

MMTC Communications - Frontiers

Vol. 13, No. 6, November 2018

CONTENTS

Message from the MMTC Chair	2
SPECIAL ISSUE ON 5G V2X AND SECURITY	3
<i>Guest Editor: Kuan Zhang</i>	<i>3</i>
<i>University of Nebraska-Lincoln.....</i>	<i>3</i>
<i>Kuan.zhang@unl.edu</i>	<i>3</i>
Standardization and Industry Promotion of C-V2X in China	4
<i>Yuming Ge, Rundong Yu, Cheng Li and Lu Li</i>	<i>4</i>
<i>Institute of Technology and Standards Research,</i>	<i>4</i>
<i>China Academy of Information and Communications Technology.....</i>	<i>4</i>
<i>geyuming@caict.ac.cn, yurundong@caict.ac.cn, licheng@caict.ac.cn,</i>	
<i>lilu@caict.ac.cn</i>	<i>4</i>
Intelligent Network Slicing in 5G for V2X Services	10
<i>Jie Mei¹, Xianbin Wang², Kan Zheng¹</i>	<i>10</i>
¹ <i>Wireless Signal Processing and Network (WSPN) Lab, Key Laboratory of Universal</i>	
<i>Wireless Communication, Ministry of Education, Beijing University of Posts and</i>	
<i>Telecommunications (BUPT).....</i>	<i>10</i>
² <i>Department of Electrical and Computer Engineering, University of Western</i>	
<i>Ontario</i>	<i>10</i>
Machine Learning Enabled Data Preprocessing in Cyber Security Applications.....	15
<i>Shengjie Xu and Yi Qian.....</i>	<i>15</i>
<i>Department of Electrical and Computer Engineering, University of Nebraska-</i>	
<i>Lincoln</i>	<i>15</i>
<i>shengjie.xu@huskers.unl.edu, yqian@unl.edu</i>	<i>15</i>
MMTC OFFICERS (Term 2018 — 2020)	19

Message from the MMTC Chair

Dear MMTC colleagues and friends,

It is my great honor to serve as MMTC Vice Chair for 2018~2020 term. I would like to wish all of you and your families a very Happy New Year 2019 on behalf of all the MMTC officers. May this year bring you all health, happiness, and prosperity.

I would like to take this opportunity to thank our editorial teams of MMTC Frontiers led by Dalei Wu. MMTC Frontiers provides a timely update on recent developments, hot research topics, and society news in the area of multimedia communications. With the great contribution of Dalei Wu, Danda Rawat, Kan Zheng, Melike Erol-Kantarci, Rui Wang and other editors, the Frontier is delivering high-quality publications.

I would also like to take this opportunity to thank those who were able to attend the MMTC meeting at Globecom 2018. The Communication Software, Services and Multimedia Applications Symposium (CSSMA) at Globecom and ICC is sponsored by MMTC and I encourage all of you to continue to be actively involved in CSSMA and submit papers there. The next CSSMAs will be at ICC 2019 in Shanghai, China in May 20-24, 2019 and we hope to see you all here.

The MMTC officers would like to encourage members to be actively involved in the TC as well as help recruit new members. Membership is open to all those who are interested, and more information can be found at the TC website, <http://mmc.committees.comsoc.org>. MMTC provides members the opportunity to actively serve the community by submitting nominations for associate editorship to journals, special issue proposals, conference chairs, and ComSoc distinguished lecturers.

I would like to take this opportunity to invite all of you to attend the MMTC meeting in IEEE ICC 2019 and ICME 2019, both of which will be held in Shanghai in May and July, respectively. I am a main organizer of IEEE ICME 2019. We will review the MMTC activities with recent updates at the meeting. I look forward to seeing you all soon in May and July 2019.

Have a wonderful holiday season. I wish you all the best!



Jun Wu

Vice Chair, Multimedia Communications Technical Committee (2018-2020)
IEEE Communications Society

SPECIAL ISSUE ON 5G V2X AND SECURITY

Guest Editor: Kuan Zhang
University of Nebraska-Lincoln
Kuan.zhang@unl.edu

With the dramatic development of the fifth generation (5G) wireless networks, a new generation of communication technology in vehicular networks named as vehicle-to-everything (V2X) promises to revolutionize the way of intelligent driving. The flourish of various V2X applications has attached a lot of attention from industry and academic sectors in recent years. The special issue follows with interest the latest progresses of V2X and the security challenges towards 5G.

In the first paper, the authors introduce the concepts of V2X and C-V2X in IoV as well as the progress of C-V2X standardization. Moreover, the development of C-V2X and its applications in China are summarized.

