submarine pipelines	Visual inspection Tension sensors, Echo sounder, contour sonar

A. *Smart monitoring instruments*

The viability of the LAROS-based smart sensing platform in wireless-based systems for predictive maintenance and management in a harsh industrial environment was demonstrated by Sachat et al. [3]. The outcome of that work was a sensing platform that was viable, low cost and of low complexity, able to be efficiently integrated in autonomous fiber-optic sensing units and capable of forming a distributed monitoring network. The selection of large core optical fibers additionally allows the use of low power light sources and photodetectors that could be integrated in the sensing unit with low power requirements (Figure 4).

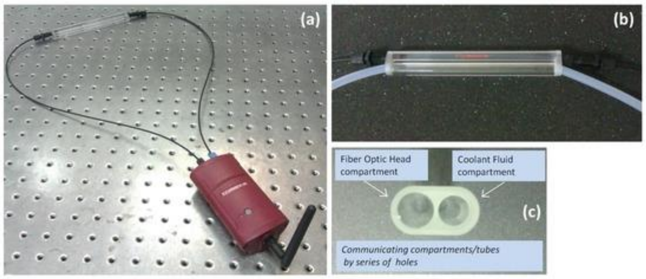

(a)
(b)
Fiber Optic Head compartment
Coolant Fluid compartment
(c)
Communicating compartments/tubes by series of holes

Figure 4. (a) Measuring apparatus with the sensing head connected to the wireless sensing node unit; (b) Photograph of the glass measuring cell (c) Photograph showing in detail the dual tube glass cell. [3]

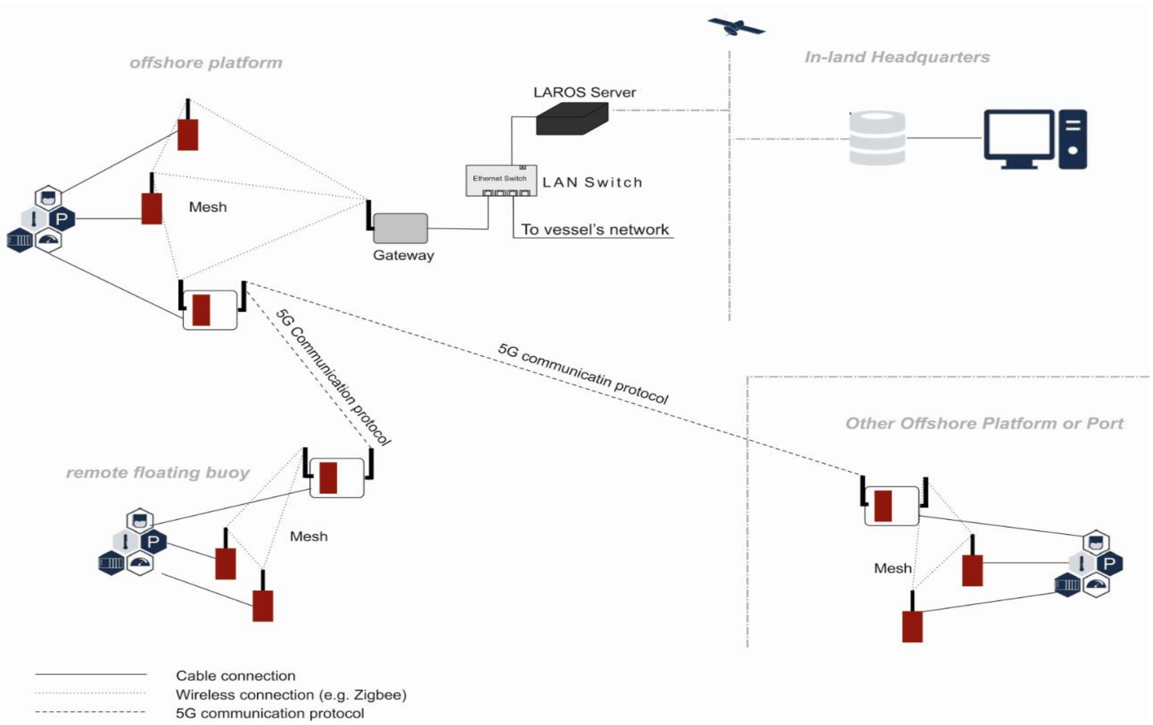


Figure 5. Decsriptive example of proposed system architecture

B. *Smart communication network*

Mesh networking combined with low power consumption (e.g., Zigbee [9]) is proposed be used for exchange of data between nodes at short distances. For communication between remote nodes or an external network, 5G protocol will be used that allows communication at long distances in an efficient and energy-efficient way.

Use of 5G mobile network has been proposed instead of other communication technologies, such as Wifi, 4G, etc, mainly due to rate (b) lower End-to-end latency, (c) large number of connection points, (d) reduced Capital and Operational Expenditures, (e) consistent Quality of Experience and (f) reduced demand for energy [10].

Following Mugen et al. [11], in order to achieve these goals, we propose a heterogeneous cloud radio access network (H-CRAN), where cloud computing is used to fulfil the centralized large-scale cooperative processing for suppressing co-channel interferences. As shown in Figure 5, central stations (on offshore platform) (Node C in [11]) act as the Base Band Unit (BBU) pool to manage all accessed Remote Radio Heads (RRHs), and the software-defined H-CRAN system architecture is presented to be compatible with Software Defined Networks (SDN). This architecture will eliminate issues of path loss and the need for line of sight due to the high operating frequency of such networks.

The development of 5G-based Low Earth Orbit (LEO) satellites will further enhance the integration of satellite and terrestrial networks in 5G [12], enabling thus even larger or distant located eco-systems (e.g., deep water oil platforms). Figure 6 schematically presents such an integrated architecture.

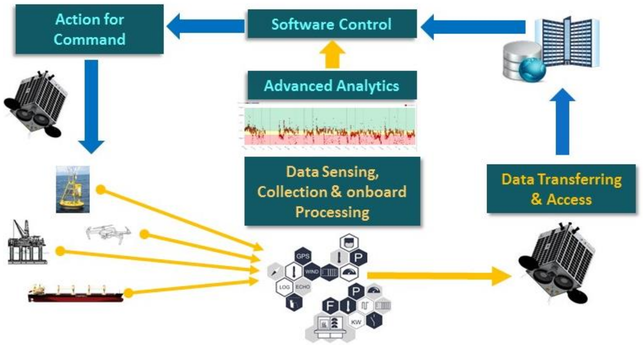

Action for Command
Software Control
Advanced Analytics
Data Sensing, Collection & onboard Processing
Data Transferring & Access

Figure 6. Schematic view of 5G Low Earth Orbit LAROS architecture

C. *Benefits of proposed monitoring platform*

An advanced monitoring system, as the one described above, enables the following core functionalities:

<u>Efficiency control</u>

The platform provides the necessary tools that allow managers and operators to measure in detail the efficiency of every asset on board. Further OPEX reduction is possible as a result of performance analysis of the facilities and corrective action plans, e.g., effective power system management, rational usage of equipment, isolation and replacement of heavy energy consumers, etc.

Condition Monitoring and Event Detection

Continuous monitoring of an asset or of a condition can only be performed by an online system onboard. The more parameters examined and analyzed simultaneously the better the monitoring performance. The platform provides alarms triggered in real time when anomalies get detected.

Centralized monitoring

Monitoring of multiple assets can be achieved using a unified dedicated dashboard that allows centralized monitoring and reporting. The operational and performance parameters of each asset can be individually tracked and analyzed using a single reporting system accessed from anywhere in the world using simple Web services.

Expandable

The platform is expandable and adaptable in order to cover any future needs and required measurements. This is easily done by connecting additional sensors to the installed LAROS Collectors or by adding extra Collectors in the existing 5G network.

V. CONCLUSION AND FUTURE WORK

To gain reliable and comparable data at low cost and in energy efficient method from a wide network of offshore assets, mesh networking combined with 5th generation network architecture is proposed. Examples from experience on vessels, for the exchange of data between nodes at short distances is presented and existing challenges for 5G mobile network are drafted for communication between remote nodes and the external network. Technology barriers and challenges for a robust operational model that drives to a competitive and integrated 5G based proposal that will make an impact in the field of Offshore “Data-Centric Operations” were further discussed. It was shown that State-of-the-art technology trends in various sectors including M2M, intelligent processing, machine learning, data agents, cyber-safe datasets, telco etc., should be part of new generation platforms to secure data science in a level that allows the Oil & Gas Industry to enter the IIOT – 5G area dynamically and support effectively decisions, safety, efficiency and interoperability. The goal for the next step would be to define the important details in order the designed platform architecture to offer a reliable and usable monitoring system even in harsh and non-accessible until now environments.

REFERENCES

[1] D. Paganie, “Assesing the digital transformation,” Offshore Magazine, vol 17:10, pp 3-5, 2018.

[2] Fugro. Fugro Complete Wellhead Monitoring Project for BP in Gulf of Mexico, 2015. (source: https://www.fugro.com/media-centre/news/fulldetails/2015/10/01/fugro-complete-wellhead-monitoring-project-for-bp-in-gulf-of-mexico).

[3] A. El Sachat et al., "Characterization of industrial coolant fluids and continuous ageing monitoring by wireless node-enabled fiber optic sensors", Sensors MDPI 17, 568, pp. 568-588,2017. doi:10.3390/s17030568.

[4] J. T. Adams, “An Introduction to IEEE 802.15.4”, IEEE Press, vol. 2, pp. 1-8, 2006.

[5] W. Peng, T. Xinliang, P. Tao, and L. Yong, A review of the state-of-the-art developments in the field monitoring of offshore structures, Ocean Engineering, Volume 147, pp 148-164, 2018.

[6] S. Yuan,“Prototype Measurement and Monitoring Technology Research of Floating Offshore Platform (Master's thesis). Da Lian Engineering University, 2013.

[7] W. Wu et al., “Design, implementation and analysis of full coupled monitoring system of FPSO with soft yoke mooring system”. Ocean. Eng. 113, pp. 255–263, 2016.

[8] J. Von Aschwege, K. Jassal, and R. Barker, “An independent remote monitoring system for Gulf of Mexico deepwater floating production systems”, Offshore Technology Conference,pp. 2007.

[9] K. Pantelaki et al., “Survey of the IEEE 802.15.4 Standard’s Developments for Wireless Sensor Networking”, American Journal of Mobile Systems, Applications and Services vol. 2, No. 1, pp. 13-31, 2016.

[10] J. G. Andrews et al., “What will 5G be?”, IEEE Journal on Selected Areas in Communications, vol. 32, no. 6, pp. 1065–1082, 2014.

[11] M. Peng,Y. Li, Z. Zhao, and C. Wang, “System architecture and key technologies for 5G heterogeneousc loud radio access networks,” IEEE Netw., vol. 29, no. 2, pp. 6–14, Mar. 2015.

[12] M. D. Sanctis, E. Cianca, G. Araniti, I. Bisio, and R. Prasad, “Satellite communications supporting internet of remote things,” IEEE Internet of Things Journal, vol. 3, no. 1, pp. 113–123, 2016.

An Augmented Reality Facet Mapping Technique for Ray Tracing Applications

Varun Kumar Siddaraju
Ingram School of Engineering
Texas State University
San Marcos, USA
varunsiddaraju@gmail.com

George Koutitas
Ingram School of Engineering
Texas State University
San Marcos, USA
george.koutitas@txstate.edu

Abstract—**This paper presents a novel spatial mapping technique that is capable of extracting the vector map of an indoor environment based on images captured from a smart phone camera. The extracted vector map follows the facet model concept and can be used as input in ray tracing algorithms for indoor wireless channel predictions. The algorithm computes the coordinates of the walls and doors of each room of the indoor environment and creates the facet model of the entire 3D space by applying edge and corner detection on the wall images of each room. The output of the algorithm is a facet model that can be used by ray tracing algorithms which are embedded in Augmented Reality (AR) applications. The overall process provides a better human-to-network interface and an improved user experience that is expected to provide a new way for indoor network planning of residential 5G systems.**

Keywords-augmented reality; spatial mapping; facet model; ray tracing; indoor networks; channel prediction.

I. INTRODUCTION

The computation of indoor vector maps and spatial mapping are active research fields for various Augmented Reality (AR) applications. Characteristic examples are gaming, interior design, property advertising, indoor security, indoor navigation, that all require information of the indoor space in order to overlay holograms. Spatial mapping requires high-end cameras like RGB depth (RGB-D) and Simultaneous Localization and Monocular (SLAM) cameras and this increases the overall cost of the system [1]. Spatial mapping is the process of analyzing the 3D space and transforming it to a set of vertices coupled to other information such as vertices normal and vertices type. In most applications, this transformation is very useful since a user can place holograms and avatars in the real space and interact with them. In some occasions, other applications may require a simplified spatial mapping where the overall objective is just to create the vector map of the walls and doors of the indoor environment without the need for indoor clutter information. This paper proposes a novel technique that can provide 3D mapping of indoor spaces utilizing the facet concept and only requires the use of a commodity smart phone cameras.

Different types of AR algorithms and limitations for real-time imaging are discussed in [2]. The presented applications mainly concern the use case of military, medical, gaming, interior designing and advertising. A survey of AR technologies and applications is also presented in [3][4]. With the increase of various AR applications, the need for more sophisticated spatial mapping and 3D indoor mapping algorithms also increases. Spatial mapping is usually performed by RGB-D cameras [5][6] and simultaneous localization and monocular (SLAM) cameras [7][8]. The RGB-D camera captures 3D RGB images with their depth details. SLAM cameras, simultaneously map the indoor environment with localization of indoor environment features and clutter. Both RGB-D cameras and SLAM cameras are integrated within expensive AR devices.

The next generation of communication networks, namely, the 5G networks, are expected to create new opportunities for mobile AR applications [9]. One characteristic application is network visualization and human-to-network interaction. For example, with the use of an AR application a user can visualize the results of ray tracing simulations that are overlaid on top of the physical space. This is very important for 5G networks where short range communications are expected to create important indoor network planning challenges [9]. To perform field strength prediction, ray tracing algorithms use the facet model where the indoor environment is represented in a vector format with facets incorporating data of the coordinates and the material structure of each facet [10]. An example of the use of indoor vector maps for ray tracing algorithms is given in [11].

The proposed algorithm uses a simple camera of a smart phone device that captures images of individual walls and is capable of constructing a simplified 3D map of the indoor space. The 3D map is a facet model that can be used by indoor channel estimation algorithms. The AR application then overlays the ray tracing results to enable a better human to network interaction. The facet model is created by identifying the coordinates and sizes of the walls and doors of the indoor environment. This process incorporates image processing techniques responsible for the edge and corner detection. The proposed solution uses the *Canny* edge detector to extract wall and door boundaries [12]. Corners on the found edges are detected using the concept of *detect minimum eigenvectors* algorithm. An interesting analysis and comparison of corner detection techniques is given in [13]-[15]. Based on the detected corners of the walls and doors of individual rooms, the entire indoor environment is synthesized to create a full 3D vector representation. A Graphical User Interface (GUI) is also developed to enable an easier interaction between the user and

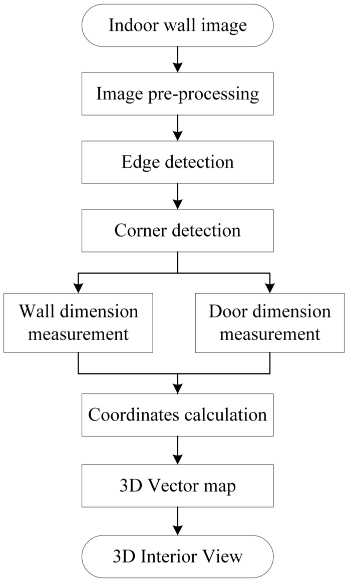


Figure 1. Flowchart of 3D indoor facet mapping.

the application. The overall objective of the proposed solution is to create the necessary foundations for the efficient network planning and positioning of femtocell stations with the use of a typical smart phone device and AR applications.

The rest of the paper is structured as follows. In Section II, we present the system description with overview of the new algorithm. In Section III, we present the facet model with 3D model construction and its data structure. In Section IV, we present the results with detailed Graphical User Interface (GUI), computed facet model and ray tracing visualization output. Finally, we conclude the paper in Section V with the algorithm applications and future work.

II. System Description

A. Overview

The algorithm processes images of the indoor environment, identifies the walls and doors positions, computes the coordinates and creates the facet model for each wall and door. For the purpose of this investigation, window detection was omitted. This process is performed for each room of the indoor space and the found facets are combined in a data structure to represent the entire indoor environment. This process can be considered as a simplified spatial mapping technique that neglects the detailed furniture clutter since it is not significantly affecting signal propagation. The input of the algorithm are the images of every wall but also the height of the ceiling. The images can be captured using a standard camera of a typical smart phone device, without the need of using an expensive depth camera. The input images are then pre-processed to enable an efficient edge and corner detection process which is important for the identification of the vertices and coordinates of the walls and doors. The 3D Cartesian coordinates of a room are calculated using the length, width and height of the room which is computed once the wall and door vertices are detected. Using these coordinates, the 3D vector map or else the facet model can be constructed and become available to third party applications such as ray tracing and AR.

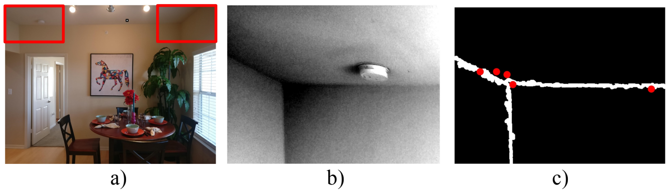


Figure 2. a) The two regions of interest on the original wall image, b) Pre-processed image of top left region, c) detected edges and candidate corners.

The detailed overview of the proposed solution is presented in Figure 1 and is analyzed in the following sections. For efficient performance of the algorithm, the following assumptions should stand:

- Capture photo of the wall from the center of the room by standing parallel to the wall
- The captured image must be clear without any clutter near the top corners of the walls
- If the wall is large, the user can use the panorama function of the smartphone device to capture the entire wall in a single image file

In practice, the aforementioned conditions are usually met in most typical residential units. It should be noted that the proposed technique cannot be used for large corporate offices, since a wall is usually large enough and cannot fit in one photo screen.

B. Image pre-processing

The image pre-processing is the first step of the overall technique and prepares the images of the room for the edge and corner detection phase. For the efficient edge and corner detection, the input image is converted into a grayscale image [5]. The second step of the pre-processing phase is to crop selected regions of interest from the gray scale image.

For the purpose of our investigation these are the top and left corners of the wall as shown in Figure 2. The regions of interest are used to minimize unwanted edge and corner detection and reduce the computational demands of the algorithm. The last part of the pre-processing phase corresponds to a down sampling of the image pixel size procedure on the cropped images that further reduces the computational demands of the process. Usually, the image can be convolved with a Gaussian filter to reduce the number of unwanted edges [9]. The smoothing process [9] is given in the following formula:

$$S[i,j] = G[i,j;\sigma] * I[i,j] \tag{1}$$

where $I[i, j]$ denotes the input image of pixel size ixj, $G[I,j;\sigma]$ denotes Gaussian smoothing filter and $S[i, j]$ denotes the array of smoothed data and σ is the gradient level of the filter. Image is down-sampled to different resolutions like 1280x768, 960x720, 640x480 and experimented for best corner detection results. Images with low resolution 640x480 help to reduce the number of false corner detection compared to higher resolution images. A 5x5 size Gaussian filter is used for efficient edge and corner detection [13]. The overall process of the image pre-processing is demonstrated in Figure 2 a) and Figure 2 b). The next phase of the proposed solution is to process those images

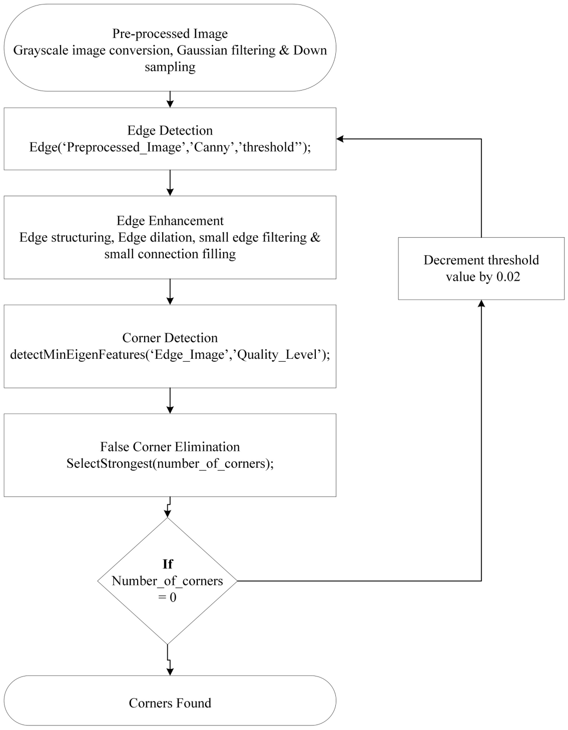

Pre-processed Image
Grayscale image conversion, Gaussian filtering & Down sampling
Edge Detection
Edge('Preprocessed_Image','Canny','threshold");
Edge Enhancement
Edge structuring, Edge dilation, small edge filtering & small connection filling
Corner Detection
detectMinEigenFeatures('Edge_Image','Quality_Level');
False Corner Elimination
SelectStrongest(number_of_corners);
If
Number_of_corners
= 0
Decrement threshold value by 0.02
Corners Found

Figure 3. Flowchart of edge detection and corner detection.

for the edge and corner detection, as shown in Fig 2. c).

C. Edge and corner detection

The edge and corner detection of the wall image is the most crucial part of the algorithm. This is because, corner detection is directly related to the coordinates of the wall of the room, and thus the development of the facet model. The edge and corner detection flowchart is given in Figure 3. The first part of the algorithm is to perform edge detection upon the preprocessed input wall images by implementing the edge *Canny* method [9]. The *Canny* method calculates the gradient using the derivative of a *Gaussian* filter and uses two thresholds to identify strong and weak edges. With this approach, the edge detection of unwanted noisy parts of the image is minimized. The *Canny* method uses a threshold to distinguish between strong and weak edges. For the purpose of our investigation, the edge is detected according to the following function.

$$EM = edge\ (S, Canny, \delta, \sigma) \qquad (2)$$

where *S*, denotes the pre-processed image, 'Canny' denotes the edge detection algorithm, δ is the threshold used and is a two element vector, σ is the standard deviation of the *Gaussian* filter and is a scalar and *EM* denotes the edge map of the wall image. The *EM* image is a binary matrix with *1s* representing the points where an edge is detected. The threshold value is a sensitivity value, and is used to ignore all edges that are not stronger than the selected threshold. It is a scalar value that specifies the standard deviation of the Gaussian filter. The initial threshold was set to $\delta=0.4$ and if no corners found, it

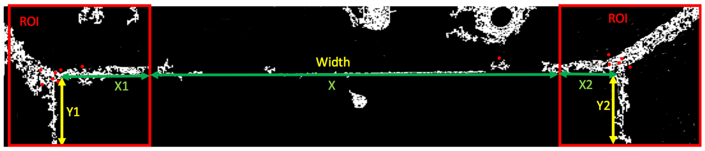


Figure 4. Corner points matching between two regions of interest of a wall image.

decrements by 0.02. The standard deviation was set to $\sigma=sqrt(2)$.

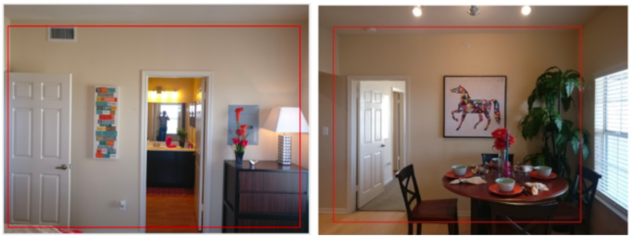
The corner detection is the second step of the process during which the edge map of the image is processed for the identification of the candidate corners. The output of the corner detection algorithm is a set of potential points that can be considered corners of the walls, as shown in Fig 2. c). The red mark corresponds to the set of potential points. It is obvious that the corner point that falls on the intersection of the three edges is the preferred wall corner. The identification of the final corner is described in the next section of the appear. For the purpose of our investigation, the *detectMinEigenFeatures* corner detection algorithm [14] was used. This is a function of MATLAB and has the following structure:

$$Corner = detectMinEigenFeatures\ (EM, q, G); \qquad (3)$$

where *EM* denotes the edge map in gray scale (binary), *q* is a scalar value between [0, 1] and denotes the corner strength and quality. Larger values of *q* are used to eliminate erroneous corner points. For the purpose of our investigation, the value was set $q=0.5$ because the pre-processing phase of the image eliminates the majority of erroneous points. The function returns an object file called *Corner* that incorporates location of corners in pixel coordinates *i*, *j* and the corner metric value, C_{metric}. Larger corner metric indicates a strongest candidate for a corner [13]. Parameter *G* is the Gaussian filter dimension and is an odd integer value in the range [3, inf]. For the purpose of our investigation, we set $G=3$. The Gaussian filter is used to smooth the gradient of the input image. The minimum Eigen values of the corner detection algorithm is computed using the following formula [14]:

$$C_{metric} = \sum \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \qquad (4)$$

where I_x denotes the horizontal gradients of the edge map, I_y denotes the vertical gradients of the edge map, I_xI_y denotes the edges on diagonal. C_{metric} denotes the matrix with two *Eigen* values λ_1, λ_2 characterized by their shape and size of the principal component ellipse inside each filter of an image were computed. According to the used parameter *q* the output of the corner detection algorithm may not provide any candidate corner points. In that case, the algorithm reduces the corner quality parameter *q* with a step of 0.05 until corner points are detected. This process is also presented in Figure 3. The corner detection phase ends with the detection of at least one or more strong candidate corner points with a corner metric value above



a) b)

Figure 5. a) Detected wall boundary. (b) Detected wall corner on a wall with an unclear top right corner

the quality level. The same procedure is performed for the bottom corners of the wall. Thus, the output of the corner detection process is a set of corner points for each region of interest of the wall. For the top left part of the wall, the output is a set of points (x_i,y_i), $i \in TL$ where TL indicates the number of found corners for this region of the wall. Respectively, for the top right part of the wall the potential corner points are (x_j,y_j), $j \in TR$. The bottom left part includes the candidate corner points (x_m,y_m), $m \in BL$. Finally, the bottom right part of the image includes the candidate corner points (x_n,y_n), $n \in BR$.

D. *Computation of wall width*

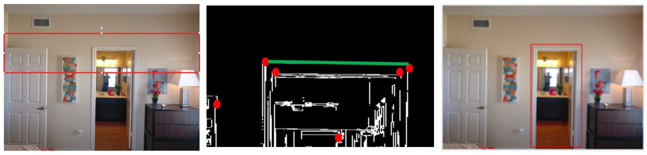
This part of the algorithm provides an estimation of the wall width according to the detected candidate corners. These corner points may include both good candidate corners but also erroneous corners. In order to avoid the negative effects of the erroneous corners in wall width measurements, the best candidates should be determined. For that reason, each corner point in all regions of interest are compared with each other. The corner points from the top part of the wall that have approximately the same *y* value y_i~y_j, where *i* and *j* are the two candidate corner points from the top right and top left part of the wall, are preferred. In addition, the corner points from the top left and bottom left part of the wall that have the same *x* value x_i~x_m, where *i* and *m* are the two candidate corner points from the bottom and top left part of the wall, are preferred. Similarly, the same procedure occurs for the bottom left and right and also for the top and bottom right part of the walls. The final corner detection is computed according to:

$$i^*, j^*, m^*, n^* = \min_{i,j,m,n} \left[|x_i - x_m| \cdot |y_i - y_j| \cdot |y_m - y_n| \cdot |x_j - x_n| \right] \qquad (5)$$

The overall process is shown in Figure 4. The width, *w*, of a room wall is calculated by measuring the pixel distance between the two final corners.

$$w = \frac{h}{|y_{i^*} - y_{m^*}|} \cdot |x_{i^*} - x_{j^*}| \qquad (6)$$

where $h/|y_{i^*} - y_{m^*}|$ is the pixel resolution r_p measured in meters/pixel. The pixel resolution can be computed according to the height of the wall, *h*, which is defined by the user and the number of pixels between the two corners. In a mathematical form, this is presented in (6). The detected wall boundary is demonstrated in Figure 5.



a) b) c)

Figure 6. (a) Input image with region of interest (ROI) selected. (b) Detected door corners over edge map. (c) Detected door on a wall

E. *Door detection*

The door detection process follows a similar approach where an edge and corner detection algorithm is used to find the location of the boundaries of the door [17]. An illustration of the overall process is given in in Figure 6. For the door detection, the region of interest is focused above the half of the wall and below the third quarter of a wall. This is because, most doors found in typical residential units have these height values. To increase the efficiency of the door corner detection algorithm, the following conditions were assumed:

- Preferred door corner should have y-axis value relatively equal to standard door height of 2.1 meter. Thus, $r_p \cdot |y_i - y_m| = 2.1$m.
- Two corner points should have relatively same y-axis values. Thus, $|y_i - y_j| \sim 0$.
- Two corner points should be separated relatively by standard door width 0.9 meters. Thus, $r_p \cdot |x_i - x_j| = 0.9$m.

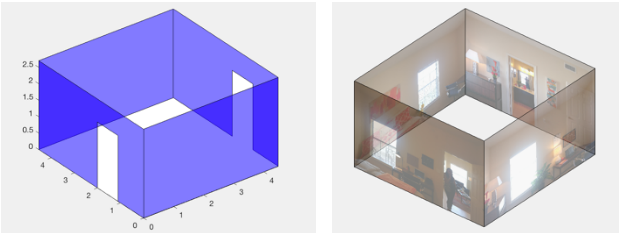
Similar to the wall detection process, the algorithm first identifies the position of the door boundaries, computes the door dimension and defines the coordinate values of its corners.

III. The Facet Model

A. *Constructing the 3D environment*

After the successful wall and door width detection, the final coordinates of the room can be stored in a facet model format. The vector map is represented by its facet where each wall and door is defined by four coordinate points *x,y,z*. These coordinates indicate the respective corners. The facet representation of a single room is presented in Figure 7. For more enhanced experience, it is possible to overlay the picture as texture on the facet as presented in Figure 7b.

