AMBIENT 2019

The Ninth International Conference on Ambient Computing, Applications, Services and Technologies

ISBN: 978-1-61208-739-9

September 22 - 26, 2019

Porto, Portugal

AMBIENT 2019 Editors

Birgit Gersbeck-Schierholz, Leibniz Universität Hannover, Germany

AMBIENT 2019

Forward

The Ninth International Conference on Ambient Computing, Applications, Services and Technologies (AMBIENT 2019), held between September 22-26, 2019 in Porto, Portugal, continued a series of events devoted to a global view on ambient computing, services, applications, technologies and their integration.

On the way for a full digital society, ambient, sentient and ubiquitous paradigms lead the torch. There is a need for behavioral changes for users to understand, accept, handle, and feel helped within the surrounding digital environments. Ambient comes as a digital storm bringing new facets of computing, services and applications. Smart phones and sentient offices, wearable devices, domotics, and ambient interfaces are only a few of such personalized aspects. The advent of social and mobile networks along with context-driven tracking and localization paved the way for ambient assisted living, intelligent homes, social games, and telemedicine.

The conference had the following tracks:

- Ambient devices, applications and systems
- Ambient services, technology and platforms
- Ambient Environments for Assisted Living and Virtual Coaching
- User Friendly Interfaces

We take here the opportunity to warmly thank all the members of the AMBIENT 2019 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 that dedicated much of their time and effort to contribute to AMBIENT 2019. 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 AMBIENT 2019 organizing committee for their help in handling the logistics and for their work that made this professional meeting a success.

We hope that AMBIENT 2019 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 ambient computing, applications, services and technologies. We also hope that Porto provided a pleasant environment during the conference and everyone saved some time to enjoy the historic charm of the city.

AMBIENT 2019 Chairs

AMBIENT Steering Committee

Yuh-Jong Hu, National Chengchi University-Taipei, Taiwan (ROC)
Naoki Fukuta, Shizuoka University, Japan
Jerry Chun-Wei Lin, Harbin Institute of Technology, Shenzhen, China

Kenji Suzuki, Illinois Institute of Technology, USA
Oscar Tomico, Eindhoven University of Technology, the Netherlands | ELISAVA Escola Universitaria, Spain
Lawrence W.C. Wong, National University of Singapore, Singapore

AMBIENT Industry/Research Advisory Committee

Evangelos Pournaras, ETH Zurich, Switzerland
Johannes Kropf, AIT Austrian Institute of Technology GmbH, Austria
Daniel Biella, University of Duisburg-Essen, Germany
Carsten Röcker, Fraunhofer Application Center Industrial Automation (IOSB-INA), Germany
Marc-Philippe Huget, Polytech Annecy-Chambery-LISTIC | University of Savoie, France
Tomasz M. Rutkowski, Cogent Labs Inc., Japan

AMBIENT 2019

COMMITTEE

AMBIENT Steering Committee

Yuh-Jong Hu, National Chengchi University-Taipei, Taiwan (ROC)
Naoki Fukuta, Shizuoka University, Japan
Jerry Chun-Wei Lin, Harbin Institute of Technology, Shenzhen, China
Kenji Suzuki, Illinois Institute of Technology, USA
Oscar Tomico, Eindhoven University of Technology, the Netherlands | ELISAVA Escola Universitaria, Spain
Lawrence W.C. Wong, National University of Singapore, Singapore

AMBIENT Industry/Research Advisory Committee

Evangelos Pournaras, ETH Zurich, Switzerland
Johannes Kropf, AIT Austrian Institute of Technology GmbH, Austria
Daniel Biella, University of Duisburg-Essen, Germany
Carsten Röcker, Fraunhofer Application Center Industrial Automation (IOSB-INA), Germany
Marc-Philippe Huget, Polytech Annecy-Chambery-LISTIC | University of Savoie, France
Tomasz M. Rutkowski, Cogent Labs Inc., Japan

AMBIENT 2019 Technical Program Committee

Shaftab Ahmed, Bahria University, Pakistan
Mohammed Alia, AL Zaytoonah University of Jordan, Jordan
Cesar Analide, Universidade do Minho, Portugal
Nazmul Arefin Khan, Dalhousie University, Canada
Maxim Bakaev, Novosibirsk State Technical University, Russia
Maria João Barreira Rodrigues, Universidade da Madeira, Portugal
Rachid Beghdad, Bejaia University, Algeria
Daniel Biella, University of Duisburg-Essen, Germany
Lars Braubach, Complex Software Systems | Bremen City University, Germany
Philip Breedon, Nottingham Trent University, UK
Ramon F. Brena Pinero, Tecnológico de Monterrey, Mexico
Valerie Camps, IRIT - Universite Paul Sabatier, France
Juan Carlos Cano, University Politécnica de Valencia, Spain
Florent Carlier, Centre de Recherche en Education de Nantes - Le Mans Université, France
Kelsey Carlson, Independent Consultant, USA
Carlos Carrascosa, Universidad Politécnica de Valencia, Spain
John-Jules Charles Meyer, Utrecht University, The Netherlands
DeJiu Chen, KTH Royal Institute of Technology, Sweden
Albert M. K. Cheng, University of Houston, USA
William Cheng Chung Chu, Tunghai University, Taiwan
Juan Manuel Corchado Rodríguez, University of Salamanca, Spain

Mauro Dragone, Heriot-Watt University | Edinburgh Center for Robotics, UK
Rachida Dssouli, Concordia University, Canada
Mauro Dragone, Heriot-Watt University | Edinburgh Center for Robotics, UK
Duarte Duque, 2Ai - Polytechnic Institute of Cávado and Ave, Barcelos, Portugal
Christiane Eichenberg, Sigmund Freud University Vienna, Austria
Ahmed El Oualkadi, Abdelmalek Essaadi University, Morocco
Larbi Esmahi, Athabasca University, Canada
Imad Ez-zazi, National School of Applied Sciences of Tangier | University of Abdelmalek Essaadi, Morocco
Biyi Fang, Michigan State University, USA
Gianni Fenu, University of Cagliari, Italy
Jesus Fontecha, University of Castilla-La Mancha, Spain
Naoki Fukuta, Shizuoka University, Japan
Claudio Gallicchio, Universita' di Pisa, Italy
Matjaz Gams, Jožef Stefan Institute - Ljubljana, Slovenia
Valentina Gatteschi, Politecnico di Torino, Italy
Musab Ghadi, Lab-STICC | University of Bretagne Occidentale, France
Marie-Pierre Gleizes, IRIT, France
Huan Gui, University of Illinois at Urbana-Champaign, USA
Maki Habib, American University in Cairo, Egypt
Ileana Hamburg, Institut Arbeit und Technik, Germany
Ibrahim A. Hameed, Norwegian University of Science and Technology (NTNU) in Ålesund, Norway
Ahmad Harb, German Jordanian University, Jordan
Jean Hennebert, iCoSys Institute | University of Applied Sciences HES-SO, Fribourg, Switzerland
Mehrdad Hessar, University of Washington, USA
Tzung-Pei Hong, National University of Kaohsiung, Taiwan
Yining Hu, University of New South Wales / Data61-CSIRO, Australia
Yuh-Jong Hu, National Chengchi University-Taipei, Taiwan (ROC)
Marc-Philippe Huget, Polytech Annecy-Chambery-LISTIC | University of Savoie, France
Reyes Juárez Ramírez, Universidad Autonoma de Baja California, Mexico
Martin Kampel, Vienna University of Technology, Austria
Viirj Kan, Primitive Labs / Samsung Research America, USA
Evika Karamagioli, NKUA, Greece
Imrul Kayes, Sonobi, Inc., USA
Dina Khattab, Ain Shams University, Cairo, Egypt
Maher Khemakhem, King Abdulaziz University, Jeddah, KSA
Kwang-Man Ko, Sang-Ji University, Republic of Korea
Johannes Kropf, AIT Austrian Institute of Technology GmbH, Austria
Adam Krzyzak, Concordia University, Montreal, Canada
Markus Kucera, Ostbayerische Technische Hochschule Regensburg (OTH Regensburg), Germany
Alar Kuusik, TalTech, Estonia
Frédéric Le Mouël, INSA Lyon, France
Fedor Lehocki, National Centre of Telemedicine Services | Slovak University of Technology, Slovakia
Lenka Lhotská, Czech Institute of Informatics, Robotics and Cybernetics | Czech Technical University in Prague, Czech Republic
Jerry Chun-Wei Lin, Harbin Institute of Technology, Shenzhen, China
José Luís Silva, Instituto Universitário de Lisboa (ISCTE-IUL), ISTAR-IUL / M-ITI, Portugal
Irene Mavrommati, Hellenic Open University, Patras, Greece

Jochen Meyer, OFFIS e.V. - Institut für Informatik, Oldenburg, Germany
Vittorio Miori, Institute of Information Science and Technologies "A. Faedo" (ISTI) | CNR - National Research Council of Italy, Pisa, Italy
El Houssaini Mohammed-Alamine, Chouaib Doukkali University, El Jadida, Morocco
Lia Morra, Politecnico di Torino, Italy
Jouini Mouna, University of Tunis, Tunisia
Keith V. Nesbitt, University of Newcastle, Australia
Gerhard Nussbaum, Kompetenznetzwerk Informationstechnologie zur Förderung der Integration von Menschen mit Behinderungen (KI-I), Austria
Brendan O'Flynn, Tyndall Microsystems | University College Cork*, Ireland*
Eva Oliveira, IPCA - School of Technology - DIGARC, Portugal
Kamalendu Pal, City University of London, UK
Pablo Pancardo, Juarez Autonomous University of Tabasco, Mexico
Ivan Pires, University of Beira Interior, Covilhã / Altran Portugal, Lisbon, Portugal
Evangelos Pournaras, ETH Zurich, Switzerland
Rafael Pous, Universitat Pompeu Fabra, Spain
Sigmundo Preissler Jr., Universidade do Contestado, Brazil,
Valérie Renault, Le Mans Université, France
Antonio M. Rinaldi, Università degli Studi di Napoli Federico II, Italy
Michele Risi, University of Salerno,Italy
Carsten Röcker, Fraunhofer Application Center Industrial Automation (IOSB-INA), Germany
Michele Ruta, Polytechnic University of Bari, Italy
Tomasz M. Rutkowski, Cogent Labs Inc., Japan
Khair Eddin Sabri, The University of Jordan, Jordan
Jose Ivan San Jose Vieco, University of Castilla-La Mancha, Spain
Peter Schneider-Kamp, University of Southern Denmark, Denmark
Floriano Scioscia, Polytechnic University of Bari, Italy
Hasti Seifi, Max Planck Institute for Intelligent Systems, Germany
Riaz Ahmed Shaikh, King Abdulaziz University, Saudi Arabia
Jingbo Shang, University of Illinois, Urbana-Champaign, USA
Sheng Shen, University of Illinois at Urbana-Champaign, USA
Ingo Siegert, Otto von Guericke University Magdeburg, Germany
Carine Souveyet, Université Paris 1 Panthéon-Sorbonne, France
Andreas Stainer-Hochgatterer, AIT Austrian Institute of Technology GmbH, Austria
Daniela Ströckl, Carinthia University of Applied Sciences - Institute for Applied Research on Ageing (IARA) in Carinthia, Austria
Mu-Chun Su, National Central University, Taiwan
Kenji Suzuki, Illinois Institute of Technology, USA
Oscar Tomico Plasencia, Eindhoven University of Technology, Netherlands
Markku Turunen, Tampere University, Finland
Alexandra Voit, University of Stuttgart, Germany
Yunsheng Wang, Kettering University, USA
Lawrence W.C. Wong, National University of Singapore, Singapore
Marcin Wozniak, Silesian University of Technology, Poland
Xiang Xiao, Google Inc., USA
Qiumin Xu, Google, USA
Kin-Choong Yow, Gwangju Institute of Science and Technology, South Korea

Chuan Yue, Colorado School of Mines, USA
Yang Zhang, Human-Computer Interaction Institute | Carnegie Mellon University, USA

Copyright Information

Table of Contents

Efficient Online Cough Detection with a Minimal Feature Set Using Smartphones for Automated Assessment of Pulmonary Patients

Md Juber Rahman
Electrical and Computer Engineering Department
The University of Memphis
Memphis, USA
e-mail: mrahman8@memphis.edu

Ebrahim Nemati, Mahbubur Rahman, Korosh Vatanparvar, Viswam Nathan, Jilong Kuang
Digital Health Lab
Samsung Research America
Mountain View, USA
e-mail: e.nemati@samsung.com

***Abstract*—An automated monitoring of chronic diseases may help in the early identification of exacerbation, reduction of healthcare expenditure, as well as improve patient's health-related quality of life. Cough monitoring provides valuable information in the assessment of asthma and Chronic Obstructive Pulmonary Disease (COPD). In this multi-cohort study, we have used every-day wearables such as smartphone and smartwatch to collect cough instances from 131 subjects including 69 asthma patients, 9 COPD patients, 13 patients with a co-morbidity of asthma and COPD and 40 healthy controls. For online cough detection we have identified the audio features suitable for resource-constrained platforms (e.g., smartphone), ranked the features and identified top 9 features to obtain an F-1 score of 99.8% in the offline classification of 23,884 cough instances from non-cough (speech/silence, etc.) events using Random Forest classifier. Finally, a power and time-efficient scheme for continuous online cough detection from the audio stream has been illustrated. The proposed model has an online cough detection sensitivity of 93.3%, specificity of 98.8% and accuracy of 98.8%. In addition, a good improvement in reducing the on-device execution (feature extraction and classification) time and power consumption has been achieved compared to the current state of the art algorithms. The proposed on-device cough detector has been implemented to meet the criteria for integration in the passive monitoring and online assessment of asthma/COPD patients.**

Keywords- cough; online detection; pulmonary disease; random forest; streaming audio.

I. INTRODUCTION

Lung diseases are among the biggest killers in the world. In the USA, lung disease is the third leading cause of death [1][2]. Many of the lung diseases are chronic conditions in nature which severely impacts the patients' health-related quality of life. As a result, the associated healthcare expenditure is substantial. The annual direct and indirect healthcare cost related to obstructive lung diseases such as asthma and COPD has been estimated to be $154 billion [3]. Early diagnosis and follow-up have the potential to reduce hospital visits, associated expenditures, and improve patients' quality of life.

Spirometry and standard questionnaires have been used extensively in the diagnosis and severity estimation of asthma and COPD. Monitoring of warning signs such as cough, shortness of breath, etc. has been proven to be useful in the detection and management of asthma and COPD [4]. Usually, cough frequency and severity are reported by the patient himself. This approach is highly subjective and not suitable for continuous passive monitoring. As an alternative, there have been attempts in developing automated cough monitors.

Audio signal from the acoustic sensor has been primarily used as the basis for automatic cough detection. Though there are multi-sensor approaches that include non-acoustic sensors for automatic cough detection, previous research efforts indicate that high sensitivity and accuracy can be achieved with audio signal solely [5]. Also, employing acoustic sensor seems to be the most suitable for 24 h ambulatory monitoring. Commonly used features for cough detection include audio spectral features, Mel-Frequency Cepstral Coefficients (MFCC), Linear Prediction Cepstral Coefficients (LPCC), Hu moments, etc. Birring *et al.* introduced an automatic cough detection system called Leicester cough monitor using Hidden Markov Model (HMM) [6]. The system incorporates 24 h ambulatory recording solely relying on the acoustic signal. They obtained a sensitivity of 91% and specificity of 99% with spectral audio features. Larson *et al.* proposed a low-cost microphone-based cough sensing system using Random forest classifier [7]. These approaches requiring a specialized device incur extra cost and burden for the user as he needs to carry that extra device all the time. Shin *et al.* investigated a hybrid model consisting of both Artificial Neural Network and HMM [8]. They used sound pressure level, cepstral coefficient, and temporal features of audio and obtained 91.3% accuracy for cough detection. However, their dataset is too small containing only 143 cough sounds and 110 environmental sounds. Also, it is based on MATLAB, which is not suitable for on-device detection. Liu *et al.* investigated a combination of deep neural network and HMM for automatic cough detection using MFCC features [9]. Amoh *et al.* investigated the use of a convolutional neural network in a wearable cough detection system [10]. It is noteworthy from these works that while deep learning imposed a high computational cost, there was no significant improvement in classification performance compared to traditional methods.

Cough detection from recorded audio is subjected to privacy compromise and hence has never been popular or widely accepted. Recent advancement in the quality of acoustic sensors and processing capacity of smartphones and wearable devices has triggered a growing tendency for online cough detection from streaming audio without recording the audio.

While automatic cough detection is well investigated, few studies addressed on device feature extraction and classification from streaming audio. Pham et al. investigated a Gaussian mixture model and a universal model for real-time cough detection using smartphones [11]. They achieved a sensitivity of 81% for subject independent training, however, they did not address system performance-related issues. Most recently, J. Alvarez et al. investigated the efficient computation of image moments for robust cough detection using smartphones [12]. While they achieved 88% of sensitivity for cough detection the app power consumption is 25% of the device power consumption for 24 hours of usage. Also, the time required for feature extraction is relatively long (5 min 28 secs, as it requires image processing) and needs to be reduced for efficient online implementation. Another limitation of previous studies is their inability to discriminate between the cough of the intended subject with the cough of other subjects, which make the cough monitoring ineffective in a social or family setting if multiple people have cough syndrome. E. C. Larson et al. reviewed the shortcomings of existing cough detection approaches in details and described the need for further investigation for smartphone-centric ambient audio sensing for effective pulmonary assessment [13]. mLung Study is our comprehensive initiative aimed to leverage the power of wearables and smartphones for early detection and continuous monitoring of asthma and COPD patients which include quality data collection, multi-layer annotation, on-device feature extraction and classification, privacy protected in-depth analysis in the cloud, etc. Previously, we reported a framework for maintaining privacy while recording audio for offline cough classification [14]. In this paper, we report a model implemented and tested on android smartphones for online cough detection from streaming audio. The model was trained on a large dataset containing both voluntary and natural coughs. Contributions of this study have been summarized below:

i) Identification of audio features, optimal window size and overlapping suitable for automatic cough detection with high sensitivity using resource-constrained devices such as a smartphone.

ii) Feature ranking and classifier optimization for computational efficiency to minimize the execution time while retaining classification performance.

iii) Enabling subject-specific cough detection and discriminating secondary subject coughs.

iv) Analysis and optimization of system overhead to achieve better performance for online processing in smartphones.

This study presents promising results for using smartphones in a privacy-preserved personalized and reliable online cough detection framework which facilitates the online assessment of asthma and COPD patients as well as healthy population.

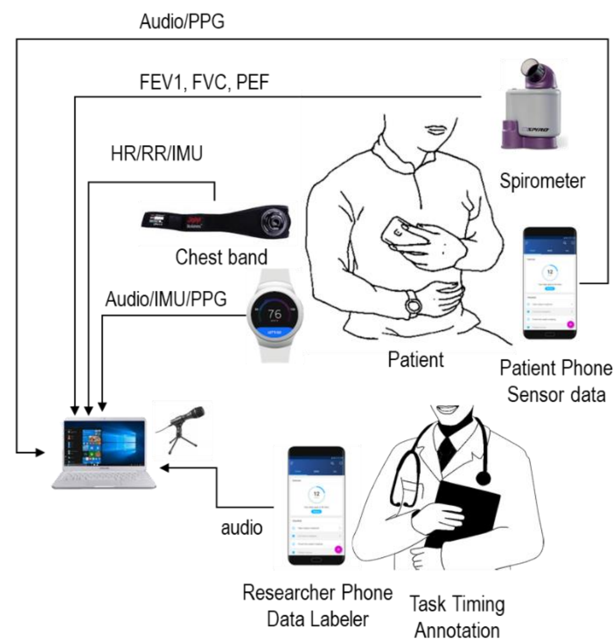


Figure 1 Study description and cough recording protocol

The remainder of this paper is organized as follows. Section 2 describes the materials and methods including the online cough detection framework and the process of system performance evaluation. Section 3 describes the result of the off-line analysis, on-line cough detection performance, and system performance for real-time processing. Finally, Section 4 presents our conclusion and future work scope.

II. MATERIALS AND METHODS

This section describes the dataset, study protocol, algorithm and framework in details:

A. *Description of Subjects and Study Protocol*

Per institutional review board (IRB) approval, a total of 131 subjects (67 males and 64 female) were recruited for this study out of which 40 were healthy controls without any diagnosed medical condition and 91 individuals were suffering from pulmonary diseases. All of the patients have been diagnosed with pulmonary diseases by medical practitioners. Out of the 91 patients 69 were diagnosed with asthma, 9 with COPD and 13 exhibited a co-morbidity of asthma and COPD.

The subjects were from different racial backgrounds including African-American, Asian, Caucasian, and Native American and had an age range from 14 to 82 years. During the recruitment process, all subjects with the history of cardiac disease i.e. arrhythmia or heart attack, pulmonary infection, vocal cord dysfunction, and inability to read or speak English were excluded. After obtaining informed written consents, spirometry-based pulmonary function test was done for all of the patients. Multimodal physiological data such as ECG, PPG, audio, and IMU were collected from the subjects in a laboratory setup using multiple wearable sensors including smartwatch (Samsung Gear Sport), chest band (Zephyr BioHarness 3.0, Medtronic plc), and smartphone (Samsung Galaxy Note 8).

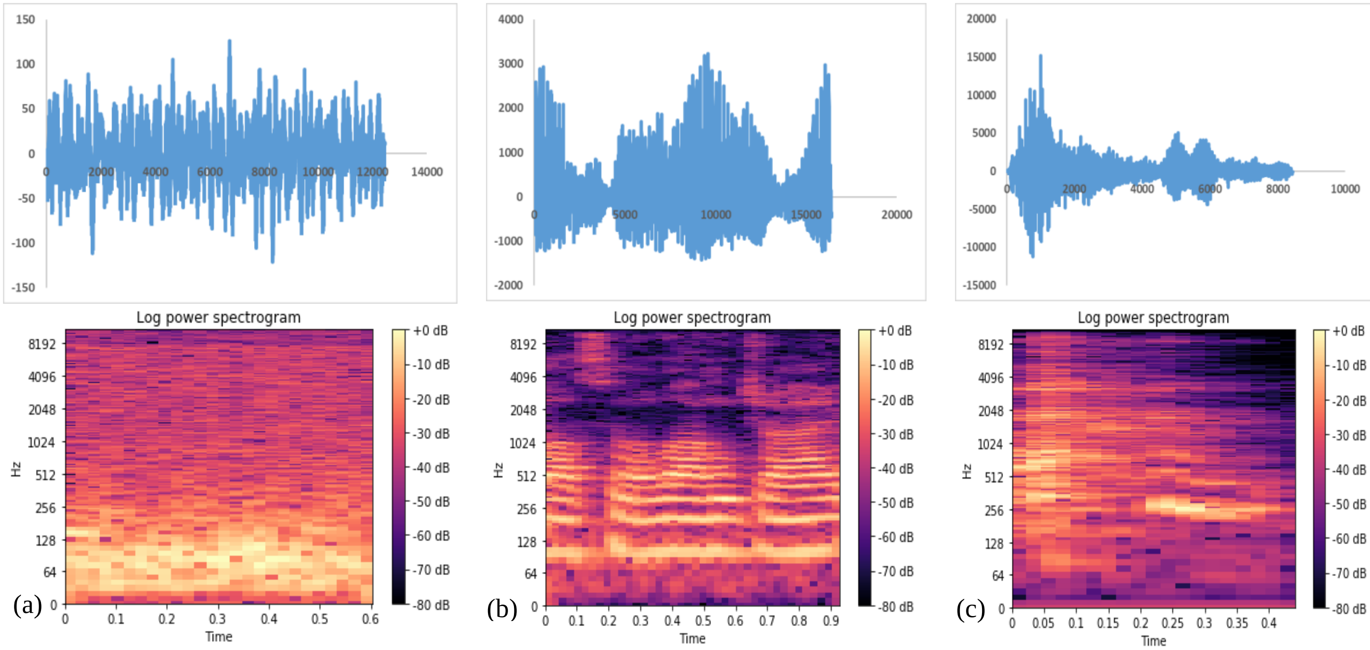


Figure 2 Typical (a) silent, (b) speech, and (c) cough episodes and their spectral representations showing the differences in frequency component and loudness.

The length of the data collection session was 30-40 minutes. During this time tasks related to the pulmonary patient assessment, were performed which included PFT at the beginning and the end. The audio was captured from both smartphone and smartwatch with a sampling rate of 44.1kHz. Participants wore a Samsung Gear S3 smartwatch on their left hand. They held the smartphone (Samsung Galaxy Note 8) on the left side of the chest to capture chest motions as well as lung sounds such as wheezes. The system architecture utilized for the data collection can be seen in Figure 1. The experiment protocol that is specifically designed for asthma and COPD patients included the following sessions:

- Pulmonary Function Test-1: standard mobile spirometry.
- Sit-Silent Breathing: sitting silently for one minute and counting breaths while keeping the phone on the chest and watch on the abdomen.
- Supine-Silent Breathing: repeat the previous task in the supine position.
- Cough: produce several voluntary coughs for up to two minutes. Also, counted natural coughs.
- A-vowel Voice: vocalizing '*Aaaa....*' sound for as long as they can.
- Speech: speaking freely about any topic of interest.
- Reading: read aloud a standard passage.
- Pulmonary Function Test-2: standard mobile spirometry.

The rationale behind the study protocol has been described earlier [15].

B. Audio Processing, Data Preparation and Feature Extraction

The entire record audio has been annotated manually for cough, speech, and silence by a crowdsourcing annotation platform, FigureEight [16]. For cough detection purpose, wheezes and other body sounds have been included in the speech category. In addition to recorded audio, spectral visualization of an audio signal has been used in the annotation process to improve the quality of annotation. Figure 2 shows the time-domain and corresponding spectral representation of silent, speech and cough episodes. The cough instances are characterized by a burst followed by a voiced part which makes them distinguishable from the speech and silence. From the spectrogram, it is clear that frequency components and loudness of the cough are very different from that of regular speech or silence. The start and stop time of each cough event has been marked in the annotation process and then the recording has been segmented into cough, speech and silent episodes and labeled accordingly. Finally, 23884 cough instances, 165948 speech instances, and 52135 silent episodes were obtained from the audio clips. For subject discriminatory cough detection, 35 coughs from one subject have been placed in one class while the other class had 170 cough instances from multiple subjects. Each wav file is a 16-bit PCM-encoded audio with the sampling rate of 44100 samples/sec.

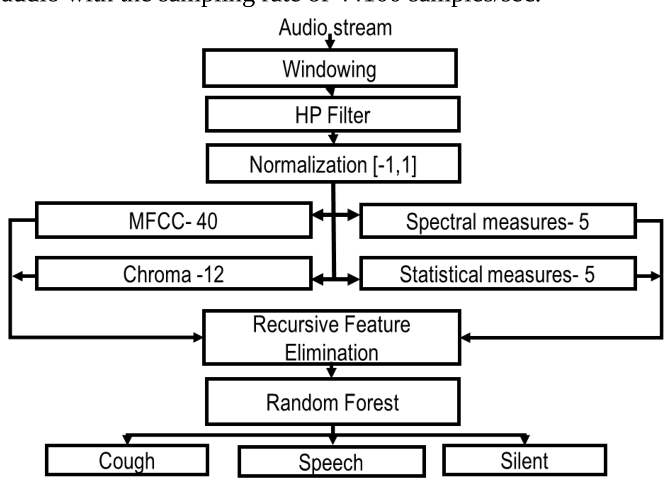


Figure 3 Method for feature extraction, feature selection, and classification.

For feature extraction, the wav file was chopped into frames of 0.6 sec, high pass filtered (200 Hz) and normalized in the range [-1,1]. An overlapping of 10% has been used between the frames for online implementation. Features were then extracted from the frames which included time-domain features such as absolute mean, absolute median, standard deviation, skewness, kurtosis, zero-crossing rate and frequency features such as spectral centroid, spectral roll-off, spectral variance, MFCC, and spectral chroma. The feature set also includes energy and sound pressure level. An open-source library, Taros-DSP, has been used to read the audio from microphone, process the audio file and extract MFCC features [17]. Another open-source library jMusic has been used for extracting the spectral features [18]. Signal pre-processing and feature extraction were done in JAVA. The process of feature extraction, feature selection and classification has been shown in Figure 3. To compute the MFCC features, Fast Fourier Transform with a hamming window has been used in estimating the magnitude spectrum. The number of Mel filters used is 50, the lower filter frequency is 300 Hz and the upper filter frequency is 8000 Hz [19].