The second paper focuses on the design of intelligent network slicing architecture for V2X service provisioning. The intelligent network slicing can exploit virtualized multi-dimension resources to meet the stringent requirements of V2X service.

The third paper proposes a machine learning enabled data preprocessing scheme in cyber security applications. The scheme preserves the critical property of dimensionality reduction for high-dimensional data, and introduces a robust detection boundary to the presence of outliers.

Thanks very much for the receiving three papers from academia and industry laboratories.



Kuan Zhang (S'13-M'17) has been an assistant professor at the Department of Electrical and Computer Engineering, University of Nebraska-Lincoln, USA, since September 2017. He received the B.Sc. degree in Communication Engineering and the M.Sc. degree in Computer Applied Technology from Northeastern University, China, in 2009 and 2011, respectively. He received the Ph.D. degree in Electrical and Computer Engineering from the University of Waterloo, Canada, in 2016. He was also a postdoctoral fellow with the Broadband Communications Research (BBCR) group, Department of Electrical and Computer Engineering, University of Waterloo, Canada, from 2016-2017. His research interests include security and privacy for mobile social networks, e-healthcare systems, cloud/edge computing and cyber physical systems.

Standardization and Industry Promotion of C-V2X in China

Yuming Ge, Rundong Yu, Cheng Li and Lu Li

Institute of Technology and Standards Research,

China Academy of Information and Communications Technology

geyuming@caict.ac.cn, yurundong@caict.ac.cn, licheng@caict.ac.cn, lilu@caict.ac.cn

1. Introduction

V2X (Vehicle to Everything) is a new generation of information and communication technology (ICT) that connects vehicles to everything, for which V stands for vehicle and X stands for everything that interacts with the vehicle, e.g., vehicles, people, roadside infrastructure, and network [1]. The modes of V2X interactions include V2V (Vehicle to Vehicle), V2I (Vehicle to Infrastructure), V2P (Vehicle to Pedestrian), and V2N (Vehicle to Network) [2]. By organizing factors involved in traffic, such as "people, vehicles, roadside infrastructure and cloud platform", V2X allows the vehicle not only to obtain more information than the unlinked vehicles, which stimulates the innovation and application of autonomous driving technology, but also to develop intelligent transportation systems (ITS) which fosters new modes and formats of automotive and transportation services. This is significant for improving traffic efficiency and management, conserving resources, reducing pollution and reducing accident incidence.

C in C-V2X stands for cellular [1], and there are two interfaces for cellular communication: the direct interface (PC5) for short-range communication among vehicles, people and roadside infrastructure and the Uu interface for reliable long-range communication between the terminal and the base station [3][4][5]. C-V2X includes LTE-V2X (Long Term Evolution Vehicle to Everything) and 5G-V2X (Fifth Generation Vehicle to Everything) with technologically achievable evolution from LTE-V2X to 5G-V2X [1][6][7].

Up to now, China has made positive progress in the formulation of C-V2X standards, product development, application demonstration, test and verification, and laid a good foundation for the industrialization of V2X. In terms of standardization, China has completed the construction of LTE-V2X standard system, which includes core standards and specifications. In terms of product development, China has built the world's largest 4G network, and initially formed an industrial chain covering LTE-V2X systems, chips and terminals. In terms of application demonstration, the Ministry of Industry and Information Technology (MIIT) and the Ministry of Transportation actively promote the construction of national demonstration zones and create favorable conditions for cross-industry coordination. In terms of test and verification, the C-V2X Working Group of IMT-2020 5G Promotion Group has completed tests of end-to-end communication function, performance and interoperability of LTE-V2X in laboratory and small-scale outfield environment, which lays the foundation for large-scale application demonstration and commercial deployment.

In this paper, we explain the concepts of V2X and C-V2X in the Internet of Vehicles (IoV) and introduce the progress of C-V2X standardization. Then we reveal the development of C-V2X application demonstration and the progress of C-V2X testing in China. Finally, the tasks of C-V2X Working Group are also introduced.

2. C-V2X Standardization

As an expansion of the LTE platform to new services in the vertical industries, 3GPP has conducted the research and development of standards to provide enhancements for vehicular communication. Currently, the C-V2X standardization can be divided into three phases. The 3GPP R14 specification for LTE-V2X services was officially published in 2017 [8][9], and the 3GPP R15 specification for LTE enhanced V2X (eV2X) services formally completed in June 2018 [10]. The 3GPP R16+ specification for 5G-V2X services has been under study since June 2018 and will complement the LTE-V2X and eV2X specifications [11].