Using the same method and principles, the 3D vector map of the remaining rooms of the indoor environment can be constructed. One difficulty for this case, is the positioning of the rooms to form a realistic indoor environment, close to the real one. For the purpose of our investigation, we assume that the user takes four pictures per room to cover the 360 space and takes the pictures in a clockwise manner. Once the user completes this process for one room, then the user takes the pictures of the adjacent room, starting from the wall that is shared with the previous room. In that way, there is always a "calibration" or orientation point that allows the algorithm to reconstruct and attach the facet of each room and form a realistic indoor environment. This process is presented in Figure 8. In this figure, the first room is marked as initial and attached to the adjacent room according to the shared wall of the two rooms. In the next iteration, the second room becomes



a) b)

Figure 7. a) 3D vector map. (b) 3D indoor environment interior view.

the reference room and the third room is attached according to their shared wall. This process is followed until the user captures images of all rooms of the indoor environment and the indoor environment is fully constructed.

B. Data structure

The data structure of the facet model is presented in Table I. The indoor environment in composed by a set of individual rooms. Each room has a number of walls and each wall may have a number of doors. The elements of the room structure store all the details of the indoor environment like wall width, room number, room position, wall image, wall coordinates and door coordinates. The room position field is used to determine the position of the room according to the previous one. The wall image is used as a texture and is overlaid on the facet model to enhance the user experience. The pixel size is used for the computation of the dimensions of the walls and doors length and width and may also be used for future applications. The wall coordinates and door coordinates represent the vector format of the facet and is the most valuable element of the structure, used by the ray tracing algorithm. Finally, each facet incorporates its constitutive parameters that are used for the computation of the diffraction, reflection and transmission coefficients of the ray tracing model. For the purpose of our investigation, the wall was assumed to be made by brick material and the doors by wood material. The constitutive parameters of these materials can be found in [12]. The details of the indoor environment can be fetched using the 'Building Details' button of the main GUI, as described in the following section of the paper.

TABLE I. DATA STRUCTURE OF THE FACET MODEL

Room Wall Position	**Room Properties**	**Description**
Room(i).Wall(j)	Room_Position	Top, down, left, right, front & back position
---------||---------	Wall_Image	Respective room wall image
---------||---------	Width_Pixel	Wall width in pixel size
---------||---------	Width	Wall width in meter
---------||---------	Height	Wall height in meter
---------||---------	Coordinates	Wall (x,y,z) coordinates as a set of four corners
---------||---------	Door	Door (x,y,z) coordinates as a set of four corners

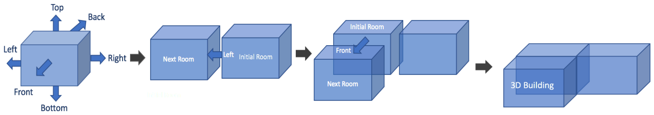


Figure 8. Integrating individual room blocks into a building based on the direction of next room with respect to initial room.

IV. Results

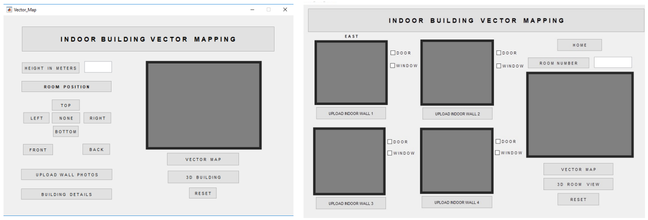
A. Graphical User Interface (GUI)

A GUI was designed to make the use of the developed app easy and user friendly. The user can enter the standard height of the ceiling that is used as reference for the pixel resolution definition. The user also enters the room position that is used as a reference point for the construction of the 3D space. Finally, the user uploads the images for each wall of the indoor environment by using a secondary GUI as indicated in Fig 9. The user can upload four individual wall images per room and indicate if there is a door in the room. The door checkbox was used to reduce the computational cost by eliminating unwanted door detection processes. Once the user uploads the data to the system, the facet model is computed. Within the GUI, there is a button to indicate if there is a window in a wall. For the purpose of our investigation, windows were not incorporated in the facet model and is something that will be integrated in future versions of the algorithm.

B. Augmented Reality to Ray Tracing

The scenario under investigation is presented in Figure 10. A two-bedroom student dorm apartment was examined that has three main rooms. The facet model of the apartment was successfully reconstructed when the user uploads the twelve images of the walls of the three rooms. A commodity smart phone device was used to capture the images. The user spent approximately 3 minutes to take the photos and upload into the system using the GUI. When the user uploads the images to the system, the algorithm performed the pre-processing phase by down sampling and applying Gaussian filters. The input images were down sampled to different resolutions, and the best performance was met when the resolution was set to 640x480 from. It was found that the most suitable corner detection technique was '*DetectMinEigenFeatures*' of MATLAB since it provided the most accurate results and is widely used by the research community. The found coordinates of all rooms were integrated together to form the facet model of the entire indoor environment. The processed images are then embedded on to their respective walls to form a 3D interior view which is as demonstrated in Figure 10 b). It should be noted that the user inputs at the GUI, such as the height of the ceilings, the position of different rooms and the existence of doors on walls, reduced the computational cost by approximately 35%-40%. This is because, the algorithm did not search for doors in case there was no door at the room and made a more efficient positioning of the rooms to form the entire indoor space.

ISBN: 978-1-61208-623-1



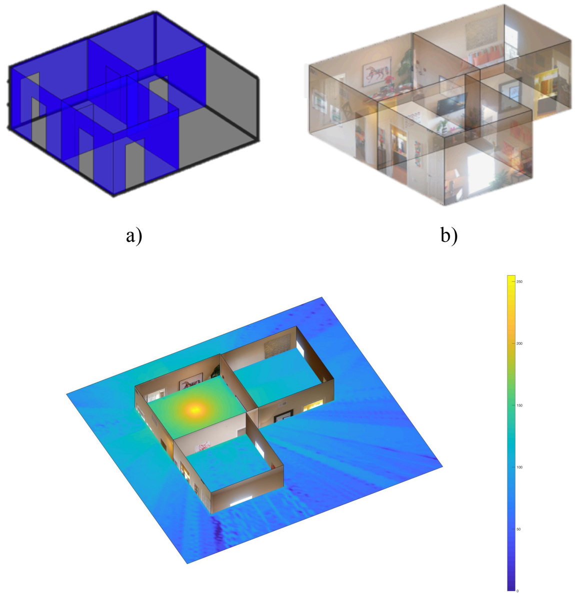
a) b)

Figure 9. a) Main GUI of indoor building vector mapping, b) Secondary GUI for uploading images of a room and mapping individual rooms.

The final step of the proposed system is to use the facet model of the indoor space as input to a ray tracing algorithm [11]. The ray tracing algorithm models the propagation of the electromagnetic waves using the Geometric Optic (GO) technique and decomposes the total field strength as sum of individual rays each one carrying a different amplitude and phase. The amplitude was computed as a combination of multiple reflection, transmission and diffraction coefficients. For the purpose of our investigation the used frequency was assumed to be of the order of the 6GHz band of 5G systems. The results are presented in Fig 10 c). It is observed that the walls and doors of the environment interact with the electromagnetic waves and change the signal strength. With the use of the proposed system, the user is able to take 12 images of the walls of the house and with just a few clicks be able to visualize the signal variation and channel condition of the 5G femtocell station inside the house. This process opens new frontiers in indoor network planning that can be performed by non-technical users and non-experts in the field. In addition, it creates new opportunities for the education of indoor channel modelling with the use of Augmented Reality (AR) devices and applications.

V. Conclusion and Future Work

This paper presented a novel image processing algorithm that is capable of creating the facet model of an indoor environment based on images captured by typical smart phone cameras. The algorithm can be considered as a simplified spatial mapping technique that leverages Augmented Reality (AR) technologies and principles. The application of the algorithm was focused on ray tracing and wireless indoor channel prediction. With the evolution of 5G networks and AR application, it is expected that there will be a great need for integrating network planning and visualization algorithms with AR technologies. It was found that the proposed solution could be used for standard indoor residential houses, but it is not efficient for large or complex indoor spaces. The proposed solution applies edge and corner detection algorithms on the images of the walls and identifies the coordinates and dimensions of the basic electromagnetic clutter, which are walls and doors. The coordinate system was based on the facet model that is used by most of the ray tracing and channel estimation algorithms. It was found that in less than 3 minutes a user could obtain signal strength estimations in a 3-bedroom house just by uploading *.jpg* images of the walls of all rooms.



a) b)

c)

Figure 10. a) 3D Vector map of a building. b) 3D Interior view of a building, c) Implementation of a Ray Tracing algorithm on the facet model.

References

[1] T. Gupta and H. Li, "Indoor mapping for smart cities — An affordable approach: Using Kinect Sensor and ZED stereo camera," International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1-8, 2017.

[2] N. I. A. M. Nazri and D. R. A. Rambli, "Current limitations and opportunities in mobile augmented reality applications," International Conference on Computer and Information Sciences (ICCOINS), pp. 1-4, 2014.

[3] D. Chatzopoulos, C. Bermejo, Z. Huang, and P. Hui, "Mobile Augmented Reality Survey: From Where We Are to Where We Go," IEEE Access, pp. 6917 - 6950, 2017.

[4] M. E. C. Santos, "Augmented Reality Learning Experiences: Survey of Prototype Design and Evaluation," IEEE Transactions on Learning Technologies, vol. 7, pp. 38-56, 2014.

[5] L. C. Chen, N. V. Thai, and H. I. Lin, "Real-time 3-D feature detection and correspondence refinement for indoor environment-mapping using RGB-D cameras," IEEE International Symposium on Industrial Electronics, Taipei, Taiwan, pp. 1-6, 2013.

[6] X. Xu and H. Fan, "Feature based simultaneous localization and semi-dense mapping with monocular camera,"Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), International Congress on, pp. 17-22, 2016.

[7] J. P. Collomosse, "Real-time environment mapping for stylised augmented reality," The 3rd European Conference on Visual Media Production (CVMP), pp. 184-184, 2006.

[8] S. Damodaran, A. P. Sudheer, and T. K. Sunil Kumar, "An evaluation of spatial mapping of indoor environment based on point cloud registration using Kinect sensor," Control Communication & Computing India (ICCC), International Conference on, pp. 545-552, 2015.

[9] S. Singh and R. Singh, "Comparison of various edge detection techniques," 2nd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, pp. 393-396, 2015.

[10] X. Ge, "Multipath Cooperative Communications Networks for Augmented and Virtual Reality Transmission," IEEE Transactions on Multimedia, vol. 19, no. 10, pp 2345-2358, 2017.

[11] G. Koutitas, A. Karousos, and L. Tassiulas, "Deployment Strategies and Energy Efficiency of Cellular Networks," IEEE Transactions on Wireless Communications, vol. 11, no. 7, pp. 2552-2563, 2012.

[12] F. S. D. Adana, O. G. Blanco, I. G. Diego, J. P. Arriaga, and M.F. Catedra, "Propagation model based on ray tracing for the design of personal communication systems in indoor environments," IEEE Transactions on Vehicular Technology, vol. 49, no. 6, pp. 2105-2112, 2000.

[13] S Singh and R Singh, "Comparison of various edge detection techniques," Computing for Sustainable Global Development (INDIACom), 2nd International Conference on, pp. 393-396, 2015.

[14] X. C. He and N. H. C. Yung, “Corner detector based on global and local curvature properties,” Optical Engineering 47, no. 5, pp. 1-4, 2017.

[15] P. Ram and S. Padmavathi, "Analysis of Harris corner detection for color images," International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES), pp. 405-410, 2016.

[16] Bastanlar, Yalin, and Y. Yardimci, “Corner validation based on extracted corner properties”, Computer Vision and Image Understanding, 112, pp. 243-261, 2008.

[17] X. Yang and Y. Tian, "Robust door detection in unfamiliar environments by combining edge and corner features," IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, San Francisco, CA, pp. 57-64, 2010.

Modular and Reconfigurable Testbed For 5G Systems And Applications

Heena Rathore, Abhay Samant

Hiller Measurements
Austin, TX
Email: [heena.rathore,abhay.samant]@hillermeas.com

George Koutitas

Texas State University
San Marcos, TX
Email: george.koutitas@txstate.edu

***Abstract*—The fifth generation (5G) cellular standards operating in various millimeter frequency bands, are the proposed next telecommunications standards beyond the current 4G standards. The specifications of 5G technology are currently being standardized by international regulatory agencies and they hold promise for a wide array of applications ranging from transportation to health. Testing of this standard across a matrix of specifications and applications presents a daunting challenge. To overcome this, a 5G testbed design which is based on reconfigurable components enabled by Software Defined Networks (SDNs) and Software Defined Radios (SDRs) has been presented in this paper. The reconfigurable measurement hardware has been designed such that it can be integrated across all the layers of TCP/IP protocol through an open-source software defined architecture. Programmability is a key feature of this architecture, and this has been addressed by a Software Development Kit (SDK). The SDK contains pre-built IP, a baseline end to end stack implementation, and an application programming interface (API) for accessing different features of the platform. The testbed has been designed with a modular hardware and scalable software architecture so that it can facilitate the development of numerous 5G applications in the long run, allowing multiple users to operate it, thus making the testbed self-sustainable over the years.**

Keywords–testbed; scalability; modular; 5G; health; transportation; energy.

I. Introduction

The fifth generation (5G) of cellular standards, holds promise for a variety of applications including, but not limited to, health, energy, transportation and public safety. 5G focuses on solving various present-day communication challenges, such as area traffic capacity, network energy efficiency, connection density and latency. It is being designed for enhanced mobile broadband applications such as streaming 4k video, augmented reality, and 3D gaming. Ultra reliable, low latency communication for autonomous driving and mission critical applications has been presented as an application area. 5G also is focused on improving spectrum efficiency and mobility, thus making it a prime candidate for massive machine to machine type applications such as in smart cities and smart factories.

While there is a lot of excitement around the promise that 5G holds, a key requirement is the need of common methodology and systems for testing these applications in real-world situations. It is extremely critical that researchers have a common test platform to validate scalability and interopertablity across these applications. 5G CHAMPION testbeds, described in [1], were designed for the 2018 Winter Olympic games, to validate how 5G-enabled mmWave wireless backhaul can provide an interoperable and seamless connection between two different access networks. 5G Hardware Test Evaluation Platform, presented in [2], deploys software defined wireless networks in the urban area, allowing academics, entrepreneurs and wireless companies to test, evaluate, and improve their hardware design and software algorithms in real-world environment. Additionally, it supports advanced wireless communications theory and technology research. An educational setup for service oriented process automation with 5G has been presented in [3]. The intended outcome is that students can obtain knowledge with emerging industrial technologies and become the actors of upcoming industrial revolution. A testbed described in [4] demonstrates SDN orchestration capabilities in adapting data paths across IoT, cloud, and network domains, based on the real-time load state of switches. This enables recovery from congestion, thereby assuring reliable data delivery services.

All of these testbeds have been designed to meet the requirements of a specific application or achieve a particular learning outcome. A key challenge is that these testbeds are not flexible to scale for different applications and evolving specifications. This paper presents a testbed architecture, based on reconfigurable Software Defined Networks (SDN) and Software Defined Radio (SDR) components. The testbed has been designed in a modular hardware fashion with a scalable software interface to allow its use for evolving 5G systems and technologies. It has been designed such that it can be integrated across all the layers of TCP/IP stack through an open-source software defined architecture containing a pre-built IP, a baseline end to end stack implementation, and an Application Programming Interface (API) for accessing different features of the platform. Section II describes the hardware architecture of the platform. Software architecture has been described in Section III. Sustainability aspects of the testbed have been described in Section IV. The testbed has been designed with the objective that it will facilitate the development of numerous 5G applications in the long run, some of which have been described in Section V.

II. Hardware Architecture

The core of the testbed hardware lies in the SDR and SDN components that interface with other essential services such as data logging and aggregation access, administrative and performance monitoring services. They have been integrated in such a way that they can be controlled through a programming interface. This allows the testbed to interface with a variety of services frameworks. Figure 1 describes the overall architecture of the 5G testbed.

Flexibility at the baseband level is enabled via the use of SDRs such as USRP [5]. The current generation USRPs support 160MHz instanteneous bandwidth with frequency coverage from 10MHz to 6GHz, referred to as sub-6GHz in

ISBN: 978-1-61208-623-1

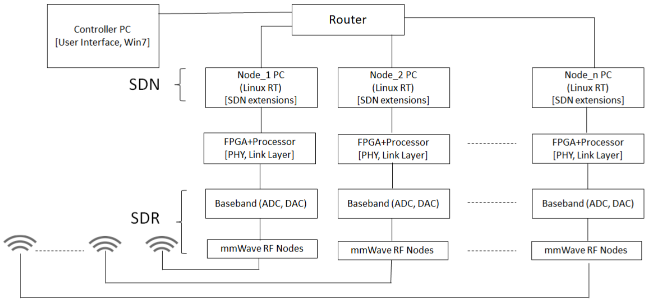


Figure 1. 5G Testbed Architecture

the rest of the paper. This serves as the baseband system of our testbed and provides sufficient frequency coverage and reconfigurability for research on topics such as Long Term Evolution (LTE)-to-5G migration, LTE-5G co-existence and/or convergence, and IoT. The modular nature of our testbed addresses one of the key 5G challenges related to the different frequency bands being considered [6], [7]. The World Radio Conference WRC-19 [8] and the designated ITU-R [9] qualifier for 5G, includes a set of bands to be considered for 5G, with direct applicability to 5G New Radio (5G NR). 5G NR is already taking shape in 3GPP with OFDM-based Unified Flexible Radio Access Technology below 40 GHz. Likewise, a non-stand-alone version was finalized in Dec. 2017 and several companies are releasing their 5G/KT mmWave spec. with a pre-standard for mmWave in 5G at 28 GHz for fixed wireless. As shown in Figure 2, the modular nature of the testbed allows the sub-6GHz system to be extended with mmWave up and down converters for different frequency bands. For example, a mmWave up/down converter for the 27.5-29.5 GHz band and direct interfaces to multi-element phased-array antenna RFICs have been added to the testbed. Some of the other spectrum bands which are under study for WRC-19 and can be added to the 5G testbed are the 37-40.5GHz 60-66GHz and 71-76GHz. To get to custom mmWave frequencies, RF daughter cards can be replaced by new commercially available upgrades or custom front ends. Local Oscillators (LO) can be used independently, in pairs or in external/shared modes.

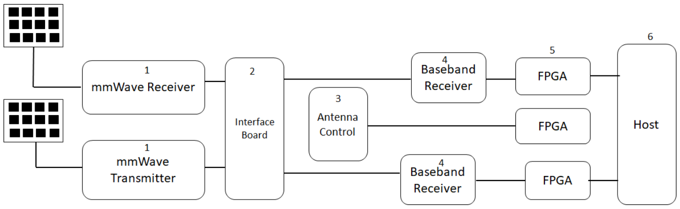


Figure 2. 5G Testbed Hardware Architecture

Earlier this year, ITU agreed on key 5G bandwidth requirements for IMT-2020 [10]; for example, the target values for downlink and uplink user experienced data rates were set at 100 Mbit/s and 50 Mbit/s respectively. These values are defined assuming supportable bandwidth as described in [11]; however, the bandwidth assumption does not form part of the requirement. Considering that the frequency and bandwidth requirements for 5G standards are still being defined and are in development, the 5G testbed described in this paper has been designed in a modular fashion to adapt to different requirements. As Analog to Digital Conversion (ADC) technology evolves, we expect to see new digitizers with wider bandwidths to become commercially available. In the meantime, signal processing techniques such as spectrum stitching [12] can be used to achieve wider bandwidths by concatenating multiple USRPs. The RF interface block contains elements such as filters, amplifiers, switches and interfaces. Testbed has been designed to include multiple sub-6GHz RAN nodes, some of which can support indoor network research, while the rest can be located outdoors using roof mount and fixed ground installations.

Since directionality will be a key feature of 5G networks, beam forming and beam steering will be essential. Hybrid beamforming has been approved for LTE release 13, Phase 1. It has been shown that by significantly increasing the number of antennas, for example 64 antennas, the narrower beams can provide three to five times capacity gains, while taking advantage of existing infrastructure. For Phase 2 of LTE release 13, it is expected that each antenna will include its own transceiver, enabling both traditional MIMO techniques and the Zero Forcing beamforming approach, achieving 10 times capacity gain. To enable testing of this feature, an external phased-array antenna RFIC (such as SiBeam 12x12 element phased-array antenna) can be easily added to the 5G testbed. Many of the new 5G implementations will require beam steering on multiple beams. So, as advanced beam steering technology is developed and integrated into new 5G designs, the 5G testbed can be easily adapted using the Antenna Control block in Figure 2, to use 4 beams and 4 phased arrays for a total of 256 elements.

III. Software Architecture

Reconfigurability of the hardware elements is a key feature of the testbed and this is enabled by the control and processing software running either on the FPGAs or host computer. This section describes the software architecture of the 5G testbed, as shown in Figure 3. The testbed software has been designed using a plug and play architecture such that researchers can easily introduce new algorithms at any of the layers of the TCP/IP stack. It aims to provide users an insight into different blocks from an application point of view, without requiring them to go into any of the implementation details. Software will integrate the new algorithms, which can be deployed either on FPGA or host PC, with the rest of the stack and allow researchers to experiment with different settings. The testbed provides a default end-to-end stack, which will be used as a baseline. Users can then replace individual components in the stack. As an example, a researcher may like to develop a MaxWeight multiuser scheduler and then plug this algorithm at the MAC layer, without the need to know anything else about the baseline wireless stack. Once the user provides a binary representation of the new algorithm through a prescribed API, the platform will generate the entire stack, run the program, and provide throughput results. This will be based off the Key Performance Indicators (KPIs) which will indicate performance metrics, such as bandwidth, energy consumption, or latency.

The novel and guiding design principle of this testbed is rooted in its programmability across all layers, pre-built IP

ISBN: 978-1-61208-623-1

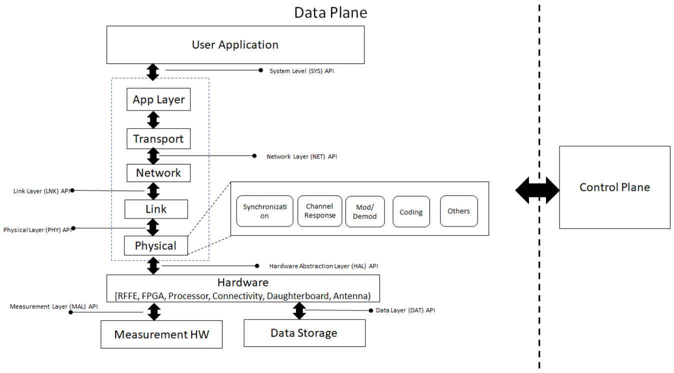


Figure 3. 5G Testbed Software Architecture

as a baseline implementation, and a hybrid approach to spectrum management that features both sub-6GHz and mmWave frequencies. This programmability has been provided through well defined interfaces such as NET API, PHY API, RF API, as shown in Figure 3 and described next.

SDN provides a breakthrough in network transformation. It decouples the software and the hardware layers by disengaging the data plane and control plane of the networking device [13]. Evolution of SDN programmability at network level will pave the way for new innovations. The testbed will leverage the principles of SDN to enable scalable, flexible and highly adaptive networking and communication layers through decoupling of the control and data planes. SDN will support efficient traffic and flow management, including resource reservations, and custom network layer protocols by abstracting the routing and networking intelligence (control layer) away from the switching hardware (data layer). The NET API will interface with the underlying radio signal processing plane abstractions using a well-defined API at each layer. It will be a superset of OpenFlow [14] and extensions for receiving information from different SDR blocks, such as modulators, coders, timing. It will also include a mechanism to define a set of actions for different network nodes. Users will be able to modify some contents of the packet, define performance indices, and deploy customized routing and switching protocols using this API. OpenFlow currently supports a limited number of protocols. This testbed will extend OpenFlow constructs for 4G and eventually 5G networks.

The PHY API is responsible for controlling the physical layer algorithm parameters such as modulation scheme, symbol rate, filter type, channel response equalization filter taps, coding parameters and such. This layer will also monitor the received signal characteristics such as RSSI and provide feedback to the upper layers. The testbed will be designed such that it allows for real-time configuration of RF layer parameters. The RF API will be responsible for controlling hardware specific parameters such as frequency, power level, and instantaneous bandwidth (specified as sampling rate). The RF API will abstract the features of the interface board to make it easier to program. Using General Purpose Input/Output (GPIO) lines on the USRP, the testbed will also facilitate real-time reconfiguration of RF hardware parameters using features such as Adaptive Gain Control (AGC).

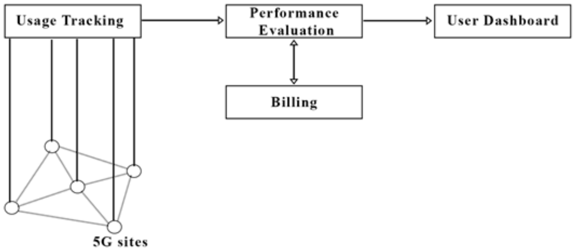


Figure 4. System Process for Usage Tracking and Billing

IV. USAGE TRACKING AND BILLING

Usage tracking and billing are important aspects for the successful implementation and monetization of the proposed 5G testbed. Usage tracking is crucial for the collection of raw experimental and platform utilization data whereas the billing mechanism provides the required monetization and the engagement of users. The proposed testbed will have a set of measurement instruments that will collect experimental and utilization data. The nature of the experimental data depends on the API layer that is used. For example, referring to Figure 3, measurements related to the testbeds Physical and Link layer API can be channel fading, time domain, frequency domain and modulation quality data. On the other hand, energy efficiency, throughput, latency and other measurements can be used for the characterization of the higher layers. Utilization data will keep track of the used platform resources of the experiment such as duration, number of APIs and 5G sites used and the overall energy consumption. The energy consumption provides information about the operational expenses of the platform for the purpose of the experiment but also can work as an incentive for researchers to design energy efficient 5G algorithms. This is because the billing will depend on the energy consumption.

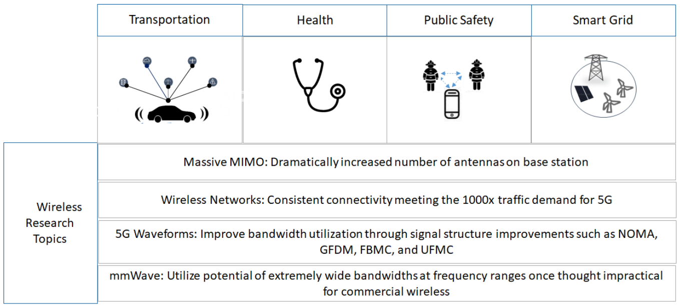
In general, an important dimension of 5G networks is the energy efficiency [15] and different Key Performance Indicators (KPIs) have been proposed to quantify the efficiency [16]. An overlaid IoT based smart metering network capable of monitoring the energy consumption of the network will be used for energy monitoring. Each 5G site is expected to have at least 4 energy consumption points to capture the consumption at the radio unit, the IT equipment, the cooling/power units and of course the entire station to provide the required KPIs such as Joule/bps, W/m2, W/user, Power Usage Effectiveness (PUE). These measurements will also allow researchers to monitor and disaggregate the energy efficiency of the developed algorithms and create new energy efficiency proxies and metrics following recent advancements in the data center sector such as the Green Grid association [17].

Usage tracking data will be available to the researchers in real time during their experiments in a cloud based dashboard connected to the usage tracking database. A cloud platform interfacing the software and hardware of the 5G testbed will empower the users with the ability to remote access the system using time/frequency sharing scheduling processes.

The usage tracking data will be used to not only evaluate the efficiency of the experiments and the performance of the newly developed algorithms, but also to provide billing. The billing algorithm will use pricing principles met in Software as a Service (SaaS) paradigms [18]. The pricing schemes can be based on an a-la-carte or a bundle approach providing the option to the end user to select according to the needs. The a-la-carte scheme allows a user to pay per experiment whereas the bundle approach creates packages of products, services and usage priorities with monthly subscription. For this case, the product is the API and the 5G site of the experiment and a user can select a number of them to be utilized for the experiment. The final cost of the pricing scheme is a function of the energy consumption, the number of 5G sites and APIs used and the performance of the algorithm. The overall process is presented in Figure 4.

V. Applications

The 5G testbed has been designed with various general purpose wireless research and application specific research in mind, as shown in Figure 5. Understanding and defending against potential cyber-security attacks on SDNs, such as denial of service, is a key. The testbed can be used to test the self-evolving and self-healing characteristics of SDNs, taking inspiration from bio-inspired techniques. Some of the significant challenges that can be tackled through the testbed are the interference of small cells and macro-cells, new interference situations, and synchronization. Algorithms focusing on ultra-low latency (less than 10ms for remote medicine applications), ultra-high throughput (several Gbps for large scale data transfers) and support for massive number of devices (greater than 1000 for IoT applications). This testbed will allow for the fast reconfiguration of network nodes to handle a dynamic mix of such applications via selection of appropriate per-packet scheduling mechanisms. The testbed will also allow for sampling performance metrics such as latency, throughput, jitter, packet loss, and communicating these back to a central controller, which will use data-analytics to choose between a set of curated per-packet mechanisms to attain optimal system performance for the current applications.



Massive MIMO: Dramatically increased number of antennas on base station

Wireless Networks: Consistent connectivity meeting the 1000x traffic demand for 5G

5G Waveforms: Improve bandwidth utilization through signal structure improvements such as NOMA, GFDM, FBMC, and UFMC

mmWave: Utilize potential of extremely wide bandwidths at frequency ranges once thought impractical for commercial wireless

Figure 5. 5G Applications

Some additional general purpose wireless research topics include analyzing new RF front end elements such as power amplifiers, filters and transceivers for mmWave, study of coexistence of sub-6GHz, 5G, and hybrid multi-spectral communication systems and the development of appropriate channel and fading models that are correlated to develop new systems concepts that can capitalize on spectral agility, new access schemes, and other resource management paradigms. While these are very important stand-alone research topics, they are particularly critical for many of the 5G application areas, as discussed next.