C. Sound Event Detection

Sound events can be detected based on different features. In this work, we have used sound pressure level and energy to detect the sound event. The mean sound pressure level of all silent episodes has been used as the threshold for sound event detection. Any episode with a sound pressure level greater than mean value has been considered as a sound event followed by classification performed to detect if it is a cough, speech or silent episode. The possibility of missing a sound event has been almost eliminated due to this dynamic thresholding. This makes the cough detection feasible even when the smartphone is relatively far from the patient. Audio episodes with mean less than the threshold, are not considered as an event and therefore skipped for feature extraction and classification which helped with the overall power consumption.

D. Feature Selection, Classification, and offline-Training

Dimension reduction is important to reduce the time and computational complexity associated with the implementation of algorithms on wearables. Also, optimal feature selection is an important step to enhance the performance of classification and ensure better generalization. In this work, we have used the recursive feature elimination technique for selecting the top-ranked features. Caret package from R has been used for the feature ranking [20].

For finding the best classification model we have explored logistic regression, support vector machines (SVM) with different kernels and random forest. The decision tree has been used as the base classifier for random forest and samples are drawn with replacement. The number of estimators used in the random forest is 100. To reduce memory consumption and the time complexity, maximum depth (=20) of the tree has been decided using a heuristic approach. 10-fold cross-validation was employed on the training data to evaluate classifier performance and adju st the hyper-parameters. WEKA has been used as the model development environment [21]. The subject discriminatory model has been trained and evaluated separately.

E. Online Cough Detection and Counting Framework

i) Design Goals and Considerations

- Passive sensing- no user effort is expected.
- Privacy-preserving- since speech and related data can reveal user identity, no audio is being recorded. Classification is done using the features generated on the device.
- Reliable detection- high sensitivity not to miss any cough instances.
- Execution Time- keep the processing time for feature extraction and classification low enough to facilitate real-time processing and prediction.
- Performance Optimization- the high emphasis has been given to keep the app power consumption, memory usage and latency low, so that device normal functionality is not drastically impacted by the app.

ii) System Overview, Implementation, and Evaluation

The online cough detection framework has been shown in. Figure 4.

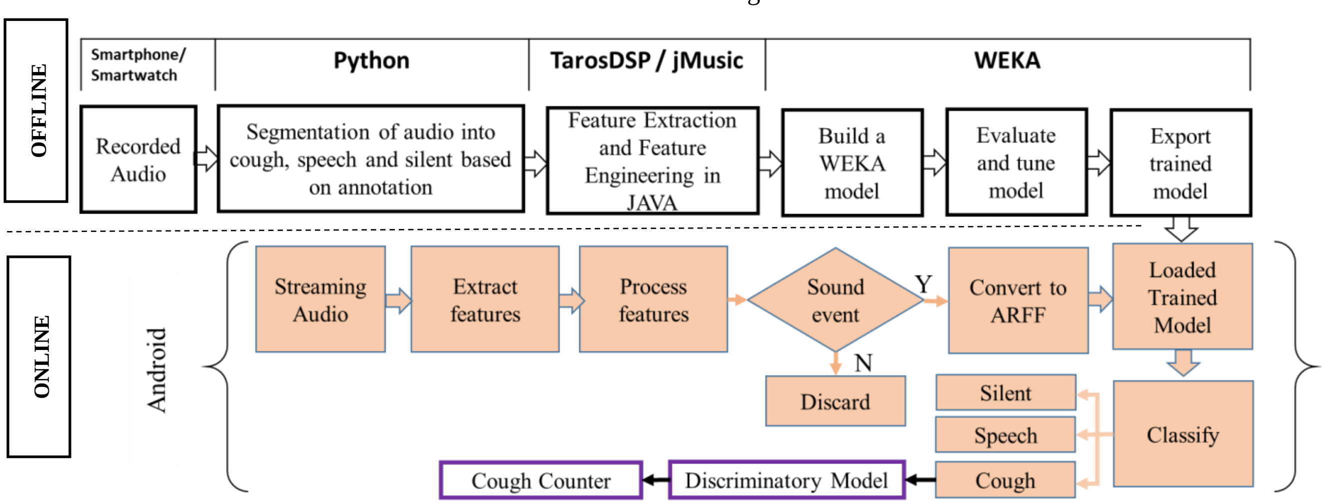


Figure 4 Online cough counter framework

Trained, evaluated and tuned model from WEKA has been exported for use in Android. In android, the trained model was stored in the asset directory and was loaded in the activity for online classification of audio frames using on-device extracted features. The audio signal was directly read from the microphone in an audio buffer and was processed as an audio event in 0.6 sec frames and prediction is made for each of these frames. Confirmed silent episodes are discarded and no further feature extraction/classification is done to reduce execution time and power consumption. The extracted features, as well as the predictions, are written to a CSV file and exported to the cloud for in-depth offline processing. For evaluating the online cough detection performance, the app has been tested for 2 days in the real-world scenario which include home environments, driving, walking in the street and social gathering.

III. Results

The boxplots for the sound pressure level of cough, speech and silent episodes have been shown in Figure 5. It is evident that the sound pressure level of silent episodes is much lower compared to cough and speech. Figure6 shows the result of feature ranking by recursive feature elimination technique. The feature ranking suggests that best performance (good accuracy and low dimension) can be achieved with 9 top-ranked features. The top-ranked features are *mfcc_0, pressure level, standard deviation, kurtosis, mean, mfcc_1, median, zero-crossing rate*, and *mfcc_2*. Figure 7 shows the boxplot comparison between cough and speech events for the top-ranked MFCC features. A good visual separation between cough and speech episodes can be observed in the boxplot comparison. The classification performance of different classifiers with the top-ranked features has been shown in Table I. Random Forest performed best with a precision of 99.8%, recall of 99.8% and an F-1 score of 99.8% for 10-fold cross-validation. The confusion matrix for 10-fold cross-validation has been shown in Table II. Only 4 cough instances have been misclassified out of 23884 cough instances. Figure 8 shows the model build time, test time and F-1 score at different depths of the forest for the Random Forest classifier. Low model build time is important for subject-specific cough detection as it requires online training. It can be seen that increasing the depth beyond 20 increases the build and test time with a minimal gain in F-1 score. Hence, the optimal depth is found to be 20 to create the final model. A precision of 94.2%, recall of 94.1% and F-1 score of 93.7% have been achieved with Random Forest in detecting cough from the intended subject while discriminating coughs from other subjects as shown in Table III.

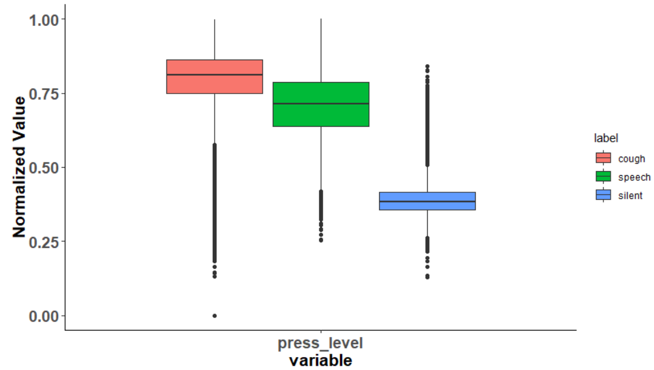


Figure 5 Boxplots showing normalized sound pressure level for cough, speech and silence instances

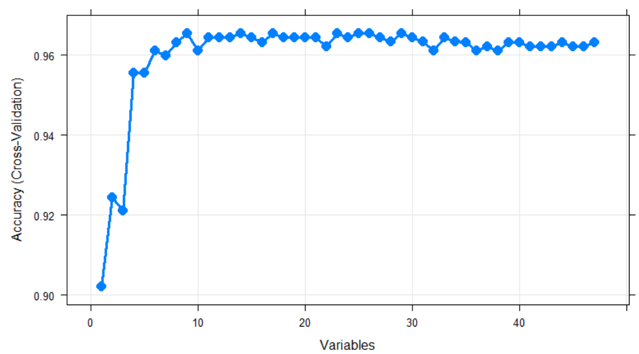


Figure 6 Optimal no. of features using recursive feature elimination technique

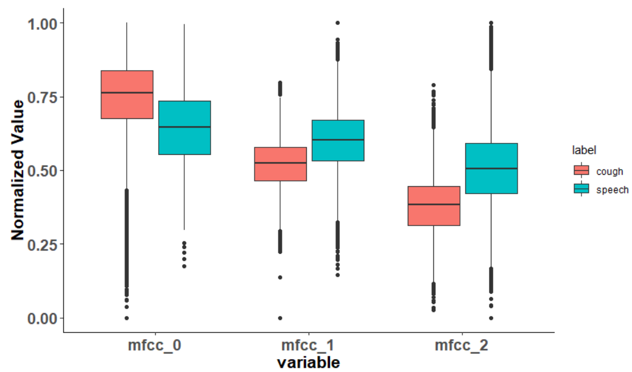


Figure 7 Boxplot comparison between cough and speech for top-ranked MFCC features

TABLE I. Off-line classification performance with different classifiers (10-fold CV)

Classifier	Cough detection		
	precision	***recall***	***F-1 score***
Logistic Regression	93.0%	92.9%	92.9%
SVM (kernel=Poly)	93.1%	93.0%	93%
Random Forest	**99.8%**	**99.8%**	**99.8%**

TABLE II. CONFUSION MATRIX FOR RANDOM FOREST CLASSIFIER (10-FOLD CV)

cough	*speech*	*silence*	
23880	4	0	*cough*
0	165944	468	*speech*
0	54	52081	*silence*

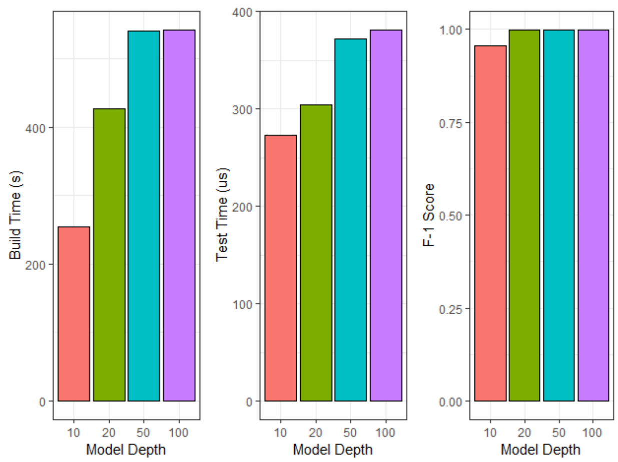


Figure 8 Build time, test time and F-1 score at different depths of the Random forest classifier

TABLE III. CLASSIFICATION PERFORMANCE FOR SUBJECT-SPECIFIC COUGH DETECTION (10-FOLD CV)

Classifier	**Cough detection**		
	precision	***recall***	***F-1 score***
Logistic Regression	89.2%	88.8%	89.0%
SVM	93.3%	93.2%	93.2%
Random Forest	94.2%	94.1%	93.7%

TABLE IV. SYSTEM OVERHEAD FOR ON-DEVICE FEATURE EXTRACTION AND COUGH CLASSIFICATION IN SMARTPHONES FROM STREAMING AUDIO

App	***Latency***	***Memory***	***Avg. CPU***	***Power Consumption***
Cough Counter	375 ms	99 MB	8.69 %	55 mAh

Using the implemented model, a sensitivity of 93.3%, specificity of 98.8% and accuracy of 98.8% have been achieved for online cough detection. The feature extraction and classification time for a 2 min audio clip is only 9.8 secs which is much lower compared to other feature set reported in previous studies [12]. The system overhead on a smartphone for running the cough detection app continuously has been shown in Table IV and compared with VoiceOver app (already available in the play store, 500K+ downloads) in Figure 9. The functionality of VoiceOver app includes audio recording, audio processing, sharing and audio storage; whereas the functionality of Cough app includes audio sampling, audio processing, feature extraction and classification and export/storage of feature values. It can

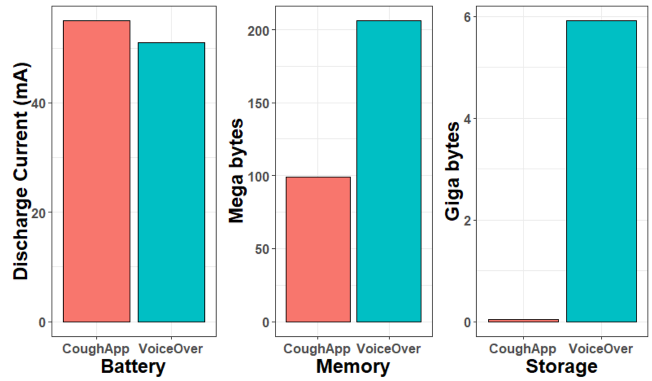


Figure 9 Comparison of the performance metrics of Cough Counter app with VoiceOver app

TABLE V. COMPARISON OF THIS WORK WITH PREVIOUS STUDIES

Ref.	Platform	Subjects (Healthy/ Patient)	Online cough detection	
			Classifier	Performance
[6]	Specialized Wearable	71 (8/65)	HMM	Sen. 91 % Sp. 99%
[8]	Specialized Wearable	84	ANN+ HMM	Sen. 91.3%
[10]	Specialized Wearable	14 (14/0)	CNN	Sen. 95.1% Sp. 99.5%
[11]	Smartphone	Not mentioned	GMM UBM	Sen. 91%
[12]	Smartphone	13 (0/14)	kNN	Sen. 88.5% Sp. 99.77%
[7]	Smartphone	17 (0/3), other-14	RF	TPR-92% FPR-0.5%
Proposed Work	Smartphone	131 (40/71)	RF	Sen, 94.3% Sp. 98.8%

be observed that storing feature values instead of audio require much lower storage space. The Cough app consumes less memory but more power than VoiceOver app. Nonetheless, the power consumption (11% of the device total power usage) is lower than previously reported (25% of the device total power usage) cough detection framework [12]. Minimal no. of features, optimal forest depth and silence removal (processing less no. of frames) have contributed to reducing the power consumption. The low latency and CPU usage of cough app are suitable for continuous operation. All testing has been performed using a Samsung Galaxy Note 8 smartphone. A comparison of this work with similar previous studies has been shown in Table V. It can be seen that other smartphone based approaches for cough detection have much lower performance compared to the proposed method [7] [11] [12]. In addition, the size of their dataset is very small which will impact the generalization capacity of the developed models.

IV. CONCLUSION

Cough pattern analysis may be helpful in monitoring asthma and COPD patients passively. However, the privacy of the users is at great risk when it comes to continuous listening if the processing has to be done on the cloud. We have proposed an on-device cough detection framework that detects the cough occurrence from the streaming audio without the need to store the audio on the device or send it to the cloud. To reduce the computational burden, we have ranked the features and identified the top 9 features to obtain a reasonable accuracy and optimized the classifier to have low complexity while providing a high accuracy of 98.8%. This approach is computationally efficient and suitable for smartphones. Our future work includes the improvement of model generalization performance and robustness. Also, we are planning to implement this cough detector as a module among other modules to provide an assessment of the severity of asthma and COPD patients.

REFERENCES

[1] Global Initiative for Chronic Obstructive Lung Disease (GOLD) 2019 "Global Strategy for the Diagnosis, Management and Prevention of COPD" Available from http://www.goldcopd.org Accessed May 20, 2019.

[2] K. D. Kochanek, S. L. Murphy, J. Q. Xu , and B. Tejada-Vera, "Deaths: Final data for 2014." *National vital statistics reports*; vol 65, no 4. pp. 1-122, Jun 2016.

[3] Lung Institute, "The Cost of Lung Disease" Available: https://lunginstitute.com/blog/the-cost-of-lung-disease/, Accessed May 20, 2019.

[4] E. R. McFadden Jr, "Clinical physiologic correlates in asthma." *Journal of allergy and clinical immunology* 77, no. 1, pp. 1-5, 1986.

[5] T. Drugman *et al*., "Objective Study of Sensor Relevance for Automatic Cough Detection," in *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 3, pp. 699-707, May 2013.

[6] S.S.Birring *et al*. ,"The Leicester cough monitor: preliminary validation of an automated cough detection system in chronic cough," Eur. Respiratory J., vol. 31, no. 5, pp. 1013–1018, May 2008.

[7] E. C. Larson, T. J. Lee, S. Liu, M. Rosenfeld, and S.N. Patel. "Accurate and privacy preserving cough sensing using a low-cost microphone." In *Proceedings of the 13th international conf. on Ubiquitous computing*, pp. 375-384. ACM, 2011.

[8] S. Shin, T. Hashimoto, and S. Hatano, "Automatic Detection System for Cough Sounds as a Symptom of Abnormal Health Condition," in *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 4, pp. 486-493, July 2009.

[9] J. M. Liu *et al.*,"Cough event classification by pretrained deep neural network." *BMC medical informatics and decision making* vol. 15 Suppl 4, 2015.

[10] J. Amoh and K. Odame, "DeepCough: A deep convolutional neural network in a wearable cough detection system," *2015 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, Atlanta, GA, 2015, pp. 1-4.

[11] C. Pham, "MobiCough: real-time cough detection and monitoring using low-cost mobile devices." *Asian Conference on Intelligent Information and Database Systems*. Springer, Berlin, Heidelberg, 2016.

[12] J. Monge-Álvarez and C. Hoyos-Barceló, "Robust Detection of Audio-Cough Events Using Local Hu Moments," in *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 1, pp. 184-196, Jan. 2019.

[13] E. C. Larson, E. Saba, S. Kaiser, M. Goel, and S. N. Patel. "Pulmonary Monitoring Using Smartphones." In *Mobile Health*, pp. 239-264. Springer, Cham, 2017.

[14] E. Nemati *et al.*, "Private Audio-Based Cough Sensing for In-Home Pulmonary Assessment using Mobile Devices" *13th International Conference on Body Area Networks*, 2018.

[15] M. Rahman *et al*., "Towards Reliable Data Collection and Annotation to Extract Pulmonary Digital Biomarkers Using Mobile Sensors" *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare*. ACM, 2019.

[16] https://www.figure-eight.com/, accessed on May 1, 2019.

[17] J. Six, O. Cornelis, and M. Leman, "TarsosDSP, a real-time audio processing framework in Java." *Audio Engineering Society Conference: 53rd International Conference: Semantic Audio*. Audio Engineering Society, 2014.

[18] A. R. Brown and A. C. Sorensen, "Introducing jmusic." *InterFACES: Proceedings of The Australasian Computer Music Conference.* Brisbane: ACMA, pp. 68-76, 2000.

[19] X. Huang, A. Acero, and H. Hon, *Spoken Language Processing: A guide to theory, algorithm, and system development*. Prentice Hall, 2001.

[20] M. Kuhn, "Building predictive models in R using the caret package." *Journal of statistical software* 28, no. 5 ,pp.1-26, 2008.

[21] M. Hall *et al*. "The WEKA data mining software: an update." *ACM SIGKDD explorations newsletter* 11, no. 1, pp.10-18, 2009.

An Efficient Message Collecting and Dissemination Approach for Mobile Crowd Sensing and Computing

Tzu-Chieh Tsai
Department of Computer Science
National Chengchi University
Taipei, Taiwan
e-mail: ttsai@cs.nccu.edu.tw

Hao, Teng
Department of Computer Science
National Chengchi University
Taipei, Taiwan
e-mail: 6105753019@nccu.edu.tw

***Abstract*—Mobile Crowd Sensing and Computing (MCSC) is an emerging technology along with the popularity of mobile devices. We utilize the concept of Delay Tolerant Networks (DTNs) and edge/fog computing to support the message collection and dissemination for the MCSC. The biggest challenge here is to design an efficient routing method to deliver messages for both "upload" (data collection to edge nodes) and "download" (data dissemination to nodes that interest) paths. We assume that the mobile crowd nodes will like to exchange data in a DTN manner based on opportunistic transmission in order to save energy and data transmission cost. We design a probability-based algorithm to upload data carried by normal mobile nodes to the edge nodes. Then, we use cosine similarity to relay specific message of attributes to users who have high interest to receive the message. We simulate the algorithm with the National Chengchi University (NCCU) real trace data of campus students, and compare it with other traditional DTN routing algorithms. The performance evaluations show the improvement of message delivery ratio and decreasing latency and transmission overhead.**

Keywords- Delay Tolerant Network*; *Mobile Crowd Sensing and Computing*; *Opportunistic Mobile Networks; Personal Interests*; *Trace Data

I. INTRODUCTION

The popularity of mobile phones has been growing dramatically recently. Each mobile phone is equipped with many sensors, varieties of wireless communication interfaces (WiFi, Bluetooth, 3G/4G) and sufficient storage space. Therefore, there is a new category of applications arising, namely, Mobile Crowd Sensing and Computing (MCSC) [1]. Unlike traditional wireless sensor networks, MCSC does not require a large number of sensors pre-installed. The mobile devices can cooperate, collect the interested information and exchange with each other. To increase the incentive of the cooperation, the common interest or social relationships may be considered because the mobiles are carried by human beings. Thus, MCSC can be seen as a good way to solve the problem utilizing the power of the human participation.

Mobile Crowd Sensing and Computing has a wide range of applications, such as temperature [2] and air quality detection in the city [3], restaurant recommendations [4] and so on. With the concept of MCSC, there are still many problems need to be solved. The computing power of mobile devices is usually limited or the required data for computation is large. The sensed data from the mobile crowd should be collected into some place such that is suitable for computing. On the other way, the computing results are somehow needed to deliver the interested users who are not restricted to a precise destination. Therefore, both an efficient data collection approach and a message dissemination method are needed to develop.

Nowadays, people can get or dissemination message by social networks, such as Facebook, Twitter. However, the data sensed by mobile crowd are usually to be processed first, and then become usable information to be sent to people who may interest the message.

Previous studies tried to solve the influence maximization problem [5] in the online social network [6][7][8], or how to do trace data processing and data exploration effectively [9]. However, most users are free to participate in MCSC environment [10]. Furthermore, the activity of MCSC should be done on the background and as transparently to users as possible. So there is not enough incentive for users to upload sensor data using their own mobile network with no cost. In addition, the hardware resources and energy of mobile devices are also quite limited. So we utilize the concept the DTN and edge computing to support the MCSC. How to keep data transmission as efficient as possible and save network resources are the key issue for MCSC applications.

The rest of this paper is organized as follows: Section II introduces the motivation using our MCSC scenario as an example. Section III addresses some important related works. Section IV will explain our approach including that the MCSC process is divided into two phases: "upload" and "download" for message collecting and dissemination. Section V validates and evaluates the system performances of our proposed scheme. Section VI concludes this paper.

II. MOTIVATION

Let us think about the following campus scenario as an MCSC example. There are all kinds of messages being spread on a campus. Students carry their mobile devices and move around the campus for attending classes or go to library or restaurant, etc. In this situation, all the students who carry mobile devices on campus are assumed to the mobile crowd nodes. They can generate or collect messages, and receive and transmit certain kinds of messages when they encounter each other. We assume the messages have relevant interest attributes: sports, arts, or social, etc. Students who carry their

 ISBN: 978-1-61208-739-9

message may leave or post the message to the building(s) which we assume to be the edge node(s) for further computing purpose. The edge node(s) gather enough necessary information and process them, and then disseminate the results to the interested people. As in Figure 1, the mobile devices can be the helpers as the routing roles of message relay for collecting and disseminating messages.

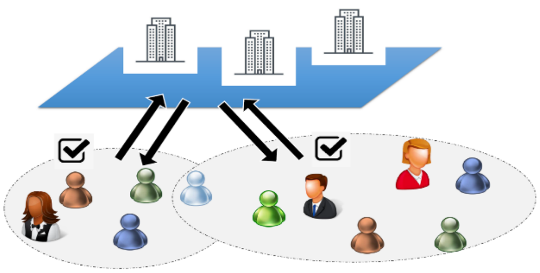


Figure 1. MCSC in our situation.

The facts here impacting the message delivery are the mobility of the students (i.e., the opportunity of encountering somebody to exchange message), and the interest attributes of the message itself and the students (i.e., the willing to carry for relaying the message). There should be a trade-off between message copies and communication overheads. Since students have their hobbies and their class schedules, the message delivery will somehow form as an opportunistic mobile social networks. Our goal of the research will be developing an efficient method for "upload" (message collecting to buildings (edge nodes)) and "download" (computing result dissemination) in this MCSC scenario. The features of MCSC are that the messages collected should be processed or computed first, and then the results could be delivered to those who are interested, but not sure which nodes are destinations in advance as traditional routing.

III. Related Works

Before further illustrating our approach, some important related works are introduced in this section.

A. *Delay Tolerant Network (DTN)*

DTN is a special network transmission concept. Compared with traditional network architectures, there may be situations such as intermittent disconnection and unable to access Internet. DTN allows nodes to store-and-carry messages, and transmit/forward the message when they get access or encounter somebody else later on to further relay. DTN only needs peer-to-peer communication with the help of infrastructure. So, some applications are suitable for using the techniques, such as disaster response networks, and vehicle-to-vehicle networks.

Since Messages can be passed between nodes and nodes by "store, carry and forward", the decision to forward which messages to the nodes they encounter become the key issues for efficiency of delivery ratio and overheads. How to design an appropriate routing method for message transmission with optimal efficiency is an issue for DTN.

B. *Edge Computing*

Edge Computing refers to the processing and operation of data more locally than cloud computing. It helps message moving closer to the data source so as to shorten the delay of network transmission, as well as to obtain the wisdom of data analysis faster.

The network concept is proposed by Cisco. Compared with traditional Internet architecture, fog computing uses layered, local processing distributed network packet transmission to calculate computing requirements. Each fog is directly linked to the local device on the local side. It is responsible for collecting sensor data, raw data and doing preliminary processing. One fog can communicate with other fogs. In addition, it uploads to the cloud so as to do the best calculation, and do regular information update. The current technology has been able to be applied in the temporary system in the smart grid and some houses.

C. *Social Trace Data File*

We need a realistic social trace data file for MCSC simulation. Social trace data include the mobility trace of the nodes and the social relationship and personal interest for the users who carry the mobile sensors. Even the attributes of the building they visited are also described in the data file.

In the DTN environment, the main purpose of social trace data file is including:

1) To simulate the movement history track of the users.

2) To make the forward policy based on the personal data

Our previous research results have completed a data file called NCCU Trace File [19][20] which includes personal mobility trace for two weeks and personal interest profiles, and already used it for evaluating our developed methods in many scenarios.

IV. Our Approach

Due to the features of MCSC, we will divide the MCSC process as two phases: "upload" and "download" as in Figure 2. Nodes will upload the message to any one of the edge nodes, and we assume some edge to be the designated node to compute the collected sensed data for some specific purpose, and to then send the result messages to the users who are interested in. In the upload phase, the message destination can be any one of the edges nodes, i.e., anycast. In the download phase, it is one-to-many transmission, however, we cannot know in advances who will be the destination until the message encounters the nodes to compare the interest attributes.

A. *Model*

Anypath routing means the final message destination can be any one of the candidate destinations.

In our model, the destinations are buildings which we assume to be fixed edge nodes. Mobile crowd may encounter one of the edge nodes directly or encounter other mobile node before visiting buildings. In the former case, the message that mobile crowd carried can be directly forwarded to the edge node (anycast destination). In the latter case, the node should

 ISBN: 978-1-61208-739-9

decide whether forward the message to the encountered node will have a better chance to faster relay message to destinations.

Since node encountering is based on "opportunity", and is not certain to meet the "right" person in the near future. We use the concept of most appropriate "forwarding set" to estimate the "cost" for relaying if we forward the message to the encountered node. That is, we determine the appropriate forwarding set by calculating the cost of the transmission between the node and the edge node set. The details will be described in the following section.

According to the above, our model has many nodes (many users in the trace data), and an edge node set (buildings in the trace data). We use the delay-tolerant network technology to transfer the message. And we use anypath routing algorithm to select appropriate nodes to become relay nodes for "upload". Messages are carried and uploaded by each node to the edge for pre-processing, and then be downloaded to the nodes that need the messages (as in Figure 2).

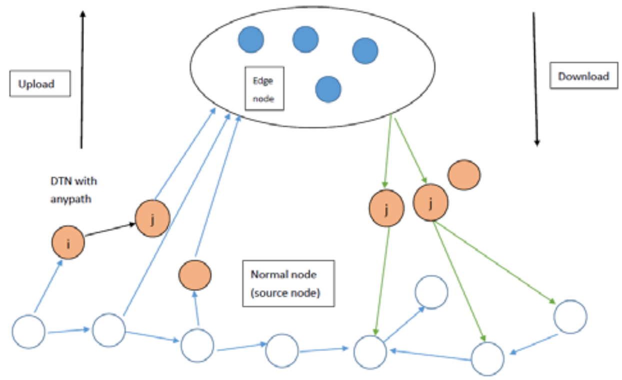


Figure 2. Our approach model.

B. Cost Probability

Due to easy implementation and low maintenance cost, Bellman-Ford-like algorithm is adopted to calculate the cost from node *i* to any destination through the forwarding set. We define the link cost C_{ij} as the reciprocal of the encounter probability between node *i* and *j* (P_{ij}). The forward set J for node *i* will be the future most likely encountering nodes to be forwarded messages to any one of destination set E. Thus, the anypath routing cost for node i to E:

$$C_{iE} = C_{iJ} + C_{JE}$$

$$P_{ij} = \frac{i\ encounter\ j}{\text{i encounter all other nodes}}$$

$$C_{iJ} = \frac{1}{P_{iJ}} = \frac{1}{1 - \prod_{j \in J}(1 - P_{ij})}$$

C_{iJ} is the cost of reciprocal of the probability that the node *i* meets at least one node in the relay set J. Message can be transferred to any node j in Set J to help relay. The cost for Set J to any edge node e_x, C_{Je_x} is:

$$C_{Je_x} = \sum_{j \in J} w_{ij} C_{je_x}$$

where, C_{je_x} represents the path cost for node j to edge node e_x, and w_{ij} the normalized probability that node *i* meets the specific node *j* in the relay node set J.

$$w_{ij} = \frac{P_{ij} \prod_{k=1}^{j-1}(1 - P_{ik})}{1 - \prod_{j \in J}(1 - P_{ij})}, \quad \sum_{j \in J} w_{ij} = 1$$

Finally, the C_{JE} can be derived as the following:

$$C_{JE} = \min\{C_{Je_x}\}$$

C_{JE} indicates the cost that relay node set J reach at least one of an edge node e_x in an edge node set E. Edge nodes will do the message preprocessing as mentioned before, as long as enough message are uploaded and concentrated to a designated node to computing results.

Whenever, node *i* encounters node *j*, the C_{iE} and C_{jE} will be compared, then make the forwarding decision based on Bellman-Ford shortest path algorithm.

C. Relay Node Set

We use the above calculation to select the appropriate nodes to form the relay node set. The basic idea is to choose relay nodes with the highest probability to future encounter and lowest cost to any edge nodes. Actually, the relay node set should use appropriate size to estimate the anypath cost better. Our solution set the size of relay node set to be three with minimum cost to edges. (as shown in Figure 3)

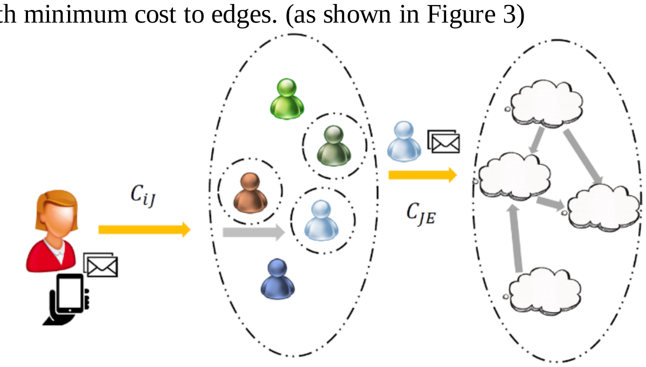


Figure 3. Selection of Relay Nodes.

D. Cosine Similarity

In the download phase, the result message will be delivered to those who are interested. To decide whether the user interests the message or not, we use "cosine similarity". There are many methods in computer science that can help us measure vector similarity. Vector similarity can help us classify different categories. Classic vector similarity includes Euclidean Distance, Cosine Similarity, Hamming distance, or Jaccard similarity, etc. Different similarity methods are used at different cases due to their characteristics. For proof of concept only, we use the cosine similarity for simplicity. Since each message has its interest attributes and every student has his/her interest profile, we define the interest *I* between node *j* and message *m* as the following:

$$I(j,m) = \frac{\vec{j}, \vec{m}}{\|\vec{j}\| \cdot \|\vec{m}\|} = \frac{\sum_{k=1}^{n} j_k \times m_k}{\sqrt{\sum_{k=1}^{n}(j_k)^2 \times \sum_{k=1}^{n}(m_k)^2}}, \ k < n$$

 ISBN: 978-1-61208-739-9

$\vec{j}$ is the interest vector of node j, and $\vec{m}$ is the interest vector of message m. The relay node j has its own interest attributes same as message attributes(i.e. m_k and j_k includes sports, art, reading, service, social like we mentioned before). The message sensed/generated by the initial node also has its own interest attributes. We compare the message and the relay node by each other's interests, and use cosine similarity to set up an association. We also set a similarity threshold. If the cosine similarity of message and node is greater than the similarity threshold, we can determine that the node is interested in this message enough. We will relay the message to this relay node, and this node will also become the destination of this message in the download phase.

E. *Influence Gain*

We calculate whether this node is appropriate to help us relay based on the history record of this node. Our NCCU trace data has two-week data. If the data of first week is calculated, the second week's data will be used as a reference. On the contrary, if the data of second week is used, the first week's data will be used as a reference. The reason is that our NCCU trace uses campus data, so we can assume that the student's mobile and interactive records may be related to their weekly schedule whenever weekdays or holidays.

We consider that node i with message m encounters node j: If the node j does not have enough interest in the message m, we then determine whether node j might meet with high possibility other nodes who are interested. If yes, then the message could be relayed to it. To do this, we introduce the influence gain that calculates the normalized "interest inherited" from those who might be met in the future. The inherited interest is sum of the interest of those who met on the same day of the other week of the record. Thus, influence gain (IG) of nod i is:

$IG(i) = \sum_{j \in J_D} P_{ij} * \text{normalized\_inherited\_interest}(j)$,where J_D is the forward set of node i on the same day D.

If $I(IG(j), m) > I(IG(i), m)$, and $I(IG(j), m) >$ Threshold, then the message m will be forwarded to node j.

V. SIMULATION

We use The One Simulator, a network simulator based on the action mass perception network developed by the Norwegian Nokia research engineer. The "ONE" is the acronym for The Opportunistic Network Environment simulator, which is an open source development tool available on GitHub.

The One is made with Java and implemented in a GUI interface. It can fully simulate the routing results of network nodes over a specific time period. Among the ONE simulator, we have built-in multi-node transmission methods including one-to-one, many-to-many, one-to-many and many-to-one. There are also built-in Epidemic, Spray & Wait, Prophet and other classical DTN routing methods for users to do simulation experiments. Since the MCSC in our cases is different from traditional DTN routing, we can only compare with epidemic which is similar to flooding as the baseline (anycast in the upload phase, and no information about destinations in the download phase at the time of route computation). Existing related works cannot be directly applied to these MCSC cases.

For the sake of fair comparisons, in our experiments, we modified the epidemic routing to be fit in our scenario for both the upload and download phases. Notice that the message received in the upload phase should exclude the duplicate; however, in the download phase there are possible many destinations for each message.

A. *Latency*

As the latency is concerned, compared with the epidemic, our method achieves better performance in most days of the week as depicted in Figure 4. This is because our method forwards the message to appropriate nodes without overloading the nodes' buffer. The epidemic method might quickly fill the buffer and takes time to "digest" the message, thus, incurring more latency.

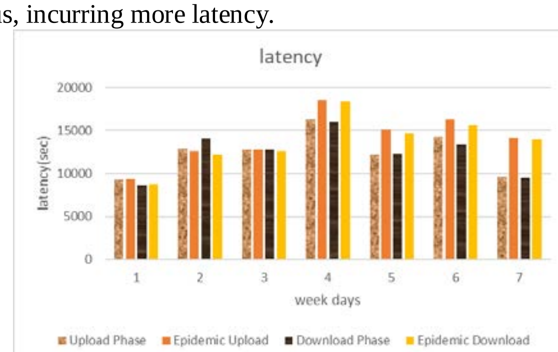


Figure 4. Latency simulation.

B. *Hop Count*

Average Hop count for the path to destinations is also one of the important factors in DTN consideration. Our method outperforms epidemic (in Figure 5). This again confirms our method picks more suitable relay node(s) to help relay messages efficiently. Especially in download, the hop count is significantly reduced.

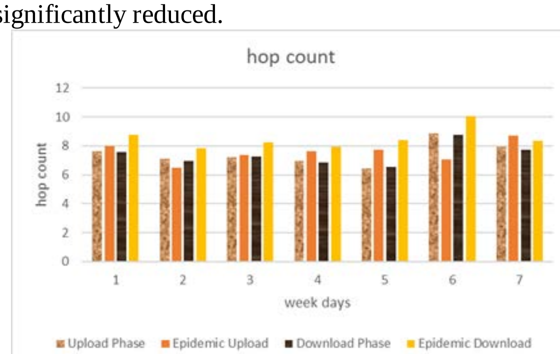


Figure 5. Hop count simulation.

C. *Delivery Ratio*

We calculate the delivery ratio from the process of generating the message to the end of the arrival. It is expressed as the proportion of number of messages successfully delivered to the number of all messages transmitted. The formula can be expressed as:

$$\text{Delivery Ratio} = \frac{\text{Number of successful msg to Destination}}{\text{Number of msg be sent}}$$

Because of efficient selection of relay node set, the number of messages that can reach the destination is much higher. So, in most cases there will be a higher delivery ratio whether it is upload or download phase in our approach model.(Figure 6)

Figure 6. Delivery ratio simulation.

D. Overhead ratio

Finally, the overhead ratio is considered. The overhead ratio here refers to the proportion of messages that are wasted among all the transmitted messages. The formula can be expressed as:

$$\text{Overhead Ratio} = \frac{\sum_{m_i}^{M} \frac{\text{Relayed} - \text{DestinationRelayed}}{\text{DestinationRelayed}}}{\text{Total Message Number}}, \forall m_i \in M$$

In the process of Epidemic, because the probability of random transmission is very high, the amount of information received by each node is likely to cause buffer problems. Because our message is composed of interests options, so it is easy for messages to be truly transmitted to destination with fully interests. (Figure 7)

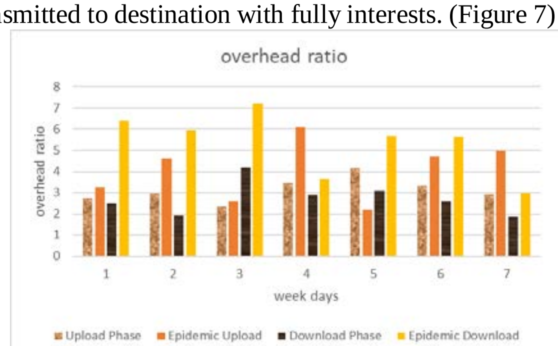


Figure 7. Overhead simulation.

E. Weekdays and Weekend

Our trace data is mainly to record the walking and interaction records of campus students to help us deliver messages in two weeks. Therefore, our data and simulation results may have more obvious differences between weekdays and weekends compared with other general environment or situation.

According to the results (Figure8-11), Weekdays have fewer hop counts than weekends because there are more choices in the case of a large number of students. Especially our approach still has a good overhead ratio. In the epidemic, there is a higher overhead in the weekdays because of the number of transmission choices.

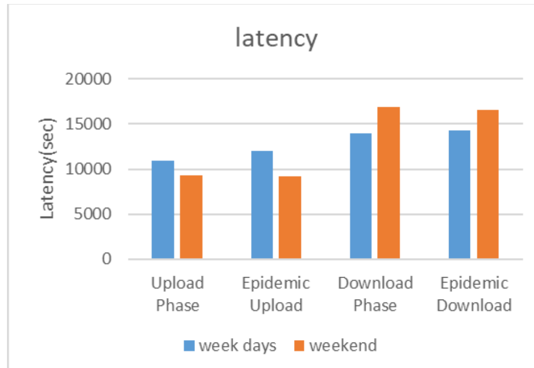


Figure 8. Latency of week days and weekend.

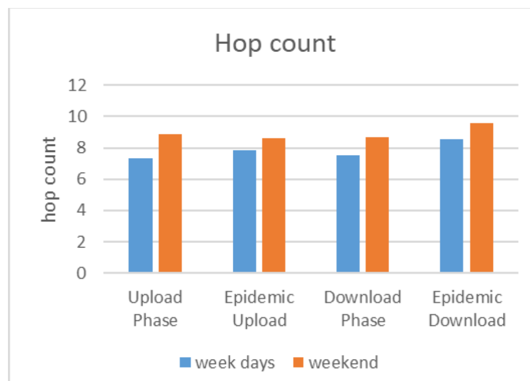


Figure 9. Hop count of week days and weekend.

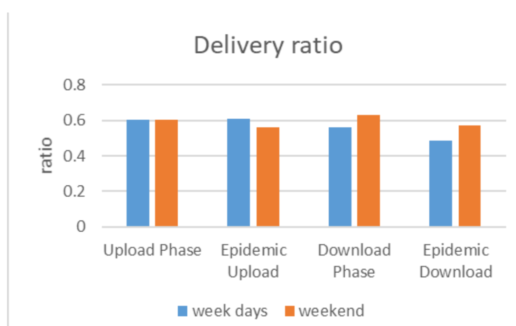


Figure 10. Delivery ratio of week days and weekend.

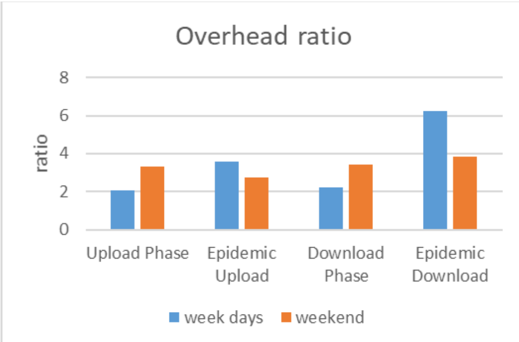


Figure11. Overhead of week days and weekend.

VI. Conclusion and Future Works

This paper presents the design of an efficient approach with upload and download phase for MCSC applications. The developed approach tries to transfer the specific message with interests based content that everyone carries to the people who really need it. We use the concept of Mobile Crowd Sensing and Computing to bring the power of the masses to the limit by the mobile device in the user's hand as the node. We use the NCCU trace data record as the benchmark to calculate the probability of encounter and reaching the edge node in the upload phase. According to the calculation result, the message will be uploaded to the main processing edge. The edges will help compute messages. The result messages will be ready to be transmitted for the download phase. We propose the inherited interest from those who are possibly met in the future to determine the relay nodes. The performance evaluations show the improvement of message delivery ratio and decreasing latency and transmission overhead.

We might consider the latency constraints for the message into our routing method in the future. Although the delay in the DTNs is not very stringent, it still needs the lifetime limit in some cases or to avoid bandwidth wastage. Another direction is the buffer size. Since DTNs use the store-carry-and-forward approach to relay messages, the mobile nodes may store too many messages and over utilize their computing capabilities. A real mobile APP can be practiced to ensure the merits of this research.

References

[1] H. Ma, D. Zhao, and P. Yuan, “Opportunities in mobile crowd sensing”, IEEE Communications Magazine, August, 2014.

[2] M. Mun, et al, “PEIR, the personal environmental impact report, as a platform for participatory sensing systems research”, Proceedings of the 7th international conference on Mobile systems, applications, and services. ACM, 2009. pp.55-68.

[3] C. Xiang, et al, “Passfit: Participatory sensing and filtering for identifying truthful urban pollution sources”, Sensors Journal, IEEE, 2013.

[4] S. Gaonkar, J. Li, R. Choudhury, L. Cox, and A. Schmidt, “Micro-blog: sharing and querying content through mobile phones and social participation”, In: Proceedings of the 6th international conference on Mobile systems, applications, and services. ACM, 2008. pp. 174-186.

[5] D. Kempe, J. Kleinberg, and E´. Tardos, “Maximizing the spread of influence through a social network”, ACM SIGKDD, 2003, pp. 137–146.

[6] W. Chen, Y. Wang, and S. Yang, “Efficient influence maximization in social networks”, ACM SIGKDD, 2009, pp. 199–208.

[7] M. Han, M. Yan, Z. Cai, and Y. Li, “An exploration of broader influence maximization in timeliness networks with opportunistic selection”, Journal of Network and Computer Applications, 2016.

[8] T.Shi, J.Wan, S.Cheng, Z.Cai, Y.Li, and J.Li, “Time-bounded positive influence in social networks”, in IIKI, 2015.

[9] B. Guo, et al, “Mobile crowd sensing and computing: The review of an emerging human-powered sensing paradigm”, ACM Computing Surveys (CSUR), 2015, 48.1: 7

[10] T. Higuchi, H. Yamaguchi, T. Higashino, and M. Takai, “A neighbor collaboration mechanism for mobile crowd sensing in opportunistic networks”, Communications (ICC), 2014 IEEE International Conference on. IEEE, 2014. pp.42-47.

[11] N. Eagle and A. Pentland, “Reality mining: sensing complex social systems”, Personal and Ubiquitous Computing, Vol 10(4):255–268, May 2006.

[12] P. Hui, “People are the network: experimental design and evaluation of social-based forwarding algorithms”, Ph.D. dissertation, UCAM-CL-TR-713. University of Cambridge, Comp.Lab., 2008

[13] R. Marin, C. Dobre, and F. Xhafa, “Exploring Predictability in Mobile Interaction”, EIDWT. 2012.

[14] J. Cai, M. Yan, and Y. Li, “Using crowdsourced data in location-based social networks to explore influence maximization”, The 35th Annual IEEE International Conference on Computer Communications (INFOCOM 2016). 2016.

[15] J. Qin, et al, “Post: Exploiting dynamic sociality for mobile advertising in vehicular networks”, IEEE Transactions on Parallel and Distributed Systems, 2016.

[16] R. Mayer, H. Gupta, E. Saurez and U. Ramachandran, “The Fog Makes Sense: Enabling Social Sensing Services With Limited Internet Connectivity”, ACM New York, 2017.

[17] K. Dolui, and S. Datta, “Comparison of edge computing implementations: Fog computing, cloudlet and mobile edge computing”, IEEE Global Internet of Things Summit (GIoTS), 2017.

[18] T. Tsai, and H. Chan, “NCCU Trace: social-network-aware mobility trace”, Communications Magazine, IEEE, 2015.

[19] T. Tsai, H. Chan, C. Han, and P. Chen, “A Social Behavior Based Interest-Message Dissemination Approach in Delay Tolerant Networks”, International Conference on Future Network Systems and Security. Springer International Publishing, 2016. pp. 62-80.

[20] R. Laufer, H. Dubois-Ferriere ,and L. Kleinrock, “Multirate Anypath Routing in Wireless Mesh Networks”, IEEE INFOCOM, 2009.

[21] H. Li, T. Li, W. Wang, and Y. Wang, "Dynamic Participant Selection for Large-Scale Mobile Crowd Sensing," in IEEE Transactions on Mobile Computing.

[22] Z. Zhou, H. Liao, B. Gu, K. M. S. Huq, S. Mumtaz, and J. Rodriguez, “Robust Mobile Crowd Sensing: When Deep Learning Meets Edge Computing”, IEEE Network Volume: 32, Issue: 4, July/August 2018.

Initial Investigation of Position Determination of Various Sound Sources in a Room

Takeru Kadokura Yuki Hashizume Shigenori Ioroi Hiroshi Tanaka
Course of Information & Computer Science
Graduate School of Kanagawa Institute of Technology
Atsugi-Shi, Kanagawa, Japan
email: {s1885004, s1621081}@cco.kanagawa-it.ac.jp, {ioroi, h_tanaka}@ic.kanagawa-it.ac.jp

***Abstract*—Special sounds, such as ultrasonic waves and diffused sound sources used to perform high precision indoor positioning. If the positioning of various ordinary sound sources in a room becomes possible, it is thought that numerous applications are likely. This paper proposes a new method to estimate with high accuracy the position of various ordinary indoor sound sources, rather than using any special sound source. We constructed an environment where positioning experiments can be performed by knowing the size of the actual sonic environment. Positioning experiments were conducted using the sound of an operating microwave oven, the ringing tone of a telephone, and a diffused sound source. Highly accurate positioning about +/- 20 cm could be realized for the diffused sound and the microwave oven. In contrast, it was confirmed that the ringtone had a positioning error exceeding 1 m.**

Keywords-Positioning; Sound Source; TDOA; CSP Analysis; Real Environment.

I. INTRODUCTION

Many methods and technologies have already been proposed for indoor positioning, including the use of radio waves, such as Wi-Fi and BLE, and those using inertial sensors, such as acceleration sensors [1]. At present, the positioning accuracy and positioning area differ depending on the principle and method, and each system is individually selected according to the user's requirements. The most widely studied methods are those using a wireless LAN [2]. There is also a positioning method using BLE [3], which has a narrower radio propagation range. Although the devices used in this method are already in widespread use and are not limited in terms of use, the positioning accuracy is about m order, and high-accuracy position detection is difficult.

The authors have been studying systems capable of highly accurate positioning by using sound, with its low propagation speed. These systems attach sound sources to a positioning target, including ultrasonic [4] and spread spectrum sound [5], and then detect the target position. In this method, although positioning accuracy within several centimeters can be secured, attachment of a dedicated sound source is essential, so its application is restricted. There is also a system that achieves high positioning accuracy (about 10 mm) by obtaining the reception time of the sound wave with high accuracy using radio waves and sound waves [6]. However, the configuration is complicated, and the application area seems to be limited to a region where some extremely accurate positioning is required.

To examine a room, there are various sound sources, such as a device that emits a notification/warning sound like a calling buzzer, the voice of a person, and home electrical appliances. Also, a drone can be a sound source. If the positions of the various sound sources are known, the utility will be greatly expanded. For example, the location of the speaker, indoor flight control of the drone, and indirect identification of a sound source from the position of the sound source can be considered. A lot of studies have been conducted to estimate the direction of the sound sources in an indoor area [7] [8]. The purpose of these investigations is to separate each sound from multiple sound sources, and the position of the sound source is not obtained.

Our positioning method is based on the Time Difference Of Arrival (TDOA) method [9], which conducts positioning calculations using the reception time differences of the sound wave between a reference point receiver and another three or more reception points. The Cross-power Spectrum Phase (CSP) analysis method has been proposed [10] as a method to detect the reception time difference of sound waves at two reception points. There are also studies in which positioning was performed using the time difference obtained by this method. However, in these papers, the arrival direction of the sound source was obtained by time difference, the position was obtained as an intersection of the sound direction [11]. Although this is a simple calculation, the position could not be determined with high accuracy. Because it didn't use the information on the reception time differences obtained from other microphone sensors existing neighbors. The sound source is only speech sound.

This paper describes the possibility of high accuracy positioning in the same area as an actual room environment for various indoor sound sources. The presentation is as follows. Section 2 describes the positioning principle and the essential method of detecting the reception time difference including the method of selecting reference points. Section 3 shows an experimental environment where the receiving points are installed on the ceiling of the laboratory to secure the same area as an actual usage environment. Section 4 shows the results of detecting reception time differences for three types of sound sources, and the positioning results. Section 5 presents a summary.

II. POSITIONING PRINCIPLE

The positioning method we have applied is based on the TDOA scheme. The positioning calculation is conducted by

 ISBN: 978-1-61208-739-9

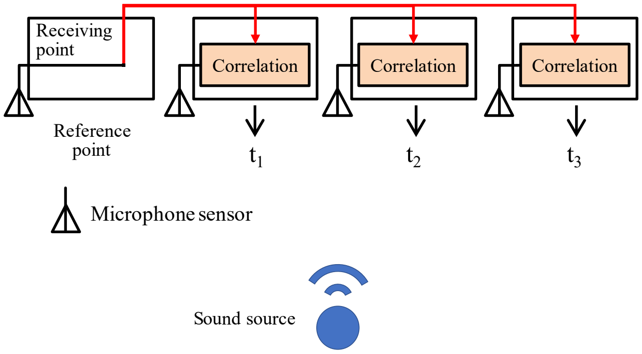


Figure 1. Principle of detection of time difference.

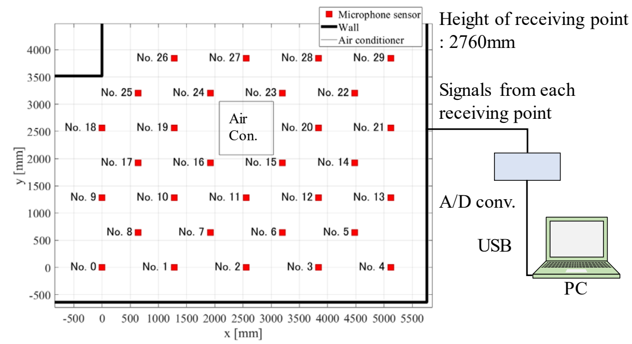


Figure 2. Microphone sensor installation position and experimental configuration.

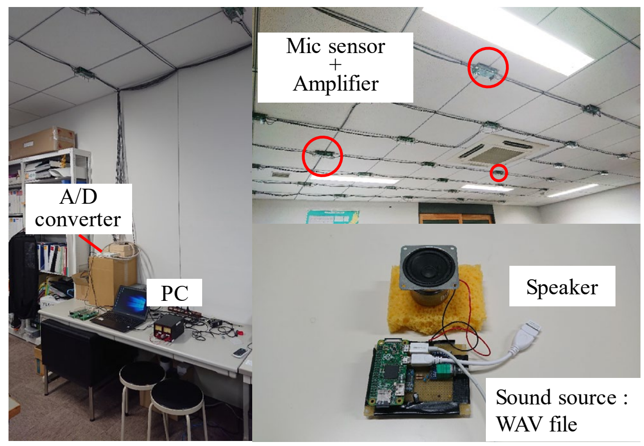


Figure 3. Experimental environment appearance.

using the reception time differences between a reference receiving point and some other points. Positioning is conducted by using (1). This equation for positioning is the same as that in a GPS/GNSS, in which radio signals are used. In previous studies, a special sound source provided a sound that had been diffused using an M sequence code. The receiving side had the same sound source data as that of the transmitting side (replica), and detected the sound reception timing by cross correlation calculation between the received signal and the replica [5].

$$\begin{aligned}
\sqrt{(x-x_0)^2+(y-y_0)^2+(z-z_0)^2} &= ct \\
\sqrt{(x-x_1)^2+(y-y_1)^2+(z-z_1)^2} &= c(t+t_1) \\
\sqrt{(x-x_2)^2+(y-y_2)^2+(z-z_2)^2} &= c(t+t_2) \\
\sqrt{(x-x_3)^2+(y-y_3)^2+(z-z_3)^2} &= c(t+t_3)
\end{aligned} \tag{1}$$

where,

t : propagation time [s]

x, y, z : position of transmitter [mm]

t_i : propagation time difference to each microphone sensor [s]

c : speed of sound [mm/s]

x_i, y_i, z_i : installation position of each microphone sensor [mm]

In this investigation, we considered the several sounds in the room. This makes it difficult to implement a replica on the receiving side as can be done when the conventional method is used. The reception time difference is obtained by cross correlation between the signal received at the reference point and the signal at other reception points as shown in Figure 1. Based on the reception time differences obtained with this configuration, the positioning calculation is performed as it is in the conventional method. In this calculation, it is necessary to determine the reception point to use as the reference point.

III. COMPOSITION OF EXPERIMENT ENVIRONMENT

To evaluate the reception time differences by the CSP method, a microphone sensor (Primo EM-258) was attached using a frame of about 1 m^2, and various sounds were produced by a speaker (Tang Band W2-858S) at the center of the frame, in the configuration used last year. In this investigation, the dimensions of the room were obtained to relate the sound detection time differences to the spatial separations, for positioning accuracy evaluation. The 30 receivers integrated with the microphone sensor and the amplification circuits were attached to the ceiling of the laboratory room. Figure 2 shows the installation positions and the experimental configuration. They were attached at intervals of 1.2 m, except for the area near the air conditioner. The height of the ceiling is 2.8 m. The appearance of the experimental environment is shown in Figure 3.

The sound of an operating microwave oven as a common home appliance, and the ring tone of a phone, were used as sound sources in the room. It was decided to evaluate using three types of sound sources, one of which was a diffused sound source from which high accuracy can be expected. The sounds were recorded in advance in WAV format, and each sound was sent out from a speaker, including the diffused sound source. The sampling frequency on the receiving side was 16 kHz, and the number of received samples was 10000.