A. Transportation

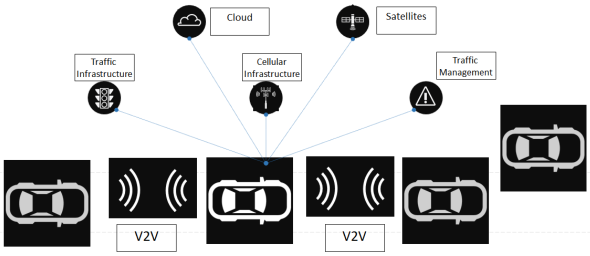


Figure 6. 5G Applications for Transportation

Wireless communication is bringing a new level of connectivity to cars. As shown in Figure 6, with wireless, cars may communicate with each other directly in Vehicle-to-Vehicle (V2V) mode, or through the infrastructure in Vehicle-to-Infrastructure (V2I) mode. There are many applications of connectivity to support safety, transportation efficiency, and internet access. Additionally, connectivity makes self-driving cars safer by increasing their sensing range, leveraging what can be seen by other vehicles in the front, in the back, or on the sides. Exchanging such information between vehicles will improve driver assist and full automation over time. Unfortunately, conventional technologies such as dedicated short-range communications, which support data rates of megabits per-second and low-latency messaging, will not be sufficient to support the exchange of high rate sensor data or exchange of data to support automatic high definition map updates. 5G networks hold the promise to support high data rates and low latency for connected vehicles, which is driving tremendous interest in transportation as a key use case. In particular, mmWave 5G is especially attractive because of very high data rates, which can be used for the exchange of raw sensor data, enhancing the safety and efficiency of automotive driving. This would allow vehicles to enhance their situational awareness by seeing many car lengths in different directions, and around corners. Additionally, 5G can support lower latency and ultra-reliability to facilitate distributed control for transportation systems. For example, vehicles can travel together with smaller gaps using platooning, or can be coordinated through an intersection at high speeds without a traffic lights. These attributes enable safe operation of connected vehicles in a variety of traditional crash hot-spot situations such as overtaking on rural roads, conflicts at urban intersections, and weaving sections on highways. There are also opportunities to co-locate sensing and communication together in 5G systems. Sensing on the base station gives a birds-eye-view of the environment and may assist in automated intersection management. This functionality is supported by edge-computing, which will be supported by 5G networks.

B. Health

5G networks hold the possibility to empower new potential avenues regarding health care including imaging, diagnostics, data analytics, and treatment [13]. This includes devices such as clinical wearables, remote sensors and numerous different gadgets that screen and electronically transmit medical information such as vital signs, physical activity, individual security, and pharmaceutical adherence [19]. These devices will provide unprecedented telemedicine diagnosis and treatment benefits, while significantly lowering health costs. These devices and capabilities generate higher fidelity data, thereby enabling precise analytics capabilities. For example, doctors typically require access to detailed information about hereditary, social condition, and way of life attributes to provide informed health care. The billions of devices and sensors connected through 5G will make collection of this information possible. Storing this information on a cloud infrastructure enables all-time accessibility. Some mission-critical medical functions require high dependability and accessibility with latency intervals that are down to a few milliseconds [20]. 5G will make this possible and predictable, thus enabling dependable client encounters to enhance medical care. Similarly, remote surgery will be possible once latency levels are reduced to small intervals. Surgeons will have the capacity to utilize virtual and augmented reality tools for certain kinds of techniques. Some other examples incorporate imaging, remote monitoring and diagnostics, and data analytics for effective treatment. Recently, wireless medical devices have enabled many hospital facilities the ability to provide continuous patient care throughout the treatment process. Standardization strategies to assist machine learning algorithms in adding to the efficiency of these devices would be the core interest. It will characterize the necessities of machine learning algorithms as they relate to network architectures and data security. Machine learning based data security will examine the information being transmitted between the patient and specialist to provide better treatment process. Utilizing the machine learning algorithms that iteratively learn from data, would enable these gadgets to discover hidden insights without being explicitly programmed to look for a specific pattern.

C. Energy

The energy sector is expected to be technologically enhanced with the direct, systematic and indirect implementations of 5G networks for energy-efficiency. The first pillar concerns the direct implementation of energy efficiency techniques in 5G access networks. By introducing new network planning and Base Station (BS) management strategies, the Joules required per offered bit will be reduced, integrating the 5G network into the paradigm of energy efficient networking, part of the ITU-R and IMT 2020 vision. 5G systems with high energy performance should be built on two design principles, a) to only be active and transmit when needed, b) to only be active and transmit where needed. The lean design architecture of 5G sites support sleep modes and the SDR capabilities enable BS on/off schemes. The most important challenges are resource allocation, resource sharing and base station management as well as integration of the latter with Renewable Energy Sources (RES). The deployment of a set of 5G off-grid sites powered by RES will provide an important testbed for experimentation and research. In the second pillar, the systematic application of 5G in the smart grid will enable a communication infrastructure which is able to support the emerging energy use-cases of 2020 and beyond. Part of the so-called Internet of energy, 5G networks support reliable data and command flow between a network of Internet of Things (IoT) such as smart meters/smart actuators and electric utilities. The third pillar concerns the indirect implementation of 5G networks to incorporate virtual spaces. This approach integrates densely-deployed IoT devices into an Augmented Reality (AR) environment for energy management. The main challenges concern the integration of high-rate, human-centric AR data with low-rate, machine-centric IoT data for effective cross-domain, real-time control. In addition, low latency end-to-end communications will be addressed, including applications of Ultra Reliable Low Latency Communications (URLLC). Such services are a key driver for the successful penetration of VR/AR services in 5G networks.

VI. Conclusion

Health, energy, public safety and transportation are some of the many applications that can benefit from 5G capabilities for latencies, massive bandwidth, and connectivity. 5G focuses on various aspects of present-day communication challenges such as area traffic capacity, network energy efficiency, connection density and latency. 5G also is focused on driving spectrum efficiency and mobility, thus making it a prime candidate for massive machine to machine type applications such as in smart cities and smart factories. This paper addresses a key gap in the long term adoption of this standard for these applications by presenting a modular and scalable testbed architecture to test interoperability and scalability of this standard across the various applications. For example, the ability of the testbed to adapt to various mmWave frequencies is crucial for self-driving cars as it has to work across the cellular, high bandwidth wireless (60G-66G) and evolving vehicular radar frequency bands (76G-82G). Likewise, the ability of the testbed to scale to multiple Transmit and Receive nodes is of big benefit for health applications. Ability to program at different layers is crucial to test machine learning enabled safety and security of 5G enabled wireless devices. Likewise, the ability to try different algorithms is of immense benefit to general purpose research and energy applications. In summary, we envision that this testbed will facilitate the development of numerous 5G applications in the long run. The testbed has some limitations which need to be addressed in future scope of work. For example, the USRP is currently not seen as a network interface card (NIC) by the operating system on the host machine. One needs to overcome this limitation by making the USRP visible as a virtual network interface. Once this is done, users can directly use the USRP to provide Internet connectivity via bridging it to an Ethernet NIC, and exploiting the existing TCP/IP stack on the host. Thus, using an Ethernet-Host-USRP combination as a fixed base station, with combinations such as USRP-Host-WiFi, USRP-Host-Bluetooth, and USRP-Host-Zigbee acting as client/mobile-access-points, connectivity to off-the-shelf handhelds and IoT devices will be possible, thereby enabling a new set of applications.

References

[1] S. H. Won, et al., “Development of 5G CHAMPION testbeds for 5G services at the 2018 Winter Olympic Games”, *IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, no. 18, pp. 1-5, 2017.

[2] Y. Yang, J. Xu, G. Shi, C. X. Wang, “5G Hardware Test Evaluation Platform”, *5G Wireless Systems. Wireless Networks*. Springer, Cham, 2018.

[3] J. Kortela, B. Nasiri, A. Smirnov, A. Lahnalammi, S. L. Jamsa-Jounela, “Educational Setup for Service Oriented Process Automation with 5G Testbed”, *FAC-PapersOnLine*, 50(2), pp. 127-132, 2017.

[4] S. Fichera, M. Gharbaoui, P. Castoldi, B. Martini, A. Manzalini, “On experimenting 5G: Testbed set-up for SDN orchestration across network cloud and IoT domains”, *IEEE Conference on Network Softwarization (NetSoft)*, pp. 1-6, 2017.

[5] www.ni.com/white-paper/12985/en/ [Accessed on 03/17/2018]

[6] P. Rost, et al., “Mobile Network Architecture Evolution towards 5G”, *IEEE Communications Magazine*, Vol. 54, Issue. 5, pp. 84-91, May 2016

[7] T. Wang, et al. , “Spectrum Analysis and Regulations for 5G”, *5G Mobile Communications”*, SpringerLink, pp. 27-50.

[8] M. J. Marcus, “5G and IMT for 2020 and beyond [Spectrum Policy and Regulatory Issues],” *IEEE Wireless Communications*, vol. 22, no. 4, pp. 2-3, August 2015. doi: 10.1109/MWC.2015.7224717

[9] International Telecommunication Union (ITU), http://www.itu.int/en/about/Pages/default.aspx [Accessed on 02/26/2018]

[10] D. Soldani, and A. Manzalini, “Horizon 2020 and Beyond: On the 5G Operating System for a True Digital Society,” *IEEE Vehicular Technology Magazine*, vol. 10, no. 1, pp. 32-42, March 2015

[11] X. Meng, J. Li, D. Zhou and D. Yang, “5G technology requirements and related test environments for evaluation,” *China Communications*, vol. 13, no. 2, pp. 42-51, N/A 2016. doi: 10.1109/CC.2016.7405721

[12] S. L. Dark, D. J. Baker, J. R. W. Ammerman, “Spectral stitching method to increase instantaneous bandwidth in vector signal analyzers”, US Patent US9326174B1

[13] H. Rathore, “Bio-inspired Software-Defined Networking”, *Mapping Biological Systems to Network Systems*. Springer, Cham, 2016.

[14] https://www.opennetworking.org/projects/open-datapath/ [Accessed on 02/26/2018]

[15] S. Buzzi, Chih-Lin I, T. E. Klein, H. V. Poor, C. Yang, A. Zappone, “A Survey of Energy-Efficient Techniques for 5G Networks and Challenges Ahead”, *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, April 2016.

[16] L. Budzisz, et al., “Dynamic Resource Provisioning for Energy Efficiency in Wireless Access Networks: A Survey and an Outlook”, *IEEE Communications Surveys and Tutorials*, vol. 16, no. 4, 2014

[17] https://www.thegreengrid.org/ [Accessed on 02/26/2018]

[18] A. Ojala, “Selection of the Proper Revenue and Pricing Model for SaaS”, *IEEE International Conference on Cloud Computing Technology and Science (CloudCom)*, no. 6, 2014.

[19] A. Weissberger, “Selected Applications/Use Cases by Industry for ITU-R International Mobile Telecommunications (IMT) 3G, 4G, 5G”, *IEEE Comsoc Technology Blog*, 2017.

[20] D. M. West,“How 5G technology enables the health internet of things”, Brookings, 2016.

A Survey of 5G Security Considerations

Michael L. Casey
Ingram School of Engineering
Texas State University
San Marcos, Texas, USA
e-mail: mc74@txstate.edu

Stan McClellan
Ingram School of Engineering
Texas State University
San Marcos, Texas, USA
e-mail: sm65@txstate.edu

***Abstract*—The 5G cellular standard is scheduled to begin the first phase of implementation in 2020. The requirements of new services and, therefore, new security requirements, architectures, and technologies mean the new standard will have a very different appearance relative to the prior standard. This paper surveys some key aspects of the 5G standard, and discusses the effect of security considerations in the context of new 5G features.**

Keywords - 5G standard; wireless communications.

I. Introduction

THE advent of the 5G cellular standard means new services will become available in addition to conventional voice, text, and data. Many of these services are forecast to be present in the first phase of implementation in 2020 [1], including support for capabilities related to vehicular communications [2], wearables, healthcare, transportation, and the Internet of Things (IoT) [3]. These "vertical services" are a new aspect of 5G networks, which bring a new dimension to the design problem, requiring additional research and pre-planning for deployment. Here, we present a review of current technology and how different aspects of the security features will impact those technologies.

Vertical services are an important aspect of the 5G network. The requirements of these dedicated or industry-specific solutions provide much of the motivation for the transition to 5G-enabled technologies. The eight key verticals addressed by the 5G architecture include the following: Manufacturing, Media/Entertainment, Public Safety, Public Transport, Healthcare, Financial Services, Automotive, and Energy/Utilities markets [4]. Previously, these industries employed dedicated, single-use networks or other industry-specific communications solutions. With the contemporary shift of most activities to data-driven commerce, it is logical that public telecommunications networks would respond with a broad-based and ubiquitous solution such as 5G. However, the disparate requirements of these vertical markets create a number of difficult challenges.

The requirements imposed on the network by the eight key verticals can be viewed in terms of Operational, Functional, and Performance categories [4]. Each of these categories has specific requirements, as listed in Table 1. The approach to achieving these often contradictory or mutually exclusive requirements is via the implementation of dynamic, programmable, segment-specific virtualized subnetworks. These isolated 5G subnetworks are known as "slices" and are implementations of the business model of Networking as a Service (NaaS).

TABLE 1: VERTICAL INDUSTRY REQUIREMENTS FOR 5G

Operational	**Functional**	**Performance**
Self Managing/Policies	Security	Latency
Programming Interfaces	Identity Management	Throughput
Service Assurance	Isolation	Reliability/Availability
Charging/Billing		Resiliency
Global Operation		Coverage

As key enabling concepts in 5G networks, network slices are a drastic paradigm shift from the management of conventional telecommunications networks. Network slices are logical networks implemented on a common, shared infrastructure. They are required to accommodate the large variety of vertical services and the disparate service requirements imposed on the network by each vertical service. In most cases, slices are viewed as an "on demand" meta-service which optimizes Operational, Functional, and Performance requirements for various use cases, service types, and business models. In their most basic form, network slices are groups of functions, resources, and connections, which enable certain types of application services, which bound certain important performance requirements, and which ensure specific service-level agreements between users and providers. In this context, it is clear that one of the most critical aspects of network slicing is the ability of the infrastructure to isolate a multiplicity of slices. Isolation is a key component of the general concept of "security," where the isolated slice benefits from (a) greatly reduced attack vectors, (b) highly segregated internal and operational data, and (c) intelligent limitations on connectivity via restricted architecture.

ISBN: 978-1-61208-623-1

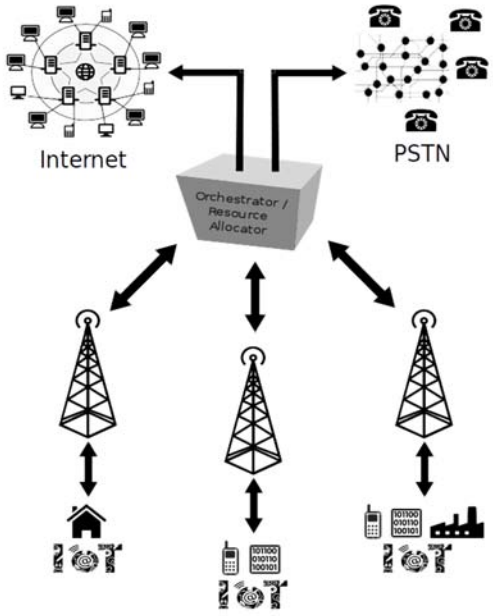


Figure 1: Illustration of network slices

Figure 1 presents a simplified perspective of three different network slices. The Internet and Public Switched Telephone Network (PSTN) are connected in the 5G network. Each tower represents a connection with different network needs. The tower on the left side of the figure needs data connectivity for IoT activities in a suburban setting. The tower in the middle of the figure needs multiple resources for mobile units. The tower on the right side of the figure has mixed needs. In all cases, the network resources needed for the subnetworks addressed by each tower are different. Each tower needs the ability to create its own Virtual Private Network (VPN) in order to serve the connected mobile units or other equipment. This VPN constitutes a slice of network resources. The slice could involve multiple service providers, e.g. a server for cloud storage, a company supplying the physical infrastructure, and a voice network. These different service providers have their own respective domains within the network structure. The subdivision of the slice among the different service providers must be designed carefully to delineate where liability and security needs for one service provider end and liability and security needs for another service provider begin. The slices and their subdivisions need to be isolated from one another since the security needs are different for different slices as well as the subdivided domains.

An important component in Figure 1 is the Orchestrator / Resource Allocator (ORA). The ORA is responsible for creating end-to-end realizations of services, which are requested by network-resident applications. Such applications request network services, which may span multiple operating domains and may be expressed in abstract fashion, via a common Application Programming Interface (API). A primary function of the ORA is to translate abstracted service requests into resource requests to be handled by controllers in the various domains. Additionally, the ORA maps SLA requirements and Quality of Service (QoS) requirements into formats to be managed by domain-specific controllers. The three slices seen in Figure 1 are likely to change with respect to time; therefore, the changing needs will mean a new instance of a slice will need to be managed by the ORA. Since the concept of "security" in the 5G network is logical rather than physical, the ORA will also have to be virtualized, and the complexity of the ORA will be quite substantial.

The remainder of this paper explores the highlights and important aspects of the Functional Requirements in 5G networks, or the extended concept of "security." The intent is to introduce the reader to tradeoffs, architectures, and considerations which may pervade ongoing implementations and standardization efforts. This discussion is undertaken in the context of requirements for the several vertical markets, and illustrates how security concerns arise in certain cases. For example, vehicular communications are an important "vertical service" in the 5G standard which contain a number of different and considerably complex scenarios [1][2]. Connected cars will be expected to interface seamlessly with the 5G network, just as many cars already connect easily to the 4G/Long-Term Evolution (LTE) network. Additionally, this discussion is undertaken in the context of technologies that are addressed by 5G implementations. For example, integration of IoT systems is an important set of technologies, which will be challenging in the development and deployment of 5G networks [3]. IoT systems will impact canonical network layers (e.g. MAC, PHY) and other vertical services, and will drastically alter the security landscape of the overall network. A brief historical perspective is discussed based upon [5], which was written for 5G Public Private Partnership (5GPPP) and [3] which was written for 3GPP. Also included in the discussion are use cases and performance evaluation models from 5GPPP, and issues related to IoT from 3GPP. While many of the 5G requirements are not globally unique, certain aspects may be designed and adapted to fit local geography, specific use cases, or regulatory requirements. From these aspects and other architectural concerns, it is clear that new security mechanisms, architectures, and technologies are required to manage various aspects of the 5G network.

Section II reviews the features of the security protocol for 5GPPP. Section III reviews the security implications for vehicular communications. Section IV review security implications for IoT. Section V reviews the beginning of the design of the security protocol from the 3GPP perspective. Section VI concludes the paper.

II. 5GPPP SECURITY LANDSCAPE

Security risks in modern network communications are of utmost importance. Developing 5G networks are no different, and security aspects of 5G include issues such as: unauthorized usage/access, weak slice isolation, traffic embezzlement, service level agreement (SLA) compliance, slicing versus neutrality, trust management, service provider lock-in, and insufficient technology readiness levels (TRL) [2]. Each of these items is briefly discussed below.

Unauthorized usage/access: Unauthorized usage/access of assets has several security risks clearly identified. One known risk is that of identity theft or cloning. Subscriber credentials may also be stolen or cloned. The desired seamless interworking between different domains, e.g. a vertical slice or a core slice, may expose the 5G security level to new threats. Another identified risk is that of allowing appropriate security measures for massive IoT deployments while still accounting for necessary security of non-IoT services. The security features must account for all of these requirements, which will likely produce a heterogeneous access security protocol.

Weak slice isolation: If the isolation of the slices is weak, then side channel attacks are a distinct possibility as a security risk. Likewise, management of sensitive data in one security domain may be exposed in another security domain due to a different set of security requirements. Monitoring and management of security protocols across all the security domains implies substantial additional complexity in the interfaces between various slices.

Traffic embezzlement: The specific security risk in traffic embezzlement lies in the weakness of third parties being able to capture or alter control plane data or user plane data without detection. The heart of this risk lies in the inconsistency between three logical segments: the Orchestrator abstraction, the software defined network (SDN) abstraction, and the physical and network resources. This weakness is of critical concern to use cases such as eHealth and lawful interception due to recursive/additive virtualization.

SLA compliance: Several security risks, which could be called vertical SLA and regulation compliance management risks, have been identified by the standards authors. One risk is encountered when an API is used to request geolocation information. This API request must be clear to the user and managed correctly so that information reaches its correct network destination as well as satisfying the requirements of the third party making the request. The third party may also be requesting access to a user’s infrastructure or assets, and the orchestrator must manage this request in conjunction with the third party. With virtual network functions (VNFs), a clear liability chain must be present to protect the user, the orchestrator, and the third party. Also, VNF life cycles must provide evidence that they will not passively introduce additional security risks to the network via updates and software evolution. Unfortunately, the management of these life cycles are outside the control of the operator.

Slicing vs. neutrality: The concepts of network neutrality and slicing are yet to be fully defined by the standards authors in [1]. While some regulations exist within the EU, the regulations do not fully define how to navigate the remaining differences between network neutrality and slicing. Delivering services via a 5G network outside of applicable regulations or in the absence of fully-formed regulations is a clear risk.

Trust management: Current trust management protocols do not account for the diversity to be found in the 5G infrastructure. Given the vertical services (as one dimension to the 5G infrastructure) and slicing between security domains and layers (as another dimension to the 5G infrastructure), trust management protocols must be able to span both dimensions simultaneously. Therefore, liability must be considered as the question of which party (a delegated third party or otherwise) is responsible for which part of the chain in a vertical service. Answering this question will be part of the design of the overall security protocols for the 5G standard, and may lead to unwanted or unsupportable system complexity.

Service provider lock-in: Each tenant/owner of a network slice must be flexible with their services and infrastructure without negatively affecting security SLAs. A tenant/owner may offer a service in one slice of the network while the supporting infrastructure spans multiple domains. If the 5G security protocol is not designed to account for these needs, then a tenant/owner would be locked in to a single domain and unable to fully exploit the 5G standard; therefore, a common standard must be designed with flexibility for migration as a defining feature.

Insufficient TRL: The final version of the security standard will not be fully available during the first phase of deployment (2020). The security requirements of the 5G standard illustrate the insufficient TRLs by exposing new vulnerabilities of the new technologies, which the technologies may not be fully able to mitigate. Designers propose using a “bridge” version of the security standard for the first phase of deployment in 2020 to allow new and non-mature technologies to begin using the 5G standard while adapting and maturing in the time leading up to the final phase of deployment. The “bridge” version may be viewed as a precursor to and primer for the fully deployed security protocol.

Furthermore, security requirements will have to consider which tasks are for which canonical network layer, or which party in the vertical service bears the burden of managing certain functional aspects of the implementation. Additionally, the 5G security protocol must interface with legacy systems. This multi-dimensional problem means new security countermeasures must be designed and standardized.

The multi-dimensional problem of vertical services and the wide range of these services, including health, transportation, and industrial automation applications means the security protocol should be logical instead of physical. This supports solutions to other problems since many network functions will be virtualized in order to support the vertical services while still working within the framework of physical infrastructure to be implemented. For example, unwanted traffic detection could be based upon an intercept-perceive-decide-execute (IPDE) model. This model is a forward-looking method of detecting (intercepting) problems (unwanted traffic) as they occur, perceiving how the unwanted traffic occurs, deciding how to counteract the unwanted traffic, and executing the chosen countermeasure. These functional components of the virtualized network service would necessarily have to be implemented in multiple canonical network layers spread among multiple physical systems.

Prior European security architectures, including TS 23.101 [6], may be modified slightly in order to account for the new security requirement as well as the context of virtualized network functions (VNFs). Likewise, access control adds a level of complexity to determining which provider in the vertical service is responsible for which aspect and level of security. A privacy-by-design approach is required to accommodate greater awareness of privacy concerns among users.

One possible solution for the three main use cases (cloud, mobile, and IoT) may include forms of attribute-based encryption (ABE). ABE extends and generalizes the concepts of public-key encryption, where users have a private (secret) key as well as a public (accessible) key, and private 1:1 communication with the holder of the private key is possible when messages are encrypted with the public key. In ABE, the encryption keys and encrypted messages may be dependent on sets of user-specific attributes, and may be associated with access policies. As a result, data is encrypted via attributes and/or policies related to groups of target users rather than via each user's public key. Thus, messages can only be decrypted by users whose attributes align with the intended requirements, and/or who satisfy the intended policies.

III. V2X: VEHICULAR CONNECTIVITY

Considering the transport vertical services [3], short and long range communications standards will be necessary [3], and they will be required to dovetail with the 5G standard. Transportation services typically are referred to as "Vehicle-to-anything" (V2X) which encompasses the four component services listed in Table 2. Primary use cases in V2X scenarios include activities such as automated driver assistance systems (ADAS), situational awareness, mobility services, and auxiliary services/comfort. Two highly desirable auxiliary service use cases include dynamic route guidance and having municipalities connect to vehicles denoting the locations of available parking, which would provide a mechanism for conserving fuel. Key risks are summarized in [7]-[9]. In the US, Europe, and China multiple projects and testing sites have been leveraged to understand the different foci of the V2X spectrum in different locations.

TABLE 2: 5G VEHICULAR SERVICES

Service	Description
V2V	Vehicle-to-Vehicle
V2I	Vehicle-to-Infrastructure
V2N	Vehicle-to-Network
V2P	Vehicle-to-Pedestrian

Contemporary communications technologies, such as 4G/LTE and dedicated short-range communications (DSRC), have shown promise in V2X applications. DSRC is a two-way, short-range wireless technology that provides high throughput for active safety applications [10] and is based on a conventional implementation of frequency-division multiplexing (FDM). In some respects, 4G/LTE V2X communications may provide operational advantages over DSRC. As new V2X use case appear, become possible, or become desirable, the Society of Automotive Engineers (SAE) J2735 [11] dictionary already has the necessary flexibility to adopt these new use cases. J2735 has a dictionary of at least 16 messages with more than 230 elements, which means LTE adaptation and adoption of LTE V2X is likely because most use cases are already included. Additionally, the connectivity/platform for road operators, certificate and certificate revocation list distribution, range extension, and roadside unit (RSU) backhaul can be done on the LTE network, which provides business value for mobile providers. However, in comparison testing, 4G/LTE has been shown to lack important characteristics for many real-time V2X scenarios. For example, the cellular handoff mechanisms required by 4G/LTE implementations resulted in long lag-times for collision avoidance, and although 4G/LTE has extended range, it is not effective when high throughput and/or point-to-multipoint connections are required [12]. As a result, and even though it may be cost-prohibitive in certain scenarios, DSRC may continue to dominate V2X communications technologies and intelligent transportation systems for near-term applications, as many manufacturers are already implementing DSRC systems in some or all of their vehicles.

While incumbent technologies such as 4G/LTE and DSRC may prove useful in V2X applications, the exploitation of the 5G standard and IEEE 802.11p [13] could solve current and future problems altogether. Unfortunately, IEEE 802.11p has not been updated to account for multiple transmit and receive antennas and other optimizations such as Multiple Input/Multiple Output [MIMO] and beamforming), or advanced modulation and channel access techniques (orthogonal frequency-division multiple-access, or OFDMA), which may become important aspects of V2X technologies in the future. And, again, security issues arise. The Security Credential Management System (SCMS) will have to be designed to account for multiple authorities across several network functions in the virtualized 5G network. The design, for the sake of privacy, will have to be such that no one authority has enough information to track a vehicle for a long period of time. Instead, a lawful intercept (LI) will bring together enough pieces of the total picture of a vehicle's data to track them, and an entity with a LI will never have all of the pieces of the picture. The disparate security requirements of the different services to be provided causes the design of the security protocol to have increased complexity. Furthermore, the security protocol design must account for how strong (or weak) the slicing must be between the different providers in this vertical service.

IV. THE INTERNET OF THINGS (IOT)

IoT is a widespread aspect of the 5G standard. IoT has two main use cases: critical and massive [4]. These use cases have key differences between them. Critical IoT must have low latency and high reliability because it provides connectivity cases such as public safety. Massive IoT requires that devices be inexpensive with multi-year battery lives; low latency and high reliability are desirable if they can be designed into the device, otherwise these features need not be present.

Enterprise applications comprise a third use-case, which will address needs serving vertical services. Typical needs may include personal digital assistants or insurance telematics. The primary market drivers include applications such as connected wearables, cars, homes, cities, and industrial IoT. Vertical requirements will depend upon the operator's perspective, and the operator will have requirements to a greater or lesser degree depending up on the services they provide. Typical functional requirements include traffic patterns, identity/security, simple installation, mobility, SLAs, reliability, sector regulations, analytics, and charging efficiency. To address these requirements, 3GPP Rel.14 [14] was enhanced to improve positioning capabilities, greater multicast downlink transmission, mobility awareness, higher data rates, and packetized voice via voice-over-LTE (VoLTE). These enhancements provide for third party and group-based communications with better support in the radio aspect.

Of special note is the use case regarding private and other networks that intend to use unlicensed or shared spectrum. In most instances, basic capabilities exist in wireless ("WiFi" or IEEE 802.11x) [13] and wired Ethernet [15] to create network partitions. For example, wireless partitions can be created in unlicensed spectrum using the Service Set Identifier (SSID or "network name"), and wired partitions can be created using Virtual LANs (VLANs). Both of these approaches create isolated traffic via a shared infrastructure, which is a foundational capability for 5G networks. However, the overlapping or simultaneous use of licensed and unlicensed wireless spectra can be more complicated.

One promising approach in this regard is the concept of Licensed-Assisted Access (LAA), which is standardized in 3GPP Rel.13 [16] and enhanced (eLAA) in 3GPP Rel.14 [14]. LAA and eLAA provide systems based on 4G/LTE the ability to operate using unlicensed spectrum. Via a combination of techniques, including dynamic channel avoidance and "listen before talk," these hybrid systems can coexist efficiently. MuLTEfire is the tradename for Qualcomm's implementation of LAA/eLAA [17]. MuLTEfire exploits parts of LAA for downlink and eLAA for uplink transmissions. In trials, MuLTEfire has been shown to coexist fairly with WiFi in a fashion which can roughly double overall system throughput. Future releases of MuLTEfire will include IoT-specific enhancements. Private networks using MuLTEfire will have to meet the new security requirements for the disparate services to be provided so that they complete the private tasks necessary to them while operating seamlessly within the new standard, within the unlicensed spectrum, and without degrading the security requirements across disparate domains of providers.

V. USE CASES & PERFORMANCE EVALUATION MODELS

Although highly preliminary, a starting point is necessary for understanding whether or not an aspect of the 5G standard will work. In [5], the authors provide a background setting of how testing was to be conducted, and whether it could be applied to almost all aspects of the 5G standard. The beginning of the roadmap denotes use cases meant to encompass the entire standard, namely: device density, mobility, infrastructure, traffic type, user date rate, latency, reliability, availability, and 5G service type (e.g. machine type communication, or MTC). Key performance indicators (KPIs) are sorted based upon their evaluation method, and those methods are inspection, analysis, or simulation. Furthermore, vertical services will have security requirements and localized needs/requirements. Vertical services have a set of use cases to which these KPIs apply. The use cases are dense urban, broadband everywhere, connected vehicles, future smart offices, low bandwidth IoT, and tactile internet/automation. These use cases are mapped to vertical services use cases, including automotive, eHealth, energy, media and entertainment, and factories of the future. The KPIs and the use cases cover most, if not all, of the needs presented by the 5G standard.