IV. EXPERIMENTAL RESULTS

In this section, the results of the experiment for detecting the reception time difference of each sound source, and the positioning experiment results obtained from the time difference are described by dividing into two subsections.

A. Time difference detection

The spectrogram of the sound sources is shown in Figure 4. It can be confirmed that the diffused sound source has a uniform frequency distribution and that the other sound sources have identifiable frequency characteristics. It was confirmed that all receivers could receive correctly.

The reception time difference between two receivers of the various sound sources was determined using the CSP method shown in (2).

$$\boldsymbol{CSP}_{x,y} = DFT^{-1}\left[\frac{DFT[\boldsymbol{x}]\overline{DFT[\boldsymbol{y}]}}{|DFT[\boldsymbol{x}]||DFT[\boldsymbol{y}]|}\right] \quad (2)$$

where,

$\boldsymbol{CSP}_{x,y}$: normalized cross-power spectrum

$\boldsymbol{x}$: received signal at reference receiving point

$\boldsymbol{y}$: received signal at target receiving point

The reception time difference between two points is given as follows.

$$\tau = \arg\max_{k}(\boldsymbol{CSP}_{x,y}) \quad (3)$$

The reception time difference is obtained from the data of two receiving points, that is, microphone sensors. The number of combinations of calculations for the reception time difference between two points becomes enormous, and the calculation time for that will greatly affect the positioning time interval. By determining a reference point for obtaining the reception time difference, it becomes a realistic calculation. The following two methods were compared in this investigation for selecting the reference point.

Method (1): the receiving point directly above the sound source (Since the sound source position is unknown, this method cannot be applied in real life circumstances.),

Method (2): the receiving point at which the signal strength is the highest, here, that strength is the sum of the sound energy received during the positioning time.

The reception time difference of various sound sources between two points, that is, a reference point and other points, was determined using the CSP method. An example of the cross correlation waveform by the CSP method is shown in Figure 5. The result from method (1) is in the upper column, and that from method (2) is the lower column. It was confirmed that although there is a difference in peak sharpness depending on the sound source, it is possible to detect the peak position indicating the reception time difference. The peak position was detected 100 times, and the mean and standard deviation were evaluated. The results are shown in Table I. Here, since the reference point changes every time in method (2) (reception strength fluctuates), evaluation method (2) was not used. Although the time difference matched with the theoretical value was obtained for the sound of the diffused and the microwave sound, it was confirmed that these values became unstable for the ringing tone of the telephone. The theoretical value is the reception time difference of the sound determined from the positions of the speaker and the two reception points.

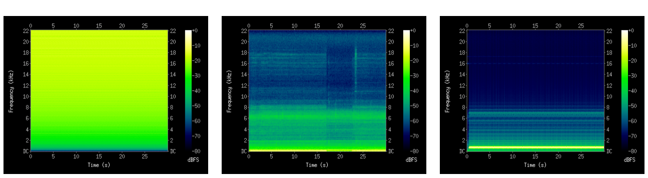


Figure 4. Spectrogram of received sound.

B. Positioning experiment result

The positioning results produced by 4 kinds of experimental conditions were examined. One sound source was set directly below reception point No.11, and another was set at an intermediate position between reception points No.14 and No.15. Two methods of reception time difference detection were applied to each case. Environmental noise was 36.9-55.6 dB at the time of the experiment.

An example of a positioning experiment result is shown in Figure 6. This is when the sound source was placed on a table 70 cm in height and below a position between reception points No.14 & No.15, and the reception time difference was detected by method (2). The positioning results of three sound sources are shown. While the positioning of the diffused sound and the operating sound of the microwave oven can be realized with high accuracy, a large error occurred in locating the ring tone of the telephone.

The positioning experiment results are summarized in Table II. Average errors and Root Mean Square (RMS) errors are shown for 100 times positioning results. Here, any solution that was out of the positioning range was excluded, that is, any that was not in the area of the receiving points, and the number of exclusions in 100 tests of positioning is also shown as a ratio in the table. The difference in positioning accuracy due to a particular sound source can be confirmed as shown in Figure 6.

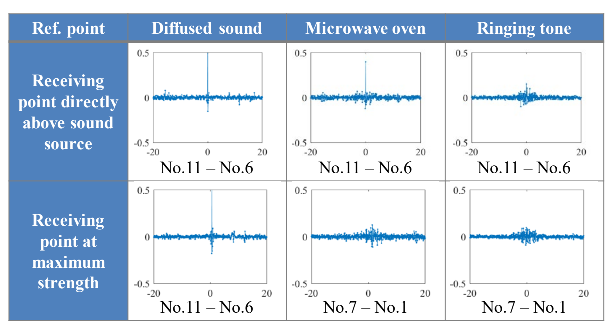


Figure 5. An example of cross correlation waveform by CSP method.

TABLE I. DETECTED RECEPTION TIME DIFFERENCES

	Reception time difference [ms]						
	Diffused sound		Microwave oven		Phone ringing		Theory
	μ	σ	μ	σ	μ	σ	μ
t1	0.56	1.1E-18	0.56	1.1E-18	0.24	7.9E-04	0.55
t2	0.56	1.1E-18	0.56	1.1E-18	0.32	7.1E-04	0.55
t3	0.56	1.1E-18	0.56	1.1E-18	0.23	5.0E-04	0.55
t4	0.56	1.1E-18	0.56	1.1E-18	0.17	7.0E-04	0.55

 ISBN: 978-1-61208-739-9

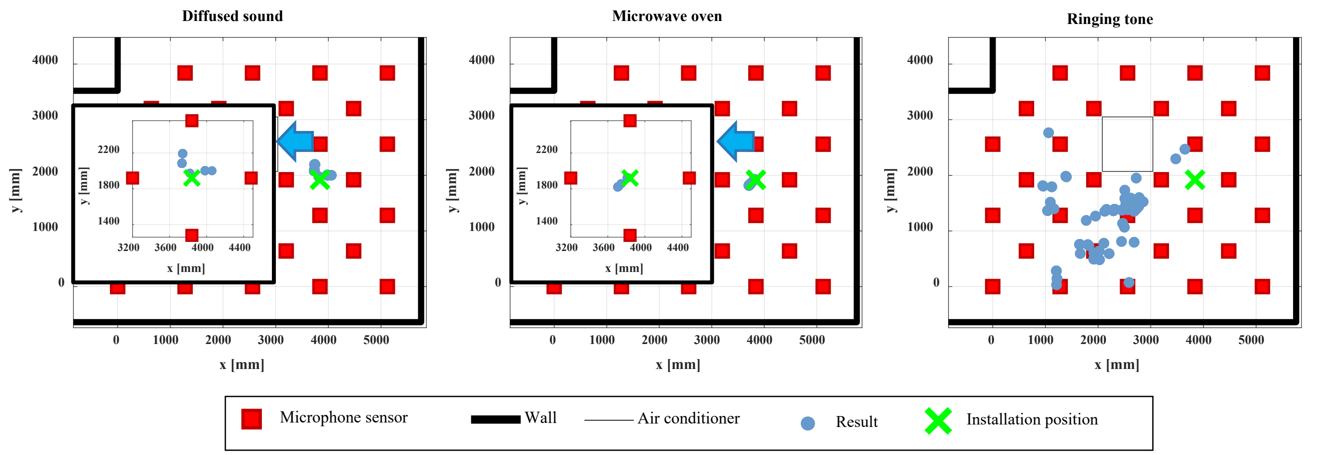


Figure 6. Example of positioning experiment result.

TABLE II. EVALUATION OF POSITIONING ACCURACY

Reference point	**Fix (upper position) [mm]**				**Maximum receiving level [mm]**			
Sound source position (2560,1280,700)								
Sound source	***X_average error***	***Y_average error***	***RMS error***	***Ratio***	***X_average error***	***Y_average error***	***RMS error***	***Ratio***
Diffused sound	0.0	0.0	0.0	100	0.0	0.0	0.0	100
Microwave oven	0.0	0.0	0.0	100	19.6	19.6	27.8	100
Phone ringing	-110.2	315.8	1118.5	87	-496.8	-143.7	1167.8	69
Sound source position (3840,1920,700)								
Sound source	***X_average error***	***Y_average error***	***RMS error***	***Ratio***	***X_average error***	***Y_average error***	***RMS error***	***Ratio***
Diffused sound	-105.5	181.7	213.9	100	-84.0	174.1	211.6	100
Microwave oven	-13.3	156.5	157.0	100	-39.1	-6.5	56.6	98
Phone ringing	-2.8	-666.6	667.2	100	-1717.4	-710.1	2006.2	98

The reason why the positioning accuracy by the ringing tone is worse than the other is that the reception time difference cannot be detected accurately as shown in Table I. The sound reception time is obtained from the peak position. The peak of ringing tone is not clear compared to those of the diffused sound and the microwave oven. The correlation between the two received points is obtained in the CSP calculation. The peak of CSP calculation becomes sharp when the sound source has a uniform frequency component over a wide frequency range. As can be seen from Figure 4, the ringing tone does not cover a wide frequency range compared to the diffused sound source and the microwave oven sound, so it is difficult to obtain a clear peak in the CSP calculation value. The error in reception time difference obtained from the peak is considered to be increased due to the unclearness of the peak because another peak might be selected as reception time difference. For this reason,-the positioning accuracy is deteriorated for the ringing tone.

V. CONCLUSION

An experimental system was constructed in the space of an actual usage environment in order to clarify a positioning method for various sound sources in an indoor environment. A method of selecting a reference point to detect the reception time difference of sound waves was investigated and the time difference was obtained by using the CSP method. The positioning experiment was carried out using the TDOA method.

Together with a diffused sound source, which is a dedicated sound source expected to allow high positioning accuracy, position estimation was performed using the sound of an operating microwave oven and the ringing of a telephone as indoor sound sources. High-precision positioning resulted with an RMS error of about 20 cm for the diffused sound and the microwave oven, but a positioning error of more than 1 meter occurred for the ringing tone. Considering the size of the sound source and the ability to discriminate between sound sources, it may be possible to use this technique if the acceptable positioning error is a few tens of centimeters. It remains as a next step to determine the cause of the lower location accuracy of the ringing tone.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Numbers JP17K00140.

REFERENCES

[1] A. Yassin, et al., “Recent Advances in Indoor Localization: A Surveyon Theoretical Approaches and Applications,” IEEE Communications Surveys & Tutorials, vol. 19, no. 2, Second Quarter, pp. 1327-1346, 2017.

[2] Z. Hailong, H. Baoqi, and J. Bing, “Applying kriging interpolation for WiFi fingerprinting based indoor positioning

systems," 2016 IEEE Wireless Communications and Networking Conference, pp. 1822-1827, 2016.

[3] M. Ji, J. Kim, J. Jeon, and Y. Cho, "Analysis of positioning accuracy corresponding to the number of BLE beacons in indoor positioning system," 2015 17th International Conference on Advanced Communication Technology, pp. 92-95, 2015.

[4] M. Akiyama, H. Sunaga, S. Ioroi, and H. Tanaka, "Composition and Verification Experiment for Indoor Positioning System using Ultrasonic Sensors," Journal of the Institute of Positioning, Navigation and Timing of Japan, vol. 3, no. 1, pp. 1-8, 2012 (in Japanese).

[5] S. Murata, C. Yara, K. Kaneta, S. Ioroi, and H. Tanaka, "Accurate Indoor Positioning System using Near-Ultrasonic Sound from a Smartphone," Int. Conference on Next Generation Mobile Apps, Service and Technologies, pp. 13-18, 2014.

[6] J. Qi and G.P. Liu, "A Robust High-Accuracy Ultrasound Indoor Positioning System Based on a Wireless Sensor Network," Sensors, 17(11), 17 pages, 2017.

[7] Z. Chen and J. Wang, "ES-DPR: A DOA-Based Method for Passive Localization in Indoor Environments," Sensors 19(11), 17 pages, 2019.

[8] B. Suksiri and M. Fukumoto, "An Efficient Framework for Estimating the Direction of Multiple Sound Sources Using Higher-Order Generalized Singular Value Decomposition," Sensors, 19(13), 27 pages, 2019.

[9] T. H. Do and M. Yoo, "TDOA-based indoor positioning using visible light," Photonic Network Communications, vol. 27, Issue 2, pp. 80-88, 2014.

[10] M. Omologo and P. Svaizer, "Use of the Crosspower-spectrum Phase in Acoustic Event Location," IEEE Trans. on Speech and Audio Processing, vol. SAP-5, no. 3, pp. 288-292, 1997.

[11] T. Nishiura, T. Yamada, S. Nakamura, and K. Shikano, "Localization of multiple sound sources based on a CSP analysis with a microphone array," 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 1053-1056, 2000.

A Neural NLP Framework for an Optimized UI for Creating Tenders in the TED Database of the EU

Sangramsing N Kayte

Department of Mathematics and
Computer Science (IMADA),
University of Southern Denmark (SDU),
Odense, Denmark
Email: sangram@imada.sdu.dk

Peter Schneider-Kamp

Department of Mathematics and
Computer Science (IMADA),
University of Southern Denmark (SDU),
Odense, Denmark
Email: petersk@imada.sdu.dk

***Abstract*—The developments in the fields of web technologies, digital libraries, technical documentation, and medical data have made it easier to access a larger amount of textual documents, which can be combined to develop useful data resources. Textmining or the knowledge discovery from textual databases is a challenging task, in particular when having to meet the standards of the depth of natural language that is employed by most of the available documents. The primary goal of this research is to implement the Machine Learning algorithms like RecurrentNeural Networks (RNN) and Long Short Term Memory networks(LSTM) techniques that is applied for text mining / text prediction in the context of creating an optimized user interface for tender creation in the Tender Electronic Daily database used for public procurement throughout the European Union (EU). This paper explains the scope and concept of Natural Language Processing(NLP) for such text mining projects. We also present a detailed overview of the different techniques involved in Natural language processing and Understanding along with the related work.**

Keywords–Natural Language Processing, Natural Language Understanding, Natural User Interfaces, Named-Entity Recognition, Logistic Regression, European Union, Tender Electronic Daily, Common Procurement Vocabulary

I. Introduction

Natural language processing (NLP) is a emerging field which involves the combination of different fields like linguistics, computer science, and artificial intelligence leveraged to build assertive devices for variants of language assistance [1]. The foundation of NLP is laid down with the version of punch cards and now evolved to the heights of computing operations with infractions. [2]. NLP is broaden area to perform a wide range of natural language-related tasks at all levels like parsing, Part-Of-Speech (POS) tagging to different forms of machine response [3].

The advancement of Machine learning algorithms and deep learning architectures in NLP research areas is demanding and increasing every day. The early experiments in machine learning for NLP was targeting NLP problems which were based on shallow models (e.g., support vector machines and logistic regression) trained on very high dimensional and sparse features. But last few years, the neural networks based on dense vector representations producing outstanding superior results on various NLP tasks [4]. The significant improvement is with the implementation of word embeddings and deep learning methods. Deep learning enables multi-level automatic feature representation learning. This has replaced the traditional machine learning techniques used in NLP systems which relies heavily on hand-crafted features, most of them are time-consuming and often incomplete [5].

The NLP task like named-entity recognition (NER), semantic role labelling (SRL), and POS tagging are the methods proving the outstanding performance to solve difficult NLP tasks [6]. We reviewed different deep learning-related models and methods applied to natural language tasks such as a convolutional neural network (CNN, or ConvNet), Recurrent Neural Networks (RNN), and recursive neural networks. We also presented memory-augmenting strategies, attention mechanisms, and unsupervised models, reinforcement learning methods applied for language-related tasks [7].

In public procurement in the European Union (EU), tenders are called online via the Tender European Daily (TED) database. Tenders in TED are called under different categories like parking allotment, factories, etc. This dataset consists of syntactically-parsed information about different types of TED tenders [8]. The TED tenders are called from different fields and information is stored in the form of text and numbers. The important features of the data are given in the following table.

The primary goal of this research is to develop an optimizing user interface for tender creation in the context of TED using text mining techniques and applying Machine Learning (ML) algorithms for automation of the entire TED process. After a deep study of the dataset, it is understood that the CPV code describes the uniqueness of each type of tender. Here, we tried to extract information about certain tenders in TED based on year-wise allotment along with their CPV codes. So when the CPV code is given it retrieves the information about the project.

In Section II of this paper, work related to the proposed method is described. Section III reviews relevant natural processing techniques. The proposed framework and corpus description is detailed in Section IV. Section V covers a preliminary experimental evaluation of the framework. We conclude in Section VI.

II. Related Work

The neurons interconnected with each other simulate the network structure of neurons in the human brain [9]. The nodes are interconnected by edges, and then nodes are organized into layers. Computation and processing are done in a feed-forward

TABLE I. FEATURES OF DATA

Features	Data type	Values
Name of Project	Text	singe line free text
Description of Project	Text	multiple line free text
Types of contract	Text	one of WORKS, SUPPLIES, or SERVICES
CPV code	Number	from predefined hierarchical catalog of codes
additional CPV codes	List of numbers	from predefined hierarchical catalog of codes

manner, and errors can be back-propagated to previous layers to adjust the weights of corresponding edges [10], [11]. The hidden nodes are randomly assigned to avoid the extra load pay-off in the structure. In a simple network, the weights are usually learned in one single step which results in faster performance [12]. For more complex relations, deep learning methods with multiple hidden layers are adopted. These methods were made feasible thanks to the recent advances of computing power in hardware, and the Graphics Processing Unit (GPU) processing in software technologies [13]. Depending on the different ways of structuring multiple layers, several types of deep neural networks were proposed, where ConvNet and RNN are among the most popular ones [14]. ConvNet is usually used in computer vision since convolution operations can naturally be applied in edge detection and image sharpening [15]. They are also useful in calculating weighted moving averages and calculating impulse responses from signals [16]. RNNs are a type of neural networks where the inputs of hidden layers in the current point of time depend on the previous outputs of the hidden layer [17].

This allows them to deal with a time sequence with temporal relations such as speech recognition [18]. According to a previous comparative study of RNN and ConvNet in NLP, RNN is found to be more effective in sentiment analysis than ConvNet [19]. Thus, we focus on RNN in this paper. As the time sequence grows in RNNs, it's possible for weights to grow beyond control or to vanish. To deal with the vanishing gradient problem in training conventional RNN, bidirectional Long Short-Term Memory (BLSTM) has been proposed to learn long-term dependency among longer time period [20]. In addition to input and output gates, forget gates are added in BLSTM [21]. They are often used for time-series prediction and hand-writing recognition. The shortfall of conventional BLSTM is that they are only able to make use of the previous context. To avoid this, BLSTM is designed for processing the data in both directions with two separate hidden layers, which are then feed forwards to the same output layer. Using BLSTM will run your inputs in two ways, one from past to future and one from future to past. This will help to improve the results and understand the context in a better way.

For NLP, it is useful to analyze the distributional relations of word occurrences in documents [22]. The simplest way is to use one-hot encoding to represent the occurrence of each word in documents as a binary vector [23]. In distributional semantics, word embedding models are used to map from the one-hot vector space to a continuous vector space in a much lower dimension than conventional bag-of-words (BoW) model [24]. Among various word embedding models, the most popular ones are distributed representation of words such as Word2Vec and GloVe, where neural networks are used to train the occurrence relations between words and documents in the contexts of training data [25]. In this paper, we adopt the Word2Vec word embedding model to represent words in short texts. Then, BLSTM classifiers are trained to capture the long-term dependency among words in short texts. The sentiment of each text can then be classified as positive or negative [19]. In this paper, we utilize BLSTM in learning sentiment classifiers of short texts [26].

III. NATURAL LANGUAGE PROCESSING

The most important application of NLP is to make machines understand, respond to, and communicate with human languages. Thus, it acts as a bridge to improve the performance of communication [27]. With the growth and extension of machine learning and deep learning modules in different fields along with NLP, it increases the performance and efficiency of the system [28]. It is divided into the following categories.

a) Parsing:
Parsing is a technique of breaking up sentences according to grammar rules. A sentence is broken into a Noun Phrase and Verb Phrase. The Noun Phrase could again be divided into Article and Noun. This helps to convert the text into required grammar formats [29].

b) Stemming:
Stemming is the process of reducing inflexion in words to their root forms, e.g. by mapping a group of words to the same stem even if the stem itself is not a valid word in the language [30]. The stemming a word or sentence may result in words that are not the actual word. In stemming the 'stem' is obtained after applying a set of rules but without bothering about the part-of-speech (POS) or the context of the word occurrence [31].

c) Text Segmentation:
Text Segmentation is the division of sentences into smaller parts called segments. These segments can be words, sentences, topics, phrases, or any information unit depending on the task of the text analysis [32].

d) Named Entity Recognition (NER):
Named entity recognition is the process of allocating names to different entities like a person, location, organization, drug, time, clinical procedure, or biological protein in text. NER systems are often used as the first step in question answering, information retrieval, co-reference resolution, topic modelling, etc, [33].

e) Sentiment Analysis:
Sentiment analysis involves the aspects covered in the text with respect to the writer's perspective. The focus is on identifying some topics or the overall sentiment polarity of a text, such as positive or negative [34].

f) Word Embedding:
Word embedding is a big part of NLP-related research works. A word embedding with word2vec determines the syntactic properties of words based on co-occurrence in a text corpus and

 ISBN: 978-1-61208-739-9

assigns to each of the words a vector in the vector embedding space. Word embedding of sentences can determine the word characteristics and context [35].

1) **Sentiment Analysis**:- The word embedding has been proven to improve the performance of sentiment classification systems [36].
2) **Sentiment-Specific**:- The generic word embedding only considers semantic relations from the words; it fails to differentiate sentiment words such as “good” or “bad”.

g) Word2vec:
Word2vec is an embedding technique which is the evaluation of a set of word vectors [37]. It is also known as distributed word representation that can capture both the semantic and syntactic information of words from a large unlabelled corpus. Words that occur in similar contexts tend to have similar meanings and are labelled accordingly with the word2vec method [38].

1) **Continuous Bag-of-Words** This is the first proposed architecture which is similar to the feed-forward neural network language models (NNLM), where the non-linear hidden layer is removed and the projection layer is shared for all words (not just the projection matrix); thus, all words get projected into the same position (their vectors are averaged). The advantage of this method is that as more gets added in future, it does not effect the order of words in history. The training complexity is then:

$$Q = ND + Dlog2(V) \tag{1}$$

We denote this model further as CBOW, as unlike the standard bag-of-words model, it uses a continuous distributed representation of the context [39].
2) **Skip-gram Model** The Skip-gram model is to find word representations that are useful for predicting the surrounding words in a sentence or a document [40]. More formally, given a sequence of training words $w_1, w_2, w_3, \ldots, w_T$, the objective of the Skip-gram model is to maximize the average log probability

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{-c\leq j\leq c, j\neq 0}\log p(w_t + j|w_t) \tag{2}$$

where c is the size of the training context, which can be a function of the centre word wt. Larger c results in more training examples and, thus, can lead to higher accuracy at the expense of the training time [41].

h) Bag-of-Words:
The Bag-of-Words (BoW) model learns a vocabulary from all of the documents and then models each document by counting the number of times each word appears. The BoW model is a simplifying representation used in NLP and information retrieval (IR). In this model, a text (such as a sentence or a document) is represented as the bag (multiset) of its words, disregarding grammar and even word order but keeping multiplicity [42].

i) Stop Words:
Text may contain stop words like ‘the’, ‘is’, ‘are’. Stop words can be filtered from the text to be processed. There is no universal list of stop words in NLP research, however, the NLTK module contains a list of stop words [18]. Consequently, such types of words can be processed separately.

IV. Proposed Framework

The proposed framework consists of preprocessing, feature extraction, word embedding, and BLSTM classification. The corpus description for the overall architecture is related to the information about the announcement and calls related to tenders in TED. The data is stored in an open XML format. We extracted it to CSV format, containing the information in form of text and numbers as indicated in Table I. Recall that the CSV file contains information like the title of the project and CPV code as an important entity. The other entities like types of contract and detailed description about a particular project are also part of the data set. In this way, we are having a dataset of approx, $40{,}000$ texts for each year from 2015 to 2018. The architecture flow is shown in Figure 1.

As shown in the Figure 1, short texts are first pre-processed and word features are extracted. Second, the Word2Vec word embedding model is used to learn word representations as vectors.

The output of the LSTM cell can also be implemented via the output gate $q_i^{(t)}$, which also uses a sigmoid unit for gating:

$$q_i(t) = \sigma\left(b_i^o + \sum_j U_{i,j}^o x_j^{(t)} + \sum_j W_{i,j}^o h_j^{(t-1)}\right) \tag{3}$$

which has parameters b^o, U^o, W^o for its biases, input weights and recurrent weights, respectively. Among the variants, one can choose to use the cell state $q_i^{(t)}$ as an extra input (with its weight) into the three gates of the $i-th$ unit, as shown in above equation. Thus the BLSTM model takes the recurrent input of weights and concurrently gives the output to the cell state.

A. Preprocessing

None of the magic described above happens without a lot of work on the back end. Transforming text into something an algorithm can digest is a complicated process. The preprocessing is divided into four different steps:

1) **Cleaning** consists of getting rid of the less useful parts of a text through stopword removal, dealing with capitalization and characters, and other minor details.
2) **Annotation** consists of the application of a scheme to texts. Annotations include structural markup and part-of-speech tagging.
3) **Normalization** consists of the translation (mapping) of terms in the scheme or linguistic reductions through Stemming, Lemmatization, and other forms of standardization.
4) **Analysis** consists of statistically probing, manipulating, and generalizing from the dataset for feature analysis.

B. Tokenization

Tokenization is the process of breaking up the sequence of characters in a text by locating the word boundaries, the points where one word ends and another begins. For computational linguistic purposes, the words thus identified are frequently referred to as tokens. In written languages where no word boundaries are explicitly marked in the writing system, tokenization is also known as word segmentation, and this term is frequently used synonymously with tokenization.

 ISBN: 978-1-61208-739-9

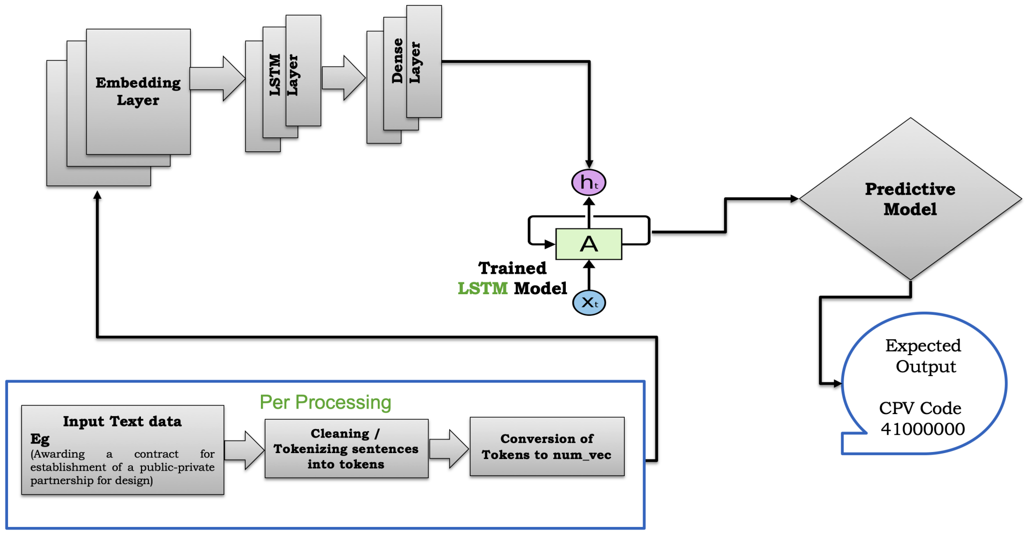


Figure 1. The system architecture of the proposed approach

C. Word Embedding

Simply put, word embeddings let us represent words in the form of vectors. But these are not random vectors, the aim is to represent words via vectors such that similar words or words used in a similar context are close to each other while antonyms end up far apart in the vector space.

D. Dense Layer

The dense layer contains a linear operation in which every input is connected to every output by weight (so there are n_inputs * n_outputs weights - which can be a lot!). Generally, this linear operation is followed by a non-linear activation function.

E. Long Short-Term Memory (LSTM)

The architecture of LSTM is a combination of recurrent neural networks and feed-forward neural network [43]. LSTM has numerous applications related to text document processing based on context solving for the given task. So it predicts the next coming term from the relative context thus, words are not treated as independent individuals, but as the units dependent on their immediate neighbourhood in the text. LSTM advantage it is that the output does not depend on the length of input because the input is entered sequentially, one input per time step [44]. The basic architecture of the LSTM recurrent neural network receipts the memory block, that consists of several memory cells with which one can communicate by the input gate, forget gate, and output gate of that cell. The training of the LSTM neural network is performed in two phases, forward pass and backward pass. The important feature to note is that LSTM memory cells give different roles to addition and multiplication in the transformation of inputs. The central plus sign in both diagrams below is essentially the secret of LSTM. Figure 2 shows the architecture of a simple recurrent network with one input and one output gate. The working architecture of LSTM is clearly described on the right-hand side of Figure 2. Instead of determining the subsequent cell state by multiplying its current state with new input, they add the two, and that quite literally makes the difference. The forget gate still relies

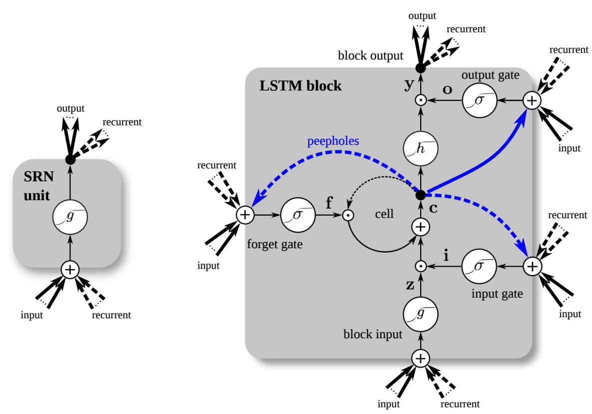


Figure 2. The Simple Recurrent Network unit (left) and a LSTM block (right) as used in the hidden layers of a recurrent neural network [45].

on multiplication, of course. Different sets of weights filter the input for input, output, and forgetting. The forget gate is represented as a linear identity function. The reasons for this is that if the gate is open, the current state of the memory cell is simply multiplied by one, to propagate forward one more time step [41].

V. Baseline Results from Experimental Evaluation

We have calculated the accuracy of our NLP model using Precision and Recall. These ways of evaluation use the concepts of true positives, true negatives, false positives, and false negatives. Precision is defined as the number of true positives divided by the number of true positives plus the number of false positives. False positives are cases the model incorrectly labels as positive that are actually negative, or, in our example, the text which is not identified is considered a false negative. Our trained model achieves a precision of 95.5% where it accurately identifies the sequence prediction for the given initial text words and identifies the respective CPV code. This gives a very good accuracy as it resolves ambiguity in predicting the sequence.

 ISBN: 978-1-61208-739-9

The research study benchmarks shows the Defence Research and Development Organisation (DRDO) lab (India) which prediction and classification technique applied to the vendor's system where it predicts and occurrences of the database [46].

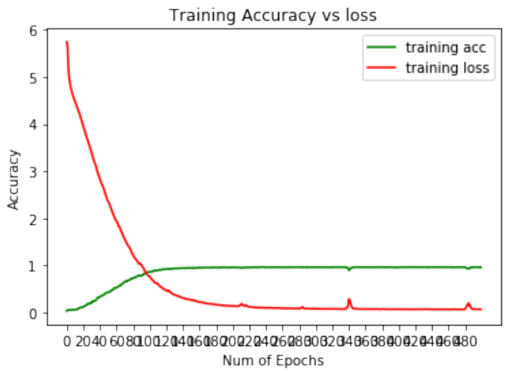


Figure 3. ROC curve against the training accuracy and loss

The recall is the number of true positives divided by the number of true positives plus the number of false negatives. True positives are data points classified as positive by the model that actually is positive (meaning they are correct), and false negatives are data points the model identifies as negative that actually are positive (incorrect). A ROC curve plots the true positive rate on the y-axis versus the false positive rate on the x-axis. The true positive rate (TPR) is the recall and the false positive rate (FPR) is the probability of a false alarm. Figure 3 shows the ROC curve of training accuracy and training loss of the system. It is observed that after the increase in a number of epochs the training accuracy increases and loss becomes constant. This makes our system more stable with improvements in performance accuracy.

VI. Conclusion

This project work focuses on the automation and optimization of tender creation in the European TED system. TED data sequence generation is a complex task. In most TED titles there is 25 %-30 % sequence of text is similar that makes more ambiguity to predict the correct sequence. In our case, the models with 85 % couldn't resolve complete ambiguity we need model accuracy more than 90 %. We have also presented the literature review with respect to NLP and text mining concepts. The implement of this algorithm is the initial but an important step towards an optimized interface for tender creation. The final goal is a natural user interface, improving user experience by automation and increasing human efficiency significantly. In future work, we can develop a GUI-based model with the proposed algorithm for making the optimized process for calling TED.

References

[1] J. Hirschberg and C. D. Manning, "Advances in natural language processing," Science, vol. 349, no. 6245, 2015, pp. 261–266.

[2] K. S. Jones, "Natural language processing: a historical review," in Current issues in computational linguistics: in honour of Don Walker. Springer, 1994, pp. 3–16.

[3] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent trends in deep learning based natural language processing [review article]," IEEE Comp. Int. Mag., vol. 13, no. 3, 2018, pp. 55–75.

[4] Y. LeCun, Y. Bengio, and G. E. Hinton, "Deep learning," Nature, vol. 521, no. 7553, 2015, pp. 436–444.

[5] S. Marcel, M. S. Nixon, J. Fierrez, and N. W. D. Evans, Eds., Handbook of Biometric Anti-Spoofing - Presentation Attack Detection, Second Edition, ser. Advances in Computer Vision and Pattern Recognition. Springer, 2019.

[6] R. Collobert and J. Weston, "A unified architecture for natural language processing: Deep neural networks with multitask learning," in Proceedings of the 25^{th} International conference on Machine learning. ACM, 2008, pp. 160–167.

[7] A. Conneau, H. Schwenk, L. Barrault, and Y. Lecun, "Very deep convolutional networks for natural language processing," arXiv preprint arXiv:1606.01781, vol. 2, 2016.

[8] C. J. Gelderman, P. W. T. Ghijsen, and M. J. Brugman, "Public procurement and EU tendering directives–explaining non-compliance," International Journal of Public Sector Management, vol. 19, no. 7, 2006, pp. 702–714.

[9] Q. V. Le, M. Ranzato, R. Monga, M. Devin, K. Chen, G. S. Corrado, J. Dean, and A. Y. Ng, "Building high-level features using large scale unsupervised learning," arXiv preprint arXiv:1112.6209, 2011.

[10] D. Ulyanov, V. Lebedev, A. Vedaldi, and V. S. Lempitsky, "Texture networks: Feed-forward synthesis of textures and stylized images." in ICML, vol. 1, no. 2, 2016, pp. 1–4.

[11] J. C. Hoskins and D. Himmelblau, "Artificial neural network models of knowledge representation in chemical engineering," Computers & Chemical Engineering, vol. 12, no. 9-10, 1988, pp. 881–890.

[12] G. G. Towell and J. W. Shavlik, "Extracting refined rules from knowledge-based neural networks," Machine learning, vol. 13, no. 1, 1993, pp. 71–101.

[13] D. Cireşan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," arXiv preprint arXiv:1202.2745, 2012.

[14] J. Wang, Y. Yang, J. Mao, Z. Huang, C. Huang, and W. Xu, "CNN-RNN: a unified framework for multi-label image classification," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2285–2294.

[15] G. Chartrand, P. M. Cheng, E. Vorontsov, M. Drozdzal, S. Turcotte, C. J. Pal, S. Kadoury, and A. Tang, "Deep learning: a primer for radiologists," Radiographics, vol. 37, no. 7, 2017, pp. 2113–2131.

[16] H. Kawahara, I. Masuda-Katsuse, and A. De Cheveigne, "Restructuring speech representations using a pitch-adaptive time–frequency smoothing and an instantaneous-frequency-based f0 extraction: Possible role of a repetitive structure in sounds," Speech communication, vol. 27, no. 3-4, 1999, pp. 187–207.

[17] M. Sundermeyer, R. Schluter, and H. Ney, "LSTM neural networks for language modeling," in Thirteenth annual conference of the international speech communication association, 2012, pp. 194–197.

[18] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, and K. J. Lang, "Phoneme recognition using time-delay neural networks," Backpropagation: Theory, Architectures and Applications, 1995, pp. 35–61.

[19] J. Wang, L.-C. Yu, K. R. Lai, and X. Zhang, "Dimensional sentiment analysis using a regional CNN-LSTM model," in Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), vol. 2, 2016, pp. 225–230.

[20] A. Graves, S. Fernandez, and J. Schmidhuber, "Bidirectional LSTM networks for improved phoneme classification and recognition," in International Conference on Artificial Neural Networks. Springer, 2005, pp. 799–804.

[21] J.-H. Wang, T.-W. Liu, X. Luo, and L. Wang, "An LSTM approach to short text sentiment classification with word embeddings," in Proceedings of the 30th Conference on Computational Linguistics and Speech Processing (ROCLING 2018), 2018, pp. 214–223.

[22] T. Joachims, Learning to classify text using support vector machines. Springer Science & Business Media, 2002, vol. 668.

[23] J. Turian, L. Ratinov, and Y. Bengio, "Word representations: a simple and general method for semi-supervised learning," in Proceedings of the 48th annual meeting of the association for computational linguistics. Association for Computational Linguistics, 2010, pp. 384–394.

[24] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using

rnn encoder-decoder for statistical machine translation," arXiv preprint arXiv:1406.1078, 2014.

[25] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 2014, pp. 1532–1543.

[26] M. Dwarampudi and N. Reddy, "Effects of padding on LSTMs and CNNs," arXiv preprint arXiv:1903.07288, 2019.

[27] A. Vinciarelli, M. Pantic, D. Heylen, C. Pelachaud, I. Poggi, F. D'Errico, and M. Schroeder, "Bridging the gap between social animal and unsocial machine: A survey of social signal processing," IEEE Transactions on Affective Computing, vol. 3, no. 1, 2012, pp. 69–87.

[28] I. Arel, D. C. Rose, T. P. Karnowski et al., "Deep machine learning-a new frontier in artificial intelligence research," IEEE computational intelligence magazine, vol. 5, no. 4, 2010, pp. 13–18.

[29] J. Kimball, "Seven principles of surface structure parsing in natural language," Cognition, vol. 2, no. 1, 1973, pp. 15–47.

[30] M. F. Porter, "Snowball: A language for stemming algorithms," pp. 1–4, 2001.

[31] I. A. Al-Sughaiyer and I. A. Al-Kharashi, "Arabic morphological analysis techniques: A comprehensive survey," Journal of the American Society for Information Science and Technology, vol. 55, no. 3, 2004, pp. 189–213.

[32] I. Pak and P. L. Teh, "Text segmentation techniques: a critical review," in Innovative Computing, Optimization and Its Applications. Springer, 2018, pp. 167–181.

[33] L. Ratinov and D. Roth, "Design challenges and misconceptions in named entity recognition," in Proceedings of the thirteenth conference on computational natural language learning. Association for Computational Linguistics, 2009, pp. 147–155.

[34] B. Pang, L. Lee et al., "Opinion mining and sentiment analysis," Foundations and Trends® in Information Retrieval, vol. 2, no. 1–2, 2008, pp. 1–135.

[35] O. Levy and Y. Goldberg, "Neural word embedding as implicit matrix factorization," in Advances in neural information processing systems, 2014, pp. 2177–2185.

[36] D. Tang, F. Wei, N. Yang, M. Zhou, T. Liu, and B. Qin, "Learning sentiment-specific word embedding for twitter sentiment classification," in Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), vol. 1, 2014, pp. 1555–1565.

[37] W.-L. Chao, S. Changpinyo, B. Gong, and F. Sha, "An empirical study and analysis of generalized zero-shot learning for object recognition in the wild," in European Conference on Computer Vision. Springer, 2016, pp. 52–68.

[38] Z.-Y. Niu, D.-H. Ji, and C. L. Tan, "Semi-supervised feature clustering with application to word sense disambiguation."

[39] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," arXiv preprint arXiv:1301.3781, 2013.

[40] P. Liu, X. Qiu, and X. Huang, "Learning context-sensitive word embeddings with neural tensor skip-gram model," in Twenty-Fourth International Joint Conference on Artificial Intelligence, 2015, pp. 1284–1290.

[41] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in Advances in neural information processing systems, 2013, pp. 3111–3119.

[42] J. Gao, Y. He, X. Zhang, and Y. Xia, "Duplicate short text detection based on word2vec," in 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS). IEEE, 2017, pp. 33–37.

[43] F. A. Gers, N. N. Schraudolph, and J. Schmidhuber, "Learning precise timing with LSTM recurrent networks," Journal of machine learning research, vol. 3, 2002, pp. 115–143.

[44] L. Skovajsová, "Long short-term memory description and its application in text processing," in 2017 Communication and Information Technologies (KIT). IEEE, 2017, pp. 1–4.

[45] K. Greff, R. K. Srivastava, J. Koutnk, B. R. Steunebrink, and J. Schmidhuber, "LSTM: a search space odyssey," IEEE transactions on neural networks and learning systems, vol. 28, no. 10, 2017, pp. 2222–2232.

[46] S. Goswami, P. Bhardwaj, and S. Kapoor, "Naïve bayes classification of drdo tender documents," in 2014 International Conference on Computing for Sustainable Global Development (INDIACom). IEEE, 2014, pp. 593–597.

Multivariate Event Detection for Non-Intrusive Load Monitoring

Alexander Gerka and Benjamin Cauchi

OFFIS e.V.
Institute for Information Technology
Oldenburg, Germany
Email: gerka@offis.de
benjamin.cauchi@offis.de

Andreas Hein

Carl von Ossietzky University,
Oldenburg, Germany
Email:andreas.hein@uol.de

***Abstract*—Due to the growing interest for in-home activity monitoring, the tracking of appliances use, usually referred to as Non-Intrusive Load Monitoring (NILM), has to address new challenges. Indeed, as NILM has long been motivated by potential energy savings, most event detectors for NILM have focused on the detection of on- or off-switches of high power devices. On the contrary, in-home monitoring typically relies on the detection of events related to low-power devices from potentially noisy signals. Additionally, approaches that apply expert heuristics to a single-variate input, often favored for their low complexity and real-time applicability, can be overly sensitive to the choice of an arbitrary defined detection threshold. This paper aims at decreasing the sensitivity of a detector based on expert heuristic by applying it to the Hotelling-T^2 statistic of a multivariate input, computed online from the current and voltage inputs. Focusing on realistic scenarios, the approach is evaluated on a dataset recorded in a real apartment using a commercially available smart-meter. The results, expressed in terms of precision, recall and F-score, show that the proposed approach can both yield higher performance and be less sensitive to the choice of the detection threshold.**

Keywords*—*NILM; Event Detection; Hotelling-T^2 statistic

I. Introduction

Non-intrusive load monitoring (NILM), introduced by George Hart in the 1980s [1], denotes the tracking of appliances use through analysis of their power consumption. The growing presence of smart meters in homes, encouraged, e.g., by a directive of the European Union in 2009 [2], has led to renewed research efforts in NILM. These efforts have been mostly motivated by the potential for energy savings and, therefore, focussed on the monitoring of high power devices. However, due to the ageing population [3], applications such as activity monitoring for disease prevention [4][5] or the surveillance of life-critical devices present in homes, e.g., for respiration support [6], have become of crucial importance. The main advantage of NILM-based monitoring system is their unobtrusiveness, as they do not require an additional installation. Additionally, a NILM-system could prevent other, more obtrusive, installations such as power plugs. Consequently, NILM approaches able to reliably monitor the use of low power devices in realistic settings are urgently needed.

Many approaches extract multiple features from the recorded power consumption signal in order to determine the active appliances during a given time sequence. It has been proposed in [1] to use both reactive and real power. However, due to the high correlation between real and reactive power consumption, these features are not sufficient for the classification of low power devices [7]. As a result, numerous features for NILM have been introduced in recent years such as the current's harmonic [8], the shape of the so-called VI-trajectory [9] or the poles and residues of the power signal's impulse response [10]. The extraction of these complex features require to record the current and voltage of the power supply at a sampling frequency much higher than 50 Hz. Additionally, the applied classifier typically relies on an event based method rather than a state-based method and necessitate to segment the recorded signal into sequences of interest. This segmentation requires an event detector, i.e., the detection of on- and off-switches of appliances. This paper focuses on event detection for NILM in realistic scenarios.

Approaches aiming at event detection for NILM can be broadly split between three categories, namely, based on expert heuristics, matched filtering, or probabilistic methods [11]. Though promising, probabilistic methods based on, e.g., generalized likelihood ratio (GLR) [12], goodness-of-fit (GOF) [13] or cumulative sum control (CuSuM) [14] can be computationaly costly and often rely on long sequences of input signal, making them impractical in many realistic settings. Approaches based on matched filtering, e.g., using Cepstrum smoothing [15] or Hilbert transformation [16] can be sensitive to mismatch between the dataset used to tune the approach and the environment in which it is tested. Though promising, the approach presented in [15] showed largely lower performance when tested on data containing low-power devices [17].

Approaches based on expert heuristics rely on the choice of a somewhat arbitrarily defined threshold or rule-based approach [1][18], on which their performance can be greatly sensitive. However, relying on features whose extraction is typically computationally inexpensive, e.g., standard deviation of the current signal envelope [19]. Approaches based on expert heuristics are easily implemented. Therefore, reducing their sensitivity on an arbitrary defined threshold could be of great interest. Contrary to most expert heuristic approaches that use a single-parameter as input, the approach proposed in

 ISBN: 978-1-61208-739-9

this paper aims at decreasing the sensitivity of a detector based on a expert heuristic by applying the Hotelling-T^2 statistic to a multivariate input.

In order to evaluate the benefit of the proposed approach, a dataset recorded in realistic conditions has to be used. Indeed, most datasets from the literature only considered low frequency systems [1][12][20], specific devices [18] or sets of (mostly) high-power devices [13][14][16][21]. Additionally, those studies were often performed under laboratory conditions [1][14], i.e., in absence of capacitive effects of the long supply line and other environmental noise factors present in a real apartment. Such signal disturbances can have a large impact. Consequently, the evaluation conducted in this paper is done using a dataset recorded in a real apartment using a commercially available smart-meter.

The remainder of this paper is structured as follows. First, the proposed approach and the expert heuristic approach that is used as benchmark is described in Section II. The recorded dataset and experimental framework are described in Section III and the results in Section IV. Section V concludes the paper.

II. Proposed Approach

In this section the proposed approach is described.

A. Threshold based NILM event detection

The signal recorded from a monitoring device, e.g., a smart meter, typically consists of $M = 2$ channels representing the current in Ampere and the voltage in Volts. In the remainder of this paper, we use $x_m(n)$ to denote the signal recorded at a sampling frequency fs, at sample index n, in the m-th channel. We arbitrary set $m = 0$ as the index of the current channel and $m = 1$ as the index of the voltage channel. Event detection methods for NILM based on experts heuristics, such as the one proposed in [19], typically rely on segmenting the input signal into overlapping frames of length L with an hop size of H samples and assigning a label $d(\ell) \in [0, 1]$ equal to 1 if an event is detected in the ℓ-th frame and equal to 0 otherwise.

The methods considered in this paper rely on computing a change quantifying value $v(\ell) \in \mathbb{R}_{\geq 0}$, whose computation is the focus of the next subsection, for each frame, and applying

$$d(\ell) = \begin{cases} 1 \text{ if } v(\ell) \geq \tau(\ell), \\ 0 \text{ otherwise}, \end{cases} \tag{1}$$

where $\tau(\ell)$ denotes a decision threshold. It can be noted that contrary to methods in which the decision threshold is computed from a complete signal utterance, e.g., [13][21], this paper considers event detection for real-time application and uses a frame dependant threshold computed as

$$\tau(\ell) = \alpha \cdot \sigma_v(\ell - \delta), \tag{2}$$

where δ denotes a decision delay and where

$$\sigma_v(\ell) = \sqrt{\frac{1}{\Delta - 1} \sum_{i=0}^{\Delta-1} \left| v(\ell - i) - \frac{1}{\Delta} \sum_{j=0}^{\Delta-1} v(\ell - j) \right|^2} \tag{3}$$

denotes the standard deviation of $v(\ell)$ computed over Δ buffered values. In practical applications, the values of δ and Δ are often hardware dependant. However, the choice of value assigned to the constant α in (2), though of critical importance for the detection performance, is typically arbitrarily defined. The sensitivity of the detection performance to the choice of α can be limited if the value $v(\ell)$ suitably quantifies the potential changes in the signal at a given time frame, i.e. $v(\ell) \approx 0$ when no change is present.

B. Change quantification

As most similar detectors, the approach proposed in [19] that we chose as benchmark (cf. Section IV) due to its low computational complexity and promising performance only uses the recorded current to quantifies the change in the input signal, i.e., computes $v(\ell)$ from $x_0(n)$ only. This computation relies on extracting for each frame, the envelope $e_\ell(n)$ of length $L_e = H \cdot (E - 1) + L$, where E denotes the number of frames used in the envelope extraction. The envelope $e_\ell(n)$ is computed as the interpolation, e.g., linear or cubic, between the maximums of $|x_0(n)|$ in each of the E considered frames and, assuming that $\mathrm{E}\{e_\ell(n)\}$ is constant in absence of event, the change quantifying value $v(\ell)$ can be computed as the Change of Mean Amplitude Envelope (CMAE)

$$v(\ell) = \frac{1}{L_e} \left| \sum_{i=0}^{L_e-1} e_{\ell-1}(i) - e_\ell(i) \right|. \tag{4}$$

Unfortunately computing $v(\ell)$ as in (4) can, as stated in [19], result in an unreliable detector in presence of noisy signals. We aim at improving the reliability of the threshold based detector from (2) by improving the computation of $v(\ell)$.

Preliminary works showed that using multiple features can be extracted from a NILM-signal. In general those features can be separated into different domains:

1) power related features in time domain
2) power related features in frequency domain
3) features related to the phase difference between current an voltage

Therefore, we propose to use a feature from each feature domain as the information about the NILM-signal contained by a feature is different for every domain. More specifically, we have shown in [22] that the variance of the current, taking advantage of both input channels, the phase of the input signal and, the current's frequency ratio were beneficial. In our proposed approach, we extract for each frame a $L_{\mathbf{v}} = 3$ elements feature vector

$$\mathbf{v}(\ell) = \left[\ \sigma_x^2(\ell)\ ,\ \phi(\ell)\ ,\ \omega(\ell)\ \right]^{\mathrm{T}}, \tag{5}$$

where $\cdot^{\mathrm{T}}$ denotes the transpose operator and where $\sigma_x^2(\ell)$, $\phi(\ell)$ and $\omega(\ell)$ denote the current variance, the phase and the current

 ISBN: 978-1-61208-739-9

frequency ratio computed as

$$\sigma_x^2(\ell) = \frac{1}{L-1}\sum_{i=0}^{L-1}\left|x_0(\ell H+i) - \frac{1}{L}\sum_{j=0}^{L-1}x_0(\ell H+j)\right|^2 \quad (6)$$

$$\phi(\ell) = \cos^{-1}\frac{\sum_{i=0}^{L-1}x_0(\ell H+i)\cdot x_1(\ell H+i)}{\sqrt{\sum_{i=0}^{L-1}x_0(\ell H+i)^2}\cdot\sqrt{\sum_{i=0}^{L-1}x_1(\ell H+i)^2}} \quad (7)$$

$$\omega(\ell) = \frac{\sum_{i=0}^{L-1}\left|\tilde{x}_f(\ell H+i) - \frac{1}{L}\sum_{j=0}^{L-1}\tilde{x}_f(\ell H+j)\right|^2}{\sum_{i=0}^{L-1}\left|\overline{x}_f(\ell H+i) - \frac{1}{L}\sum_{j=0}^{L-1}\overline{x}_f(\ell H+j)\right|^2}, \quad (8)$$

where $\tilde{x}_f(n)$ denotes the output of a bandpass filter applied to $x_0(n)$ and centred around the carrier frequency f of the input signal (cf. Section III-B) and $\overline{x}_f(n) = x_0(n) - \tilde{x}_f(n)$.

We consider each vector $\mathbf{v}(\ell)$ to be a single independent realisation of a random process with an $L_\mathbf{v}$-dimensional F-distribution and define the sequences of vectors $\mathbf{V}_0(\ell)$ and $\mathbf{V}_1(\ell)$ of length $L_\mathbf{0}$ and $L_\mathbf{1}$, respectively, as

$$\mathbf{V}_0(\ell) = \{\mathbf{v}(\ell - L_\mathbf{0})\ ,\cdots,\mathbf{v}(\ell-2)\ ,\ \mathbf{v}(\ell-1)\}, \quad (9)$$

$$\mathbf{V}_1(\ell) = \{\mathbf{v}(\ell)\ ,\mathbf{v}(\ell+1)\ ,\cdots,\ \mathbf{v}(\ell+L_\mathbf{1}-1)\}. \quad (10)$$

An event is considered to occur at frame ℓ if the $\mathbf{V}_0()$ and $\mathbf{V}_1()$ are composed of realisations of significantly distinct distributions. This significance of can be expressed by the Hotelling-T^2 statistic [23]. The computation of the Hotelling-T^2 statistic depends on the homogeneity of the partial covariance matrices ($\mathbf{\Gamma}_0(\ell)$ and $\mathbf{\Gamma}_1(\ell)$) computed separately from the vectors $\mathbf{V}_0(\ell)$ and $\mathbf{V}_1(\ell)$. If, using the box test [24], these matrices are determined to be homogeneous, the Hotelling-T^2 statistic is computed as

$$\begin{aligned}T^2(\ell) =& \frac{L_\mathbf{0}\cdot L_\mathbf{1}}{L_\mathbf{0}+L_\mathbf{1}}\cdot\big(\boldsymbol{\mu}_0(\ell)-\boldsymbol{\mu}_1(\ell)\big)^\mathrm{T}\cdot\\ &\mathbf{\Gamma}(\ell)^{-1}\cdot\big(\boldsymbol{\mu}_0(\ell)-\boldsymbol{\mu}_1(\ell)\big),\end{aligned} \quad (11)$$

otherwise,

$$\begin{aligned}T^2(\ell) =& \big(\boldsymbol{\mu}_0(\ell)-\boldsymbol{\mu}_1(\ell)\big)^\mathrm{T}\cdot\left(\frac{\mathbf{\Gamma}_0(\ell)^{-1}}{L_\mathbf{0}} - \frac{\mathbf{\Gamma}_1(\ell)^{-1}}{L_\mathbf{1}}\right)\cdot\\ &\big(\boldsymbol{\mu}_0(\ell)-\boldsymbol{\mu}_1(\ell)\big),\end{aligned} \quad (12)$$

where $\boldsymbol{\mu}_0(\ell)$ and $\boldsymbol{\mu}_1(\ell)$ denote the average of the vectors in $\mathbf{V}_0(\ell)$ and $\mathbf{V}_1(\ell)$, respectively, and $\mathbf{\Gamma}(\ell)$ denotes the $L_\mathbf{v}\times L_\mathbf{v}$ covariance matrix computed using the vectors in both sequences. Finally, the label $d(\ell)$ can be assigned by substituting $v(\ell)$ by $T^2(\ell)$ in (1) and (3).

III. Experiment

In this section the experiment is presented.

A. Collected Dataset

We evaluated the approach proposed in Section II on a dataset constructed to evaluate the performance of NILM approaches applied to activities monitoring for health applications. The data was recorded in a three-room apartment occupied by two elderly people who agreed to take part on the study. The signals were recorded using a commercially available smart-meter placed on the main power permitting to record a continuous 2-channel stream, i.e., current and voltage, sampled at the sampling frequency fs =4800 Hz. To pass the data to a measurement computer an optical interface on the backside of the smart meter was used. The recorded current and voltage stream was stored with a resolution of 16 bits. Due to the recording conditions, the recorded signals contain the disturbances to be expected in such a realistic setting, e.g., noise generated by the supply line itself. Therefore, the dataset is particularly valuable to assess the performance to be realistically expected from NILM approaches.

A total of 36 individual appliances were present in the apartment and each one was switched on/off at least once during the recording session. Without storing the phase to which the appliances were connected, a total of 142 events to be detected. The distribution of these appliances in terms of both types and location is summarised in Figure 1. It can be seen that, as expected in a real apartment, most appliances were low power and a large number of lamps were present. It can be noted that, e.g., the monitoring of lamp usage can be a good indicator of activity and location of the user in an apartment. Additionally, aiming at health applications, a respiratory support machine (Resp. support) was among the considered appliances and is an example of appliance whose reliable monitoring could potentially be life critical.