Analysis methods have been developed and have been applied to measure such details as control plane latency ([5], Table 3), user plan latency ([5], Table 4), massive MTC (mMTC) device energy consumption improvement ([5], Table 5), inter-system handover, interruption time, mobility interruption time, and peak data rate. Although these measurements and calculations are simple to complete, they provide benchmarks regarding device performance with respect to the new network.

These benchmarks need to be measured in the different contexts of the use cases even though not all use cases occur in all contexts. A context is a specific configuration for a BS, and contexts being considered for the 5G standard include indoor hotspot, urban macro, outdoor small cells, and rural macro/long distance configurations.

VI. CONCLUSION

This paper set out to explore several aspects of the 5G cellular standard with respect to security issues as the focus. The paper explores the general security protocol design as written by 5GPPP; the V2X communications standard results from research completed by the 3GPP; the IoT results from research completed by the 5GPPP; and the general design of 5G cellular standard with respect to the use cases and how to measure, via KPIs, when the use cases were being met. Security is a common thread among the use cases as well as the vertical services to be used by the 5G cellular standard. Security concerns are noted in each of the aspects. The general security protocol design written by the 5GPPP provides an introduction to the issue itself. The V2X and IoT aspects highlight how the general security protocol could or does impact implementation in these specific vertical services. Both V2X and IoT aspects will result in enormous numbers of additional network nodes, each of which presents numerous threat vectors. Additionally, Network as a Service (NaaS) or slicing is one approach to reconciling competing priorities. However, slicing produces a host of additional issues related to virtualization, automation, and guarantees of isolation. Whether the discussion is about network and infrastructure or vertical services, security is a concern affecting both the vertical services and use case dimensions at all levels, and security is a concern that arises even when security is not the specific focus.

Acknowledgment

The lead author would like to acknowledge the contribution of Dr. George Koutitas, Texas State University.

References

[1] 5GPPP (2017 June). "5G PPP Security Landscape." [Online] https://5g-ppp.eu/wp-content/uploads/2014/02/5G-PPP_White-Paper_Phase-1-Security-Landscape_June-2017.pdf [accessed Sept. 2017]

[2] 5G Americas (2017 Oct.). "V2X Cellular Solutions." [Online] http://www.5gamericas.org/files/2914/7769/1296/5GA_V2X_Report_FINAL_for_upload.pdf [accessed Sept. 2017]

[3] 5G Americas (2017 Dec.). "LTE Progress leading to the 5G Massive Internet of Things." [Online] http://www.5gamericas.org/files/8415/1250/0673/LTE_Progress_Leading_to_the_5G_Massive_Internet_of_Things_Final_12.5.pdf [accessed Sept. 2017]

[4] Global Mobile Suppliers Assoc. "5G Network Slicing for Vertical Industries." [Online]. https://gsacom.com/paper/5g-network-slicing-vertical-industries/ [accessed Sept. 2017]

[5] 5GPPP (2016 April). "5G PPP Use Cases and Performance Evaluation Models." [Online] https://5g-ppp.edu/wp-contents/uploads/2014/02/5G-PPP-use-cases-and-performance-evaluation-modeling_v1.0.pdf [accesed Sept. 2017]

[6] 3GPP TS 23.101, "General Universal Mobile Telecommunications System (UMTS) architecture" [Online] http://www.3gpp.org/dynareport/23-series.htm [accessed Feb. 2018]

[7] National Transportation Library (2016 Sept.). "Connected Vehicle Pilot Deployment Program Phase 1 – Deployment Readiness Summary – New York City." [Online] https://rosap.ntl.bts.gov/view/dot/31731 [accessed Sept. 2017]

[8] National Transportation Library (2016 Sept.). "Connected Vehicle Pilot Deployment Program Phase 1 – Deployment Readiness Summary – Tampa (THEA)." [Online] https://rosap.ntl.bts.gov/view/dot/31721 [accessed Sept. 2017]

[9] National Transportation Library (2016 Sept.). "Connected Vehicle Pilot Deployment Program Phase 1 – Deployment Readiness Summary – ICF/Wyoming." [Online] https://rosap.ntl.bts.gov/view/dot/37124 [accessed Sept. 2017]

[10] S. Sill. "DSRC: The future of safe driving." Intelligent Transportation Systems Joint Program Office. [Online] https://www.its.dot.gov/factsheets/dsrc_factsheet.htm [accessed Feb. 2018]

[11] SAE J2735. DSRC Message Set Dictionary. [Online] https://www.sae.org/standards/content/j2735_200911/ [accessed Feb. 2018]

[12] Z. Xu et.al. "DSRC versus 4G-LTE for Connected Vehicle Applications: A Study on Field Experiments of Vehicular Communication Performance," Journal of Advanced Transportation, Volume 2017 (2017), Article ID 2750452, Aug. 2017. [Online] https://doi.org/10.1155/2017/2750452 [accessed Feb. 2018]

[13] IEEE 802.1, Wireless Local Area Networks, The Working Group for WLAN Standards [Online]. http://www.ieee802.org/11/ [accessed Feb. 2018]

[14] 3GPP Release 14. [Online]. http://www.3gpp.org/release-14 [accessed Feb. 2018]

[15] IEEE 802.3, Ethernet Working Group [Online]. http://www.ieee802.org/3/ [accessed Feb. 2018]

[16] 3GPP Release 13. [Online]. http://www.3gpp.org/release-13 [accessed Feb. 2018]

[17] D. Malladi, "Best Use of Unlicensed Spectrum," Qualcomm, Feb. 3, 2016. [accessed Feb. 2018]

Centrality Based Routing Protocol with Message Relay Control in Opportunistic Networks

Abineza Claudia
Computer Science
University of Rwanda
Kigali, Rwanda
e-mail: abineza1@gmail.com

Santhi Kumaran
African Center of Excellence in IoT
University of Rwanda
Kigali, Rwanda
e-mail: santhikr69@gmail.com

Chomora Mikeka
School of Applied Sciences
Chancellor College, UNIMA
Malawi
e-mail:chomora@gmail.com

Abstract— **Opportunistic Networks are a subclass of Delay Tolerant Networks (DTN), which aim at wireless data delivery in severely partitioned networks. There exist several protocols that route messages on a best effort basis. In most cases, the nodes copy and forward messages to nodes that are more likely to meet the destination. But, the major challenge is to design a routing protocol that offers the best tradeoff between cost (number of message replicas) and rate of successful message delivery. In this research paper, the tradeoff is being efficiently handled by using the concept of Google Pagerank like centrality to rank nodes in a network using social information. Unlike other nodes in the network, central nodes act as influential nodes to facilitate the message forwarding. Furthermore, to the centrality routing, a mechanism of message relay control is designed and linked to keep the network overhead ratio low. The proposed Centrality Based Routing Protocol (CBRP) with Message Relay Control algorithm was evaluated by simulations using the Opportunistic Network Environment (ONE) simulator. The results show that CBRP outperforms other typical routing protocols in Opportunistic Networks.**

Keywords- Routing protocol; Centrality; Opportunistic networks; Overhead; message delivery.

I. INTRODUCTION

In recent years, incoming of smartphones and advent of wireless technologies make a seamless and cheaper communication between wireless devices anytime and anywhere. In this setting, Opportunistic Networks (OppNets) are considered as specialized ad hoc networks characterized by frequently intermittent connections, which operate without any assistance to any infrastructure, such as Access points, Routers etc. Communications in this type of network is made possible by mobile self-configurable devices, with no infrastructure assistance feature, exploiting direct contacts among nodes with a message Store - Carry and Forward way and incentive way to guarantee information exchange, as some nodes in a network tend to refuse to share their private resources, such as buffer space. Summarily network topology is not known a-priori, at the message sending. Therefore, routing protocols in such environment rely much on network assumptions, such as mobility patterns, node capacity, scheduling knowledge, estimation and on prediction on the likelihood of future network topology [1].

Centrality is one way, among other routing protocol metrics used to forward the message in social opportunistic network. As the name expresses, centrality relates to action to identify central nodes in a network. Therefore, centrality definition should derive from various means, including social criteria. Due to the dynamics of node mobility, the influence that a node may have over the spread of information in relation to how many other nodes (encounters) this node may have been in contact with, has a significant implication in defining a centrality metric, in this work.

Based on this implication, each node in a network is assigned a centrality value using Google Pagerank like algorithm. In the proposed Centrality Based Routing Protocol (CBRP), the nodes with highest centrality values are more likely to act as the best message forwarders. The algorithm is simulated using the ONE simulator.

The rest of the paper is organized as follows: Section 2 describes a summary of the related works. Section 3 gives explanation on assumptions made on the CBRP. Section 4 deals with the design of algorithm and parameter metrics used for the evaluation of the proposed protocol. Section 5 concerns the analysis of the simulation results. Finally in Section 6, conclusion is made and the future works are recommended.

II. REVIEW OF THE RELATED WORKS

A common characteristic of routing protocols in opportunistic network is that they are replication-based, as the network topology is intermittent and not known a priori at the message sending. Consequently, the efficiency of any protocol relies much on what extent the protocol restricts message replication while maximizing a message delivery guarantee. Furthermore, most routing protocols are context-aware routing protocols, where the knowledge of the context in which nodes operate is used to identify the best next hop of a given node. Context aware based routing protocols are in turn, classified into mobility-based routing protocols, and social context-based routing schemes as the most of mobile devices are carried by humans. Furthermore, various social-based routing protocols have been proposed [2][3],

ISBN: 978-1-61208-623-1

exploiting various social characteristics, such as community and centrality.

From the replication-based to context-aware based routing protocols, various routing schemes in OppNets evolved. Initially, Vahdat and Becker [4] proposed the epidemic routing protocol, a totally replicating and flooding-based, as nodes continuously replicate and transmit messages to newly discovered contacts to eventually reach the destination. It follows the variations of epidemic routing, such as Spray and Wait (SnW) routing protocol proposed by T. Spyropoulos, K. Psounis and C. S. Raghavendra [5] to impose the limit on the number of possible replications of a message and to maximize the aggregate resource consumption (for example, bandwidth and energy) in the network. In the latter, a particular message is spread to at most *L* different relay nodes. The nodes then perform a direct delivery when they come in contact with the corresponding destination of the message. In [6], A. Lindgren, A. Doria E. Davis and S. Grasic proposed the Probabilistic Routing Protocol using History of Encounters and Transitivity (PRoPHET) by ranking nodes encounters with a set of probabilities as the greater chance of encountering the destination. A flooding-based MaxProp [7] imposes the priority on a message by maintaining an ordered-queue based on the destination of each message and on the estimated likelihood of a future transitive path to that destination. In [8] the Resource Allocation Protocol for Intentional DTN routing (RAPID) is proposed, by exchanging the expected contact time with other nodes, list of messages delivered, and average size of past transfer events are exchanged.

Few social based forwarding that exploit the interplay between the structural properties of social networks and mobility aspects are pointed out: SimBet [9] that uses social network properties, such as betweenness centrality and social similarity to inform the routing strategy and (BUBBLE Rap) [10] that targets nodes with high centrality as well as members of the communities, yielding delivery ratios similar to flooding approaches with lower resource utilization.

It has been an age since various researches on network centralities in different domains, such as in sociology, biology, physics, applied mathematics and computer science. The computations of centrality in a DTN social network forwarding, the idea is related [11]-[16]. For example, when calculating the betweeness and closeness centralities of all the vertices in a graph involves calculating the shortest paths between all pairs of vertices. Therefore, many algorithms evolved to calculate the betweenness centrality, including Floyd-Warshall algorithm [17], and Baoqiang 's algorithm [18]. However, when investigating into the above centrality computing algorithms are centralized and rely on global information of the network as they rely on the knowledge of the network size.

Recently, distributed algorithms have been proposed in [19]-[21] for computing the betweenness and closeness centralities and other centrality measures are proposed in [22][23], to adapt the algorithm to dynamic characteristics of mobile wireless network. Different approaches were adopted for computing Pagerank like centrality in mobile wireless environments, namely eigenvector centralities [24] and Peoplerank algorithm [25]. The latter, computes centrality inspired from Google Pagerank, both in a centralized and distributed way. Motivated by the above-mentioned works, this work proposes iterative algorithm to compute the distributed centrality adapted to Google Pagerank concept and opportunistic mobile network dynamics.

In the proposed CBRP algorithm, every node computes and evaluates its own centrality by using local interactions with only its current encounter without knowing the network size, and the network topology. Consequently, this fits well the opportunistic networks characteristics where network topology is frequently intermittent and change, especially when the network size becomes larger; it is usually very difficult to compute centrality measures. In addition to that, the CBRP inspired by Peoplerank algorithm, takes advantage of the fact that, a time- varying social graph, is iteratively built to reflect the dynamic of the opportunistic network, by the inference of social nodes from a node's encounters. The assumptions made from Peoplerank algorithm, as the interpretation towards this assumption inference will be given in the third Section. Assuming that only neighbors in the social graph have an impact of the popularity (i.e., the ranking), as the nodes meet, the node's centrality is updated and the number of neighbors is incremented by one to reflect the impact of a new social node. Consequently, the same idea in PeopleRank, is applied to tag people as "important" when they are linked (in a social context) to many other "important" people. The main concept originated from Google's Pagerank [26].

III. EXPLANATION OF THE ASSUMPTIONS MADE FROM PEOPLERANK ALGORITHM

PeopleRank is a social distributed routing algorithm measuring opportunistically the importance of a node in a social graph based on the social interaction between nodes and their contact frequencies using real human mobility. In other words, it tried to determine the optimal forwarding paths given the mobility patterns and their connectivity properties, to compute for the success rate as the delivery probability.

When investing the Peoplerank algorithm, 3 observations are noteworthy:

Firstly, the impact of the damping factor on the Peoplerank. It is used to control the amount of randomness forwarding in Peoplerank. Its value can be chosen and well adapted to social forwarding, according to whether the stated social relation is implicitly or explicitly declared, even though some randomized forwarding might be a little beneficial. As one of Peoplerank observation, in the 2 previous situations, the optimal value of d is around 0.87 and 0.8. Consequently, an assumption was made of using the damping factor of 0.86 to reflect the application of CBRP in a likely high social interaction and connection environment, as in the closed campus, where social rate is implicitly high.

Although a shared common interest is not an optimal social property to rely on, when selecting a best message forwarding node, social based on being in a geographic location, as it was stated by Peoplerank ("Geographic location helps user to socialize more often and meet with each other more frequently") and on implicit or explicit friendship are optimal, the CBRP will be then applied into a closed environment, such as in a campus.

Secondly, Peoplerank is compared to the following social based algorithms: Centrality, that forwards a message from u to v if, and only if, C(u) = C(v). Here, C(u) denotes the betweenness centrality of node u measured as the occurrence of this node in all shortest paths connecting all other pairs of nodes and degree that forwards a message from u to v if, and only if, d(u) = d(v). Here, d(u) denotes the degree of node u in the social graph (in a friendship graph, the degree is the number of friends of node u). Both Peoplerank and Centrality achieve a comparable result while they outperform the degree based, with a comparable success rate of a flooding-based Epidemic routing. Furthermore, Peoplerank is much preferred over the centrality-based, as the latter requires centralized computation, which is more complex to compute.

Therefore, as a higher impact and factor of the meeting event (to the Peoplerank performance) has been evaluated better (rather than above-mentioned social criteria) to improve the social patterns and node position in a social graph. An assumption is made to infer social nodes from the meeting event. This validates well the high implication of a meeting event into a social, as Peoplerank centrality update and increment as the nodes meet. Consequently, the distributed CBRP algorithm has been developed, by assuming that when the nodes meet, the node's centrality is updated and the number of neighbors (inferred encounters) is incremented by one to reflect the impact of a new social node.

Thirdly, when Peoplerank is compared to well known contact-based algorithms (namely Last Contact, Destination Last Contact, Frequency, Spray & Wait and Wait-destination), it outperforms them. This is due to the social aspect of the algorithm that delivers the messages with higher probability to the destination. This validates the importance of a dynamic and distributive social to the selection of message forwarding.

Finally, the message forwarding of CBRP does not rely on network global structure, as it does not require the known size of the network, as it is the case in Peoplerank.

IV. ANALYSIS AND DESIGN OF THE PROPOSED ALGORITHM

To design the CBRP protocol, a model of an imaginary social graph is adopted, where a node itself in a network forms a graph vertex, and its predecessors and successors are its encounters. Extracting from meeting event, implicit social node attributes, such as being in a geographic location, being friends, the protocol aggregates this imaginary node's social attributes into a contact graph. Therefore, the CBRP protocol computes for popularity of each node in a network, based on number of its encounters, inferred social nodes.

The CBRP protocol works as follows: When two nodes meet, one is considered as forming an incoming link to another node, and the nodes encounters as forming outbound links on the involved nodes. Then, the meeting nodes calculate and update their tables: Encounters table whose current number of encounters is increased by one and their centrality table get updated with newly calculated Pagerank like centrality. The Google Pagerank like centrality is calculated according to the following equation (delivered from Google PageRank):

$$C(i) = (1-d) + d(C(j)/T(j)) \qquad (1)$$

where C(i) is centrality of current node, C(j) Centrality of encountered node and Tj is a number of encounters of node j and d which depend on how much the social relationship between nodes can help improving their centrality values.

It is clearly noted that Centrality in the network is calculated dynamically in time and in space, using a distributed algorithm, for which the total number of nodes are not initially known, therefore, much suitable to the dynamic wireless mobile environment. For simulation purposes, d has been set to 0.86.

More importantly, a message is delivered from node i to node j, if the centrality value of j is greater than or equal to that of i or j is the destination node.

Finally, each node maintains an acknowledgement table in the form <Message ID, Source ID, Destination ID>, that contains information on message delivered to destination and that should be flooded among nodes in network. In fact, when two nodes meet, they should check for any new acknowledged messages in acknowledgement table of the encountered node, then update their buffer by removing a copy of it and update their acknowledgement table to spread the update information to other nodes in network.

A. Parameter metrics of the Algorithm

The performance metrics used to evaluate the developed protocols are:

• Delivery probability is defined as the number of successfully delivered messages divided by the number of created messages

$$\mathit{Delivery\ ratio} = \frac{\text{Number of Packets received}}{\text{Number of Packets sent}} \qquad (2)$$

• Overhead ratio: This is a metric used to estimate the extra number of packets needed by the routing protocol for actual delivery of data packets. It can be defined as:

$$\mathit{Overhead\ ratio} = \frac{\text{Nr. of Packets relayed} - \text{Nr. of Packets delivered}}{\text{Nr. of Packets delivered}} \qquad (3)$$

B. Algorithm for the CBRPprotocol

Notations:

- i: Current node
- j: Encounter node
- E(i): Number of encounters of current node i
- C_i: Centrality value of node i
- Buffer(i): The buffer at node I
- M: Message currently being sent
- DN: Destination node
- Ack_table: Acknowledgement table
- Ack_M: Acknowledgement message

1) Algorithm1

```
Step 1: select the next Encounter node j
Step2: If j is busy then go to Step1
Step3: Repeat all messages (M) of current node i
        3a:  If j has M then go to step 3 to select the next
             Message
        3b:  check Ack_table of j for M
             If Ack_Table of j has M then
                 Remove M from buffer of i
                 Update Ack_table of i
                 Go to step 3 for next M
             End if
        3c:  If Cj >=Ci or j is equal to DN then
                 Forward M to j
             End if
```

Algorithm related to Acknowledgement:

2) Algorithm2

When a destination node is receiving a message:

```
Step1: Receive message (M) from the Last Sender
       Node(LSN)
Step2: If Destination Node(DN) of M is current
       node i  then
           Create and Send Ack_M to LSN
           Update Ack_Table of i with Ack_M
           Remove M from the buffer of  i
       End if
```

3) Algorithm3

When the last sender is receiving the acknowledgement:

```
Step 1:Receive Message (M) from the Destina-
             tion Node (DN)
Step 2: If M contains ACK_M then
             Update Ack_Table of i with ACK_M
             Remove M from the buffer of  i
        End if
```

C. Flowchart for the CBRP protocol

The flowchart is given in Figure 1.

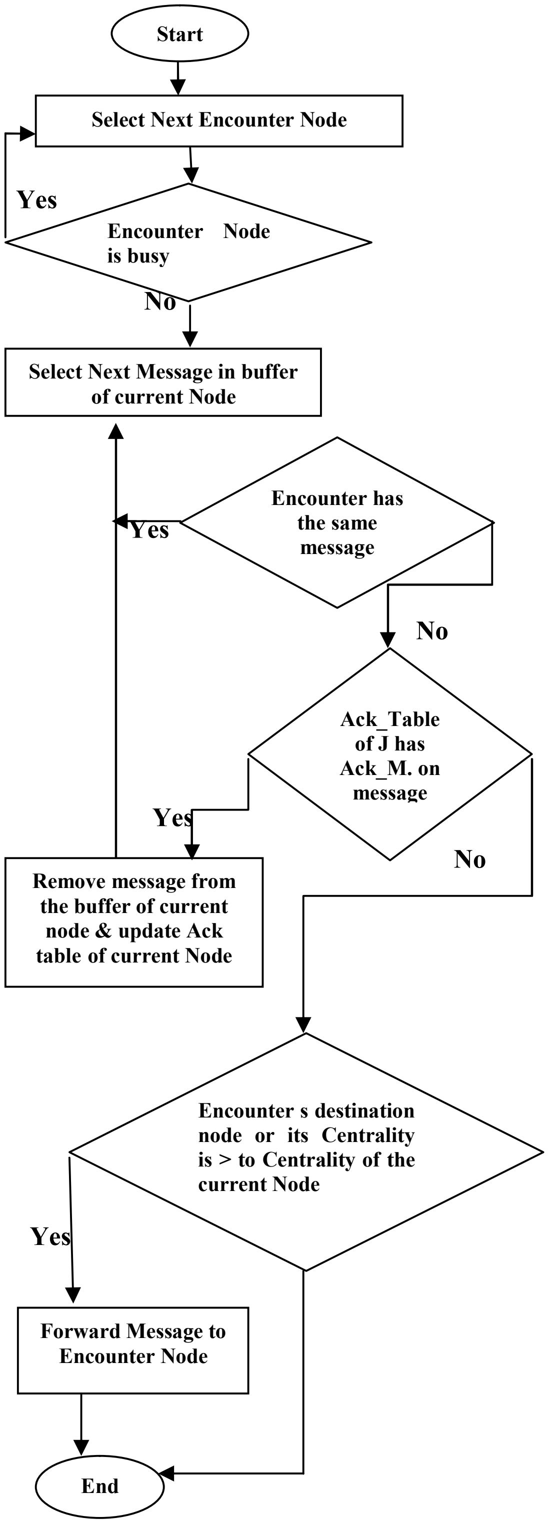

Start
Select Next Encounter Node
Yes
Encounter Node is busy
No
Select Next Message in buffer of current Node
Yes
Encounter has the same message
No
Ack_Table of J has Ack_M. on message
Yes
No
Remove message from the buffer of current node & update Ack table of current Node
Encounter s destination node or its Centrality is > to Centrality of the current Node
Yes
Forward Message to Encounter Node
End

Figure 1. Flowchart for the CBRP protocol.

V. SIMULATION RESULTS AND ANALYSIS

The proposed protocol CBRP is evaluated by Random Scenario simulations and compared against the standard routing algorithms including: Epidemic, due to its potentially high message delivery, PRoPHET due to its probabilistic routing and MaxProp due to its predictability routing with acknowledgement. The simulations focused on the performance metrics: Delivery Probability and Overhead Ratio. The simulator used is the ONE simulator version 1.6.

A. Simulation setup

The simulation parameters used are shown in TABLE I.

TABLE I. SIMULATION PARAMETERS

Simulation parameter	**Value**
Routing Protocol	CBRP, Epidemic, Prophet, Maxprop
Number of nodes groups	1
Number of nodes	Variable(36,72,100,130, 170)
Mobility Model	RandomWaypoint
Simulation Time	43200 secs
Simulation Area	(4500X3500)m
Time-To-Live(TTL)	Variable(100,150,200,250,300) minutes
Node speed	0.5-1.5(meters/sec)
Scenario.updateInterval	0.1
Interface type	Bluetooth/Simple broadcast interface
Transmit speed	2Mbps(250kBps)
Interface transmit range	10 meters
Size of the message buffer	5MB
seed for movement models	1
Number of event generator	1
Class of event generator	MessageEventGenerator
Creation interval of event	A new message every 25-35 seconds
Message size	500KB-1MB

B. Simulation running

Figure 2 shows the simulation environment where, as the nodes are moving, messages created, relayed, dropped and delivered are calculated to generate the report file at the end of the simulation. The latter contains the standard results, such as the Message Delivery Ratio, Average Latency, the network overhead, the Average Number of Duplicate Messages, and many other network statistics, used to produce the following plots:

-Plots of message delivery ratio as a function of number of nodes

-Plots of message delivery ratio as a function of message TTL

-Plots of network overhead ratio as a function of number of nodes

-Plots of network overhead ratio as a function of message TTL

Each simulation includes 20 scenarios, run for once and under the same values for parameters, to be able to compare 4 routing protocols with five values for a variable TTL/ number of nodes. This validates and maintains the performance assessment.

C. Simulation analysis

1) Analyzing the delivery ratio

Although flooding is controlled, which normally is the technique to achieve a higher delivery ratio, CBRP is comparable to flooded-based Epidemic and Maxprop, whereas it outperforms a probabilistic Prophet. This is attributed to the fact that CBRP is likely to produce a better forwarding path to deliver the messages to the destination; as the path is formed with nodes that are likely characterized by high centrality values compared to nodes that are not part of the routing paths. Therefore, CBRP has a better metric to select a relay node that is likely to meet the destination, than Prophet routing. Additionally, for the 4 routing protocols compared, as TTL increases, the delivery ratio increases. This is attributed to the fact that when TTL increases, the message remains in the buffer for a longer period of time, leading to a higher chance to meet the destination.

2) Analyzing the Overhead ratio

For the 4 protocols, it is observed that when TTL increases, the overhead ratio decreases. This is due to the fact that the message remains for a long period in the buffer. This way, messages are not dropped on the way to their destination and consequently no need to be replicated which results to a low overhead ratio. Therefore, the lesser relay messages, the better the overhead. It is observed that under varying number of nodes and varying TTL, the CBRP outperforms Prophet, while the flooded Epidemic and Maxprop achieve higher overhead ratio.

The CBRP is compared against 4 aforementioned protocols. The results are depicted in Figures 2, 3, 4 and 5.

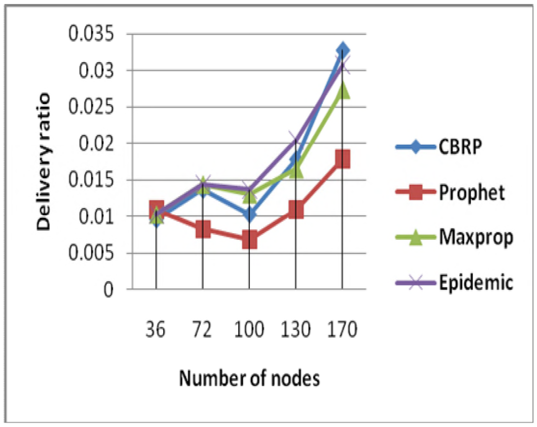


Figure 2. Comparison of CBRP, Epidemic, MaxProp and Prophet for varying number of nodes.

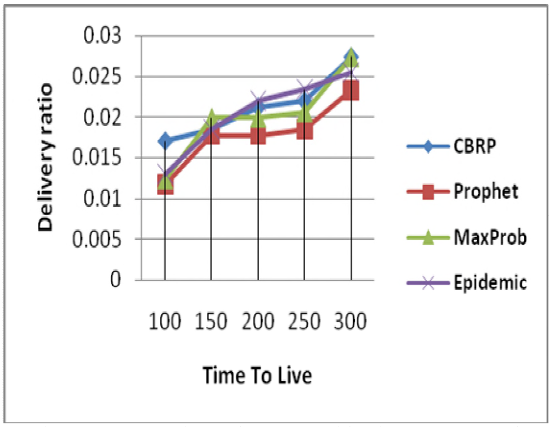


Figure 3. Comparison of CBRP, Epidemic, MaxProp and Prophet for varying message TTL.

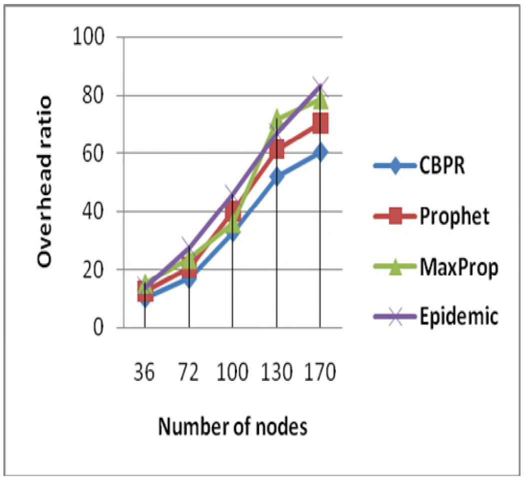


Figure 4. Comparison of CBRP, Epidemic, MaxProp and Prophet for varying number of nodes.

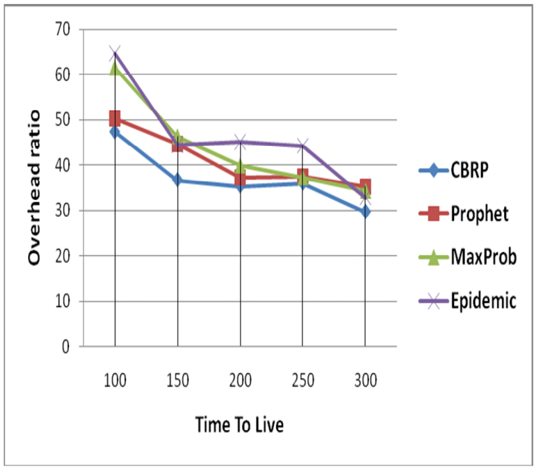


Figure 5. Comparison of CBRP, Epidemic, MaxProp and Prophet for varying message TTL.