All signals were recorded in a session of over an hour during which the timestamp of each on- off-switch event has been manually annotated. Events were setup to occur at large interval from one-another in order to avoid simultaneous events that could hinder the clustering used prior to the computation of the performance of the considered approaches, as described in the next subsection.

B. Evaluation

In order to evaluate the performance of the proposed approach, i.e., the use of Hotelling-T^2 statistic as input of the threshold based detector, and of the considered benchmark, i.e., using CMAE, in realistic framework and to avoid the detrimental effect of repeated approach initialisation, the entire dataset was processed as a single stream. The parameters were extracted using a 40 ms window, $L = 192$, and a 50 % hopsize, $H = 96$. The envelope used in the case of our benchmark CMAE was extracted using $E = 4$ blocks, similarly as in [19]. However, contrary to [19], a linear interpolation was used instead of a cubic one. This choice reduced the number of false positive obtained on the considered dataset. The frequency ratio $\omega(\ell)$ was computed by using a bandpass filter, designed as a second order butterworth filter, centred around the carrier frequency f =50 Hz with lower and higher cutoff set at 35 Hz and 65 Hz, respectively. We fixed $\Delta = 240$ (50 ms) and $\delta = 144$ (30 ms), cf. (2)-(3), and focused on the influence of the arbitrary chosen threshold, considering the

 ISBN: 978-1-61208-739-9

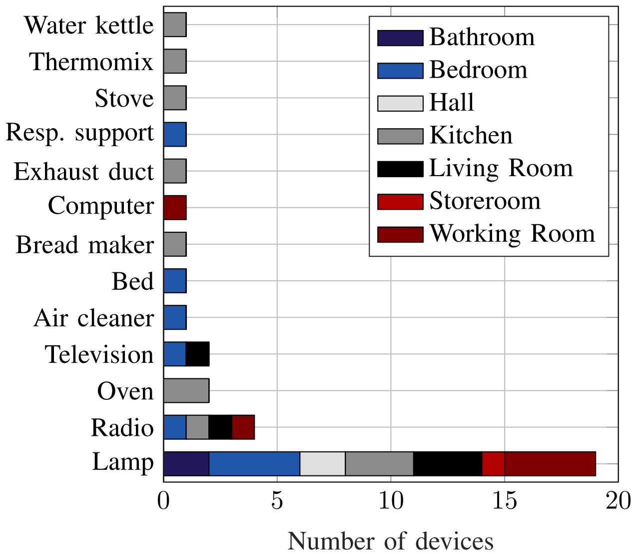


Figure 1. Distribution of the appliances used to record the evaluation dataset.

values $\alpha \in \{1 \cdots 30\}$. It can be noted that the exact values of Δ and δ seemed to have a limited impact on the performance of the considered approaches. The box test and computation of $T^2(\ell)$, cf. (11) and (12), were based on the implementation provided in [25] and according to its recommendation we used $L_\mathbf{0} = L_\mathbf{1} = 100$.

The performance of the considered approaches was determined using the Precision, Recall and F-score defined as [26]

$$\text{Precision} = \frac{\text{True positives}}{\text{True positives} + \text{False positives}}, \tag{13}$$

$$\text{Recall} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}}, \tag{14}$$

$$\text{F-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{15}$$

where the number of true positives and false negatives were determined as follows. First, events detected within the same 8 s window were considered as a single-event. The same tolerance was used to determine if a detected event should be assigned to an annotated event, i.e., true positive, or to none, i.e., false positive. Annotated events with no assigned detection were considered false negatives.

IV. Results

An example of the extracted features from the smart meter input signal and the resulting Hotelling value that represents the combination of these features is shown in Figure 2. This figure shows that relatively small changes in the variance (i.e., at the 15th second) are accompanied by relatively big changes in the phase. As a result, the Hotelling value shows a clear peak around the 15th second which should be easier to detect by the event detection algorithm than the pure variance. Therefore, this observation suits the intuition that the usage of multiple non-related features may improve the detection of events in a NILM-signal.

The observed precision and recall as function of α are depicted in Figure 3. It appears that in the case of both CMAE and Hotelling-T^2, the precision increases with the value α, until it reaches a plateau, which in both cases corresponds to a precision of about 0.6. This behaviour is to be expected as increasing the value of α reduces the likelihood of false positives.

On the other hand, high values of α would increase the likelihood of false negative. This behavior can be noticed by observing the recall (Figure 3), for which the proposed approach exhibits a large advantage compared to the use of CMAE. Using CMAE, the recall decrease sharply for values of α ranging from 3 to 7 and ultimately reaches a plateau with a recall of 0.2. On the contrary, using Hotelling-T^2 statistic, recall decreases slowly with increasing values of α with a plateau corresponding to a recall of 0.8. This shows that the proposed approach is less likely to introduce false negative, even for overly large values of α.

The advantage of the proposed method over the use of CMAE is best noticed by observing the F-score depicted in Figure 4. As a logical consequence of the behaviour observed in terms of precision and recall, using CMAE requires an accurate setting of α in order to yield the optimal F-score. Indeed, F-score obtained using CMAE decreases sharply for α values different than 3 or 4. On the other hand, not only does the use of Hotelling-T^2 statistic yields a higher maximum F-score with a value of 0.7, but its performance is much less sensitive to the setting of α value.

Further improvement of the presented approach could be achieved by using the multivariate approach with other event detectors or by improving the multivariate statistic itself, i.e. using multivariate likelihood detectors [27].

V. Conclusion

This paper proposes to use Hotelling-T^2 statistic, computed from a multivariate input, in order to reduce the sensitivity of an event detector for NILM based on expert heuristics to the value of an arbitrary defined detection threshold. The multivariate input is computed online from a recorded signal of current and voltage. The proposed approach is compared to the use of CMAE, which, as many similar approaches, does not take advantage of the available voltage input or frequency-related features.

A dataset recorded in real environment, i.e., containing appliances relevant to activities monitoring and noisy signal, was used to evaluate the proposed approach. The results, expressed in terms of precision, recall and F-score, show that not only does the proposed approach yield better performance than the considered benchmark, this performance is much less sensitive to the setting of the detection threshold.

Even if using multiple features for event detection is computationally more expensive, we think this approach improves the event detection in NILM systems in future implementations, epecially for the detection of low-power devices in high-frequency load signals. In future work, we will evaluate the

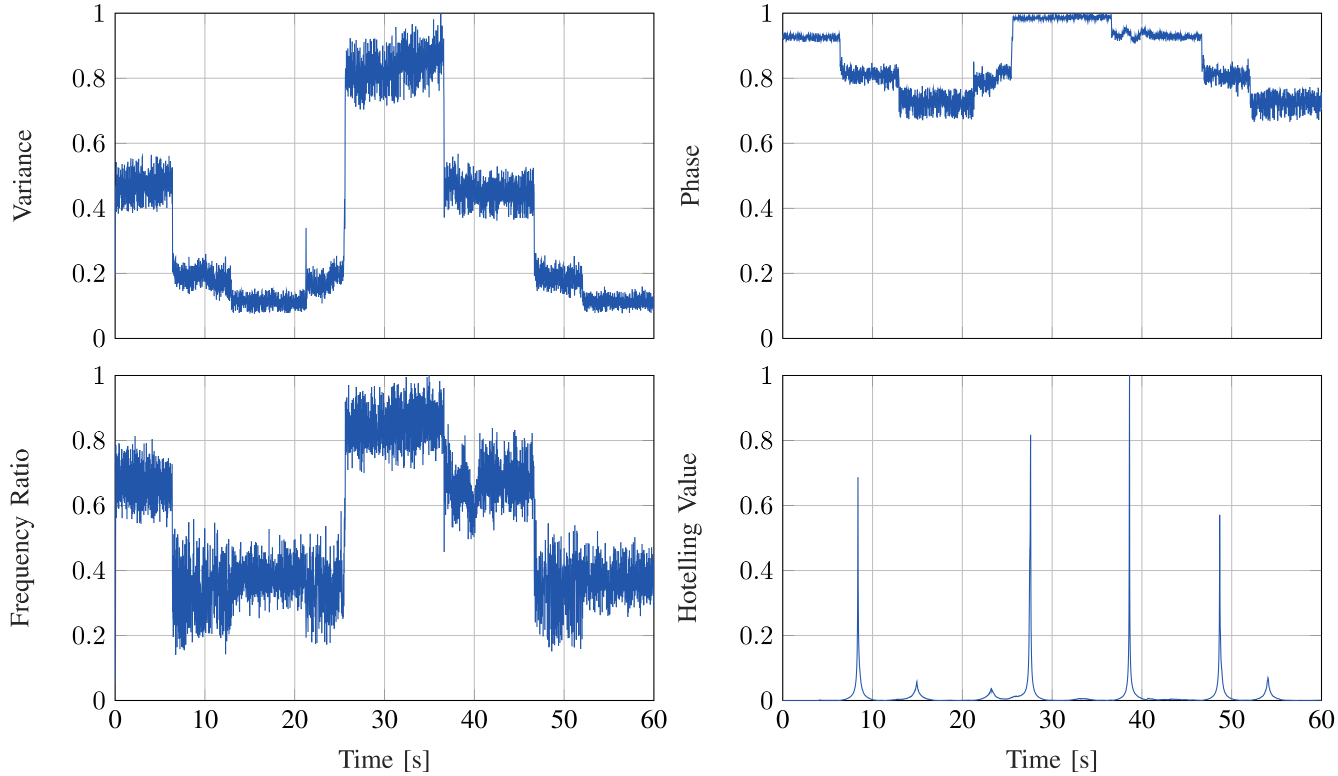


Figure 2. Extracted features and resulting Hotelling value from one phase. Values are normalized between 0 and 1 for readability.

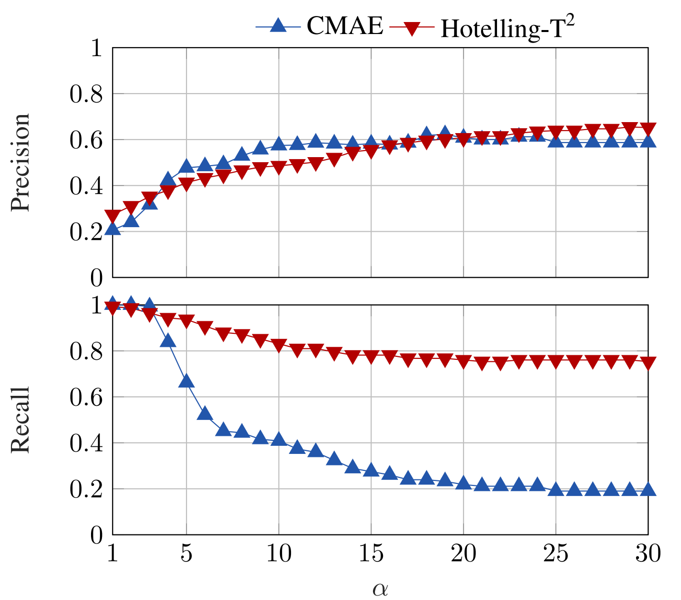


Figure 3. Precision and recall as function of applied threshold.

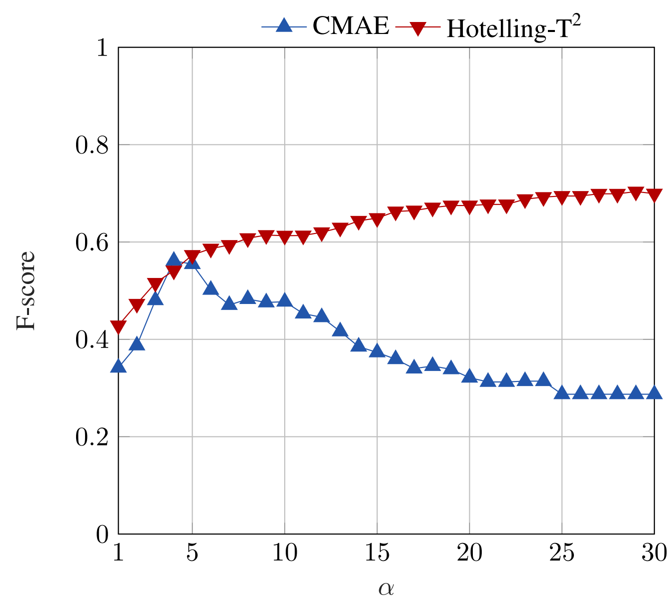


Figure 4. F-score as function of applied threshold.

presented approach with other high-frequency datasets and event detection algorithms.

ACKNOWLEDGMENT

This work is partially funded by the Central Federal Association of the Health Insurance Funds of Germany (GKV-Spitzen-verband) in the context of the QuoVadis research project and by the German Ministry for Education and Research (BMBF) within the joint research projects MeSiB (grant 16SV7723) and Audio-PSS (grant 02K16C201).

REFERENCES

[1] G. W. Hart, “Nonintrusive appliance load monitoring,” *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870–1891, 1992.

[2] Union, European, “Directive 2009/28/ec of the european parliament and of the council of 23 april 2009 on the promotion of the use of energy from renewable sources and amending and subsequently repealing directives 2001/77/ec and 2003/30/ec,” *Official Journal of the European Union*, vol. 5, p. 2009, 2009.

[3] World Health Organization, “World report on ageing and health,” 2015.

[4] C. Chalmers, W. Hurst, M. Mackay, and P. Fergus, “Smart health monitoring using the advance metering infrastructure,” in *International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM)*. IEEE, 2015, pp. 2297–2302.

[5] J. M. Alcalá, J. Ureña, Á. Hernández, and D. Gualda, “Assessing human activity in elderly people using non-intrusive load monitoring,” *Sensors*, vol. 17, no. 2, p. 351, 2017.

[6] A. Gerka, C. Lins, C. Lüpkes, and A. Hein, “Zustandserkennung von Beatmungsgeräten durch zentrale Messung des Stromverbrauchs,” in *German Medical Science GMS Publishing House*, 09 2017.

[7] A. Zoha, A. Gluhak, M. A. Imran, and S. Rajasegarar, “Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey,” *Sensors*, vol. 12, no. 12, pp. 16 838–16 866, 2012.

[8] T. Onoda, Y. Nakano, and K. Yoshimoto, “System and method for estimating power consumption of electric apparatus, and abnormality alarm system utilizing the same,” Nov. 9 2004, uS Patent 6,816,078.

[9] H. Y. Lam, G. Fung, and W. Lee, “A novel method to construct taxonomy electrical appliances based on load signatures,” *IEEE Transactions on Consumer Electronics*, vol. 53, no. 2, pp. 653–660, 2007.

[10] A. Diop, T. Jouannet, K. E. K. Drissi, and H. Najmeddine, “Method and device for the non-intrusive determination of the electrical power consumed by an installation, by analysing load transients,” Oct. 8 2013, uS Patent 8,554,501.

[11] K. D. Anderson, M. E. Bergés, A. Ocneanu, D. Benitez, and J. M. Moura, “Event detection for non intrusive load monitoring,” in *IECON 38th Annual Conference on Industrial Electronics Society*. IEEE, 2012, pp. 3312–3317.

[12] L. Norford, S. Leeb, D. Luo, and S. Shaw, “Advanced electrical load monitoring: A wealth of information at low cost,” *Diagnostics for Commercial Buildings: from Research to Practice, Pacific Energy Institute, San Francisco, CA*, 1999.

[13] A. R. Rababaah and E. Tebekaemi, “Electric load monitoring of residential buildings using goodness of fit and multi-layer perceptron neural networks,” in *International Conference on Computer Science and Automation Engineering (CSAE)*, vol. 2. IEEE, 2012, pp. 733–737.

[14] Z. Zhu, S. Zhang, Z. Wei, B. Yin, and X. Huang, “A novel cusum-based approach for event detection in smart metering,” in *IOP Conference Series: Materials Science and Engineering*, vol. 322, no. 7. IOP Publishing, 2018, p. 072014.

[15] L. De Baets, J. Ruyssinck, D. Deschrijver, and T. Dhaene, “Event detection in nilm using cepstrum smoothing,” in *3rd International Workshop on Non-Intrusive Load Monitoring*, 2016, pp. 1–4.

[16] J. M. Alcalá, J. Ureña, and Á. Hernández, “Event-based detector for non-intrusive load monitoring based on the hilbert transform,” in *Emerging Technology and Factory Automation (ETFA)*. IEEE, 2014, pp. 1–4.

[17] K. Anderson *et al.*, “Blued: A fully labeled public dataset for event-based non-intrusive load monitoring research,” in *Proceedings of the 2nd KDD workshop on data mining applications in sustainability (SustKDD)*, 2012, pp. 1–5.

[18] L. Farinaccio and R. Zmeureanu, “Using a pattern recognition approach to disaggregate the total electricity consumption in a house into the major end-uses,” *Energy and Buildings*, vol. 30, no. 3, pp. 245–259, 1999.

[19] M. N. Meziane, P. Ravier, G. Lamarque, J.-C. Le Bunetel, and Y. Raingeaud, “High accuracy event detection for non-intrusive load monitoring,” in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017, pp. 2452–2456.

[20] A. U. Rehman, T. T. Lie, B. Vallès, and S. R. Tito, “Event-detection algorithms for low sampling nonintrusive load monitoring systems based on low complexity statistical features,” *IEEE Transactions on Instrumentation and Measurement*, 2019.

[21] L. Jiang, S. Luo, and J. Li, “Automatic power load event detection and appliance classification based on power harmonic features in nonintrusive appliance load monitoring,” in *8th IEEE Conference on Industrial Electronics and Applications (ICIEA)*. IEEE, 2013, pp. 1083–1088.

[22] A. Gerka, F. Lübeck, M. Eichelberg, and A. Hein, “Electricity metering for dementia care,” in *VDE-Kongress – Internet der Dinge. Technologien / Anwendungen / Perspektiven.*, V. e. V., Ed. VDE Verlag, 11 2016, inproceedings, p. Paper 54.

[23] H. Hotelling, “The generalization of student’s ratio,” in *Breakthroughs in statistics*. Springer, 1992, pp. 54–65.

[24] J. P. Stevens, *Applied Multivariate Statistics for the Social Sciences*. Routledge Member of the Taylor and Francis Group, 2009, vol. 5.

[25] A. Trujillo-Ortiz and R. Hernandez-Walls, “Hotellingt2: Hotelling t-squared testing procedures for multivariate tests. a matlab file,” 2002.

[26] S. Makonin and F. Popowich, “Nonintrusive load monitoring (nilm) performance evaluation,” *Energy Efficiency*, vol. 8, no. 4, pp. 809–814, 2015.

[27] L. I. Kuncheva, “Change detection in streaming multivariate data using likelihood detectors,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 5, pp. 1175–1180, 2013.

 ISBN: 978-1-61208-739-9

Performance Isolation of Co-located Workload in a Container-based Vehicle Software Architecture

Johannes Büttner, Pere Bohigas Boladeras, Philipp Gottschalk, Markus Kucera and Thomas Waas
Faculty of Computer Science and Mathematics
Regensburg University of Applied Sciences
Regensburg, Germany
e-mail: {johannes2.buettner, pere.bohigas, philipp1.gottschalk, markus.kucera, thomas.waas}@oth-regensburg.de

***Abstract*—As the development in the automotive sector is facing upcoming challenges, the demand for in-vehicle computing power capacity increases and the need for flexible hardware and software structures arises, allowing dynamic managament of resources. In this new scenario, software components are to be added, removed, updated and migrated between computing units. To isolate the software components from each other and allow its orchestration, a container-based virtualization approach is being tested throughout this research. The analysis focuses on the question if this virtualization technology could be an option to ensure an interference-free operation. Four different sample applications from the automotive environment are tested for their susceptibility to resource contention. The research on the one hand shows that CPU and memory used by an application can be largely isolated with this technology, but on the other hand, it becomes apparent that support for I/O-heavy usage is currently not implemented sufficiently for container engines.**

Keywords*—*container-based virtualization; resource management; resource isolation; stress testing

I. Introduction

The automotive sector is facing an upcoming unprecedented redesign of vehicles. Besides the introduction of new electrical propulsion systems, the achievement of higher levels of driving automation and the development of the connected car are driving the conversion of vehicles into computers on wheels. Thus, on the one side, the use of deep learning and AI to drive a vehicle requires a massive computer power combined with safety redundancy and high availability. On the other side, the use of a huge quantity of sensors and the interconnectivity with a variety of IoT-based devices demands new flexible and dynamic architectures.

The desired flexibility and processing capacity therefore requires a fundamental revision and redesign of the actual in-vehicle system architectures. In the research project A^{3}F (*Ausfallsichere Architekturen für Autonome Fahrzeuge* – fail-safe architectures for autonomous driving vehicles), in which the Regensburg University of Applied Sciences (*Ostbayerische Technische Hochschule Regensburg*) takes part, evaluates new concepts for future vehicle system architectures [1]. The focus is on well-known distributed computing architectures in current enterprise IT and data centers that have similar issues and face similar challenges. Dynamic and flexible application architectures in this area have been developed decades ago by means of technologies, such as virtual machines and containers.

Within the A^{3}F project, new architectures for the in-vehicle computation system are conceived through the analysis of multi-node homogeneous computer cluster structures. This new approach to a flexible and dynamically managed hardware platform, on which distributed performance-intensive applications can be executed and orchestrated, employs high performance server nodes and reconfigurable Ethernet switches, as shown in Figure 1. The generic server nodes execute

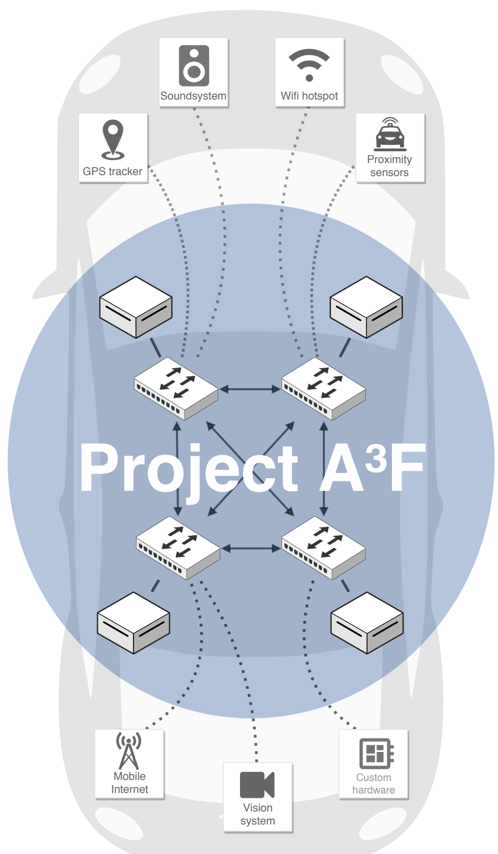


Figure 1. New architecture based on a computer cluster with redundant network connection, investigated within the project A^{3}F.

 ISBN: 978-1-61208-739-9

performance-demanding applications and provide for a flexible and extensible Software architecture. The reconfigurable Ethernet switches are responsible for maintaining and optimizing the network interconnectivity inside the cluster by rerouting the connections in real time. Other complementary hardware, such as real-time control functions, actuators, sensors and gateways for bus systems, are executed on their own dedicated, custom-tailored hardware, which are connected to the cluster and expose their signals via software services. This hardware architecture, however, falls out of the scope of this paper, and will thus not be discussed further. Instead, in the present pages the complementary designed software architecture, its framework and implementation challenges will be covered.

The rest of the paper is structured as follows: in Section II, the overall approach to the proposed system architecture is described and the formulation for its implementation is provided. The underlying important role of the isolation for the implementation of such architecture will be conferred in Section III. Section IV outlines the setup of the conducted tests, the results of which are depicted and analyzed in Section V. Finally, in Section VI, the findings of this investigation will be discussed.

II. Background

In recent decades, the size and complexity of the available software components in vehicles have been increasing dramatically [2]. So far, however, vehicles have been equipped with software configured in a static manner and dependent on dedicated hardware. Dependencies between software components are configured and validated at design-time. Subsequent changes, such as software updates or even new software components have therefore been cumbersome and expensive. The current, statically developed and configured ECU topology does not offer any practicable possibilities for dynamic changes at runtime.

In the years to come, in order to achieve the upcoming challenges of the automotive industry the size and complexity of both existing and new software components will have to grow exponentially. In addition, the safety in the vehicle will rely more and more on the efficient and uninterrupted performance of these components, whose role will be gaining importance increasingly. Thus, to be able to optimize the processing capacity available in hardware and to provide fail-safety features, the software architecture has to be reconceptualised.

A. *Flexible Software Architecture*

Today, many algorithms of modern in-vehicle functions primarily require high processing speeds, but are not dependent on specific surrounding hardware and can be executed on generic processors. Examples of these algorithms are multimedia applications, algorithms for image processing and geolocation services or calculations for optimal vehicle speeds and routes. Furthermore, dedicated hardware such as sensors and actuators may be exposed by a service-oriented architecture and are thus available to the software components on every computing unit.

In addition, today's users expect software in the vehicle to be easy to update and upgrade, the same way as in their mobile devices. A simpler software update capability, as illustrated in Figure 2, also ensures that vehicle manufacturers will no longer have to pay costly workshop visits or recalls, and enables them to integrate security-relevant, error-repairing or simply function-enhancing software updates with little effort. In the same way, software update capability would enable the possibility to have an application market, like those from the mobile world in which the user could download third party software, allowing also to import its business models.

To address the new demands, the project A^3F approaches the inclusion of flexibility and dynamism to the automotive software architecture by the adoption of available technologies used for distributed computing systems in data centers. On the one side, container-based virtualization is being tested to provide an encapsulation of the different software components and to introduce an abstraction layer between them and the hardware. Software components are thereby largely independent of the hardware used, being executed isolated inside a minimal virtualized operating system, called container, which in turn is run on a container engine. On the other hand, an orchestration tool is employed to provide container management, automating the deployment process and besides enabling the implementation of additional features that ensure fail-safe operation.

Within the research of the project A^3F the technologies *Docker* [3], for the container-virtualization, and *Kubernetes* [4], for the orchestration, are currently being tested. These have been chosen due to the widespread acceptance in the IT sector and the vast amount of compatible tools. Since Kubernetes is based on Docker, the adequacy test of these technologies have to be firstly and mainly focussed on the adequacy of Docker to meet the challenges. This work is

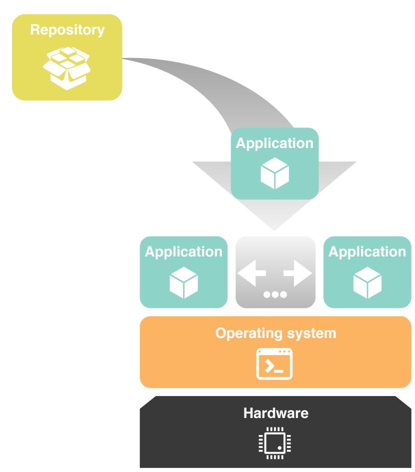


Figure 2. Layered diagram of a flexible software architecture with live software update capability.

focused in the analysis of one crucial aspect of the adequacy of the Docker technology for the automotive field.

B. Formal Description

One of the central aims of the research in the A^3F project is to determine which computing resources have to be taken into account when orchestrating multiple software components in a cluster-type architecture. In order to simplify the analysis, in this study every application is considered to be deployed once.

Below, a mathematical formulation for this problem (cf. [5]) is provided.

Given:

- A set of applications A:

  $A := \{a_0, a_1, \dots, a_{N-1}\}$, $|A| = N$
- A set of computing nodes B:

  $B := \{b_0, b_1, \dots, b_{M-1}\}$, $|B| = M$
- Each application requires a certain amount of resources:

  a_i^{CPU}, a_i^{RAM}, $\dots \qquad \forall\, a_i \in A$, $0 \le i \le N-1$
- Each computing node has a certain resource capacity:

  b_j^{CPU}, b_j^{RAM}, $\dots \qquad \forall\, b_j \in B$, $0 \le j \le M-1$

Find:

- Allocation matrix $M_{ij} \in [0, 1]$ in which $M_{ij} = 1$ if application a_i is allocated to computing node b_j.