VI. CONCLUSION AND FUTURE WORK

In this paper, a CBRP protocol for OppNets has been developed which uses the Centrality concept for the message forwarding process. Simulation results on the performance of CBRP in comparison with Epidemic, Prophet and MaxProp under the same experimental conditions have revealed improvement that cannot be ignored, both in terms of a message delivery ratio and overhead ratio. Therefore, a developed model could be a choice to follow for a message forwarding in opportunistic networks. However, as for recommendation, two things are mentioned. On one hand, the tests performed here were limited to the random scenario simulation as the Randomwaypoint movement model of nodes was used. Due to this, it can be recommended to test the developed protocol under human scenarios and other recurring pattern based structures to explicitly show how well CBRP routes message under realistic mobility scenarios. On other hand, the developed CBRP protocol can be enhanced with new capabilities that deserve attention in a network routing, such as security and privacy.

REFERENCES

[1] Chander Prabha, Surender Kumar and Ravinder Khanna, “Analysis of routing and forwarding protocols in opportunistic networks”, Procedia Computer Science, Vol. 85, pp. 891-898, 2016, doi.org/10.1016/j.procs.2016.05.279.

[2] Tie Qiu, Arun Kumar Baochao Chen and Runhe Huang, “A survey of mobile social networks: applications, social characteristics, and challenges,” IEEE Systems Journal, pp. 1-16, Nov. 2017, doi: 10.1109/JSYST.2017.2764479.

[3] Radosław O. Schoeneich and Rafał Surgiewicz, “SocialRouting: The social-based routing algorithm for delay tolerant networks”, International Journal of Electronics and Telecommunications, Vol. 62, no. 2, pp. 167–172, June 2016, doi: 10.1515/eletel-2016-0023.

[4] Amin Vahdat and David Becker, “Epidemic routing for partially connected ad hoc networks”, Technical Report CS-2000-06, Dept. of Computer Science, Duke University, Durham, NC 27708, 2000.

[5] Thrasyvoulos Spyropoulos, Konstantinos Psounis and Cauligi S. Raghavendra, “Spray and wait: An efficient routing scheme for intermittently connected mobile networks”, Proc. ACM SIGCOMM Workshop on Delay-Tolerant Networking (WDTN ’05), Philadelphia, PA, USA, Aug. 2005, pp. 252–259, doi: 10.1145/1080139.1080143.

[6] Anders Lindgren , Avri Doria, Elwyn Davies and Samo Grasic, “Prophet routing protocol for intermittently connected networks,” IETF RFC 6683, August 2012, ISSN 2070-1721.

[7] John Burgess, BrianGallagher and David Jensen, “”Maxprop: routing for vehicle-based disruption- tolerant networks,” the 25th International Conference on Computer Communications, IEEE INFOCOM, April 2006, pp.1-11, doi:10.1109/infocom.2006.228

[8] Aruna Balasubramanian, Brian Neil Levine and Arun Venkataramani, “DTN routing as a resource allocation problem,” SIGCOMM, Aug 2007, doi: 10.1145/1282380.1282422.

[9] Elizabeth Daly and Mads Haahr, “Social network analysis for routing in disconnected delay- tolerant MANETs,” Proc. Symp. ACM international on Mobile ad hoc networking and computing (MobiHoc '07), USA, Sept. 2007, pp. 32-40, ISBN: 978-1-59593-684-4, doi:10.1145/1288107.1288113.

[10] Pan Hui, Jon Crowcroft and Eiko Yoneki, “ Bubble rap: social- based forwarding in delay- tolerant networks,” IEEE Transactions on Mobile Computing, Vol. 10, No. 11, Nov. 2011, pp. 1576-1589, doi:10.1109/TMC.2010.246

[11] Santiago Segarra and Alejandro Ribeiro, “Stability and continuity of centrality measures in Weighted Graphs,” IEEE Transactions on Signal Processing , vol. 64, no. 3, Feb. 2016, pp. 543-555, doi: 10.1109/TSP.2015.2486740.

[12] Deepak Kumar Sharma, Amit Gupta, Ashray Anand and Sanjay K. Dhurandher , “Centrality based congestion controlled routing protocol for social opportunistic networks,” 1st India International Conference on Information Processing (IICIP),Aug. 2016, doi: 10.1109/IICIP.2016.7975388.

[13] Pavlos Nikolopoulos, Therapon Papadimitriou, Panagiotis Pantazopoulos and Ioannis Stavrakakis “How much off-center are centrality metrics for routing in opportunistic networks”, ACM MobiCom 2011 Workshop on Challenged Networks (CHANTS'11), Las Vegas, USA, Vol. 6, October 2011, doi: 10.1145/2030652.2030657.

[14] Paolo Boldi and Sebastiano Vigna, “Axioms for centrality,” Internet Mathematics, vol. 10, No.3-4, 2014, pp. 222-262, https://doi.org/10.1080/15427951.2013.865686.

[15] Zhenxiang Gao1, Yan Shi1, and Shanzhi Chen, “Identifying influential nodes for efficient routing in opportunistic networks,” Journal of Communications, Vol. 10, No. 1, pp. 1-53, Jan. 2015, doi: 10.12720/jcm.10.1.48-54.

[16] Muhammad Arshad Islam, Muhammad, Azhar Iqbal Zahid Halim and Muhammad Aleem,“Analysing Connectivity Patterns and Centrality Metrics for Opportunistic Networks,” Proc. International Conference on Communication, Computing and Digital Systems (C-CODE), pp. 64-70, 2017, doi.org/10.1109/c-code.2017.7918903.

[17] Weisstein, Eric W. "Floyd-Warshall Algorithm."[Online]. Available from: http://mathworld.wolfram.com/Floyd-WarshallAlgorithm.html, 2008.

[18] Baoqiang Li, Guangya Si , Jianfei Ding and Fei Wang, “A faster algorithm to calculate centrality based on Shortest Path Layer,” IEEE Control And Decision Conference(CCDC), Chongqing, China , May 2017,pp. 6283 - 6290 ISSN: 1948-9447 doi: 10.1109/CCDC.2017.7978302.

[19] Wei Wangu and Choon Yik Tang ,“Distributed computation of node and edge betweenness on tree graphs,” The 52nd Annual Conference on Decision and Control (CDC), 2013 , IEEE, 2013, pp. 43–48.

[20] Wei Wangu and Choon Yik Tang, “Distributed estimation of closeness centrality, “Proc. IEEE Conference on Decision and Control,” Osaka, Japan, 2015, pp. 4860–4865.

[21] Wei Wangu and Choon Yik Tang, “Distributed computation of classic and exponential closeness on tree graphs,” Proc. American Control Conference, Portland, OR ,2014, pp. 2090–2095.

[22] Keyou You, Roberto Tempo, and Li Qiu, “Distributed algorithms for computation of centrality measures in complex networks,” IEEE Transactions on Automatic Control, vol. 62, no.5, Aug. 2016, pp. 2080–2094, doi: 10.1109/TAC.2016.2604373.

[23] Michel Raynal and Franck Petit, “Special issue on distributed computing and networking,” Theoretical Computer Science, vol. 561, pp. 87–144, 2015.

[24] Robert John D.Souza and Johny Jose, “Significance of Eigenvector Centrality for Routing in a Delay Tolerant Network,” Journal of Computations & Modelling, vol.1, no.1, 2011, pp. 91-100, ISSN: 1792-8850.

[25] Abderrahmen Mtibaa, Christophe Diot, Martin May and Mostafa Ammar, “PeopleRank: social opportunistic forwarding,” conference Paper, Proc. IEEE INFOCOM , March 2010, doi: 10.1109/INFCOM.2010.5462261.

[26] Vince Grolmusz: “A note on the pageRank of undirected graphs,” Information Processing Letters, Vol. 115, No. 6-8, pp. 633-634, June 2015, doi: 10.1016/j.ipl.2015.02.015.

Adaptive Detection of Transients by the Complex Cepstrum of Higher Order Statistics Based on Givens Rotations

Christos K. Papadopoulos, George Ch. Ioannidis,
Constantinos S. Psomopoulos
Piraeus University of Applied Sciences, Dept. of Electrical Engineering, Egaleo, Greece
e-mails: {cp26041960, gioan, cpsomop}@puas.gr

Konstantinos Ch. Papadopoulos
National Technical University of Athens
School of Mechanical Engineering
Athens, Greece
e-mail: mc16066@central.ntua.gr

Abstract— **In this paper, the problem of detecting transient signals of unknown waveforms and arrival times embedded in white Gaussian noise is addressed. The use of the cepstrum coefficients of the 4th order correlations of the transient signal for forming a detection statistic is demonstrated. It is considered an adaptive approach for the detection of the signal which is assumed to satisfy a linear constant coefficient difference equation. The adaptive approach is a least squares realization based on Q-R decomposition of the 4th order statistics matrix involved in the computation of the cepstrum coefficients. It is shown that the adaptive approach allows for detection of short length transients which are of unknown arrival times using a single data record even before the whole amount of data becomes available.**

Keywords-Detection; Transient signals; Complex Cepstrum; Q-R decomposition; Givens Rotations.

I. Introduction

Detection of transient signals of unknown waveforms and unknown arrival times is a common problem in several signal processing areas. Some applications include detecting targets by radar and sonar. Another application is in hydraulic and power systems where monitoring sudden changes protects the system. In the detection of seismic waves and in biomedicine, the signal carries important information of the disease and an early detection is essential for the treatment. Transients can be either deterministic or stochastic signals, are short in duration, and are embedded in long periods of background noise. In both cases we have a highly non-stationary problem. Classical signal detection theory has been applied to this problem mainly using the autocorrelation or data domain.

If the deterministic signal waveform is unknown, but the arrival time is known, and the signal is embedded in additive white Gaussian noise, a generalized likelihood ratio test is discussed in [1], where the signal is the impulse response of a proper rational transfer function. However, some a-priori knowledge for the signal is required. Furthermore, the detector is not of Constant False Alarm Rate (CFAR). A similar approach is presented in [2] where the noise is colored Gaussian Autoregressive of order M (AR(M)) process. For the same transient problem, but for unknown arrival times, the Gabor representation of the signals is used in [3].

Higher order statistics have been used for spectrum estimation of stochastic signals [4]-[11]. For detection problems, their use has not been very extensive.

Here we propose a new detection scheme for the detection of transient signals based on the computed cepstrum coefficients of the fourth-order statistics of the signal [12]. Cepstrum coefficients are appropriate for representation of transient signals because they contain all the information of the signal. Since they also peak around the origin, they are suitable for signal detection. The proposed method does not require knowledge of the noise variance or skewness and it is also able to detect the signal in the presence of non-Gaussian white noise as long as it is of zero mean independent, identically distributed (i.i.d.)

Higher order (3rd, 4th, etc.) cumulants are zero for Gaussian i.i.d. process [12]. This means that cumulants have the ability to suppress the noise. This fact is one of the reasons that we present herewith a detection statistic based on cepstrum coefficients and particularly the ones based on the tricepstrum sequence. However, the same detection statistic still works when noise is not Gaussian i.i.d. but non-skewed (e.g. symmetrically distributed).

The paper is organized as follows. In Section II, the adaptive approach is presented for the proposed detector. In Section III, its performance is demonstrated by means of simulation examples. Finally, conclusions are drawn in Section IV.

II. Adaptive Q-R Decomposition of the Trispectrum Cepstral Equation

A. Problem Definition

The following detection problem is considered

$$\begin{aligned} &H_0: x(n) = w(n) \\ &H_1: x(n) = m(n) + w(n)\ n = 0, \dots, N-1 \end{aligned} \qquad (1)$$

where $w(n)$ is a stationary, zero mean, white, Gaussian noise of unknown variance σ_w^2 and $m(n)$ is a deterministic transient signal of unknown waveform. The complex cepstrum of the 4th order statistics of a random process $\{x(n)\}$ is known to satisfy the following identity [12],

$$\sum_{k=1}^{p} A(k)\left[f_x(-(m+k),-m,-m) - f_x\big(-(m-k),-(m-k),-(m-k)\big)\right] + \sum_{k=1}^{q} B(k)\left[f_x\big(-(m+k),-(m+k),-(m+k)\big) - f_x(-(m-k),-m,-m)\right] =$$

$$m \cdot f_x(-m,-m,-m) = c_x(-m,-m,-m) \quad p,q \to \infty \qquad (2)$$

where the minimum phase $\{A(k)\}$ and maximum phase $\{B(k)\}$ parameters are given by,

$$g_x(k,0,0)\begin{cases} -\frac{1}{k}\cdot A(k), & k=1,\dots,p \\ \frac{1}{k}\cdot B(-k), & k=-1,\dots,-q \end{cases} \qquad (3)$$

and $g_x(k,l,n)$ is the tricepstrum of the 4th order statistics $f_x(k,l,n)$ of the signal. In this paper, the cepstrum coefficients in (2) are being used, for the detection problem given by (1). The following assumptions are being made.

1) Under H_0 it is assumed that $\{A(k)\},\{B(k)\}$ for all k are equal to zero.
2) Under H_1 since the process $\{x(n)\}$ is not stationary, it is assumed availability of many data records, i.e, $x^{(i)}(n) = m(n) + w^{(i)}(n), i = 1,\dots,M$ is the given ensemble data set, where $\{w^{(i)}(n)\}$ are different noise realizations of identical statistical properties then,

$$f_x(k,l,m) = E\left\{\cdot \sum_n x(n)x(n+k)x(n+l)x(n+m)\right\} \qquad (4)$$

3) Since $\{A(k)\},\{B(k)\}$ are decaying sequences they can be truncated (2) and $\mathrm{p,q}$ finite integers can be used [12].

By choosing $p=q$ then (2) can be written,

$$c_x(m,n) = \sum_{k=1}^{p} f_x(m,k,n)\cdot A(k,n) + \sum_{k=1}^{p} f_x'(m,k,n)\cdot B(k,n),$$

$$m = -p,\dots,-1,1,\dots,p \qquad (5)$$

where $\{A(k,n)\},\{B(k,n)\}$, $f_x(m,k,n)$, $f_x'(m,k,n)$, denote the estimates of the corresponding values of (2) at time instant n based on N samples. In a matrix form,

$$\mathbf{F}(p,n)\cdot\mathbf{T}(p,n) = \mathbf{C}(p,n) \qquad (6)$$

where the elements of $\mathbf{T}(p,n)$ are

$$T(k,n) = \begin{cases} A(k,n), k=1,\dots,p \\ B(k-p,n), k=p+1,\dots,2p \end{cases} \qquad (6.1)$$

and

$$\mathbf{F}(p,n) = \begin{pmatrix} f_x(p,1,n) & \cdots & f_x(p,p,n) & f_x'(p,1,n) & \cdots & f_x'(p,p,n) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ f_x(-p,1,n) & \cdots & f_x(-p,p,n) & f_x'(-p,1,n) & \cdots & f_x'(-p,p,n) \end{pmatrix}, (6.2)$$

$$\mathbf{C}(p,n) = \big(c_x(p,n),\dots,c_x(-p,n)\big)^T, \qquad (6.3)$$

"T" denotes the transpose operation. Under $\mathrm{H_0}$, the matrix $\mathrm{F(p,n)}$ is of full, 2p rank. Asymptotically under H_0 the estimates of the tricepstrum coefficients, $\{\mathrm{A(k)}\},\{\mathrm{B(k)}\}$, are Gaussian random variables of zero mean and constant covariance matrix. It is also assumed that if $\mathrm{N} \to \infty$ $\mathrm{f_x(m,k,n)}$,$\mathrm{f_x'(m,k,n)}$ become their true values $\mathrm{f_x(m,k)}$, $\mathrm{f_x'(m,k)}$. Therefore, the following variable can be used as a detection statistic:

$$L_T = \sum_{k=1}^{l} (A(k,n)^2)/(\sigma_{a_k}^2) + (B(k,n)^2)/(\sigma_{b_k}^2), \qquad (7)$$

where, $\sigma_{\mathrm{a_k}}^2$, $\sigma_{\mathrm{b_k}}^2$ are the variances of $\mathrm{A(k,n)}$, $\mathrm{B(k,n)}$. This is a central quadratic form $(l \le p)$. For fixed probability of false alarm $\mathrm{P_{FA}}$, the threshold can be computed using the cumulative distribution $\mathrm{F_0}$ of $\mathrm{L_T}$ under $\mathrm{H_0}$,

$$\mathrm{t_h} = \mathrm{F_0^{-1}}(1-\mathrm{P_{FA}}) \qquad (8)$$

Instead of using 4th order statistics, 3rd order statistics can be used. However, in this case the noise cannot be Gaussian, i.i.d. The Additive White Non-Gaussian Noise (AWNGN) with zero mean assumption is enough to guarantee asymptotically under $\mathrm{H_0}$ rank 2p, for the matrix $\mathrm{F(p)}$.

Summarizing the algorithm for detecting deterministic transient signals embedded in additive white Gaussian noise, we have the following:

1) Estimate the 4th order statistics of $\mathrm{f_x(k,l,n)}$ [12].
2) Estimate $\mathrm{A(k)}$, $\mathrm{B(k)}$ using a least squares solution to the overdetermined system of equations (2) when $\mathrm{p=q, m=p,\dots,p-W}$. W≥2p
3) Compute L_T and compare it with a threshold chosen according to (8).

B. The Recursive Approach of the Higher Order Cepstrum Based Detector

Instead of estimating the cepstrum coefficients $\{\mathrm{A(k)}\}$ and $\{\mathrm{B(k)}\}$ in one step when the whole data record is available we seek for a recursive solution of (6) which will allow for fast updating of the coefficients when new data arrive and the amount of the data is large. Another reason for developing a recursive approach is when the arrival times of the transients are unknown. It is assumed the following partition for $\mathrm{f_x(m,k,n)}$, $\mathrm{f_x'(m,k,n)}$.

$$\begin{aligned} f_x(m,k,n) = \sum_{i=n_0+1}^{n} & \lambda^{n-i} x(i)(x^2(i-m),1) \\ & \cdot \left(x\left(i-(m+k)\right), -x^3\left(i-(m-k)\right)\right)^T \\ & +\lambda^{n-n_0} f_x(m,k,n_0), \end{aligned} \tag{9.1}$$

$$\begin{aligned} f_x'(m,k,n) = \sum_{i=n_0+1}^{n} & \lambda^{n-i} x(i)(x^2(i-m),1) \\ & \cdot \left(-x\left(i-(m-k)\right), x^3\left(i-(m+k)\right)\right)^T \\ & +\lambda^{n-n_0} f_x'(m,k,n_0), \end{aligned} \tag{9.2}$$

where, λ is a weight constant, $0 < \lambda \leq 1$ and $\mathrm{f}(\mathrm{m},\mathrm{k},\mathrm{n}_0)$, $\mathrm{f}_\mathrm{x}'(\mathrm{m},\mathrm{k},\mathrm{n}_0)$ are computed values from the initialization period which will be explained in the sequel. Also note that a time recursion for $\mathrm{C}(\mathrm{p},\mathrm{n})$ is

$$\mathbf{C}(p,n) = \lambda \cdot \mathbf{C}(p,n-1) + \mathbf{a}^T(n), \tag{10}$$

$$\mathbf{a}(n) = \left(px^3(n-p)x(n), \ldots, (-p)x^3(n+p)x(n)\right) \tag{11}$$

To realize this solution matrix $\mathrm{F}(\mathrm{p},\mathrm{n})$ is decomposed into two sub matrices, V and U and their Q-R decomposition is updated at each time instant that new information is present,

$$\mathbf{F}(p,n) = \mathbf{V}^T(p,n) \cdot \mathbf{U}(p,n) \tag{12}$$

Both V^T and U start with,

$$\mathbf{Q}(1) \cdot \widehat{\mathbf{A}}(1) = \begin{pmatrix} \mathbf{F}(1) \\ 0 \end{pmatrix} \tag{13}$$

where $\widehat{\mathrm{A}}$ matrix will represent either V^T or U, $\mathbf{F}(1) = x(0)$ and

$$\mathbf{Q}(1) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \widehat{\mathbf{A}}(1) = \begin{pmatrix} 0 \\ x(0) \end{pmatrix} \tag{14}$$

Given the above initial values, new data at each iteration i, for both V^T and U, namely $\mathrm{u}_{\mathrm{in1}}(\mathrm{i})$ and $\mathrm{u}_{\mathrm{in2}}(\mathrm{i})$ are obtained. The general notation $\mathrm{u}_{\mathrm{in}}(\mathrm{i})$ is used here. Then, Q-R decomposition is applied on $\widehat{\mathrm{A}}(1)$

$$\widehat{\mathbf{A}}(2) = \begin{pmatrix} \widehat{\mathbf{A}}(1) | \mathbf{0}_{(2x1)} \\ \mathbf{u}_{in}(i) \end{pmatrix} \tag{15}$$

In the subsequent steps of the initialization period, we input new data until we finally obtain Q-R decompositions for V^T and U.

It remains to describe the orthogonal transformations that are used to compute the Givens rotation parameters and then the Givens transformation matrix G(i). In particular for each step, the partially triangularized matrix is assumed,

$$\mathbf{F}(i) = \begin{pmatrix} \mathbf{D}_{ixj}^{1/2}(i) \cdot \widehat{\mathbf{F}}_{jxj}(i) \\ \mathbf{0}_{1xj} \end{pmatrix}, \tag{16}$$

$$j = \begin{cases} i, & i \leq 2p \\ 2p, & i > 2p \end{cases}, \tag{17}$$

where $\mathrm{D}^{1/2}(\mathrm{i})$ is diagonal matrix and $\widehat{\mathrm{F}}(\mathrm{i})$ is a unit upper triangular matrix. The requirement is to find rotation parameters so that we can annihilate the new input vector $\mathrm{u}_{\mathrm{in}}(\mathrm{i})$. Thus, a sequence of Givens rotations [13] is used, described by (13.1)

$$G_m(i,k,l) = \begin{cases} c_m(i), k = l = m \\ s_m(i), k = i, l = i+1 \\ -s_m(i), k = i+1, l = m \\ c_m(i), k = l = i+1 \\ 1, k = l, k \neq m, 1 \leq k < i \end{cases} \tag{18}$$

and the Givens transformation matrix itself is

$$\mathrm{G}(\mathrm{i}) = \textstyle\prod_{\mathrm{m}=1}^{\mathrm{l}_\mathrm{n}} \mathrm{G}_\mathrm{m}(\mathrm{i}) \tag{19}$$

$$\mathrm{l}_\mathrm{n} = \begin{cases} \mathrm{k}_\mathrm{n} + 1, \mathrm{k}_\mathrm{n} < 2p \\ 2\mathrm{p}, \mathrm{k}_\mathrm{n} = 2\mathrm{p} \end{cases} \tag{20}$$

where k_n is the dimension of the input vector.

After the end of the initialization period, initial Q-R decompositions for $\mathrm{V}^\mathrm{T}(\mathrm{p},\mathrm{n})$ and $\mathrm{U}(\mathrm{p},\mathrm{n})$ are available. It is denoted again in general each one of them by $\widehat{\mathrm{A}}(\mathrm{p},\mathrm{n})$. For $\mathrm{U}(\mathrm{p},\mathrm{n})$ the new data sets for the next iteration are its last two lines, i.e, per iteration its Q-R decomposition is updated twice. For $\mathrm{V}^\mathrm{T}(\mathrm{p},\mathrm{n})$, per iteration its Q-R decomposition is updated $4\mathrm{p} + 2$ times, using equal number of new data sets, $\mathrm{u}_{\mathrm{in}}(\mathrm{n},\mathrm{i})$, which are described as follows,

$$\mathbf{u}_{in}(n,i) = \begin{cases} \mathbf{u}_{in1}(n,i) = x(n-p+i) \cdot \mathbf{u}_{in2}(n,i), \\ \mathbf{u}_{in2}(n,i) = \left(\overbrace{0,\ldots,0}^{i-1}, x(n+1), \ldots, x\left(n-(2p-i)\right) \right) \\ i = 1, \ldots, 2p+1 \end{cases}, \tag{21}$$

where the step $\mathrm{i} = \mathrm{p} + 1$ (which corresponds to $\mathrm{m} = 0$, row of $\mathrm{F}(\mathrm{p},\mathrm{n})$ in (6)) is ignored. The decomposition of the first step $\mathrm{i} = 1$ is stored for each time instant n and is used as initial for the iterations, $\mathrm{i} = 1, \ldots, 2\mathrm{p} + 1$ of the next time instant $\mathrm{n} + 1$ for $\mathrm{V}^\mathrm{T}(\mathrm{p},\mathrm{n})$. I.e, summarizing the update process for both $\mathrm{V}^\mathrm{T}(\mathrm{p},\mathrm{n})$, $\mathrm{U}(\mathrm{p},\mathrm{n})$ for every time instant n we have the following,

$$\mathrm{Q}_{jxj}(p,n) \cdot \widehat{\mathrm{A}}_{jx2p}(p,n) = \begin{pmatrix} \mathrm{F}_{2px2p}(p,n) \\ 0_{(j-2p)x2p} \end{pmatrix}, \tag{22.1}$$

$$\mathrm{F}_{2\mathrm{px}2\mathrm{p}}(\mathrm{p},\mathrm{n}) = \mathrm{D}_{2\mathrm{px}2\mathrm{p}}^{1/2}(\mathrm{p},\mathrm{n}) \cdot \widehat{\mathrm{F}}_{2\mathrm{px}2\mathrm{p}}(\mathrm{p},\mathrm{n}), \tag{22.2}$$

where $\mathrm{j} = 2(\mathrm{n} - \mathrm{n}_0) + 8\mathrm{p} + 2$ for $\mathrm{U}(\mathrm{p},\mathrm{n})$ and $\mathrm{j} = 2(\mathrm{n} - \mathrm{n}_0) + 4\mathrm{p} + 2$ for $\mathrm{V}^\mathrm{T}(\mathrm{p},\mathrm{n})$ and it was assumed that $\mathrm{n}_0 = 2\mathrm{p}$. At the next time instant $\mathrm{n} + 1$, new information is available (note that at time instant n, samples of a growing rectangular window are available up to time instant $\mathrm{n} + 2\mathrm{p}$, i.e, when it is said new available information it is meant that this window is moved one position forward) and $\widehat{\mathrm{A}}(\mathrm{p},\mathrm{n})$ and $\mathrm{F}(\mathrm{p},\mathrm{n})$ matrices are updated,

$$\widehat{\mathbf{A}}(p,n+1,i)=\begin{pmatrix}\widehat{\mathbf{A}}(p,n,i)\\ \mathbf{u}_{in}(n,i)\end{pmatrix}, \qquad (23.1)$$

$$\mathbf{F}^{*}(n+1,i)=\begin{pmatrix}\mathbf{F}_{2px2p}(p,n,i)\\ \mathbf{0}_{(j-2p)x2p}\\ \mathbf{u}_{in}(n,i)\end{pmatrix}, \qquad (23.2)$$

The index i is used here to indicate the iterations for every time instant n. Note that, $\widehat{\mathrm{A}}(\mathrm{p},\mathrm{n})=\widehat{\mathrm{A}}(\mathrm{p},\mathrm{n},1)$ and $\mathrm{F}(\mathrm{p},\mathrm{n})=\mathrm{F}(\mathrm{p},\mathrm{n},1)$ also for $\mathrm{U}(\mathrm{p},\mathrm{n}), \mathrm{i}=1$, where as for $\mathrm{V}^{\mathrm{T}}(\mathrm{p},\mathrm{n}), \mathrm{i}=1,\ldots,2\mathrm{p}+1$.

The rotation parameters are computed and the sequence of the square root Givens rotations are applied at time instant n on $\mathrm{F}(\mathrm{p},\mathrm{n})$ to annihilate all 2p elements of the last row. Then,

$$\mathbf{G}(n+1,i)\cdot\mathbf{F}^{*}_{(j+1)x2p}(n+1,i)=\begin{pmatrix}\mathbf{F}_{2px2p}(p,n,i)\\ \mathbf{0}_{(j+1-2p)x(2p)}\end{pmatrix}, \qquad (24.1)$$

$$\mathbf{Q}_{(j+1)x(j+1)}(p,n+1,i)=\mathbf{G}(n+1,i)\cdot\begin{pmatrix}\mathbf{Q}_{jxj}(p,n,i) & \mathbf{0}_{jx1}\\ \mathbf{0}_{1xj} & 1\end{pmatrix}, \qquad (24.2)$$

which gives the following Q-R decomposition at time instant $\mathrm{n}+1$,

$$\widehat{\mathbf{A}}_{jx2p}(p,n+1)=\mathbf{Q}^{T}_{jxj}(p,n+1)\cdot\begin{pmatrix}\mathbf{F}_{2px2p}(p,n+1)\\ \mathbf{0}_{(j-2p)x(2p)}\end{pmatrix}, \qquad (25)$$

where now $\mathrm{j}=2(\mathrm{n}+1-\mathrm{n}_0)+8\mathrm{p}+2$ for both $\mathrm{V}^{\mathrm{T}}(\mathrm{p},\mathrm{n})$, $\mathrm{U}(\mathrm{p},\mathrm{n})$. If indexes u, v are used to denote the corresponding Q, F matrices for $\mathrm{V}^{\mathrm{T}}(\mathrm{p},\mathrm{n})$, $\mathrm{U}(\mathrm{p},\mathrm{n})$ then the least squares solution (6) can be realized as,

$$\mathbf{F}^{(u)}\cdot\mathbf{T}=\left(\mathbf{Q}^{(v)}\left(\mathbf{Q}^{(u)}\right)^{T}\right)^{-1}\cdot\left(\mathbf{F}^{(v)}\right)^{-T}\cdot\mathbf{C}, \qquad (26)$$

The two square matrices on the right hand side of (26) are invertible because of the way that they were constructed using the Givens rotations and the above linear system of equations can be solved using back substitution.

III. SIMULATION EXAMPLES

Test Case 1 (Minimum phase transient, unknown arrival time). The z-transforms of the infinite duration signal is given by, example 1,

$$F(z)=\frac{1}{z^{2}-(1.35)z+0.75} \qquad (27)$$

and we assume that it is of unknown arrival time at 200 samples. The signal plus noise records for AWGN of variance, $\sigma_{\mathrm{w}}^{2}=3.162\mathrm{x}10^{-1}$, 0.1, for 15 sample signal are shown in Figure 1a, 1b. In Figure 2a, 2b, we plot the detection statistics for the tricepstrum method and for the Infinite Impulse Response (IIR) adaptive algorithm versus time (the same way as for the definition of L_{T}, we use the sum of the squares of the estimated coefficients of the recursive IIR model as a detection statistic for the comparison algorithm). For the tricepstrum $\mathrm{p}=\mathrm{q}=2$ and $\mathrm{l}=2$ and for the IIR model order, 2 were the choices for this experiment. For all the experiments below, including this one we keep l equal to the order of the model. Both methods under $\mathrm{H}_0(0,\ldots,199)$ samples remain in the zero state and when the transient appears they jump and slowly converge again to the zero state. The weighting constant λ was for both methods 0.99. In Figure 2c, we show operation of the algorithms for noise variance $\sigma_{\mathrm{w}}^{2}=1$ and $\lambda=0.98$.