Constraints:

- The application's resources allocated on each node may not exceed the node's capacity:

  $\sum_{i=0}^{N-1} (a_i^{CPU}) \cdot M_{ij} \le b_j^{CPU}, \dots \qquad \forall\, b_j \in B$
- Each application has to be allocated exactly once on each nodes:

  $\sum_{j=0}^{M-1} M_{ij} = 1 \qquad \forall\, a_i \in A$

This problem is known to be NP-hard [6]. There are several well-researched sub-optimal solutions, such as bin packing heuristics [7]. For a list and comparison of some of the algorithms, the reader is directed to [5]. However, one issue remains somewhat unanswered throughout the literature, namely which resource metrics need to be taken into account.

Most of the contributions in the literature dealing with the problem described above focus on only one type of resource (such as CPU [8] or memory [7]). In this sense, this contribution addresses the question of how strongly interferences between software components with respect to different resource types affect the runtime of these. And subsequently, the question will be examined as to how well mechanisms for performance isolation function.

III. Performance Isolation

When investigating the effectiveness of performance isolation mechanisms, the first thing to do is to find out which interferences already can be avoided with today's technologies. Therefore, two different available performance isolation mechanisms for contention of the CPU in container-based virtualization are analyzed and their impact with regard to five metrics (CPU, RAM, Cache, I/Os and Network) is measured. In the following section, first the theoretical background is explained, what we understand by interference, where it comes from and why and how to avoid it. Thereafter, the technologies that will be investigated in this research are presented together with our expectations and hypotheses for our experiments.

A. Interference

An essential aspect when operating multiple software components on shared hardware is to ensure that they are free of interferences. Interferences could be caused by, for example, contention among co-located workload for shared physical resources (such as CPU, network, and cache). It must be ensured at design-time that each application can access the required resources with the necessary frequency, with the necessary amount of time in order to guarantee an unobstructed operation. As for a single application, this might be ensured by isolating the application as if it were running on dedicated hardware. This is also known as performance isolation.

Efforts have been made to model and predict this interference by various means [7]–[9]. However, these approaches focus mainly on IT and data centre environments. For example, in [7], a distinction is made between "latency-sensitive" software components and "batch", whereby the "latency-sensitive" applications are directly user-facing and therefore have high QoS requirements, which manifests themselves in a low latency tolerance. This differentiation is mainly based on the consideration that there is a trade-off between QoS requirements and resource efficiency.

This trade-off must of course be assessed differently in vehicles than in IT due to the much more catastrophic effects that would result if QoS were not met. Although it is difficult to quantify these effects in general terms, especially at this early stage of development, one can qualitatively say that the effects of non-compliance with QoS are more catastrophic than in IT. The situation in vehicles is not foreseeable at this stage. At the same time, our development strategy is to first design a system that works conceptually. The system can then be optimized for resource utilization at a later point in time. These are reasons why this paper follows the approach that interferences between software components should never be tolerated. Instead, applications should be isolated in such a way that interferences cannot occur in the first place. In this way, latencies of software components can be reliably predicted.

To meet this goal, the use of a container engine is analyzed in this project. Container-based virtualization provides an easy way to limit the resource consumption of an application and

 ISBN: 978-1-61208-739-9

also meets the requirements of our proposed flexible software architecture. The interference between applications running inside such a resource-constrained environment are measured. For this, we run the applications while the system is put under stress by overloading certain resources. Resource contention has to be deliberately brought in to see if the limitation works. This is done through an application that is referred to here as a “stressor”.

B. Container-based Isolation

Within the research of the project A^{3}F the platform Docker is being used for isolation between software components. It allows different mechanisms for the limitation of the resources allocated to a container. As for resource isolation, Docker builds upon two mechanisms: *cgroups* and *namespaces*. Both are native Linux kernel services. Namespaces allows to create isolated virtualized system resources such as process IDs, network access, file system, etc. Cgroups provide a mechanism to limit processed resources. However, since all containers share the same host OS kernel and rely on functions within the kernel, the overhead for managing the containers increases with the number of containers. This affects and degrades the performance of the containers itself, especially for I/O-intensive workloads [10]. This problem is not limited to containers, but affects all general purpose operating systems [11]. This already highlights one big problem regarding the consolidation of containers on a computing node.

In addition, as stated in [10], any resources that do not support concurrent use will cause a major bottleneck in the host OS kernel, because concurrent access to these resources is enabled and managed by the mentioned host OS. Consequently the generation of a very large number of interrupts under such loads results in the other processes being more frequently preempted. This activity manifests itself in high CPU load and many context switches for the host OS in order to serve interrupts. Input-output (I/O) bound applications have significantly higher overhead, particularly network-intensive applications. With such workloads, the resource requirements of the host OS kernel must also be taken into account [10].

C. Hypotheses

Regarding the above effects, we had the following hypotheses for limitation:

1) The limitation should work well for applications that aren’t very I/O-intensive and less well for applications with high I/O-requirements.
2) The limitation should work less well for applications that use a lot of resources that do not support concurrent access (I/O, CPU, Network).

In order to render the above mentioned effects quantifiable and thus to support the hypotheses with real data, a test battery is performed on real automotive applications. This is described in the following section.

IV. Experimental setup

To evaluate the interference of the test applications, mentioned in Section IV-B, their execution time is measured in different cases with respect to different resource metrics. The general idea is to measure the execution time of each application while overloading certain resources with our stressor applications. After the completion of the application’s operations, the stressor is terminated and the execution time of the application is noted. Each of the different combinations between the four test cases and the five stressor configurations was tested 10 times by each of the four test applications, in order to obtain sufficient quantitative data.

A. Hardware Setup

For the experimentation and testing, several Intel NUC-Kits were used as generic server nodes, as well as Marvell Ethernet switches especially designed for automotive requirements. The NUCs are often employed as examples of homogeneous, powerful but generic hardware units. They are equipped with a recent processor (x86-64, 4 cores, 8 MB L3) and 32 GB of RAM and are connected to each other via redundant Ethernet network. As operating system they run a distribution of GNU/Linux, kernel version 4.15.0. This configuration shall allow individual applications to be run on any node in the cluster, independently to a great extent of specific hardware.

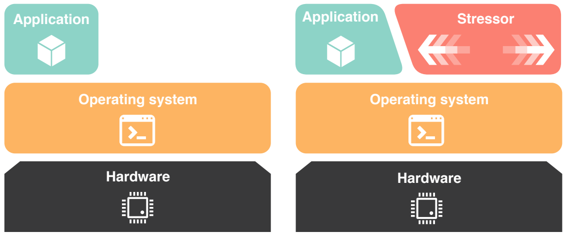
B. Application Types

In order to get an overview as close to reality as possible, four software modules from the open platform *Apollo* [12] employed to achieve the autonomous driving were tested:

1) Perception: An image-processing application to identify road signs.
2) Planning: A GPS application to plan routes.
3) Prediction: An application predicting the trajectory of an object.
4) Controlling: An algorithm, which takes decisions based on a combination of the outputs of the above applications.

C. Test cases

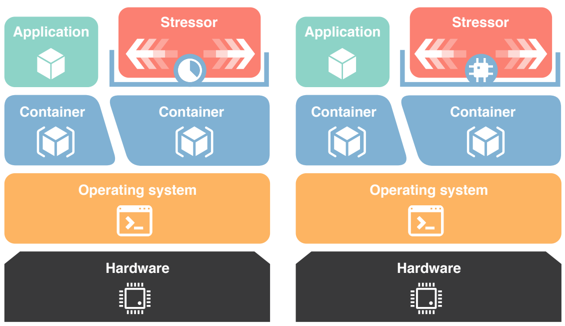
The software stacks for the four different test cases are depicted in Figure 3 and 4. In case 1, as shown in Figure 3(a), the application is run without any limitations and without the stressor being executed in parallel. In case 2, the application is run without limitation, but with the stressor being executed in parallel. This is shown in Figure 3(b). In cases 3 and 4, a container virtualization layer is introduced, namely Docker, providing some means of limitation. The capabilities of limitation of workloads incorporated in the Docker engine at this time include only the CPU and the memory. However, there are two different CPU limitation mechanisms that are worth a distinct consideration. On the one side, one may specify the *CPU quota* which a container is allowed to use within one second. This kind of limitation is used in case 3, which is depicted in Figure 4(a). On the other side, one may bind a container to one or more specific *CPU core*, which is used in case 4 and shown in Figure 4(b).



(a) Case 1: test application running without stressor.

(b) Case 2: test application running with stressor.

Figure 3. Experimental cases without CPU resource limitations.



(a) Case 3: test application running besides stressor, with CPU quota limitation.

(b) Case 4: test application running besides stressor, with CPU core limitation.

Figure 4. Experimental cases with CPU resource limitations.

D. *Stress generation*

To evaluate the impact of the CPU containment into other resources, each test case was examined under five types of external stress, each one affecting a different resource between CPU, RAM, Cache, IO and Network. The application used to generate stress (the stressor) is based on *stress-ng* [13], with the exception of the one employed to generate network stress, which uses *iperf* [14]. In order to provoke the different stress environments, the stressor application was configured as it is shown in Table I.

E. *Source Code*

The source code for the four test applications can be found at [12]. The source code for all experiments conducted throughout this paper can be found at [15].

V. RESULTS

The data collected in the tests is presented using box plots in Figure 5. The distribution of the measured time values is colored by each experimental case and classified by each stressor configuration.

The analysis of the data obtained in each of the case studies points towards the following statements:

a) Comparing the results obtained from case 1 and 2 it becomes evident that running with contention among shared resources has a severe impact on application performance.

TABLE I. CONFIGURATION OF THE STRESSOR WITH *stress-ng* AND *iperf*

Name	Stressor Configuration	Description
CPU	`stress-ng -c 25`	25 workers spinning on `sqrt(rand())`
RAM	`stress-ng --malloc 8 --malloc-bytes 2G`	8 workers exercising `malloc()/realloc()/free()`
Cache	`stress-ng -C 25`	25 workers trashing CPU Cache
I/O	`stress-ng -d 8`	8 workers spinning on `write()/unlink()`
Network	`iperf3 -c $IP -w 510M -n 510M`	send 510MB of data to a remote host (`$IP`)

b) Contrasting the non-limited cases with case 3 and 4, is clearly visible how the two analyzed limitation options offered by the platform Docker have a noticeable impact on the execution time of the applications (with the exception of the case 3 for the I/O metrics).

c) For all examined metrics, except for I/O, the CPU limitation is the first and foremost valuable kind of limitation, as all other metrics depend on the CPU resources.

d) Comparing the results obtained from case 3 and case 4 it becomes evident that running with CPU quota limitation can indeed mitigate the performance impact, but neither completely nor satisfactorily.

e) In contrast, the CPU core limitation (case 4) shows a better performance, but its execution values are still widely divergent from those of the single execution (case 1).

With regard to statement c), it is also noteworthy that I/O stress has hardly any performance impact. At this point there are two potential causes for the lower impact of the I/O stressor, which have to be considered.

- The stress generated for the I/O metric was not of the same range as the one caused by other resources. This may be due to a certain limitation of the *stress-ng* tool.
- The execution of the analyzed applications do not have intensive requirement in the I/O resources.

A preliminary analysis of the different impact into performance of the CPU quota limitation when compared to the CPU core limitation, as mentioned in the statement d), suggest two potential causes to be considered.

- The logic behind the functionality of the CPU quota limitation is probably creating in runtime a bottleneck, when the kernel / Host OS is managing CPU resources for many concurrent accesses. This confirms what we described in Section II.
- The weak performance of this CPU limitation option is caused by a bug in a quota option of the process scheduler *Completely Fair Scheduler (CFS)*, which is the default scheduler in the Linux kernel [16]–[18]. This bug appears

 ISBN: 978-1-61208-739-9

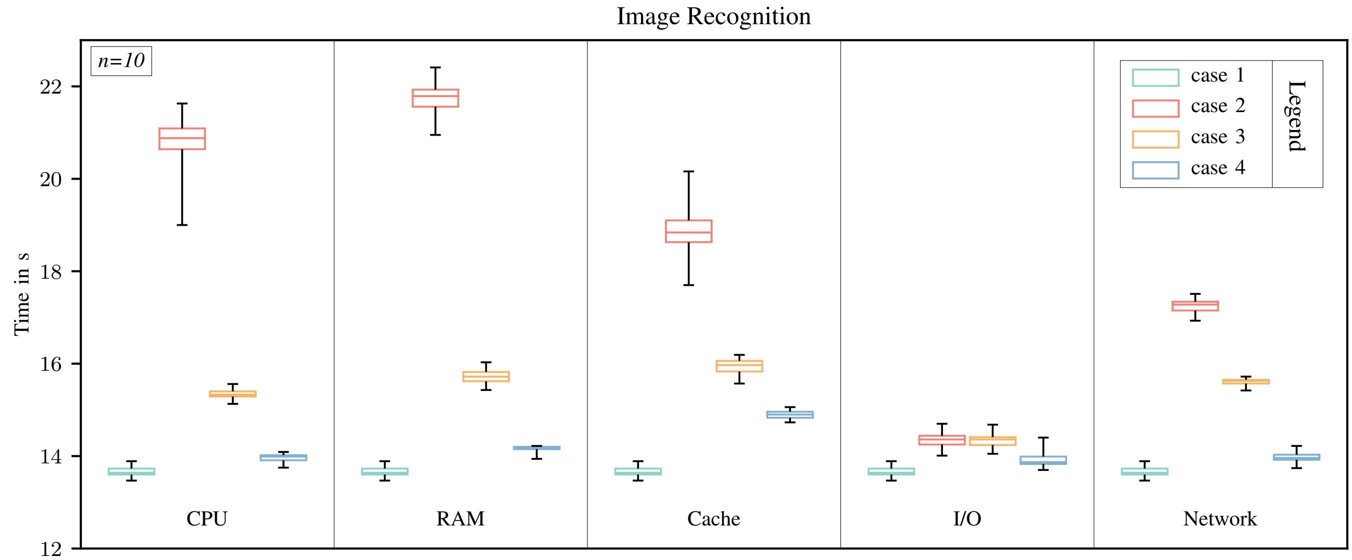


Figure 5. Test results for the Perception Application

to be fixed in more recent versions of the Linux kernel [19], [20].

These potential causes for the observed behaviour of the CPU quota limitation are uncorrelated between them, which means that despite the bug was fixed in recent kernel versions, the quota limitation could still not match the degree of isolation offered by the core limitation.

VI. Conclusions

The introduction of cluster-type architectures in vehicles demands an appropriate encapsulation of the different software components, which should abstract these components from the computing unit in which they are running on. This level of abstraction will enable flexibility and dynamism to the whole system, allowing to manage applications at runtime according to their needs. To provide this level of encapsulation the project A^{3}F studies the applicability of container-based virtualization, in particular the use of the technology Docker.

To consider the applicability of this type of virtualization in the automotive field, its fulfillment of the appropriate restrictive safety requirements has to be ensured. Safety critical systems have to be able to run uninterrupted getting access to all the resources they need.

In order to avoid having to distinguish between critical and non-critical applications and to have to proceed differently for each case, it was considered in this work that interferences between software components should never be tolerated.

As observed in test results, to be able to avoid interference when using container-based virtualization some sort of limitation of resources must be implemented. The limitation of the CPU seems to be a good choice, due to the fact that it has a direct impact on the other metrics. This appears not to be the case for the I/O resources, although it seems this could depend on test conditions. Nevertheless, a complete and fully effective containment of the available resources in the host OS cannot be provided by the Docker technology, or at least not by the time of this research.

Among the analyzed options, the usage of the CPU core limitation appears to be the best option to minimize the interferences between containers, nearing the execution times to the values obtained without stress. The CPU quota limitation, on its behalf, can only slightly reduce the interferences affecting the application but not even closely to the values obtained with the CPU core limitation. A first analysis points towards that this may be due to a bug in the internal scheduler of the Linux kernel, fixed in more recent versions. Therefore forthcoming researches in this direction to verify this hypothesis are suggested.

Although Docker seems not to be currently prepared to provide interference-free performance isolation, it becomes evident that the introduction of dynamical and distributed vehicle architectures requires container-based virtualization solutions to be implemented. Therefore, it will be necessary that the automotive industry develops its own container-based platform in order to ensure the complete isolation between the different software components, key aspect where the fulfilment of its particular safety requirements will rely on.

Acknowledgment

This research has taken place within the framework of the project A^{3}F, which is funded by the *Bayerisches Staatsministerium für Wirtschaft, Landesentwicklung und Energy* (Bavarian Ministry of Economic Affairs, Regional Development and Energy) under the funding programme *Informations- und Kommunikationstechnik Bayern* (Information and Communication Technology Bavaria). The authors would like to acknowledge the colaboration of Ms. Cristina Alonso Villa for her work on editing this paper.

REFERENCES

[1] J. Büttner, M. Kucera, and T. Waas, "Opening up new fail-safe Layers in distributed Vehicle Computer Systems," in *Proceedings of the 8th International Joint Conference on Pervasive and Embedded Computing and Communication Systems, PECCS 2018, Porto, Portugal, July 29-30, 2018.*, 2018, pp. 236–240. [Online]. Available: https://doi.org/10.5220/0006903602360240

[2] D. Reinhardt, W. Kühnhauser, U. Baumgarten, and M. Kucera, *Virtualization of embedded real-time systems in multi-core operation for partitioning safety-relevant vehicle software (Virtualisierung eingebetteter Echtzeitsysteme im Mehrkernbetrieb zur Partitionierung sicherheitsrelevanter Fahrzeugsoftware).* Ilmenau: Universitätsverlag Ilmenau, 2016, oCLC: 951392623.

[3] "Docker," [retrieved: July, 2019]. [Online]. Available: https://www.docker.com/

[4] "Kubernetes," [retrieved: July, 2019]. [Online]. Available: https://kubernetes.io/

[5] J. Xu and J. A. B. Fortes, "Multi-objective Virtual Machine Placement in Virtualized Data Center Environments," in *2010 IEEE/ACM Int'l Conference on Green Computing and Communications Int'l Conference on Cyber, Physical and Social Computing*, Dec. 2010, pp. 179–188.

[6] A. Hegde, R. Ghosh, T. Mukherjee, and V. Sharma, "SCoPe: A Decision System for Large Scale Container Provisioning Management," in *2016 IEEE 9th International Conference on Cloud Computing (CLOUD)*, Jun. 2016, pp. 220–227.

[7] J. Mars, L. Tang, R. Hundt, K. Skadron, and M. L. Soffa, "Bubble-Up: increasing utilization in modern warehouse scale computers via sensible co-locations," in *Proceedings of the 44th Annual IEEE/ACM International Symposium on Microarchitecture - MICRO-44 '11.* Porto Alegre, Brazil: ACM Press, 2011, p. 248.

[8] S. Votke, S. A. Javadi, and A. Gandhi, "Modeling and Analysis of Performance under Interference in the Cloud," in *2017 IEEE 25th International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS).* Banff, AB: IEEE, Sep. 2017, pp. 232–243.

[9] S. Govindan, J. Liu, A. Kansal, and A. Sivasubramaniam, "Cuanta: quantifying effects of shared on-chip resource interference for consolidated virtual machines," in *Proceedings of the 2nd ACM Symposium on Cloud Computing - SOCC '11.* Cascais, Portugal: ACM Press, 2011, pp. 1–14.

[10] S. K. Garg and J. Lakshmi, "Workload performance and interference on containers," in *2017 IEEE SmartWorld, Ubiquitous Intelligence Computing, Advanced Trusted Computed, Scalable Computing Communications, Cloud Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, Aug. 2017, pp. 1–6.

[11] G. Banga, P. Druschel, and J. C. Mogul, "Resource Containers: A new Facility for Resource Management in Server Systems," in *Third Symposium on Operating System Design and Implementation (OSDI-III)*, New Orleans, LA, Feb. 1999, pp. 45–58.

[12] "Apollo Auto," [retrieved: July, 2019]. [Online]. Available: http://apollo.auto/

[13] C. I. King, "stress-ng: a tool to load and stress a computer system," [retrieved: July, 2019]. [Online]. Available: https://kernel.ubuntu.com/~cking/stress-ng/

[14] "iPerf," [retrieved: July, 2019]. [Online]. Available: https://iperf.fr/

[15] "OTH Gitlab Projekt A3f," [retrieved: July, 2019]. [Online]. Available: https://gitlab.oth-regensburg.de/IM/projekt_a3f/veroeffentlichungen/ambient2019-regular-paper

[16] I. Babrou, "Linux-Kernel Archive: Unexpected CFS thottling," 2017, [retrieved: July, 2019]. [Online]. Available: http://lkml.iu.edu/hypermail/linux/kernel/1712.0/07072.html

[17] ——, "Overly aggressive CFS - GitHub Gist," 2017, [retrieved: July, 2019]. [Online]. Available: https://gist.github.com/bobrik/2030ff040fad360327a5fab7a09c4ff1

[18] ——, "CFS quotas can lead to unnecessary throttling - GitHub," 2018, [retrieved: July, 2019]. [Online]. Available: https://github.com/kubernetes/kubernetes/issues/67577

[19] D. Chiluk, "Fix low cpu usage with high throttling by removing expiration of cpu slices - LKML," May 2019, [retrieved: July, 2019]. [Online]. Available: https://lkml.org/lkml/2019/5/17/581

[20] X. Pang, I. Molnar, P. Zijlstra, L. Torvalds, T. Gleixner, and B. Segall, "sched/fair: Fix bandwidth timer clock drift condition - GitHub," Jun. 2019, [retrieved: July, 2019]. [Online]. Available: https://github.com/torvalds/linux/commit/512ac999

 ISBN: 978-1-61208-739-9

Effect of Heart Rate Feedback Virtual Reality on Cardiac Activity

A preliminary study with heart rate feedback "fire flame" scene

Shusaku Nomura, Rei Sekigawa, and Naoki Iiyama
Department of Engineering
Nagaoka University of Technology
Nagaoka, Japan
e-mail: nomura@kjs.nagaokaut.ac.jp

***Abstract*— In this study, the psycho-physiological effect of a heart rate (HR) feedback virtual reality (VR) scene on a human was investigated. The VR scene was a living room with a fire place, where a fire flame flickered in synchronisation with the viewer's own HR in a real-time manner. Fifteen male university students underwent this interactive installation for 60 s, repeatedly. In comparison with the control condition in which the fire flickered at a constant rhythm, the HR significantly decreased and exhibited a positive correlation with the subjective score of "comfortable." Thus, the change in the cardiac sympathetic nervous activity may be attributed to the psycho-physiological effect.**

Keywords-biofeedback; virtual reality; heat rate; ambient feedback.

I. Introduction

VR is used not only for the purpose of entertainment, but also for clinical purposes, such as treatment of phobias [1] and pain management [2]. Majority of these studies presented patients with a relaxing virtual world, and reported the alleviation of anxiety, a positive mood, and an improvement in physiological states of the patients. Therefore, immergence into the VR world has a significant influence on the viewer's psycho-physiological functioning.

Recently, attempts have been made to integrate VR with biofeedback training. "Virtual Meditative Walk" enables users to perform meditation training by walking in a virtual forest where the weather conditions change according to the user's galvanic skin response; thus, the users can train to control their sympathetic nervous activity by themselves [3].

The aforementioned studies demonstrate the possibility of expanding the use of bio-signal interactive VR, which positively induces the state of the human mind and body. However, to the best of our knowledge, no study has directly investigated the psycho-physiological effects on humans experiencing an interactive VR system using bio-signals. The present preliminary experimental study aims to investigate the effect of an interactive heart rate (HR) virtual reality (VR) scene on the human mind and body.

The remainder of this paper is organised as follows. The configuration of our developed bio-signal feedback VR system and the experiment conducted to test the system efficacy is described in Section 2. The experimental results and their discussion are presented in Sections 3 and 4, respectively.

II. Method

As a preliminary challenge to explore the psycho-physiological effect of the bio-signal feedback VR system, we focused on the HR signal because of its accessibility.

A. Configuration of heart rate feedback virtual reality scene

The HR feedback VR system comprises a VR head mounted display (HMD) (Oculus Rift DK 2, Facebook Technologies, LLC.), bio-signal amplifier (BIOPAC MP150, BIOPAC Systems, Inc.), game engine (Unity 5.2), and personal computer (controller). The electrocardiogram (ECG) signal was acquired by the bio-signal amplifier at a sampling rate of 200 Hz. The beat-to-beat interval (known as "RR interval") of the ECG signal was determined and conveyed to a control PC to generate the HR feedback VR "fire flame" using the game engine. Then, the VR image was exposed to a viewer via the HMD.

The VR scene was a living room with a fire place where a fire flame flickered. In the HR feedback regulation, the size and brightness of the fire flame was varied to synchronise with the viewer's own ECG signal; eventually, the size and brightness of the flame synchronised with the viewer's own heartbeat. The degree of change in the size and brightness was normalised based on the ECG signal of the viewer.

B. Experiment

The psycho-physiological effect of the HR feedback VR scene was investigated in 15 healthy university male students using a within-subject experimental design. None of the subjects had any visual disorders, such as near- or far-sightedness. To directly compare the effect of the bio-signal feedback, we prepared three different "fire flames," which were changed to synchronise with the subject's HR (SYN condition), in a manner of a sine wave at a constant rhythm (CST condition), and at random in terms of size and brightness at a constant rhythm (RND condition). The constant rhythm in CST and RND was set to the average HR of all the subjects (75 bpm), which was measured in advance. The degree of change in the size and brightness of the flame in CST and RND was set to the same range as that in SYN.

In the experiment, we used a pairwise comparison and made a subject compare two conditions sequentially, which was repeated twice. For example, a subject experienced the VR scene in SYN and RND alternatively. Each condition was experienced for 60 s with a 5 s interval (255 s in total)

 ISBN: 978-1-61208-739-9

between the two conditions. Scenes presented to the subjects during this procedure were repeated for three possible combinations of input signals, i.e., SYN-RND, SYN-CST, and RND-CST; therefore, each subject experienced nine trials in total. The sequence of the presentation of the nine conditions was randomised and counter-balanced to compensate the possible order effect.

Subjective scores, namely, “spacious” (neutral item), “comfortable,” and “natural,” were assessed using the 11-point Likert scale of +5 (strongly agree) to –5 (strongly disagree). The HR and the high frequency component (0.15–0.40 Hz) of the HR variability (HF) were calculated from the ECG data. No information regarding the system configuration was provided to the subjects. Instead, they were instructed to just watch the presented VR scene and respond to the questionnaire.

Nakaya’s variation of the Sheffé’s ANOVA for paired comparisons [4] was employed for the statistical analysis of the subjective scores. The studentised range statistic (q) was further used for multiple comparisons [5]. Pair-wised student’s t-test was employed to evaluate the HR and HF, and the level of statistical significance was set to 0.05.

III. RESULTS

According to a verbal interview conducted after the experiment, no subject complained about any malaise, including VR sickness, during the experiment, or abandoned the repetitive exposures in the experiment. Additionally, no subject was aware of the regulation manner of the fire flame. Moreover, no subject could differentiate between the SYN and RND VR scenes.

In the SYN condition, subjective scores of “comfortable” and “natural” were significantly higher than those in the CST condition (all for $p < 0.01$) and were marginally higher than those in the RND condition, as depicted in Figure 1.

The HR value in the SYN condition was smaller than that in the CST and RND conditions, and a significant difference was observed in the HR between the SYN and RND conditions ($p < 0.05$), as depicted in Figure 2. Furthermore, there was a moderately positive correlation between HR in the SYN condition and the score of “comfortable” ($r = 0.48$). No difference was observed in HF in any of the conditions, and no correlation was observed between HF in any condition and any of the psychological scores.

IV. DISCUSSION

The HR feedback VR fire flame that was implemented in this study demonstrated the effect on cardiac activity in terms of decline of HR. Moreover, the subjects viewed the bio-signal interactive VR as “comfortable” and “natural,” and the HR was found to be associated with “comfortable,” even though the subjects were not distinctly aware of the difference among the conditions.

It is well known that the sound of the heartbeat has an alleviative effect, as people say “listening your own heart beat makes you relax.” The heartbeat feedback fire flame scene may induce a similar effect on the human mind. Moreover, because the HR reduced in the SYN condition, such a psychological benefit may be mediated by the suppression of the cardiac sympathetic nervous system.

Further research with a wider population range may lead to the development of a new concept of “ambient bio-feedback system,” in which the ambient stimuli (such as lightning, sound, and smell) are regulated by synchronizing with the human bio-signal, thereby fostering a positive psycho-physiological effect.

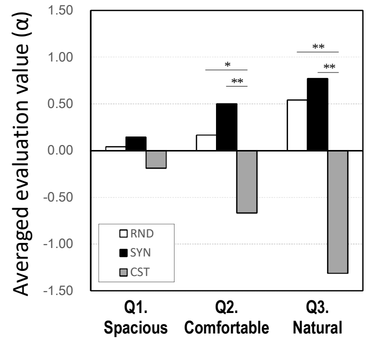


Figure 1. Subjective socres for virtual reality (VR) scene in each condition.