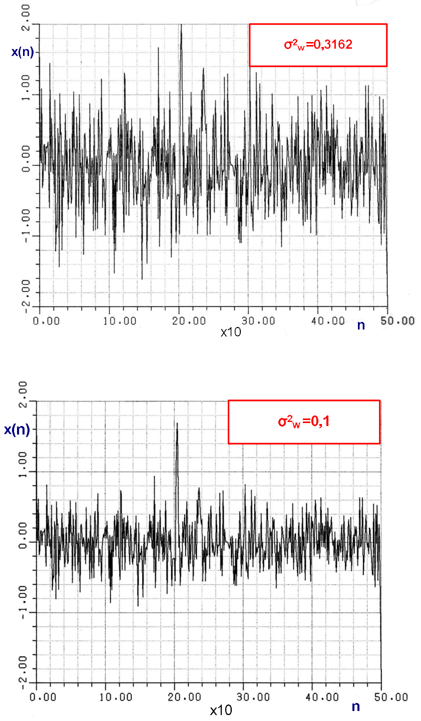
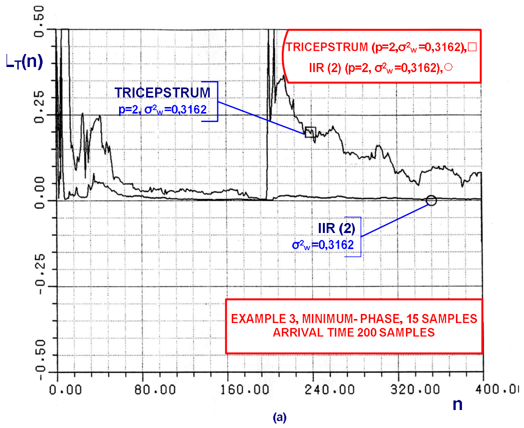


Figure 1. Signal plus noise records for the minimum-phase signal, example 1, 15 samples, arrival time, 200 samples: (a) $\sigma_{\mathrm{w}}^{2}=3.162\mathrm{x}10^{-1}$, (b) $\sigma_{\mathrm{w}}^{2}=0.1$.



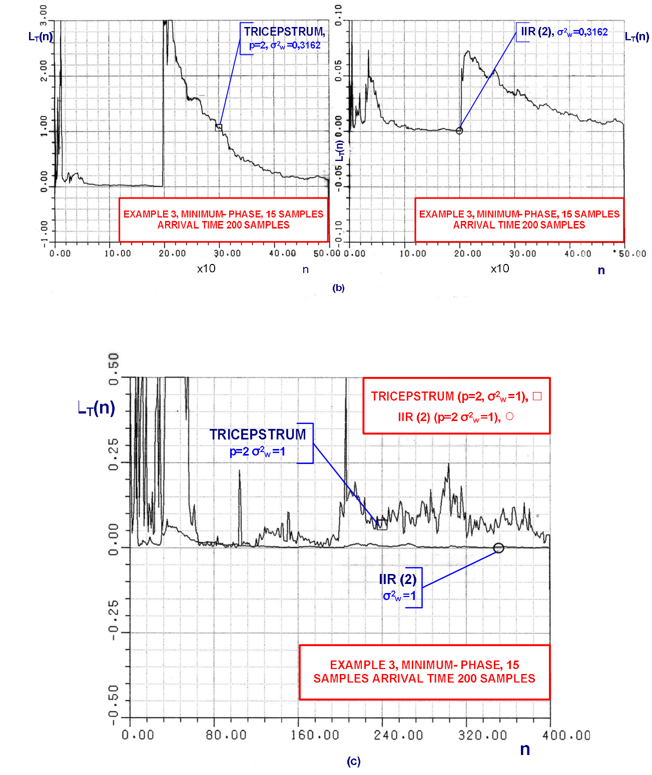


Figure 2. Additive White Gaussian Noise, minimum-phase signal, example 1, 15 samples, arrival time, 200 samples, L_T versus time $\lambda = 0.99$, (a) Tricepstrum, $p = 2$ and IIR(2), $\sigma_w^2 = 3.162x10^{-1}$, (b) $\sigma_w^2 = 0.1$, (c) $\sigma_w^2 = 1, \lambda = 0.98$.

Test Case 2 (Mixed phase transients, unknown arrival times),

example 2,

$$F(z) = \frac{(1-0.5z)\left(z^2-(1.09754)z+(0.3012)\right)}{z^2-(0.6098)z+0.3012} \qquad (28.1)$$

example 3,

$$F(z) = \frac{(1-0.5z)(1-0.2z)\left(z^4-(2.1481)z^3+(1.8221)z^2-(0.7202)z+0.1108\right)}{z^4-(1.7092)z^3+(1.3395)z^2-(0.5555)z+0.1108} \qquad (28.2)$$

For example 2, we use 15 sample signal and arrival time at 150 samples. The tricepstrum and IIR detection statistics versus time are shown in Figures 3a, 3b. The noise variance is $\sigma_w^2 = 0.1, 3.162x10^{-2}$ and we choose $p = 2$ for the tricepstrum and order 6 for the IIR, $\lambda = 0.99$ for both. It is clear that because of the inability of the IIR model to catch the non-minimum phase character of the signal its performance becomes much worse. This becomes more apparent if we compare Figures 3a and 3b, where the SNR values are 13.8 db and 12 db correspondingly.

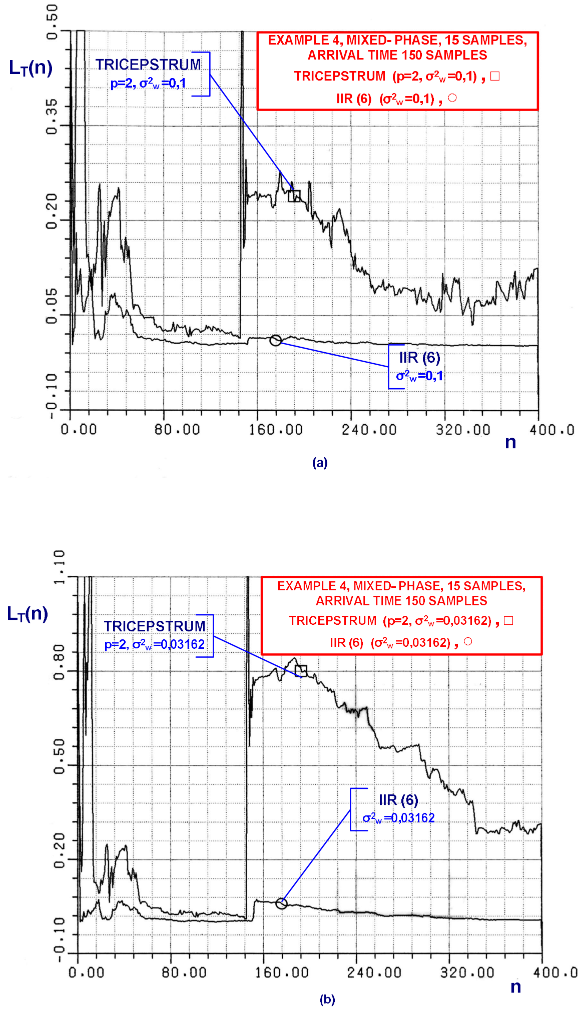


Figure 3. Additive White Gaussian Noise, mixed-phase signal, example 2, 15 samples, arrival time, 150 samples:(a) L_T versus time, $\lambda = 0.9$, Tricepstrum, $p = 2$ and IIR(6), $\sigma_w^2 = 0.1$, (b) L_T versus time, $\lambda = 0.99, \sigma_w^2 = 3.162x10^{-2}$.

In example 3, we make the non-minimum phase character of the signal even stronger and we plot also the detection statistics for both methods in Figures 4a-4c, for 25 sample signal and corresponding noise variances, $\sigma_w^2 = 3.162x10^{-1}, 0.1, 3.162x10^{-2}, 10^{-2}$. The orders of the tricepstrum and IIR methods were 2 and 10 with $\lambda = 0.99$. The performance of the first improves and for the second becomes worse with respect to examples 2 and 1.

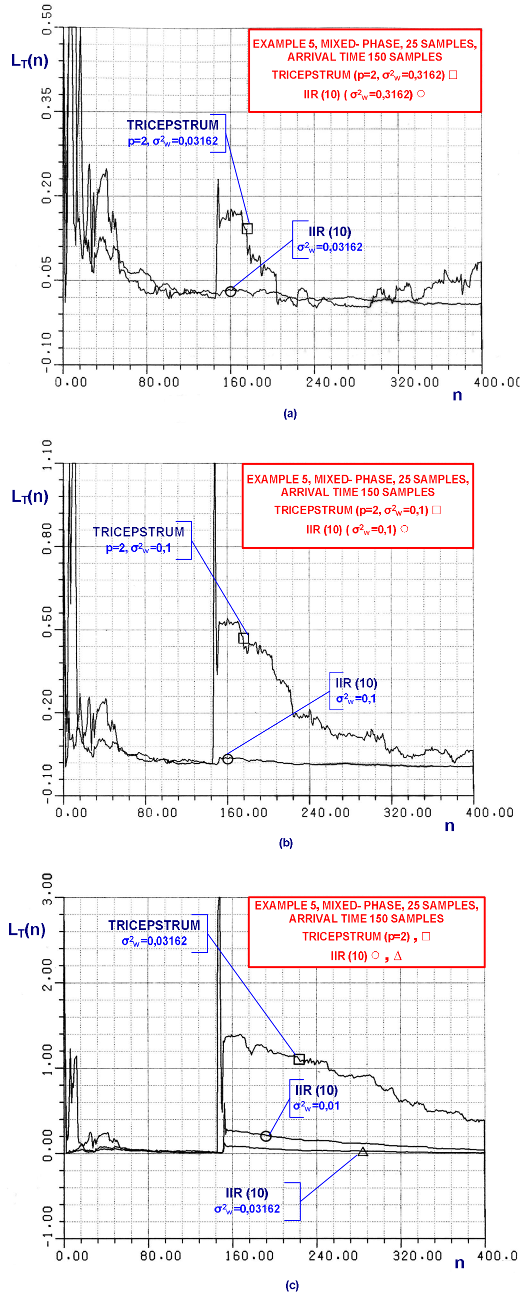


Figure 4. Additive White Gaussian Noise, mixed-phase signal, example 3, 25 samples, arrival time 150 samples:(a) L_T versus time, $\lambda = 0.9$, Tricepstrum, $p = 2$ and IIR(10), $\sigma_w^2 = 3.162x10^{-1}$, (b) L_T versus time, $\lambda = 0.99$, $\sigma_w^2 = 0.1$, (c) L_T versus time, $\lambda = 0.99$, $\sigma_w^2 = 3.162x10^{-2}, 10^{-2}$ for IIR(10) and $\sigma_w^2 = 3.162x10^{-2}$ for tricepstrum.

In Figure 5, we plot for $p = 3, \lambda = 0.99$ and $\sigma_w^2 = 3.162x10^{-2}, 0.1$ the tricepstrum detection statistic for examples 2, 3 to demonstrate performance with increased order. Note that varying the order of the IIR method does not change the situation shown in Figures 2-4.

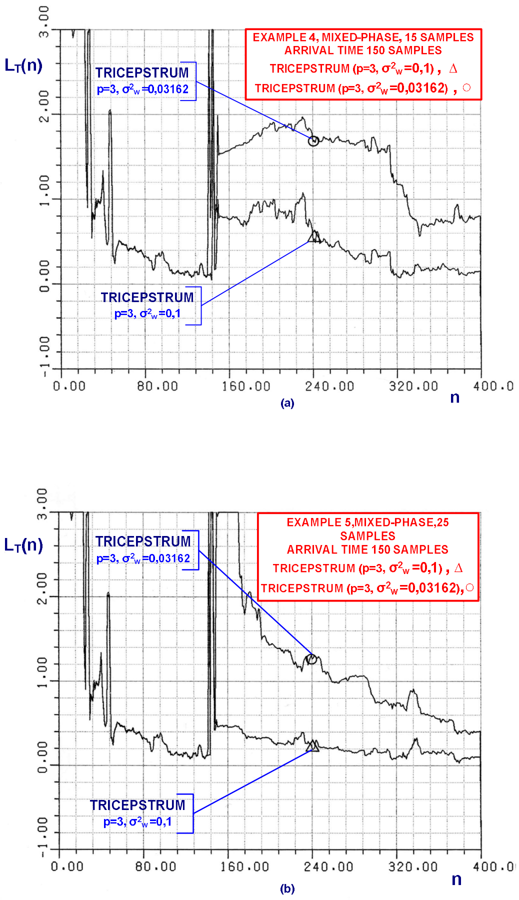


Figure 5. Additive White Gaussian Noise, mixed-phase signals, examples 2, 3, 15, 25 samples, arrival time 150 samples, L_T versus time, $\sigma_w^2 = 0.1$, $3.162x10^{-2}$,Tricepstrum $p = 3, \lambda = 0.99$, (a) example 2, (b) example 3.

IV. CONCLUSION AND FUTURE WORK

Using a suitable partition of the 4th order statistics involved in (2), a recursive solution for the cepstral equation was formulated. Because of the high variance in the estimation of the 4th order statistics, the recursive approach was based on orthogonal Q-R decompositions of the partioned data matrices which consist the cepstral equation. By means of simulation examples, it was demonstrated that the proposed algorithm is capable to detect transients of unknown arrival times. Comparing this technique with a fast adaptive algorithm based on an IIR model for the signal, significant improvement in terms of signal detection capability was demonstrated. Future work could investigate the performance of the proposed algorithm in the presence of i.i.d. noise with probability density function which follows non Gaussian distribution either symmetric (for example non-skewed) or asymmetric.

REFERENCES

[1] B. Porat and B. Friendlander, “Adaptive Detection of Transient Signals”, IEEE Trans. Acoust. Speech, Signal Processing, vol. ASSP-34, pp. 1410-1418, December 1986.

[2] P. Nicolas and D. Kraus, “Detection and Estimation of Transient Signals in Colored Gaussian Noise”, ICASSP’88, New York, pp. 2821-2824, April 1988.

[3] B. Friendlander and B. Porat, “Detection of Transient Signals by the Gabor Representation”, IEEE Trans. Acoust. Speech, Signal Processing, vol. ASSP-37, pp. 169-179, February 1989.

[4] M. Sanaullah, “A Review of Higher Order Statistics and Spectra in CommunicationSystems”, Global Journal of Science Frontier Research Physics and Space Science, vol.13, Issue 4 Version 1.0, 2013.

[5] J. L. Caillec and R. Garello , “Asymptotic Bias and Variance of Conventional Bispectrum Estimates for 2-D Signals”, Multidimensional Systems and Signal Processing, vol.16, pp. 49-84, 2005.

[6] B. Liang, S.D. Iwnicki and Y. Zhao, “Application of power spectrum, cepstrum, higher order spectrum and neural network analyses for induction motor fault diagnosis”, Mechanical Systems and Signal Processing, vol. 39, pp. 342-360, 2013.

[7] F. Gu. et al., “Electrical motor current signal analysis using modified bispectrum for fault diagnosis of downstream mechanical equipment” Mechanical Systems and Signal Processing, vol. 25, pp. 360–372, 2011

[8] M.A.Hassan, D.Coats, K.Gouda, Y.J.Shin, and A. Bayoumi,“Analysis of Nonlinear Vibration-Interaction Using Higher Order Spectra to Diagnose Aerospace System Faults”, IEEE Aerospace Conference, Paper #1345, Version 4, March 2012.

[9] B. Liang, “The higher order spectrum analysis for fault pattern extraction of induction motors”, Power Electronics and ECCE Asia (ICPE-ECCE Asia), 9th International Conference on, June 2015.

[10] A. Gudigara, S.Chokkadia,U.Raghavendraa, U.R.Acharyab, “Local texture patterns for traffic sign recognition using higher order spectra” Pattern Recognition Letters, vol. 94, pp. 202-210, July 2017.

[11] U.R. Acharya, V.K. Sudarshana, J.E.W. Koha, R. J. Martisd, J. H. Tana, S.L.Oha, A. Muhammada, Y. Hagiwaraa, M. R. K. Mookiaha, “Application of higher-order spectra for the characterization of Coronary artery disease using electrocardiogram signals”, Biomedical Signal Processing and Control, vol. 31, pp. 31–43, January 2017.

[12] R. Pan and C.L. Nikias, “The Complex Cepstrum of Higher-Order Cumulants and Nonmiminum Phase System Identification”, IEEE Trans. Acoust. Speech, Signal Processing, vol. ASSP-36, pp. 186-205, February 1988.

[13] S. Haykin, “Adaptive Filter Theory”, Englewood Cliffs, NJ:Prentice-Hall 1986.

Application of Data Mining in the 5G Network Architecture

Alexandros Kaloxylos
Department of Informatics and Telecommunications
University of Peloponnese
Tripoli, Greece
email: kaloxyl@uop.gr

***Abstract*— Data mining is considered to be one of the key enablers for the next generation of mobile networks. The building of knowledge models is expected to tackle the complexity of these networks and enable their dynamic management and operation. Recently, this research area has attracted a lot of interest and several models have been proposed by the research community. This paper provides a brief survey of these efforts and captures the latest status in 3GPP. It also provides a detailed description of which information needs to be collected by network components, so as to be analyzed by a data mining scheme. Finally, it quantifies the amount of information that is required to be reported to the data mining engine.**

Keywords- 5G cellular networks; data mining; control functions optimizations.

I. INTRODUCTION

During the past years, a tremendous effort has been made for the design of the 5th Generation of mobile networks (5G). Research and standardization activities worldwide are in the process of finalizing the first release, while all major vendors are preparing for the first commercial showcases and large-scale deployments. 5G mobile networks target the provision of tailor-cut solutions not only for the telecommunications sector but also for the so called "vertical industries" (e.g., intelligent transportation systems, smart factories, the health sector, etc.). This will be achieved by deploying multiple logical networks (a.k.a. network slices) over the same network infrastructure. Thus, 5G networks will be considerably more complex than the previous generations [1].

At the same time, the scientific community has identified that big data solutions can significantly improve the operation and management of existing and future mobile networks [2]. Data mining is used to discover patterns and relationships between variables in large data sets. Towards this end, several mechanisms that include statistical analysis, artificial intelligence and machine learning are applied in the data set to extract essentially knowledge from the examined data.

Figure 1 illustrates how data mining can be integrated as a process with the mobile networks and where the extracted knowledge can be used. More specifically, data are collected from a number of network components. These data may include a variety of information fields such as the quality of the wireless channel, the network load, accounting information, configuration and fault indications, the profile of the subscribers, etc. These data are stored and updated regularly. When collected, they are passed through a pre-processing phase. During this phase transformation, discretization, normalization, outlier detection and dimensionality reduction is executed. The outcome of this phase is then passed to a data analysis phase where a model is built to extract knowledge from the processed data. For example, the result of this process will be the identification of situations where the occurrence of some specific events (e.g., a significant increase of the number of high moving users) causes some specific result (e.g., increase of the handover blocking probability). The knowledge model may also include some solutions for specific situations (e.g., force the network components to place high moving users to macro cells). The list of the knowledge discovery results can then be communicated to either policy, management or control modules. These modules can use this information in order to optimize the operation of the network and improve the performance. Note that for selecting the best configuration or optimization action from the list of the knowledge results, the abovementioned communication modules may require also real-time information related to the current performance indicators of a network.

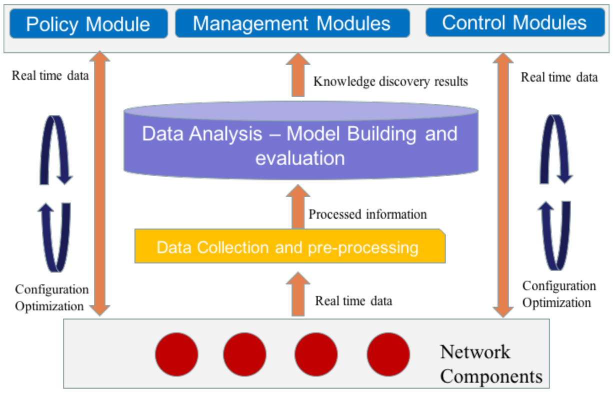

Policy Module
Management Modules
Control Modules
Real time data
Knowledge discovery results
Real time data
Data Analysis – Model Building and evaluation
Processed information
Data Collection and pre-processing
Configuration Optimization
Real time data
Configuration Optimization
Network Components

Figure 1. Big data analysis for cellular networks

As it will be presented in the next section, it is currently widely accepted that data mining will be an integral part of 5G networks. Currently, proposals focus on how data mining can feed knowledge on management and control modules. Although this work is quite valuable, the researchers focus on the data mining algorithms to be used and they do not always provide detailed examples of which exactly information needs to be collected, how often this collection has to take place and how one can minimize the data that has to be exchanged among network components and functions.

At the same time, 3GPP has included a dedicated function, called NWDAF (NetWork Data Analytics Function), in the latest specifications [3]. This function has currently limited functionality. It is used to provide information to influence, in real time, the policies that an operator is using.

The purpose of this paper is to provide a short survey for data mining solutions in 5G networks and extend the work presented in [4], where a data mining framework, called Context Extraction and Profiling Engine (CEPE), was introduced to improve the performance of control functions in a mobile network. The CEPE extensions include the identification of the 5G network components, as specified by 3GPP, that can be used to collect the necessary information. Also, the paper identifies which 5G functions can use the outcome of CEPE. The paper quantifies the amount of data that must be exchanged between network components for the operation of CEPE.

The rest of the paper is organized as follows. Section II provides an analysis of the state of the art for data mining in 5G networks and the latest status of 3GPP for the specification of NWDAF. Section III briefly presents the CEPE framework and maps its functionality CEPE in the latest 5G architecture. Section IV quantifies the amount of data that needs to be exchanged and suggest an approach on how this can be further reduced. Finally, Section V concludes the paper.

II. State of the art analysis

This section discusses the existing state of the art proposals presented by the research community and the current status of 3GPP activities.

A. Research literature survey

As mentioned previously, many academic researchers have focused on the use of data analytics mechanisms in 5G communications. The authors in [5] explain how random matrix theory and machine learning can be used to enable the adoption of big data schemes for mobile cellular networks. Moreover, they provide a survey of solutions on how big data can be used to analyze signaling information in cellular networks as well as the traffic of user plane data. This analysis is able to reveal certain traffic and user behavior characteristics and even waveform related data to estimate the mobility of users. The work in [6] presents a generic extraction and correlation framework that targets to reduce the vast data set through randomization and a coarse preservation of statistical relationship among data records. This scheme is quite valuable but is generic and the authors do not provide detailed examples on the information be used and how the outcome can be used by 5G networks. Unfortunately, the removal of unrelated and non-useful data remains an open question. This is significant, because collection and transmission of useless information is burdensome on the network.

[7] presents the findings of the SELFNET H2020 project. Its target is to provide an autonomic network management framework for 5G mobile network infrastructures. The paper focuses on the analyzer module that infers data from a set of collected metrics. It describes in detail the operation of the analyzer module but it does not provide a detailed discussion on the information that needs to be collected to support specific 5G use cases. The authors of [8] provide a new framework, named Big Data SON (BSON) that takes into consideration subscribers' level data (i.e., network related performance on a per user basis such as throughput, delay, blocking and drop rates, etc.), cell level data (received signal strength of serving and neighboring cells, number of active users, etc.), core level data (alarms, configuration, security data, Call data records, etc.). By applying data mining schemes the authors suggest that these data can be used to improve SON mechanisms and essentially transform SON to be proactive instead of reactive. The paper provides an exhaustive list of information that can potentially be used but it does not analyze the traffic volume that needs to be collected in order for the scheme to perform the desired results. Although this is a holistic solution that can be used for all SON cases the authors only demonstrate its application for a simple scenario. Following the main principles presented in Section I, the authors of [9] present the key features of user and mobile network data that can be potentially collected and processed by a data mining scheme. They also discuss how resource management, planning, interference coordination and cache server deployment is done nowadays. They suggest that these can be greatly improved if data analytics is adopted by network operators. Although the paper discusses extensively main operation principles, there are no detailed examples.

In relation to the optimization of radio resources using data analytics, the authors of [10] present a scheme that analyzes historical information gathered from an operational wireless network. Their model uses a weighted k-Nearest Neighbors model and can predict future network load levels and optimize the network accordingly. An interesting idea is presented in [11] where it is proposed that big data can be used to optimize the performance of the protocol stack in RAN (e.g., reduce the overhead in Radio Header Compression in PDCP, or the minimization of signaling during the execution of a handover, etc.).

In [12], a system that can handle 4.2 Tbytes of traffic data from 123 Gbs links in the core network of a 2G/3G operator. By analyzing application layer information, they are able to identify the exact model of an end device as well characteristics of users' behavior. The authors of [13] discuss how the call detail records collected from a legacy mobile wireless network can be used to identify how a large fraction of people are moving inside a city. Using an end-to-end Hadoop system, they are able to identify the hangouts and trajectories of users with different interests. The goal of this work is for the operators to be able to deliver such data and insights to other enterprises.

The work presented in [4], provides a framework where data mining on user related information can provide to a number of control functions the needed extra context information to improve their performance.

Finally, the work in [14] aims to improve the personalized QoE for end users by following a two-step modelling approach, combined with big-data analysis, to identify the relationship between users and services. More specifically, this approach requires to obtain real-time information about the application the users are using (online part) and adopt a data mining (offline training part) scheme to predict the users' preferences and expectations. Then, the network resources are managed accordingly to support a satisfactory QoE.

Some of the abovementioned solutions require user data traffic analysis (e.g., [5] [10] [12]). This however requires the transfer and processing of a huge amount of data some of which (e.g., video streams) may be encrypted. Some of the solutions try to address a holistic approach covering from network management process (e.g., healing, optimization, fault detection) to the support of personalized QoS for the users. These solutions identify a plethora of parameters that have to be taken into consideration (from radio measurements, preambles, link utilization, subscriber data, customer retention management data, as well as application data, etc.). Such solutions have of course a huge complexity and the current literature does not provide detailed examples, in terms of which data have to be collected and from which entities. Equally importantly, none of the abovementioned solutions elaborates on how to minimize the required information to be collected. Finally, some of the solutions like [10] and [11] are addressing the management of the resources in a coarse level for all users. Only [14] provides a personalized solution for end users, but since the paper focuses on QoE it requires that the User Equipment (UE) should collect a lot of information and transfer it regularly to the network for further processing. This requires a lot of processing power in the UE as well it may affect the battery level consumption. Moreover, the exchange of a significant amount of data over the wireless link, by a large number of UEs, may cause performance issues to the overall network. In the next section, we will describe how [4] can provide improved and personalized services to end users while at the same time eliminating the burden on the end devices and the exchange of data over the wireless link.

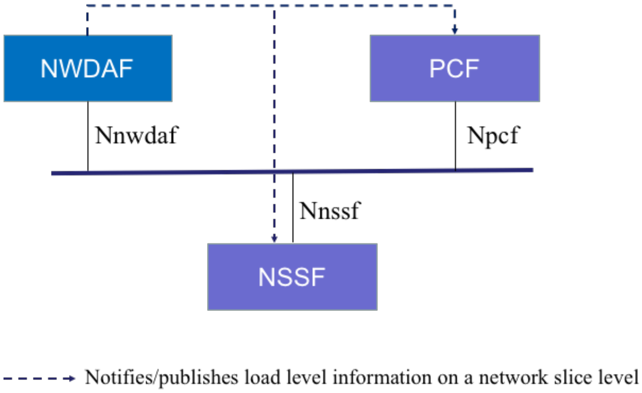

NWDAF
PCF
Nnwdaf
Npcf
Nnssf
NSSF
Notifies/publishes load level information on a network slice level

Figure 2. Data Analytics in 3GPP

B. *3GPP's NetWork Data Analytics Function (NWDAF)*

Quite recently, 3GPP has identified the need to incorporate a dedicated data analytics function (NWDAF) in the latest specifications [3] [15] [16]. This entity represents an operator's managed network analytics logical function.

As shown in Figure 2, NWDAF provides slice specific network data analytics to the Policy Control Function (PCF) and the Network Slice Selection Function over their newly specified interfaces (i.e., Nnwdaf, Nnssf and Npcf). Interestingly enough the latest specifications describe that NWDAF will provide only load level information on a network slice level and that it is not required to be aware of the current subscribers using a slice. PCF uses the received information to select policy rules whereas NSSF may use the same information for selecting the most suitable slice for a UE. Note here that this type information can be collected from the network management system. Also, in Release 14, the RAN Congestion Awareness Function (RCAF) was used to inform the PCF about the congestion in RAN. Thus, 3GPP essentially is currently using NWDAF as a placeholder for future releases. In these, additional information will be provided to PCF and NSSF. This information will be related to the type of UEs or even for specific UEs, since PCF and NSSF are used to control the placement and treatment of UEs in the appropriate cell and radio access technology. Already, the output of NWDAF is considered to feed new network components related to access traffic steering, switching and splitting schemes (ATSSS) as reported in [17].

III. Integrating the CEPE framework with the 5G Architecture

A. *The Context Extraction and Profiling Engine (CEPE)*

The work presented in [4], is able to automatically build a user profile that can be used to predict the future behavior of a subscriber. This information is used to improve the performance of network control operations. More specifically, static and dynamic information is collected about:

1. **User profile related information (static):** gender, device type characteristics (e.g., cpu, memory, os, device type)
2. **UE and device dynamic characteristics (dynamic):** location, transmission power, amount of transmitted and received data, experienced delay, loss of packets, associated cells identifiers
3. **Network related measurements (static and dynamic):** type of cell (e.g., macro, femto, etc.), power transmission level, available resource blocks, amount of transmitted and received, data, delay, packet loss, number of connected devices

Based on these measurements, the authors use data mining to build a knowledge model, the outcome of which is essentially a dynamic profile for end users. This profile predicts their future behavior based on their location, time and day, the battery level of their devices and their monetary charging status. This way the network can use this information to place users to the appropriate cells and radio access technologies during the execution of a handover or a new session establishment. Also, this information is used by the end devices to select the most suitable cell to camp on, when they are in an idle state. Extensive simulations demonstrate significant performance improvements both for the network operator as well the end users. Note here that this is exactly the information (i.e., the dynamic profile of users that captures their future behavior in terms of mobility and service consumption), that the newly introduced NWDAF can report to the PCF and other network components to improve the performance of a 5G network. Thus, CEPE is essentially potential future evolution of NWDAF.

The next subsection presents in detail how the CEPE framework can be mapped in the latest 5G architecture by extending the interfaces connecting today NWDAF to multiple network functions. It also presents which information is collected from the network functions as well and the control functions that will receive the outcome from the data mining model.

B. Mapping the CEPE framework in the 5G Architecture

The introduction of NWDAF in the 5G architecture and its interfacing with the PCF clearly indicates that in the future, its output will be used to select the most appropriate policies for UEs or types of UEs (e.g., high/low moving terminals, terminals involved in high/low data rate exchange, etc.).

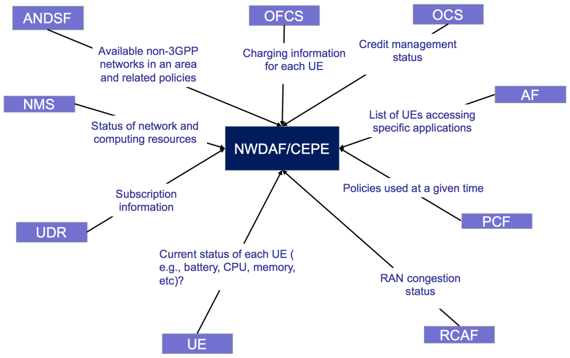

ANDSF
OFCS
OCS
Available non-3GPP networks in an area and related policies
Charging information for each UE
Credit management status
NMS
AF
List of UEs accessing specific applications
Status of network and computing resources
NWDAF/CEPE
Subscription information
Policies used at a given time
UDR
PCF
Current status of each UE (e.g., battery, CPU, memory, etc)?
RAN congestion status
UE
RCAF

Figure 3 Provision of information to NWDAF

In the state of art research efforts (e.g., [4], [13]), it has been identified that users tend to have the same behavior that is dependent on the terminal they use, their monetary charging status and preferences, the status of their battery and obviously their location (e.g., home office, on the road, etc.) during specific dates and hours. Their behavior also depends on the status of network components (e.g., the load of the network, the received signal strength, the experienced delay or packets losses). All this information is required for a data mining function like NWDAF/CEPE to create a realistic model and enable the selection of appropriate policies or fine tune network control functions. In Figure 3, the network components that have to feed information to the NDWAF/CEPE as well as the type of this information are illustrated. These 5G network components are:

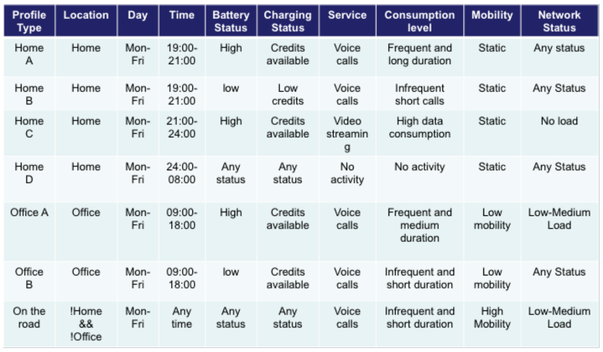
- **Unified Data Repository (UDR):** provides subscription information about a UE.
- **Network Management System (NMS):** reports performance indicators (e.g., bandwidth utilization, packet loss, latency, alerts, etc.).
- **Access Network Discovery and Selection Function (ANDSF):** reports the current policy rules shared with a UE to help it decide to which available WiFi it may connect to.
- **Offline Charging System (OFCS):** passes information contained in Charging Data Records (CDRs) that are related to the resource usage of a UE.
- **Online Charging System (OCS):** informs about the current credit management status of a UE.
- **Application Function:** is able to inform which services (even from those not being owned by the operator) are being used by a UE. As specified in [3], the communication of AF with the network operator's components takes place via the Network Exposure Function (NEF).
- **Policy Control Function (PCF):** provides information about the current session management policies being used in the network.
- **Radio Congestion Awareness Function (RCAF):** provides RAN user plane congestion information.
- **User Equipment (UE):** provides real time information about the current battery level of a UE.

Based on this information, the NWDAF/CEPE can collect information that is related to the current behavior of users and create their dynamic profile that essentially is a prediction of their future actions. Table I presents some examples of behavioral profiles that can characterize a single user or a set of users that have the same behavior. To create such a list the NWDAF/CEPE requires mainly information from the UE, the OFCS the OCS, the UDR and the AF (optionally). It also requires information from the NMS and the RCAF.

At the same time, the NWDAF/CEPE should have the information about the current policies being used by the operator as it receives the related information from the PCF and ANDSF. This way it is able to correlate the received input and identify a) the best policies to be used, b) the estimated bandwidth required for a future period of time in an area and c) what is the optimum placement of UEs in cells and radio access technologies. This is doable since the network is aware of the type of users in an area, their number (from those that are connected or those that have recently performed a location update process) and the available capacity of the network. Thus, NWDAF/CEPE can provide rules to PCF in the form of

- Profile (Home C) ^ # of users (high) ^ Network load (high) → place users in femto cells
- Profile (On the road) ^ # of users (any) ^ Network load (any) → place users in macro cells

TABLE I. EXAMPLES OF BEHAVIORAL PROFILES



Profile Type	Location	Day	Time	Battery Status	Charging Status	Service	Consumption level	Mobility	Network Status
Home A	Home	Mon-Fri	19:00-21:00	High	Credits available	Voice calls	Frequent and long duration	Static	Any status
Home B	Home	Mon-Fri	19:00-21:00	low	Low credits	Voice calls	Infrequent short calls	Static	Any Status
Home C	Home	Mon-Fri	21:00-24:00	High	Credits available	Video streamin g	High data consumption	Static	No load
Home D	Home	Mon-Fri	24:00-08:00	Any status	Any status	No activity	No activity	Static	Any Status
Office A	Office	Mon-Fri	09:00-18:00	High	Credits available	Voice calls	Frequent and medium duration	Low mobility	Low-Medium Load
Office B	Office	Mon-Fri	09:00-18:00	low	Credits available	Voice calls	Infrequent and short duration	Low mobility	Any Status
On the road	!Home && !Office	Mon-Fri	Any time	Any status	Any status	Voice calls	Infrequent and short duration	High Mobility	Low-Medium Load

The behavioral profiles of users can be communicated to any network functions that are responsible for managing the user mobility or the establishment and management of user sessions. Such entities are:

- **Access and Mobility Management Function (AMF):** supports mobility management, access authentication and authorization, security anchor functions and context management. The behavioral profile can be used to fine tune the location update procedure (e.g., its frequency, the tracking area list, etc.).
- **Session Management Function (SMF):** supports session management, selection and control of UP functions, downlink data notification and roaming. The behavioral profile can be used to select the most appropriate user plane path.
- **Traffic Steering Support Functions (TSSF):** receives traffic steering control information from the PCF, to steer traffic towards specific WiFis. The behavioral profile can affect these steering decisions.
- **5G Base Station (gNB):** provides user plane and control plane protocol terminations towards a UE. The behavioral profile can be used to fine tune the admission control and handover procedures as well as the information broadcasted to the UEs to assist them selecting the best cell to camp on.

The next section presents an approximation of the amount of data that is required to be exchanged and suggests how this can be minimized.

IV. QUANTITATIVE ANALYSIS

The analysis in Section III indicates that the sources providing data to the NWDAF can be distinguished in four categories. The first one consists of sources that provide static or rarely changed information. In this category belongs ANDSF and PCF that provide the list of active policies in a network and UDR that contains user related subscription information. Since this information does not change often limited actions can take place in order to minimize their communication to NWDAF/CEPE.

The second category contains information from the NMS and the RCAF that have to be reported to the NWDAF either at regular intervals or whenever there is a specific event (e.g., a threshold violation). This information is required for NWDAF to create a knowledge model that correlates the number of specific types of UEs in a specific area, the applied policies in this area and the performance of the network. Again, this information cannot be easily avoided or minimized.

The third group of information is related to the battery level of UEs (that influence the usage from the users) and the accessed services from the application servers. This information can be considered as optional since the behavior based on the remaining battery level can be inferred by information available in the fourth category, whereas the accessed services information is useful for the operator to fine tune the support of the services (e.g., video caching schemes). The final category contains information available at OFCS and OCS [18]. The CDRs contain essentially all the information needed to create the user behavioral profiles while avowing any extra communication of network components with the UEs. Table II contains the required parameters and their size. Note that this is a simplified version of the overall list since as specified in [19], several parameters are duplicated based on the network used (GSM, UMTS, LTE) as well as the specific services (circuit switched, packet switched, IP Multimedia Subsystem – IMS, etc.)

TABLE II. CDR PARAMETERS

Parameter	Size
1. International Mobile Subscriber Identity (IMSI)	64 bits
2. International Mobile Equipment Identity (IMEI)–*Which device a user is using*	64 bits
3. Timestamp	32 bits
4. Call duration (CS)	16 bits
5. Cell Identifier	20 bits
6. RAT Type-*which Radio access technology was used*	8 bits
6. Duration (PS)	16 bits
7. Data Volume Downlink	16 bits
8. Data Volume Uplink	16 bits
9. Record Opening time	32 bits
9. Change Condition – *e.g., user location change*	5 bits
10. QoS Profile – *requested/negotiated*	128 bits
11. Service Identifie	32 bits
13. Traffic Steering Policy Uplink	8 bits
14. Traffic Steering Policy Downlink	8bits
15. User location information	8 bits
16. User location information time	32 bits
Total	505 bits

Note that to identify the level of mobility of a user NWDAF/CEPE can simply process the information about cell identifiers, and user location that is contained in the parameters of Table II. This way neither the UE has to monitor and report its mobility level, nor the network components perform any complex functions to estimate it.

The analysis above indicates that based on this small amount of data per UE, the behavioral profile of end users can be created and improve the control performance of a mobile network [4]. Also, this process is transparent to the UEs and is based on information already available to the operators. Overall, the data that has to be collected from all network components is captured the following equation:

$$\text{Total} = \sum_{k=1}^{n} \text{CDR}_{\text{k}} * \text{T}_{\text{k}} + \sum_{l=1}^{\mu} NMS_{\text{NCl}} * \text{T}_{\text{l}} + RCAF_{\text{data}} * T_{\text{r}}$$

Where CDR_k and T_k are the information of Table I for every UE and their transmission rate, NMS_NCl and T_l are the network related information for the different network components and its correspondent transmission rate and RCAF_data and T_r indicates the information transmitted by RCAF. From the above discussion only the first part of the equation can be minimized. This can be achieved only if the behavioral profiles of the UEs are communicated to the OFCS. When the received profile of the users is consistent with their current recorded behavior, no additional information needs to be transmitted back to NWDAF/CEPE. The gain in efficiency of such a scheme can be deduced from an example with basic parameters, as presented in Figure 4. In the figure, the performance gain is compared in a network of 1-10 million users using scenarios where data is sent every 30’ to the NWDAF/CEPE without optimization versus optimization via several assumptions on “profile consistency”. The assumptions on “profile consistency” include scenarios where data is transmitted only when user behavior is inconsistent with the received profile 30%, 60%, and 90% of the time.

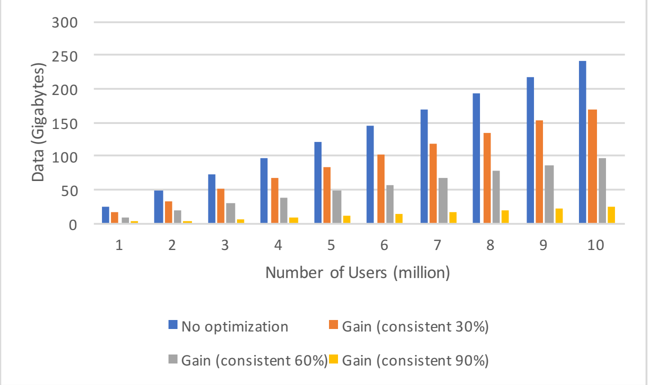


Figure 4. Comparison of data transfer schemes

V. Conclusions

This paper discusses why data mining is going to be a key enabler for 5G networks by providing a short survey of existing proposals. Also, based on the latest progress of 3GPP it analyses why the newly introduced NWDAF is not adequate, in the current version, to support data mining.

Moreover, it presents in detail which information is required to be collected and reported to NWDAF/CEPE and quantifies the amount of information that is required. It also, reports which entities of the 5G architecture could exploit of the knowledge created by a data mining framework.

References

[1] 5G-PPP Architecture WG, “View on 5G Architecture”, Version 2.0, July 2017, https://5g-ppp.eu/wp-content/uploads/2017/07/5G-PPP-5G-Architecture-White-Paper-2-Summer-2017_For-Public-Consultation.pdf, (accessed 15-12-2017).

[2] M. De Sanctis, I. Bisio and G. Araniti, “Data Mining Algorithms for Communication Network Control: Concepts, Survey and Guidelines”, IEEE Network Magazine, January/February 2016.

[3] 3GPP, TS 23.501 “System Architecture for the 5G System; Stage 2 (Release 15)”, Version 2.0.1, December 2017.

[4] P. Magdalinos, S. Barbounakis, P. Spapis, A. Kaloxylos et. al., “A context extraction and profiling engine for 5G network resource mapping”, Computer Communications, 108, pp. 184-201, 2017.

[5] Y. He, F. R. Yu, N. Zhao, H. Yin, H. Yao, and R. Qiu, “Big Data Analytics in Mobile Cellular Networks”, IEEE Access, Vol. 4, 2016.

[6] S. Robitsch, F. Zaman, S. van der Meer, J. Keeney, G. Muntean, “Magnet: Real-Time Trace Stream Analytics Framework for 5G Operations Support”, IEEE Network Magazine September/October 2017.

[7] L. I. B. Lopez, J. M. Vidal, L. J. G. Villalba, “An Approach to Data Analysis in 5G Networks”, Entropy Vol. 19, Issue 2, 2017.

[8] A. Imran, A. Zoha, and A. Abu-Dayya, “Challenges in 5G: How to Empower SON with Big Data for Enabling 5G”, IEEE Network Magazine, November/December 2014.

[9] K. Zheng, Z. Yang, K. Zhang, P. Chatzimisios, K. Yang, and W. Xiang, “Big Data-Driven Optimization for Mobile Networks toward 5G”, IEEE Network Magazine, January/February 2016.

[10] Z. Feng, X. Li, Q. Zhang, W. Li, “Proactive Radio Resource Optimization with Margin Prediction: A Data Mining Approach”, IEEE Transactions on Vehicular Technology, Vol. 66, Issue 10, October 2017.

[11] S. Han, C. Lin, G. Li, S. Wang, and Q. Sun, “Big Data Enabled Mobile Network Design for 5G and Beyond”, IEEE Communications Magazine, Vol 55, Issue 9, July 2017.

[12] J. Liu, F. Liu and N. Ansari, “Monitoring and Analyzing Big Traffic Data of a Large-Scale Cellular Network with Hadoop”, IEEE Network Magazine, July/August 2017.

[13] V. Kolar, S. Ranu, A. Prabhu et. al., “People In Motion: Spatio-temporal Analytics on Call Detail Records“, Sixth International Conference on Communication Systems and Networks (COMSNETS), 2014.

[14] Y. Wang, P. Li, L. Jiao, Z. Su, et. al., “A Data-Driven Architecture for Personalized QoE Management in 5G Wireless Networks”, IEEE Wireless Communications, Vol. 24, Issue 1, February 2017.

[15] 3GPP, TS 23.502, “Procedures for the 5G System”, Stage 2 (Release 15), Version 1.3.0, November 2017.

[16] 3GPP, TS 23.503, “Policy and Charging Control Framework for the 5G system”, Stage 2 (Release 15), December 2017.

[17] 3GPP, TR 23.793, “Study on Access Traffic Steering, Switching and Splitting support in the 5G System Architecture”, Release 15, Version 0.1.0, August 2017.

[18] 3GPP, TS 32.240, “Charging architecture and principles”, Release 14, Version 14.4.0, June 2017.

[19] 3GPP, TS 32.298, “Charging Data Record (CDR) parameter description”, Release 15, Version, 15.0.0, September 2017.

BTOOLS: Trusted Transaction Generation for Bitcoin and Ethereum Blockchain Based on Crypto Currency SmartCard

Pascal Urien
LTCI
Telecom ParisTech
France
Pascal.Urien@telecom-paristech.fr

Mesmin Dandjinou
Ecole Supérieure d'Informatique
Université Nazi BONI
Burkina Faso
Tmesmin.dandjinou@univ-bobo.bf

Kodjo Edem Agbezoutsi
Ecole Supérieure d'Informatique
Université Nazi BONI & LTCI
Burkina Faso
Kodjo.agbezoutsi@telecom-paristech.fr

Abstract— **This paper presents an innovative and open software framework whose goal is to increase the trust of blockchain transactions. Transactions are signed by the Elliptic Curve Digital Signature Algorithm (ECDSA) associated with a 32 bytes secret private key. We designed a Javacard application used for key generation, storage and cryptographic procedure dealing with the secp256k1 elliptic curve. Our open software BTOOLS generates Bitcoin and Ethereum transactions whose trust is enforced by the support of a Crypto Currency SmartCard (CCSC).**

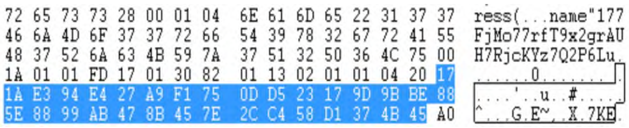
Keywords-. Blockchain; Bitcoin; Ethereum; Trust.

I. INTRODUCTION

The Bitcoin crypto currency was introduced in 2008 [1], in a famous paper written by an anonymous author Satoshi Nakamoto. This paper proposes "*a solution to the double-spending problem using a peer-to-peer distributed timestamp server to generate computational proof of the chronological order of transactions...The steady addition of a constant of amount of new coins is analogous to gold miners expending resources to add gold to circulation in our case, it is CPU time and electricity that is expended*"

Satoshi Nakamoto also wrote the win32 software Bitcoin.exe [3], about 16,000 lines of C++ code, and 6 MB binary size. This software realizes all the functions needed by the Bitcoin blockchain [2]. It manages four major tasks:

- Generation of transactions which are signed according to the *Elliptic Curve Digital Signature Algorithm*, dealing with elliptic curve private keys;
- Communication with other Bitcoin nodes running the Bitcoin application;
- Block mining;
- Blockchain management.



```
72 65 73 73 28 00 01 04  6E 61 6D 65 22 31 37 37  ress(...name"177
46 6A 4D 6F 37 37 72 66  54 39 78 32 67 72 41 55  FjMo77rfT9x2grAU
48 37 52 6A 63 4B 59 7A  37 51 32 50 36 4C 75 00  H7RjcKYz7Q2P6Lu.
1A 01 01 FD 17 01 30 82  01 13 02 01 01 04 20 17  .......0........
1A E3 94 E4 27 A9 F1 75  0D D5 23 17 9D 9B BE 88  ....'..u..#.....
5E 88 99 AB 47 8B 45 7E  2C C4 58 D1 37 4B 45 A0  ^...G.E~,.X.7KE.
```

Figure 1. The wallet.dat, a database file from Bitcoin.exe PrivateKey: 171AE394E427A9F1750DD523179D9BBE885E8899AB478B457E2CC458D1374B45. BtcAdr: 177FjMo77rfT9x2grAUH7RjcKYz7Q2P6Lu

The Bitcoin application maintains a set of data files managed by a non Structured Query Language (SQL) database, the *Berkeley Database (Berkeley DB)* [14]. In particular, the private keys are stored in the file named wallet.dat. As illustrated in Figure 1, private keys are stored in clear text in the database file.

Because all crypto currency legitimate transactions rely on private keys, their secure storage and trusted use is a major prerequisite for blockchain operations. As an illustration, the Korean Exchange *Youbit* declared bankruptcy in December 2017 after the hacking of 17% of its Bitcoin reserves, about 4,700 Bitcoins [17].

Our researches attempt to increase trust of blockchain operations, by using secure elements, enforcing secure key storage and trusted ECDSA signature. In order to reach this goal, we developed the BTOOLS (*Blockchain Tools*) open software [12], able to generate *Bitcoin* or *Ethereum* transactions, whose signature is computed by a dedicated *Crypto Currency SmartCard*, i.e. a Javacard running a Java application.

BTOOLS uses OPENSSL library and smartcard, for cryptographic operations. It provides the following services:

- Bitcoin address generation (mainnet and testnet);
- Ethereum address generation;
- Bitcoin transaction generation;
- Ethereum transaction generation;
- Simple Bitcoin node client;
- Bitcoin transaction (via the Bitcoin client or WEB APIs);
- Ethereum transaction (via WEB APIs);
- Crypto Currency SmartCard scripts for key generation and transaction signature.

The paper is constructed according to the following outline. Section 2 recalls basic notions for the generation of ECDSA signatures over elliptic curves. Section 3 details Bitcoin transactions and dedicated BTOOLS scripts. Section 4 describes Ethereum transactions and BTOOLS dedicated scripts. Section 5 introduces Crypto Currency SmartCard and its use with BTOOLS software. Finally, Section 6 concludes this paper.

ISBN: 978-1-61208-623-1

II. About the ECDSA Signature

Most crypto moneys (Bitcoin, Ethereum...) use the secp256k1 elliptic curve, whose parameters are as follow [4]:

- The p prime characteristic of the field Z/pZ, defined as:

$$p = 2^{256} + 2^{32} + 2^{9} + 2^{8} + 2^{7} + 2^{6} + 2^{4} + 1$$

- The elliptic curve E defined as the set of points (x,y) satisfying the relation:

$$y^{2} = x^{3} + 7,\ x,y \in Z/pZ$$

- The generator G uncompressed (i.e. x and y) form which is:

```
04
79BE667E F9DCBBAC 55A06295 CE870B07
029BFCDB 2DCE28D9 59F2815B 16F81798
483ADA77 26A3C465 5DA4FBFC 0E1108A8
FD17B448 A6855419 9C47D08F FB10D4B8
```

- The n order (i.e. the number of group elements) of the curve defined as:

```
FFFFFFFF FFFFFFFF FFFFFFFF FFFFFFFE
BAAEDCE6 AF48A03B BFD25E8C D0364141
```

- Finally, the cofactor is equal to 1, which means that there is only one group in E whose order is the prime n.

A. ECDSA Signature

An ECDSA signature over E [5] is a couple of two integers (r, s) such as :

Given x Є [1, n-1] the private key i.e. a 32 bytes random number, P= xG is the public key.

k is an ephemeral key, k Є [1, n-1]

kG= (x_R, y_R), and r = x_R mod n

size = number of bytes of n (size = 32)

H a hash function, e = H(M) is the hash of a message M, i.e. a set of bytes to be signed.

If H(M) has more bytes than size, then take e as the size leftmost bytes.

The couple (r, s), with $s = k^{-1}\ (e + x\ r)\ mod\ n$ is the signature.

B. ECDSA Signature Verification

Let the signature being (r, s).

Given the message M and H, compute e = H(M).

size = number of bytes of n (size =32). If e has more bytes than size, take the size leftmost bytes.

1) Compute $u_1 = es^{-1}$ mod n and $u_2 = rs^{-1}$ mod n.

2) Compute R = (x_R, y_R) = $u_1G + u_2P$.

3) Set v = x_R mod n.

Compare v and r, and if v = r the signature is valid.

C. Canonical Signature

For a given ECDSA signature, (r, s), the signature (r, n-s) is also valid. The canonical signature is computed according to the following algorithm:

1) Compute n-s =t.

2) If s < t, then (r, s) is canonical signature.

3) Otherwise, (r, t) is the canonical signature.

Bitcoin and Ethereum blockchains request canonical signature.

D. Public Key Recovery from ECDSA Signature

Given the ECDSA [5] signature (r, s).

Find the *"positive"* point R(x=r, y=y+) on the E: $y^2=x^3+7$ curve,

Given the message M and H, compute e= H(M).

size = number of bytes of n (32). If e has more bytes than size, take the size leftmost bytes.

Compute the candidate public key $Q = r^{-1}(sR - eG)$.

Check the signature (r, s), and if valid set recovery to 27 in Ethereum.

If not verified, try with the *"negative"* point –R= (x=r, y=y-), and if valid, set recovery to 28 in Ethereum.

III. Bitcoin transactions

A. Bitcoin Address

Bitcoin addresses (BA) are computed from ECDSA public key. A private key, i.e. a 32 byte number x, is generated, according to a true random number generator (TRNG). Thereafter, a public key is computed according to the relation P = xG. The uncompressed form uF(P) is a set of 65 bytes {4, x_P, y_P}, a prefix (one byte 0x04) and a point (x_P, y_P) of the curve (2x32 bytes, in Z/pZ).

The Bitcoin address [2] is computed according to the following procedure:

1) a1 = SHA256(u_F(P)), 32 bytes

3) hash160= a2 = RIPEMD160(a1), 20 bytes

3) a3 = Network-ID || a2, 25 bytes

4) a4 = SHA256(SHA256(a3)), 32 bytes

5) a5= checksum = 4 rightmost bytes of a a4

6) a6 = a4 || a5, 25 bytes

7) Bitcoin address = a7 = encoding of a6 in base 58

The base 58 encoding uses the following digits {1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F, G, H, J, K, L, M, N, P, Q, R, S, T, U, V, W, X, Y, Z, a, b, c, d, e, f, g, h, i, j, k, m, n, o, p, q, r, s, t, u, v, w, x, y, z}.

A BA is protected by a four bytes checksum; although the hash160 parameter has no checksum, it is used in transactions as payee's address.

Figure 2 illustrates the generation of Bitcoin address by BTOOLS.

```
btools -genmain
PrivateKey:
CE1DBAFD7D2E8983ED60E0E081632EB062737B1B1627AAAB276F2E037
A74A081
PublicKey:
04CFD7A542B8C823992AF51DA828E1B693CC5AB64F0CACF0F80C31A1E
CA471786E285BDD3F1FE0A006BD70567885EF57EB149C8880CB9D5AF3
04182AC942E176CC
Hash160: CB643DD608FB5C323A4A6342C1A6AC8048B409EB
BTC-Adr: 1KYSFr6CyTDMruu8wna981M4ziVyMwftcg
Double SHA2 Check OK
ID: 00
Hash160: CB643DD608FB5C323A4A6342C1A6AC8048B409EB
BTC-WIF:
5KP4YMxDzfv9P1WVAPZqHRSfi5FydGqqqRjr5oPvskpwTq59wiX
```

Figure 2. Generation of a Bitcoin address by the BTOOLS software

B. Bitcoin Transaction

A transaction is a list of inputs, associated to Bitcoin amounts (i.e. coins), and a list of outputs to which are transferred the totally of inputs. A fee is allocated to the miner if the sum of outputs is less than the sum of inputs.

A fee is usually expressed in satoshi per byte (1 satoshi = 10^{-8} Bitcoin (BTC)); since July 2017 it is expressed in weight units/byte. At the time of writing, the fee was ranging between 50,000 and 100,000 satoshi (0,0005 to 0,001 BTC).

The structure of a transaction is detailed in Figure 3.

In every input, a coin value for *Unspent Transaction Output* (UTXO) is identified by a previous transaction identifier and its output index (starting from 0). A transaction ID is equal to the double SHA256 hash of the binary content of the transaction. A signature script (*sigScript*) contains the ECDSA signature and public key of the payer's transaction.

Every output comprises an amount expressed in satoshi, and a public key script (*pubKeyScript*) including the payee's hash160 address.

Parameter	**Type**	**Comment**
version	integer 32 bits	always 1
number of inputs	var_int 1 byte or more	
One or more inputs		
transactionID	32 bytes	coin transaction
index	integer 32 bits	coin index >=0
sigScript length	1 byte	
sigScript	contains the signature and the public key	
sequence	integer 32 bits FFFFFFFF=ignore	transaction version
End of input		
number of outputs	var_int 1 byte or more	
One or more outputs		
value	integer 64bits	satoshi amount
pubKeyScript length	1 byte	
pubKeyScript		
locktime	integer 32 bits 00000000=ignore	transaction locktime

Figure 3. Structure of a Bitcoin transaction

```
01000000 // Version
01 // number of inputs
DE2D211EF429909B0AB8D2E7D25826A0 //TransactionID
EDD6281EC6DEDF2B822CE5014A349E72
01000000 // index
8A // length of the signature Script
47 // ECDSA Signature length
30 44 // Sequence of (r, s) integer values
02 20 // integer r value
0772ABD5D37D0CAAB881DBC8912628F9
3461839CC8D4BC007A355831A6061ED7
02 20 // integer s value
4CCCC34B34A9075FC09C9777EAB7A6F5
612DA2130C1FF1C0E376AD9B2209D51D
01 41 // Public key length
04 // uncompressed format
CFD7A542B8C823992AF51DA828E1B693
CC5AB64F0CACF0F80C31A1ECA471786E
285BDD3F1FE0A006BD70567885EF57EB
149C8880CB9D5AF304182AC942E176CC
FFFFFFFF // sequence
01 // number of outputs
D418040000000000 // amount in satoshi
19 // Public Key Script
76 // OP_DUP
A9 // OP_HASH160
14 // hash160 length
CB643DD608FB5C323A4A6342C1A6AC8048B409EB
88 // OP_EQUALVERIFY
AC // OP_CHECKSIG
00000000 // Locktime
```

Figure 4. Binary encoding of a Bitcoin transaction

Figure 4 presents a binary dump of a transaction using a *pay-to-pubkey-hash* script; it should be noticed that all values are encoded according the a *little endian* format.

The *pay-to-pubkey-hash* script is defined as:

OP_DUP [76] OP_HASH160 [A9]
<length=14><hash160>
OP_EQUALVERIFY[88] OP_CHECKSIG[AC]

The ECDSA signature is encoded using the following ASN.1 structure (see for example RFC 3279 [15]):

Ecdsa-Sig-Value ::= SEQUENCE {
r   INTEGER,
s   INTEGER }

```
01000000 // Version
01 // number of inputs
DE2D211EF429909B0AB8D2E7D25826A0 // Transaction ID
EDD6281EC6DEDF2B822CE5014A349E72
01000000 // index
00 // vi= length of the Signature Script
FFFFFFFF // sequence
01 // number of outputs
D418040000000000 // amount in satoshi
19 // Public Key Script (Pk script)
76 // OP_DUP
A9 // OP_HASH160
14 // hash160 length
CB643DD608FB5C323A4A6342C1A6AC8048B409EB
88 // OP_EQUALVERIFY
AC // OP_CHECKSIG
00000000 // Locktime
01000000 // hash Type
```

Figure 5. Binary dump of a raw Bitcoin transaction

The signature computing is performed according to the following procedure:

1) Build a raw transaction (see Figure 5), in which, for every input, the sigScript is removed, i.e. the length value (vi) is set to zero.