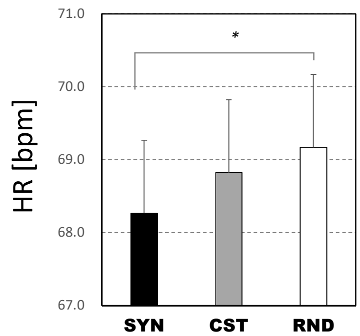


Figure 2. Mean heart rate during VR exposure.

REFERENCES

[1] D. Hartanto, I. L. Kampmann, N. Morina, P. G. M. Emmelkamp, M.A. Neerincx, and W-P. Brinkman, “Controlling Social Stress in Virtual Reality Environments,” PLoS ONE, vol. 9, e92804, pp.1-17, 2014.

[2] J. L. M. Vázquez, B.K. Wiederhold, I. Miller, D. M. Lara, and M. D. Wiederhold, “Virtual Reality Assisted Anesthesia (VRAA) During Upper Gastrointestinal Endoscopy,” Surgical Research Updates, vol. 5, pp.1-11, 2017.

[3] M. Nazemi and D. Gromala, “VR Therapy: Management of Chronic Pain Using Virtual Mindfulness Training,” Proc. CHI 2014, ACM, pp.1-2, 2014.

[4] S. Nakaya, “A Modification of Scheffe's Method for Paired Comparisons,” in 11th Meeting of Sensory Test, pp.1-12, 1970.

[5] H. Takagi, “Practical Statistical Tests Machine Learning---III---Signi,” Systems, Control and Information, vol. 58, pp. 514-520, 2014.

Introducing SAM.F: The Semantic Ambient Media Framework

David Bouck-Standen
Kingsbridge Research Center
Hamburg, Germany
email: dbs@kingsbridge.eu

***Abstract*—In our digital society, any user can become a producer of media. With the heterogeneity of devices, which vary in their capabilities or hardware resources, the question of accessing these media in a meaningful and usable context grows more important, as devices and users are more and more interconnected through digital services. To encounter the various challenges these observations present, with the Semantic Ambient Media Framework, this article proposes a framework, in which media, devices, and services are extended and interconnected through semantic models for various contexts. Providing a Web-based API, the framework allows applications and devices to access media, which are provisioned through the framework's services. These services are automatically tailoring the media, depending on the semantic information on the context they are used in, their semantic interconnection with other media, and the specific application, device, and context they are accessed from. This contribution illustrates the concept and system architecture of the Semantic Ambient Media Framework from a developer's perspective and describes a practical scenario, in which the framework is already utilized, concluding with an outline of future work.**

Keywords-Semantic Media; Semantic Repository; Cross-Platform Media Provisioning.

I. INTRODUCTION

Today, interconnected and feature-rich multimedia systems allow users to produce high amounts of user-generated content. The contexts technology is used in also shift towards mobile and ubiquitous computing. The users utilize their personal mobile devices, such as smartphones, tablets, and other devices, to connect to interconnect with other systems through the Internet [1], [2].

As each multimedia system uses technologies with different interaction paradigms, they offer different capabilities for presentation, processing, and storing information in their own content repositories [3].

Focusing a vision of a convergence of personal or social information, at least the interconnection of multimedia systems, or at best a single multi-purpose multimedia repository system would be required [4]. The latter observation would also solve the problem of media being isolated for use in a single application or on a single device.

These challenges have been researched in various context-specific domains, as related work (cf. Section 2) indicates. With the *Semantic Ambient Media Framework* (SAM.F), this contribution presents a general context-independent approach. SAM.F is a framework that semantically interconnects (a) *media*, (b) *devices and applications*, and (c) *services*, which are enriched by digital properties in the form of semantic annotations. For both client application development, as well as the extension of framework functionality, SAM.F offers interfaces for developers.

In SAM.F, media consists of, e.g., text, photos, audio, videos, animations, or 3D objects. These are extended by digital properties, e.g., by classifying the media's content in the internal model of SAM.F.

Digital properties also include Meta data from the original file, such as Meta information on MIME type or encoding. For devices, in SAM.F, we model digital properties reflecting, e.g., the devices' capabilities', location, capacity, screen size, or screen resolution.

All digital properties are utilized by the services in SAM.F. Client applications running on users' devices access the services of SAM.F through Web-based interfaces. Each service serves a dedicated purpose, interconnecting devices and applications through the shared use of devices and media.

To be able to interconnect services and devices through media, SAM.F features an extendable service-based architecture, providing developers with dedicated interfaces and the means to develop new modularized services for SAM.F, as described in detail in this article. SAM.F is accessible for devices and applications through a Web-based API.

To illustrate the use of SAM.F, Section 2 outlines a practical scenario from a developer's perspective, referencing actual work [5]. In Section 3, related work is regarded. In Section 4, Semantic Media used in SAM.F, the system's architecture, as well as the modular service-based structure is illustrated in detail, followed by a summary and outlook in Section 5.

II. RELATED WORK

Semantic media comprises the integration of data, information and knowledge. This relates to the Semantic Web [6] and aims at allowing computer systems as well as humans to make sense of data found on the Web. This research field is of core interest since it yields naturally structured data about the world in a well-defined, reusable, and contextualized manner. The field of metadata-driven digital media repositories is related to this work [7] as well. Apart from the goals of delivering improved search results

 ISBN: 978-1-61208-739-9

with the help of Meta information or even a semantic schema, SAM.F distinguishes itself from a pure repository by containing and using multiple repositories as internal components, as illustrated below. As Sikos [8] observes, semantic annotations feature unstructured, semi-structured, or structured media correlations. Sikos outlines the lack of structured annotation software, in particular with regard to generating semantic annotations for video clips automatically. SAM.F offers means for both structured and semi-structured semantic annotations. Through an interface, the functionality of SAM.F can be extended to, e.g., automatically annotate media as outlined below, but is not limited to video clips. By these means, SAM.F delivers even more sophisticated features.

In general, SAM.F facilitates collecting, consuming and structuring information through device-independent interaction with semantically annotated media, whereas the linked data research targets sharing and connecting data, information and knowledge on the Web [9]. The concept originally developed by the author [10] was already used in different contexts, e.g., the automatic reconstruction of 3D objects from photo and video footage based on semantically compiled sets of media [11]. However, with SAM.F, in this contribution the original concepts of the author [10] are presented in their revised version in order to expurgate over-weight services previously tailored for a narrow project-specific and less transferable use. Thus, SAM.F does not share code with or reuse code from any related work.

Blumenstein et al. [12] outline a technical concept in museum context, that relies on a server-based architecture to provide museum content in a multi-device ecology. SAM.F could be used in similar contexts, but is not limited to the use in museums.

Ambient systems can provide a platform for displaying

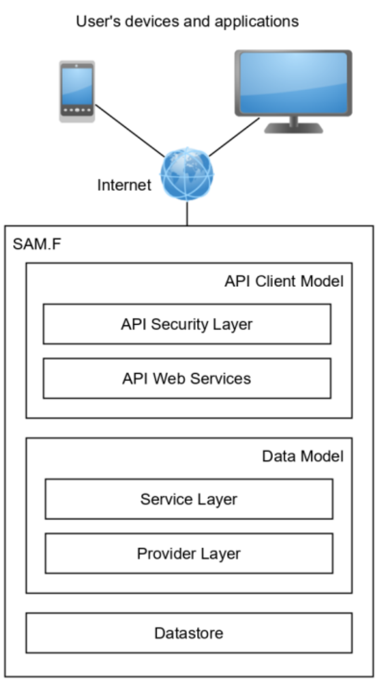


Figure 1: Layered architecture of SAM.F.

of and interaction with media [13]. In this context, the delivery of content on different devices is an important issue in SAM.F, e.g., with respect to the devices' capabilities or their context of use, and SAM.F addresses this challenge by provisioning media depending on applications and devices specifications or capabilities. SAM.F also addresses the issue of limited bandwidth of mobile devices.

The Social Web is related to this work, as it makes it easy for people to publish media online. Yadav et al. [14] propose a framework interconnecting Social Web and Semantic Web by semantically annotating and structuring information people share. SAM.F could be used in this way, but focuses on semantically enriched or described instances of media, devices and services. Semantic frameworks are used in various contexts, such as multimodal representation learning, as proposed by Wang et al. [15]. In their approach, Wang et al. use a deep neural framework to capture the high-level semantic correlations across modalities, which distinguishes this approach from SAM.F.

III. System Concept and Architecture

SAM.F is a smart media environment, which provides a device-independent access to and interaction with media through devices and applications.

The system's architecture of SAM.F is based on a system concept following these three considerations:

1. Web-based access provides platform-independent use of the services and access to media inside SAM.F and its repositories from the users' devices and applications.
2. a service-based modular architecture features extendibility, which provides developers with a framework to develop their own applications, which can be based on or reference to existing services within SAM.F.
3. the concept of Semantic Media regards media independently of their encoding or modality and automatically transcodes or converts media, where necessary and possible, to meet contexts, applications, and devices specifications or criteria.

In the following sections, we focus on the concept of Semantic Media fundamental to SAM.F. We illustrate the system's architecture and the service concept of SAM.F. Following, the application and device-specific media provisioning is outlined. In addition, technical details on the current implementation of SAM.F are given.

A. Semantic Media

In SAM.F, apart from services delivering media, media themselves are central. Semantic Media consist of plain media, such as text, audio, video, pictures, and 3D media, which are enriched by a dynamic set of semantic annotations. Together, plain media and semantic annotations form *Semantic Media* in SAM.F.

The dynamic set of semantic annotations stored in SAM.F for each media element consist of:

- the original Meta-data of the plain media file. For example, for photos taken with digital cameras, metadata usually contains information on the picture's location, and camera data such as camera make and model, or camera settings, such as camera capture settings. This data might be useful for SAM.F services and adding it to the set of annotations improves accessibility and performance when further processing media.

- data received from automated algorithms. Pictures for example are submitted to a Computer Vision algorithm by SAM.F automatically and in a background process in order to determine semantic annotations describing the media's content.
- data received from client applications. As the main user interaction with media through SAM.F is carried out through client applications, in which the context of use is known, this information is stored in additional semantic annotations. This information is collected automatically in a background process through the use of the SAM.F API Web Services, which implicitly reveal the context of use.
- data received from manual user interactions, such as manual annotations or correcting automatic annotations.

It should be noted that the semantic annotations of Semantic Media may not be complete or available for each media element at all times. This is, e.g., due to the context the media is created in, a foreign source the media is accessed from, or incomplete data entered by the user [16].

The set of annotations described above is not final and can be extended in context of client applications, devices or services.

In SAM.F, the complete set of semantic annotations are abstracted into the Data Model (cf. Figure 1) in order to be (i) accessible for all services running inside the framework and (ii) accessible independently of the underlying media repository in the Datastore layer (cf. Figure 1).

Not all annotations are made available for every client application or device through the API Web Services (cf. Figure 1), as the API Client Model only contains those properties that are required in the corresponding context. This way, overhead in the access of media through client applications is assumed to be significantly reduced. The effects on performance or bandwidth have however not been measured as part of this work.

It is one of the hypotheses of this work that the quality of semantic annotations as well as the interconnection of media will be a key issue for realizing appealing scenarios using SAM.F, as, e.g., described in the final Section of this article. An approach to achieve this is to gather additional sematic annotations through automated algorithms. As illustrated in Figure 2 and mentioned above, pictures, for example, are submitted to a Computer Vision algorithm. In the current implementation, SAM.F interfaces with Microsoft Cognitive Services. Thus, in the background, SAM.F computes additional semantic annotations, which are then stored in the internal Datastore (cf. Figure 1 and 2).

B. *System Architecture*

The architecture of SAM.F consists of a layer-based system concept, as illustrated in Figure 1. Client applications and devices utilized by users connect to the SAM.F *API Web Services* through the *API Security Layer* via the Internet in order to access media stored in SAM.F or interact with services in the *Service Layer* (cf. Figure 1).

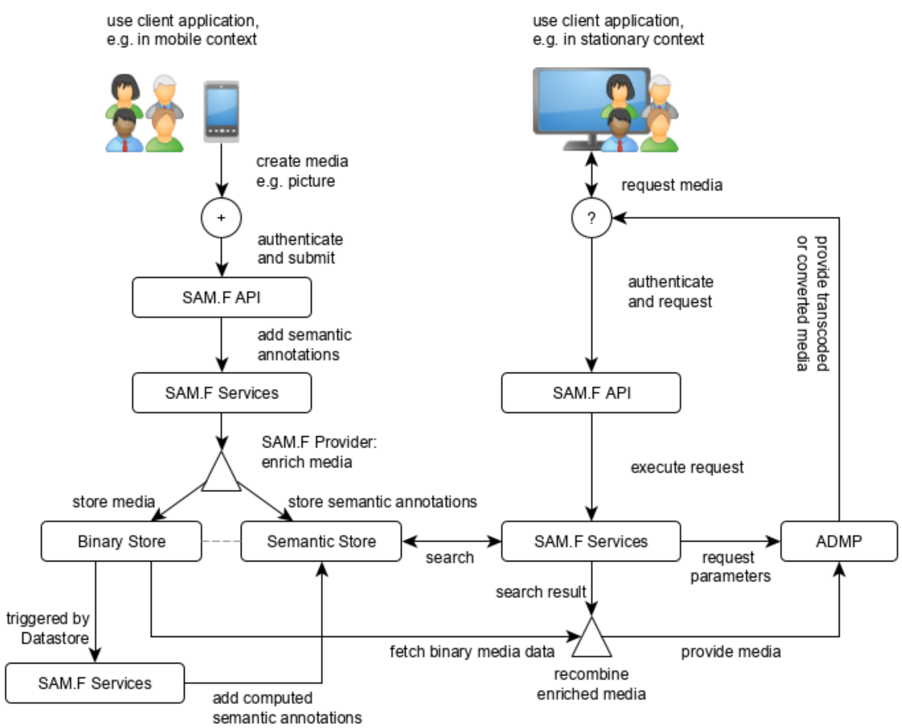


Figure 2: Media creation, enrichment through semantic annotations and retrieval. The Datastore consists of both Binary Store and Semantic Store.

When interacting with SAM.F, client applications as well as devices exchange information with the framework (cf. Figure 2) using a defined data model. Thus, for any context, the *API Client Model* can be extended to exactly match the needs of the application, device, or context, if necessary. API Web Services offer access to dedicated services provided by SAM.F, as the scenario described above outlines. Internally, SAM.F works with a dedicated *Data Model*, as illustrated in Figure 1. Any data is mapped from the *Datastore*, which includes external (semantic) databases as well as binary data stores, to the internal Data Model, which applies a homogenous model to potentially heterogenic data. Thus, SAM.F features the integration of different repositories and provides a combined access to Semantic Media. For simplification purposes, and in order to reduce the learning curve when implementing client applications accessing SAM.F, the internal Data Model is only used in the *Provider Layer*, which contains, e.g., authentication or data providers to be accessed by the upper Service Layer, and in the *Service Layer*, as shown in Figure 1. Any Semantic Media, together with semantic annotations, provided by a service to a client is mapped to the specific API Client Data Model, as outlined above, and being served through the API Web Services and the API Security Layer to the client application (cf. Figure 2).

With the Data Model only used internally, SAM.F accommodates different models used when storing media in digital repositories. A museum database for example differs significantly from, e.g., the DbPedia's semantic database. To be able to use heterogenous sources simultaneously, different data models are homogenized though the Data Model in SAM.F: by applying the data mapping techniques, the framework uses its own model internally, into which all other models are mapped. Applying data mapping in SAM.F produces constant overhead. However, services and applications, as well as their developers, benefit from only working with data models that are specific to the requirements of the services' or applications' context. This also reduces overhead when loading large sets of Semantic Media.

The range of functions of SAM.F is defined by the functionality provided by services residing in the Service Layer, as illustrated in Figure 1. In the scenario outlined below the developers extend SAM.F by implementing a custom service in order to realize the desired functionality. Thus, in the next section, the SAM.F services are regarded.

C. SAM.F Services

Following the implementation principles of SAM.F, a service features a dedicated set of functions in order to provide a certain functionality, e.g., for a use-case or scenario, as outlined above.

Utilizing the Data Model, through the Provider Layer, any service might access Semantic Media from the repositories included in SAM.F's Datastore layer. As a result, services may interchange information in a well-defined context.

SAM.F comes with a set of services that are useful to the developer in a Web-based environment and for developing applications in context of mobile use and the use of Semantic Media, explained in more detail below. In this article, we focus on the basic features the SAM.F services consist of:

- an authentication service to identify and authenticate sessions of applications, devices and users.
- a general media service that allows to retrieve or modify Semantic Media elements for a given keyword in a given general context. Media is retrieved both from the internal datastore, as well as external semantic databases housed in the Datastore layer (cf. Figure 1) and made available through SAM.F.
- the *Application and Device-specific Media Provisioning* (ADMP) service, which transcodes media based on different settings on client retrieval, as outlined below.

In the scenario outlined above, the developers extend the Service Layer of SAM.F (cf. Figure 1) and add their service to authenticate users on public displays. This service utilizes the modularized architecture of SAM.F and interfaces with the adjacent upper and lower layers. It also makes use of the default user authentication service. Service execution may either be triggered (i) on demand per request, or (ii) internally. This allows services of SAM.F to automatically run in the background without the necessity of user interactions.

D. Application and Device-specific Media Provisioning

Semantic Media in SAM.F can contain various types of plain media. However, their use is determined by the client applications. The devices running these applications are usually limited in their capabilities.

To address these challenges, SAM.F offers an *Application and Device-specific Media Provisioning* (ADMP) for any Semantic Media element retrieved through the API Web Services layer (cf. Figure 1).

In general, ADMP transcodes or converts Semantic Media due to specifications given. Trivial examples are the conversion of large photos into thumbnails, including cutting and cropping, if necessary.

ADMP is designed to work in two ways:

- on a per-request basis, in which the application submits the desired parameters (e.g., format, encoding, size, resolution) with every request, or
- on an application or device capability basis. As devices and applications are also represented in the Data Model (cf. Figure 1) of SAM.F, their capabilities are known to SAM.F. Thus, using per-request parameters can be omitted, if application or device capabilities can be generally set or are valid for multiple requests.

Especially in context of the Web-use of SAM.F and the heterogeneity of devices potentially accessing SAM.F, ADMP's usefulness can be illustrated through these examples, in which the correct parameter settings are presupposed: A video can be retrieved in different encodings or in matching screen size for the device's resolution. For example, ADMP can provide just the audio track of the video or just the textual transcript. The transcript can also be used to subtitle the video. More challenging 3D objects, which may not be viewed on any device, can be retrieved as

 ISBN: 978-1-61208-739-9

a video of the 3D object rotating around the y-axis, or just as a picture in the form of a screenshot of the 3D object.

Reviewing key event-based multimedia applications, Tzelepis et al. [17] observe an enormous potential for exploiting new information sources by, e.g., semantically encoding relationships of different informational modalities, such as visual-audio-text. SAM.F provides these means by transcoding and converting Semantic Media in the background by automated processes.

As a side-effect, using ADMP reduces the use of bandwidth, which is of special interest in mobile contexts.

As these examples indicate, this way of provisioning media though SAM.F provides the means for a vast amount of use-cases. However, the author admits that not all possibilities have been implemented. The ADMP module, which also extends the Service Layer (cf. Figure 1), can be expanded, as it features an interface with an extendable list of parameters.

IV. SCENARIO

In context of a research project related to public displays, Josefine Kipke is part of a team developing a solution that features user authentication on public displays [5].

The motivation of this project is to provide access to private and sensitive information or functionalities in cases, where, e.g., the limited capabilities of a smartphone need to be extended, information or functionality is not to be made available on the user’s device, or simply the use of a public display is intended by nature of the context.

This project presents serious challenges, as the user authentication on public displays in general is subject to vulnerability with regard to different sorts of attacks.

The team develops an interaction pattern, in which the users first authenticate themselves on their smartphone, and then enter a graphical code on the public display shown on their smartphone. Afterwards, the users confirm the logon using their smartphone again. Once the session has been confirmed, the users can start using the public display in a private context. Up to this point, the interaction pattern described is just a theoretical approach, which, to the developers, seems to cover the challenges with regard to a secure authentication on public displays. A prototype has yet to be implemented, not to mention to be validated and evaluated, as Josefine ascertains.

This rather simple approach is made possible through the interconnection of the user’s smartphone and the public displays through the Internet and, most importantly, SAM.F.

The technical challenge presents itself in the fact that a public display, in practice, although connected to the Internet, for security and other reasons cannot and should not be remotely controlled by a smartphone that belongs to any random user passing by. These foreign devices are also not accessing the same network as the public display, which implies that any direct connection between a smartphone and a single public display is generally prohibited.

However, the team observes that public displays receive the media displayed from a server or a media framework, such as SAM.F and therefore can connect to SAM.F. Thus, SAM.F is able to identify a single public display through its registered session.

In addition, the team observes that connecting a user’s device to the services of SAM.F poses no additional security issue, as SAM.F has been designed to interconnect devices and applications through services and Semantic Media, as outlined below.

After taking the observations mentioned above into account, the team develops a service for SAM.F using the interfaces provided by the framework. This service handles all the necessary steps to implement the interaction pattern for authentication the team developed. To be able to trigger and manage the authentication from a smartphone, the team also develops an application for smartphones, in this case a Web-based application that accesses SAM.F and the newly implemented service.

The team validates the functionality of their new extension of SAM.F under laboratory conditions. In the next step, they plan to evaluate the system under real conditions and with real users.

In summary, Josefine and the team have utilized the architecture of SAM.F, which interconnects the user’s smartphones and public displays through the Internet, in order to provide an authentication mechanism for public displays. They have successfully extended SAM.F with a new service by implementing the corresponding interfaces.

V. REALIZATION AND DISCUSSION

A first prototype implementation of SAM.F has been realized at the *Kingsbridge Research Center* (KRC). On the basis of a Windows Server system and its Internet Information Services (IIS) Web server, SAM.F is implemented in C# and runs as IIS Web application. Web services are provided using the Active Server Method File (ASMX) technology. Semantic annotations used in SAM.F are represented as RDF triples. For performance reasons analyzed under laboratory conditions in experimental settings, SAM.F’s internally used RDF data is stored in a NoSQL database for performance reasons, although quantitative performance measurements are future work. SAM.F is compatible to semantic media repositories, e.g. using SPARQL to execute queries. Additionally, other required annotations for external media are stored in the internal datastore of SAM.F. In these terms, external media are media that are made available through SAM.F, but are stored in semantic datastores that are not managed by, but connected to SAM.F.

The approach of combining the automated enhancement of semantic annotations for media and delivering media in a device- or context-specific modality or encoding presents a technical novelty and distinguishes SAM.F from other media frameworks or repositories.

The current prototype has been validated under laboratory conditions. Computations are implemented to be carried out in a complexity of *O(n)*. Together with our project partner, as outlined below in more detail, we will integrate SAM.F for use in context of research projects. This will provide the opportunity to evaluate the system under real conditions with regard to functionality and performance.

VI. SUMMARY AND OUTLOOK

With the Semantic Ambient Media Framework (SAM.F), this contribution presents a framework that semantically interconnects (a) *media*, (b) *devices and applications*, and (c) *services*. The practical scenario illustrated describes the use of SAM.F to provide means of a secure method of authentication for public displays [5] by developers. SAM.F provides Web-based access for devices and applications and features a service-based architecture, which allows for interaction with media, such as, e.g., text, pictures, audio, video, or 3D objects. The concept of SAM.F regards Semantic Media independently of their encoding and automatically transcodes or converts media, where necessary and possible, to meet contexts, applications, and devices specifications or criteria. Using SAM.F also solves the problem of media being isolated for use in a single application or on a single device, as SAM.F interconnects users and their devices through its services and Semantic Media.

SAM.F can be used in any context where interaction with Semantic Media is intended. Through technological means, SAM.F especially supports mobile contexts, e.g., through the application and device-specific provisioning of Semantic Media. Thus, SAM.F offers an enormous potential for exploiting new information sources, e.g., by the relationships of different informational modalities encoded semantically. As mentioned above, we have already used SAM.F in context of providing a multi-factor authentication method for public displays [5].

In future work, together with our project partner, the *Society for Audiovisual Archive of German-language Literature* based in the Hanseatic City of Bremen, we will utilize SAM.F as technical foundation to digitally enrich a cultural center for German literature. In this research project, SAM.F will interconnect media from various archives or libraries focusing on German literature. At the cultural center, the physical space will be enriched with digital media served provided by SAM.F. It is our hypothesis that providing meaningful digital content in a body- and space related environment fosters mindful knowledge.

The Kingsbridge Research Center (KRC) is a non-profit research company based in Hamburg, Germany. With our research and the systems, we develop, we have the goal to strengthen the meaningful use of digital technology in public environments, among others. We achieve this through our scientific and project-oriented work in an interdisciplinary team, by additionally achieving research funds through business returns to fund our non-profit activities, and the development of new future-oriented projects. At a time when many are confronting digitization with skepticism and uncertainty, we are committed to communicating security in the mindful use of these technologies.

REFERENCES

[1] A. Whitmore, A. Agarwal, and L. Da Xu, ‘The Internet of Things—A survey of topics and trends’, Inf. Syst. Front., vol. 17, no. 2, pp. 261–274, Apr. 2015.

[2] Eurostat, ‘Internet use by individuals’, 260/2016, Dec. 2016.

[3] S. Vrochidis, B. Huet, E. Y. Chang, and I. Kompatsiaris, Big Data Analytics for Large-Scale Multimedia Search. John Wiley & Sons Ltd, 2019.

[4] C. Vassilakis et al., ‘Interconnecting Objects, Visitors, Sites and (Hi)Stories Across Cultural and Historical Concepts: The CrossCult Project’, in Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection, Cham, 2016, pp. 501–510.

[5] D. Bouck-Standen and J. Kipke, ‘Multi-Factor Authentication for Public Displays using the Semantic Ambient Media Framework’, In Press, ADVCOMP’19, Porto, 2019.

[6] T. Berners-Lee, ‘The Semantic Web’, Sci. Am., pp. 30–37, 2001.

[7] F. Nack, ‘The future in digital media computing is meta’, IEEE Multimed., vol. 11, no. 2, pp. 10–13, 2004.

[8] L. F. Sikos, ‘RDF-powered semantic video annotation tools with concept mapping to Linked Data for next-generation video indexing: a comprehensive review’, Multimed. Tools Appl., vol. 76, no. 12, pp. 14437–14460, Jun. 2017.

[9] C. Bizer, T. Heath, and T. Berners-Lee, ‘Linked data - the story so far’, Int J Semantic Web Inf Syst, vol. 5, no. 3, pp. 1–22, 2009.

[10] D. Bouck-Standen, ‘Construction of an API connecting the Network Environment for Multimedia Objects with Ambient Learning Spaces’, Master Thesis, DOI: 10.13140/RG.2.2.12155.00804, University of Luebeck, Luebeck, Germany, 2016.

[11] D. Bouck-Standen et al., ‘Reconstruction and Web-based Editing of 3D Objects from Photo and Video Footage for Ambient Learning Spaces’, Int. J. Adv. Intell. Syst., vol. 11, Jul. 2018.

[12] K. Blumenstein et al., ‘Bringing Your Own Device into Multi-device Ecologies: A Technical Concept’, in Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces, New York, NY, USA, 2017, pp. 306–311.

[13] P. J. Denning, Ed., The Invisible Future: The Seamless Integration of Technology into Everyday Life. New York, NY, USA: McGraw-Hill, Inc., 2002.

[14] U. Yadav, G. S. Narula, N. Duhan, and B. K. Murthy, ‘An overview of social semantic web framework’, in 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), 2016, pp. 769–773.

[15] C. Wang, H. Yang, and C. Meinel, ‘A deep semantic framework for multimodal representation learning’, Multimed. Tools Appl., vol. 75, no. 15, pp. 9255–9276, Aug. 2016.

[16] P. Oliveira and P. Gomes, ‘Instance-based Probabilistic Reasoning in the Semantic Web’, in Proceedings of the 18th International Conference on World Wide Web, New York, NY, USA, 2009, pp. 1067–1068.

[17] C. Tzelepis et al., ‘Event-based media processing and analysis: A survey of the literature’, Image Vis. Comput., vol. 53, pp. 3–19, 2016.