2) For every input:

2.1) Copy the *pay-to-pubkey-hashScript* in the *sigScript* location, and modify the length (initially set to 0) accordingly (length =25 in decimal).

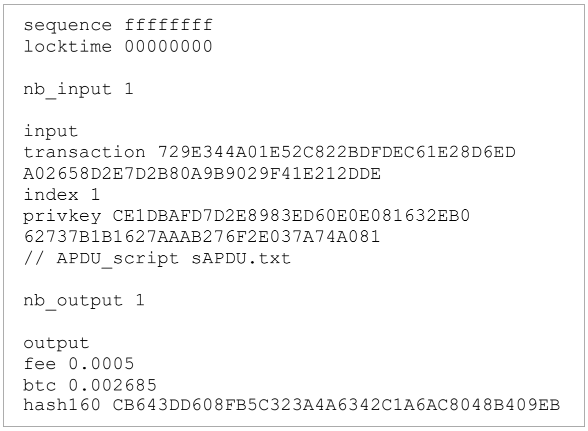
2.2) The hash160 inserted in *pay-to-pubkey-hash* is computed from the payer's public key.

2.3) Compute the double SHA256 of the modified transaction.

2.4) Generate the ECDSA signature with the payer's private key.

2.5) Insert the final *sigScript* in the input, and modify the length accordingly.

C. BTOOLS Bitcoin Script



```
sequence ffffffff
locktime 00000000

nb_input 1

input
transaction 729E344A01E52C822BDFDEC61E28D6ED
A02658D2E7D2B80A9B9029F41E212DDE
index 1
privkey CE1DBAFD7D2E8983ED60E0E081632EB0
62737B1B1627AAAB276F2E037A74A081
// APDU_script sAPDU.txt

nb_output 1

output
fee 0.0005
btc 0.002685
hash160 CB643DD608FB5C323A4A6342C1A6AC8048B409EB
```

Figure 6. A Bitcoin transaction script in BTOOLS

Bitcoin transactions are generated thanks to a script; the Crypto Currency SmartCard can be used to compute the ECDSA signature.

A script is a set of lines. A comment line begins by the '/' or '*' character. It defines *sequence* and *locktime* values (in hexadecimal Most Significant Bit (MSB) encoding).

The number of inputs is specified by the *nb_input* field. Each input must begin by the input field; it comprises:

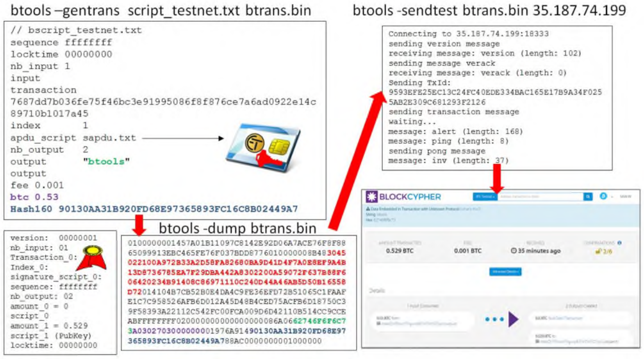
- a transaction identifier (32 bytes, hexadecimal MSB encoding);
- an index (decimal encoding);
- and a choice between the following fields :
  - privkey [private key hexadecimal MSB encoding],
  - wif [WIF],
  - APDU_script [the name of a smartcard script].

The number of outputs is specified by the *nb_output* field. Each output must begin by the output field; it comprises:

- an optional fee in decimal format, to be subtracted from the BTC (i.e. UTXO in most case) value of the current output; the character '.' is used as decimal separator;
- a BTC amount in decimal format, the character '.' is used as decimal separator;
- and a choice between the following fields:
  - adr [Bitcoin address],
  - hash160 [hash160, hexadecimal MSB encoding].

A Bitcoin transaction script is detailed in Figure 6.

D. Sending transaction to the Bitcoin blockchain



btools –gentrans script_testnet.txt btrans.bin

```
// bscript_testnet.txt
sequence ffffffff
locktime 00000000
nb_input 1
input
transaction
7687dd7b036fe75f46bc3e91995086f8f876ce7a6ad0922e14c
89710b1017a45
index       1
apdu_script sapdu.txt
nb_output   2
output      "btools"
output
fee 0.001
btc 0.53
Hash160 90130AA31B920FD68E97365893FC16C8B02449A7
```

btools -sendtest btrans.bin 35.187.74.199

```
Connecting to 35.187.74.199:18333
sending version message
receiving message: version (length: 102)
sending message verack
receiving message: verack (length: 0)
Sending TxId:
9593EFE25EC13C24FC40EDE334BAC165E17B9A34F025
5AB2E309C681293F2126
sending transaction message
waiting...
message: alert (length: 168)
message: ping (length: 8)
sending pong message
message: inv (length: 37)
```

btools -dump btrans.bin

```
version:  00000001
nb_input: 01
Transaction_0:
Index_0:
signature_script_0:
sequence: ffffffff
nb_output: 02
amount_0 = 0
script_0
amount_1 = 0.529
script_1 (PubKey)
locktime: 00000000
```

```
0100000001457A01B11097C8142E92D06A7ACE76F8F88
65099913EBC465FE76F037BDD8776010000008B483045
022100A972B33A2D58FA82680BA9D41D4F7A0E8EF9A4B
13D8736785EA7F29DBA442A8302200A59072F637B88F6
06420234B91408C86971110C240D44A46AB5D50B16558
D72014104B7CB52B0E4DA4C9FE36EFD72B51065C1FAAF
E1C7C958526AFB6D012A45D48B4CED75ACFB6D18750C3
9F58393A22112C542FC00FCA009D6D42110B514CC9CCE
ABFFFFFFFF0200000000000000000086A0662746F6F6C7
3A030270300000000001976A91490130AA31B920FD68E97
365893FC16C8B02449A788AC00000000001000000
```

Figure 7. Using BTOOLS for sending Bitcoin transaction

1) Bitcoin protocol

The Bitcoin blockchain supports a protocol running over the TCP port 8333. Some Web sites list the Bitcoin nodes available over the world, for example:

https://bitnodes.earn.com/

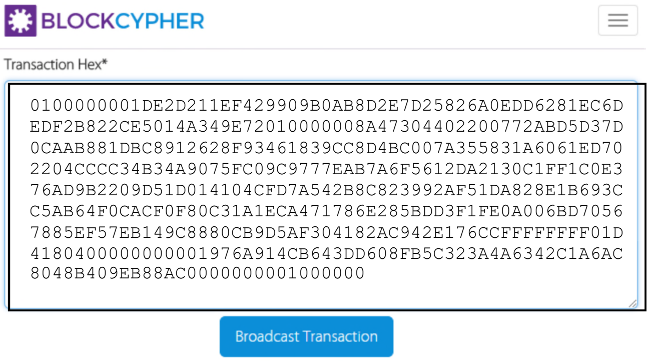
The structure of Bitcoin messages is detailed in [7][10]. The connection to a Bitcoin node requires a four way handshake, client and server exchange two *version* messages and their acknowledgment (*verack*). Afterwards, the transaction is forwarded thanks to the *tx* message.

As illustrated in Figure 7, BTOOLS realizes these operations according to the command line:

btools -sendmain transaction.bin BitcoinNode

2) Web APIs

Many full Bitcoin nodes support WEB interfaces and associated APIs. As illustrated in Figure 8 the hexadecimal representation of the transaction can be simply cut and paste in a dedicated HTML form.



Transaction Hex*

```
0100000001DE2D211EF429909B0AB8D2E7D25826A0EDD6281EC6D
EDF2B822CE5014A349E72010000008A47304402200772ABD5D37D
0CAAB881DBC8912628F93461839CC8D4BC007A355831A6061ED70
2204CCCC34B34A9075FC09C9777EAB7A6F5612DA2130C1FF1C0E3
76AD9B2209D51D014104CFD7A542B8C823992AF51DA828E1B693C
C5AB64F0CACF0F80C31A1ECA471786E285BDD3F1FE0A006BD7056
7885EF57EB149C8880CB9D5AF304182AC942E176CCFFFFFFFF01D
4180400000000001976A914CB643DD608FB5C323A4A6342C1A6AC
8048B409EB88AC00000000001000000
```

Figure 8. Sending Bitcoin transaction thanks to the WEB interface https://live.blockcypher.com/btc/pushtx

IV. ETHEREUM

A. *Ethereum Address*

Ethereum addresses (EA) are computed from ECDSA public keys [8][9] . A private key, i.e. a 32 byte number x, is generated, according to a true random number generator (RNG). Thereafter a public key is computed according to the relation P=xG. The uncompressed form u'F(P) is a set of 64 bytes {x_P,y_P}, i.e. the point (x_P,y_P) of the curve (2x32 bytes, in Z/pZ).

The Ethereum address is computed according to the following procedure:

- Compute a1= Keccak(u'F(P)), a 32 byte value. SHA3 is this a subset the Keccak [13] algorithm.

- Extract a2, the 20 rightmost bytes of a1; a2 is the Ethereum address

Figure 9 illustrates the generation of Ethereum address by BTOOLS.

```
btools -geneth

PublicKey:
0477AAA9AE8ADCAAA26F930D6022E470BBC16E10AF22A5482DAB0798A
5A2C2AF52581076023A8B33D8BA6F8E7E89EC1C5F0D66B1EFFC744582
AF063187297592F6
PrivateKey:
E49344BD32802138C9A250FCEA13F6AE30E17BC945F107F05618AFC0E
D523042
Ether Address: 777A07BAB1C119D74545B82A8BE72BEAFF4D447B
```

Figure 9. Generation of Ethereum address by the BTOOLS software

B. *Ethereum Transaction*

A transaction encodes the transfer of ethers or data between two entities, identified by their address. It includes the following fields:

- The **recipient**'s address of the message
- **nonce**, a scalar value equal to the number (>=0) of transactions generated by the sender.
- **value**, a scalar value equal to the number of Wei (1 Wei=10^{-18} Ether) to be transferred to the message recipient, or in the case of contract creation, as an endowment for the newly created account.
- A **gasLimit** value, representing the maximum number of computational steps that the transaction execution is allowed to take.
- A **gasPrice** value, representing the fee the sender pays per computational step. A scalar value equal to the number of Wei to be paid per unit of gas.
- An optional **data** field. A contract creation transaction contains an unlimited size byte array specifying the EVM (*Ethereum Virtual Machine*) code for the account initialization procedure. A message call transaction contains an unlimited size byte array specifying the input data of the message.
- The ECDSA **signature**, used to identify the sender.

C. *RLP encoding*

All transaction attributes are encoding according [11] to the RLP (*Recursive Length Prefix*) syntax, which supports *string* and *list* items.

1) *String encoding*

A *string* is a byte array, it is encoded according to the following rules :

- for one byte ϵ [0x00 0x7F] : a byte value
- if the string length ϵ [0,55] : 0x80 + Length ϵ [0x80, 0xb7] || ByteArray[Length]
- 0x80: = NULL String
- if the string Length >55 : 0xb7 + Length-of-Lengh ϵ [0xb8, 0xbf] || Length-value || ByteArray[Length]

2) *List encoding*

A list is a set of items, either *list* or *string*.

- if the list Length <=55 : 0xc0 + Length ϵ [0xc0, 0xf7] || ListItems.
- if the list Length > 55 : 0xf7 + Length-of-Length ϵ [0xf8, 0xff] || Length-value || ListItems.

D. *Example of transaction*

```
F8 6B  // list length= 107 bytes
80 // nonce = null (zero value)
85 04E3B29200 // gazPrice= 21,000,000,000 Wei)
82 9C40 // gazLimit= 40,000 Wei
94 777A07BAB1C119D74545B82A8BE72BEAFF4D447B //Recipient
87 2386F26FC10000 // value= 10,000,000,000,000,000 Wei
80 // data = null
1C // signature recovery parameter = 28
A0 F1DD7D3B245D75368B467B06CAD61002 // r value
67031935B7474ACB5C74FE7D8C904097 // 32 bytes
A0 772D65407480D7C45C7E22F84211CB1A // s value
DF9B3F36046A2F93149135CADBB9385D // 32 bytes
```

Figure 10. Binary dump of an Ethereum transaction

Transaction values are expressed according to a *Big Endian* scheme. A transaction (illustrated in Figure 10) is a list of strings, encoded with the RLP syntax. The six transaction items (*nonce*, *gasPrice*, *gasLimit*, *recipient address*, *value*, *data*), are followed by the ECDSA signature dealing with a recovery value. The recovery value is used for the recovery of the sender's public key.

E. *Ethereum Raw Transaction*

```
E8 80 // list length = 40 bytes
80 // nonce = null (zero value)
85 04E3B29200 // gazPrice= 21,000,000,000 Wei)
82 9C40 // gazLimit= 40,000 Wei
94 777A07BAB1C119D74545B82A8BE72BEAFF4D447B //Recipient
87 2386F26FC10000 // value= 10,000,000,000,000,000 Wei
80 // data = null
```

Figure 11. Example of a raw Ethereum transaction

A raw transaction (see Figure 11) is the list of six items (*nonce*, *gasPrice*, *gasLimit*, *recipient address*, *value*, *data*), without the signature elements. The ECDSA signature is performed over this structure. The recovery parameter (either 0 or 1) is added to the 27 decimal value, and is needed for the extraction of the sender public key.

F. *BTOOLS script for Ethereum transaction*

A transaction script is a set of lines (see Figure 12). A comment line begins by the '/' or '*' character.

The file (see Figure 12) comprises the following elements:

- the *private key* (**privkey**) or the name of a *smartcard script* (**APDU_script**) The Bitcoin and Ethereum smartcard scripts follow the same syntax.
- the ***nonce*** field. The nonce is expressed in decimal format.
- the ***gasPrice*** field. The gasPrice, in WEI unit.
- the ***gasLimit*** field. The gasLimit, in WEI unit.
- the **to** field indicates the ether destination address. It is a 20 bytes hexadecimal value.
- the ***value*** field indicates the transaction amount, in WEI unit.
- the ***data*** field. Three options are available:
  - **data**, text (ASCII) data field
  - **datab**, hexadecimal data field
  - **dataf**, a binary file

```
privkey E49344BD32802138C9A250FCEA13F
6AE30E17BC945F107F05618AFC0ED
523042
// APDU_script sAPDU.txt

nonce     0
gasPrice  21000000000
gasLimit  40000
to 777A07BAB1C119D74545B82A8BE72BEAFF4D447B
value     1000000000000000
data
```

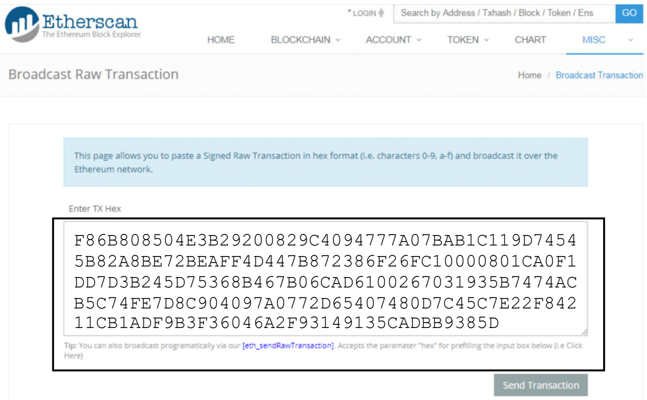
Figure 12. Illustration of an Ethereum transaction script in BTOOLS

G. *Sending a transcation to the Ethereum blockchain*

The Ethereum blockchain supports a protocol running over the TCP port 30303. Some Web sites list the Ethereum nodes available over the world, for example:

https://www.ethernodes.org

The today BTOOLS software doesn't implement the Ethereum protocol. Nevertheless many full Ethereum node support WEB interfaces and associated APIs. As illustrated in Figure 13 the hexadecimal representation of the transaction can be simply cut and paste in a dedicated HTML form.


Etherscan
The Ethereum Block Explorer
*LOGIN
Search by Address / Txhash / Block / Token / Ens
GO
HOME
BLOCKCHAIN
ACCOUNT
TOKEN
CHART
MISC
Broadcast Raw Transaction
Home / Broadcast Transaction
Send Transaction

This page allows you to paste a Signed Raw Transaction in hex format (i.e. characters 0-9, a-f) and broadcast it over the Ethereum network.

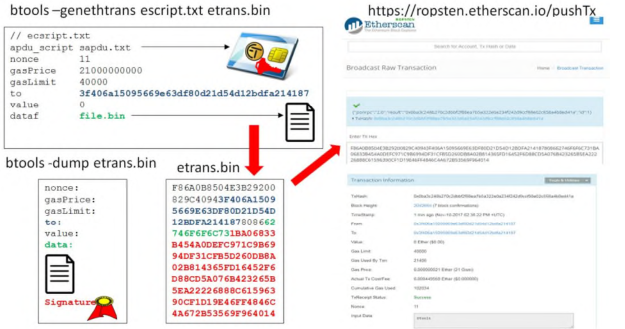
Enter TX Hex

```
F86B808504E3B29200829C4094777A07BAB1C119D7454
5B82A8BE72BEAFF4D447B872386F26FC10000801CA0F1
DD7D3B245D75368B467B06CAD6100267031935B7474AC
B5C74FE7D8C904097A0772D65407480D7C45C7E22F842
11CB1ADF9B3F36046A2F93149135CADBB9385D
```

Tip: You can also broadcast programatically via our [eth_sendRawTransaction]. Accepts the paramater "hex" for prefilling the input box below (i.e Click Here)

Figure 13. Sending an Ethereum transaction thanks to a Web API on the website https://etherscan.io/pushTx

Figure 14 illustrates an Ethereum transaction generation and forwarding thanks to BTOOLS software facilities.


btools –genethtrans escript.txt etrans.bin
https://ropsten.etherscan.io/pushTx
Etherscan
Broadcast Raw Transaction
btools -dump etrans.bin
etrans.bin
Transaction Information

```
// ecsript.txt
apdu_script sapdu.txt
nonce       11
gasPrice    21000000000
gasLimit    40000
to          3f406a15095669e63df80d21d54d12bdfa214187
value       0
dataf       file.bin
```

```
nonce:
gasPrice:
gasLimit:
to:
value:
data:
Signature
```

```
F86A0B8504E3B29200
829C40943F406A1509
5669E63DF80D21D54D
12BDFA214187808662
746F6F6C731BA06833
B454A0DEFC971C9B69
94DF31CFB5D260DB8A
02B814365FD16452F6
D88CD5A076B423265B
5EA22226888C615963
90CF1D19E46FF4846C
4A672B53569F964014
```

Figure 14. Illustration of an ether transaction generation with BTOOLS

V. CRYPTO CURRENCY SMARTCARD (CCSC)

The Crypto Currency SmartCard application (CCSC), illustrated in Figure 15, is written in Javacard, a subset of the Java language. It has three PINs: administrator, user, and user2. The default values are 8 zeros (3030303030303030) for administrator and 4 zeros (30303030) for user and user2. It is able to generate or to import elliptic curve keys (up to 8), used for the generation of ECDSA signatures used by Bitcoin and Ethereum crypto currencies. A Read/Write non volatile memory, protected by a dedicated PIN (User2), is available for the storage of any sensitive information.

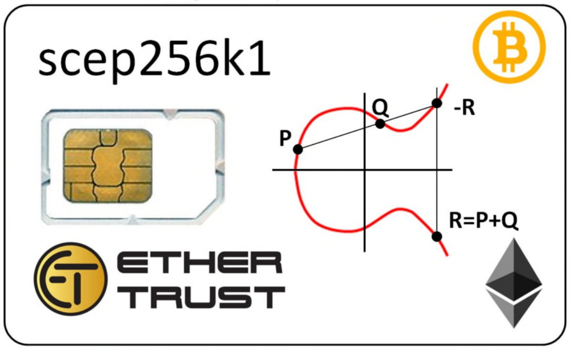


Figure 15. Illustration of a Crypto Currency SmartCard (CCSC)

The CCSC application main ISO7816 services are the following: Init Curve, Clear Key Pair, Generate Key Pair, Get Key Parameters, Set Key Parameters, Sign ECDSA.

A. *The CCSC ISO7816 interface*

According to the ISO7816-4 standard [16], a smartcard command, also called *Application Protocol Data Unit* (APDU), comprises at least five bytes named CLA, INS, P1, P2, P3; P3 is the length of data to be written or the length of information to be read. The response comprises an optional payload (up to 256 bytes) and two status bytes (SW1, SW2). The available commercial version of Javacard is 3.0.4, which API framework supports elliptic curve facility, in particular the *secp256k1* curve, and the ECDSA signature. The ISO7816 interface of the CCSC application is detailed in Figure 16.

Copyright (c) IARIA, 2018. ISBN: 978-1-61208-623-1

Command	ISO7816 encoding	Comment
Select	00A4040006<AID> AID= Application IDentifier=010203040500	Start the CCSC application
Verify	0020000004<UserPIN> 0020000204<User2PIN> 0020000108<AdminPIN>	Check PIN
InitCurve AdminPIN is required	008900$P_2$00 P2 is the key index	Init curve parameters
ClearKeys AdminPIN is required	008100$P_2$00 P2 is the key index	Clear public and private keys,
GenKeys AdminPIN is required	008200$P_2$00 P2 is the key index	Generate the keys
SetKey AdminPIN is required	008800$P_1P_2P_3$<value> P2 is the key index P1=6 for the public key P1=7 for the private key P3 is the key length Value is the key value	Set public or private key The public key is in the uncompress format
GetKey UserPin is required	008400$P_1P_2P_3$<value> P2 is the key index P1=6 for the public key P1=7 for the private key AdminPIN is required for the private key	Get public or private key, return the length (16bits) of the key and its value
SignECDSA UserPIN is required	008000P_2P_3<value> P2 is the key index P3 is the length of the hash to be signed (32) value is the hash (e)	Return the length (16bits) of the ECDSA signature and its ASN.1 encoding

Figure 16. ISO7816 interface of the CCSC application

The cryptographic keys can be generated and optionally exported, or imported.

- The procedure for key generation and export deals with the following commands: Select(AID), Verify(AdminPIN), InitCurve, ClearKeys, GenKeys, GetKeys.
- The procedure for key import uses the following commands: Select(AID), Verify(AdminPIN), ClearKeys, InitCurve, SetKey(PublicKey), SetKey(PrivateKey).

The ECDSA signature is performed according to the following sequence: Select(AID), Verify(UserPIN), GetKey(PublicKey), SignECDSA(HashValue).

B. *BTOOLS APDU script*

The *BTOOLS* software manages APDU script in order to communicate with Crypto Currency SmartCards. It is a set of lines. A comment line begins by the '/' or '*' character.

The main script token are as follow:

- ***start <optional AID>*** which initializes the ISO7816 context, and detects the first available smartcard;

- **APDU <hexadecimal value>** which sends an ISO7816 request to the smartcard. For error free operation, the response should end by the 9000 status;

- ***pub <offset>*** which MUST be specified before the APDU command used to collect the public key. It is the offset in the response of the public key (after the byte 04);

- ***signature <offset>*** which MUST be specified before the APDU command used to collect the signature. It is the offset in the response of the ASN.1 encoding of the ECDSA signature;

- ***hash <offset>*** which MUST be specified before the APDU command used to collect the signature. It is the offset in the ISO7816 request of the hash (or data) to be signed.

Figures 17 and 18 give an example of APDU script, dealing with a pair of keys identified by the index 5.

```
// script file name: sAPDU.txt
start
// Select CCSC
APDU 00A4040006 010203040500
// Verify UserPIN= 0000
APDU 0020000004 30303030
// Get PublicKey index=5
pub 3
APDU 0084 0605 43
// ECDSA Signature, index=5
signature 2
hash 5
APDU 0080 0005 20
```

Figure 17. An APDU script use for the generation of ECDSA signature

```
// start
Opening the APDU script sAPDU.txt
Reader: Broadcom Corp Contacted SmartCard 0
T=0  - ATR
// Select(CCSC)
Tx: 00 A4 04 00 06 01 02 03 04 05 00
Rx: 90 00
// Verify(UserPIN)
Tx: 00 20 00 00 04 30 30 30 30
Rx: 90 00
// GetKey(PublicKey)
Tx: 00 84 06 05 43
Rx: 00 41 04 A6 FC 0C 5F 46 7C 3D B8 C1 58 18 05 E7
C6 2C 5F AE A1 90 63 B0 1F 58 45 AD 68 DE 9D 84
38 5F 32 1E BF 3A 26 B2 99 12 41 89 92 DC DC 1F
E6 9C 28 2E FF 65 86 0E 10 9F 53 AD 27 A2 96 24
98 4B 6A 90 00
// SignECDSA(hash)
Tx: 00 80 00 05 20 DC AF B4 6D 7F 57 1D 87 C2 34 B3
20 8E 68 86 AD F4 85 AC 98 20 EA A5 67 7C 6D 37 6A
32 13 6F 34
Rx: 61 48
Tx: 00 C0 00 00 48
Rx: 00 46 30 44 02 20 65 A3 1E 14 88 20 61 82 1E A8
B7 27 C4 A8 D1 E2 CB 59 29 20 88 6B DD 70 84 B9
C1 C5 D6 6F 7D 30 02 20 5B 83 A4 69 E5 6D 3B B1
C2 77 6B 16 A3 7B C1 19 0F 6A C9 85 F7 03 54 B6
58 1B 6F 46 21 C7 63 3B 90 00
```

Figure 18. An APDU script used by a transaction script

In Figure 18, the public key is in blue characters, the value to sign in bold characters, and the ASN.1 signature encoding in red characters.

BTOOLS also provides an option that starts APDU scripts, typically used for used key generation.

Figure 19 gives an example of such a script, and Figure 20 illustrates its execution.

```
start
// select
APDU 00A4040006 010203040500
// Verify PinAdmin
APDU 0020 0001 08 3030303030303030
// ClearKeys Key 0
APDU 0081 00 00 00
// InitCurve, Key 0
APDU 0089 00 00 00
// Generate KeysPair Key 0
APDU 0082 00 00 00
// GetPublicKey Key0
APDU 0084 06 00 00
// GetPrivateKey Key 0
APDU 0084 07 00 00
```

Figure 19. Example of a script used for key generation

```
// select
Tx: 00 A4 04 00 06 01 02 03 04 05 00
Rx: 90 00
//Verify(AdminPIN)
Tx: 00 20 00 01 08 30 30 30 30 30 30 30 30
Rx: 90 00
Tx: 00 81 00 00 00 // Clear Key index 0
Rx: 90 00
Tx: 00 89 00 00 00 // Init curve index 0
Rx: 90 00
Tx: 00 82 00 00 00 // Generate Keys index 0
Rx: 90 00
Tx: 00 84 06 00 43 // Get Public Key index0
Rx: 00 41 04 BA 5A 71 A8 0E 90 76 9E DD D2 B9 6C B4
BA 47 0B 45 C6 3B 01 F5 A9 FB FC 3F 95 37 43 23
18 15 5D 59 F3 F1 75 26 08 4E 5A CC 7D 17 4D 68
AB 39 57 C4 F6 D8 5D 38 43 95 EF 8D F4 7D 05 3B
FE E6 F9 90 00
Tx: 00 84 07 00 00 // Get Private Key index 0
Rx: 6C 22
Tx: 00 84 07 00 22
Rx: 00 20 85 1F 6D 62 0B 87 FC 27 FC 9A 00 42 8F C6
01 37 D8 6B 14 07 E4 B6 8F 77 30 A4 BF AC CE 7D
A3 91 90 00
```

Figure 20. Illustration of a key generation script at run time. The public key is in blue characters. The private key is in red characters.

VI. CONCLUSION

In this paper we present the BTOOLS open software [12] that targets the generation of trusted blockchain transactions, based on smartcard cryptographic services. BTOOLS is available for Win32, Linux or Raspberry PI environments. Our future projects will address the definition of innovative services based on this trusted platform.

REFERENCES

[1] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system", www.bitcoin.org, 2008, [retrieved: June, 2017]

[2] A. M. Antonopoulos, "Mastering Bitcoin", O'REILLY, 2015

[3] P. Huang, "A Dissection of Bitcoin", ISBN 9781329754812, January 2016

[4] Standards for Efficient Cryptography "SEC 2: Recommended Elliptic Curve Domain Parameters", Certicom Research, January 27, 2010 Version 2.0

[5] Standards for Efficient Cryptography, "SEC 1: Elliptic Curve Cryptography", Certicom Research, May 21, 2009, Version 2.0

[6] https://github.com/Bitcoin, [retrieved: July, 2017]

[7] https://en.Bitcoin.it/wiki/Protocol_documentation#Message_structure [retrieved: July, 2017]

[8] V. Buterin, "Ethereum: A Next-Generation Smart Contract and Decentralized Application Platform", 2013, http://Ethereum.org/Ethereum.html, [retrieved: July, 2017]

[9] G. Wood, Ethereum Yellow Paper, "Ethereum : a Secure Decentralized Generalized Transaction Ledger", EIP-150, 2015, http://yellowpaper.io/, [retrieved: July, 2017]

[10] https://qbitninja.docs.apiary.io/#reference/transactions/retrieve-a-transaction/get, [retrieved: June, 2017]

[11] https://github.com/Ethereum/wiki/wiki, [retrieved: June, 2017]

[12] "BTOOLS, blockchain tools", https;//github.com/purien/btools, , [retrieved: November 2017]

[13] G. Bertoni, J. Daemen, M. Peeters, G. Assche, "The Keccak SHA-3 submission", 2011, http://keccak.noekeon.org/Keccak-submission-3.pdf, [retrieved: July, 2017]

[14] M. A. Olson, K. Bostic, M. Seltzer, "Berkeley DB", Proceedings of the FREENIX Track: 1999 USENIX Annual Technical Conference, Monterey, California, USA, June 6-11, 1999.

[15] W. Polk, R. Housley, L. Bassham, "Algorithms and identifiers for the Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) profile", RFC 3279, April 2002.

[16] ISO/IEC7816-4:2013, "Identification cards – Integrated circuit cards – Part 4: Organization, security and commands for interchange", 2013

[17] https://www.reuters.com/article/us-bitcoin-exchange-southkorea/south-korean-cryptocurrency-exchange-to-file-for-bankruptcy-after-hacking-idUSKBN1ED0NJ, [retrieved: December, 2017]

ISBN: 978-1-61208-623-1