In order to play the role of the top-level design and regulation in the construction of the eco-environment of IoV, MIIT and Standardization Administration of P.R.C jointly developed Guideline for Developing National Internet of Vehicles Industry Standard System, whose sub-volumes are general requirements, intelligent connected vehicle, information communication, electronic products and services. The information communication and intelligent connected vehicle sub-volume clearly defined the technical standard selection of LTE-V2X and 5G-V2X from perspectives of communication technology evolution and intelligent connected vehicle application.

Many China industry associations and standardization organizations have paid enough attention to the promotion of C-V2X standards in China. China Communications Standards Association (CCSA), Telematics Industry Application Alliance (TIAA), China Society of Automotive Engineers (China-SAE), the Technical Committee for National on ITS (TC/ITS), China ITS Industry Alliance (C-ITS), and China Industry Innovation Alliance for the Intelligent and Connected Vehicles (CAICV) have actively carried out C-V2X related research and standardization work. As shown

IEEE COMSOC MMTc Communications - Frontiers

in Figure 1, a standard system that covers all the layers of the C-V2X standard protocol stack has been already initially established.

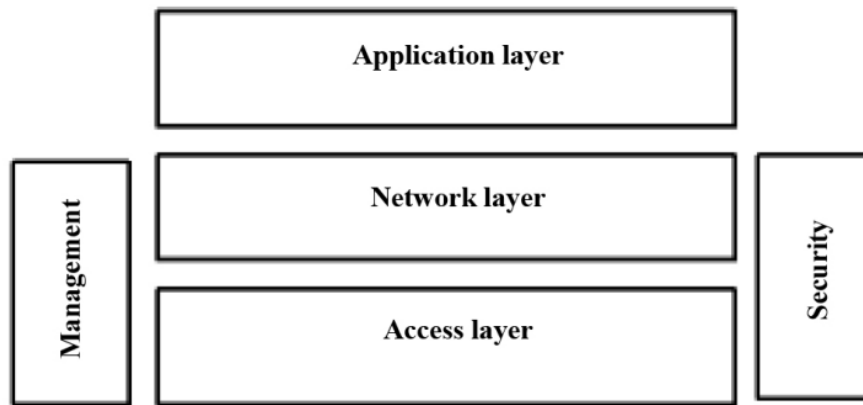


Figure 1: China's C-V2X standard system

The relevant standardization work of China's standard organizations has supported China's system of C-V2X standards, covering application definitions and requirements, general technical requirements, key technologies, and information security. Industry associations and standardization organizations in the fields of information communication, transportation and automobile have been actively engaged in the formulation of technical standards at all levels of the LTE-V2X standard stack, which include overall technical requirements, air interface technical requirements, application message set, information security and so on, and can guide upstream and downstream enterprises to carry out related product researches and developments. However, the majority of these standards has been studied and developed by different groups of organizations or industry standardization committees and still need coordination to accelerate the formation of full-fledged unified national standards.

Therefore, there is a greater synergy among national cross-industry standards associations recently. In Nov 2018, the National Technical Committee of Auto Standardization (SAC/TC 114), the National Technical Committee 268 on Intelligent Transport Systems of Standardization Administration of China (SAC/TC 268), the National Technical Committee 485 on Communication of Standardization Administration of China (SAC/TC 485) and the National Technical Committee on Road Traffic Management of Standardization Administration of China have jointly signed the "Framework Agreement on strengthening Cooperation in C-V2X Standards for Automobile, Intelligent Transportation, Communication and Traffic Management". Four parties will establish an efficient and smooth communication mechanism, support each other and participate in the development of standards, and jointly promote the application of C-V2X and other new-generation information and communication technologies in automobiles, intelligent transportation and traffic management.

In order to further promote synergy and cooperation among the various standardization organizations, this agreement points out: 1) The relevant C-V2X standards formulated by SAC/TC 485 and CCSA should fully consider the practical application requirements of the automobile, transportation and traffic management industries, and provide the support for the communication and interconnection between vehicles and related facilities; 2) To promote the application of C-V2X technology in various scenarios of automobiles The contents related to communications in the relevant standards formulated by SAC/TC 114 should adopt the communication industry standards and consider the compatibility with standards of transportation and traffic management facilities; 3) To promote the application of C-V2X technology in transportation infrastructure, the transportation infrastructure standards formulated by SAC/TC 268 should consider the compatibility with the standards of communication, automobile and traffic management facilities; 4) In order to promote the application of C-V2X technology in traffic management, the standards of traffic management facilities formulated by the Committee on Road Traffic Management of Standardization should adopt standards in the field of communication industry and consider the compatibility with standards of automobile and transportation infrastructure.

3. Development of C-V2X Industry

3-A: Development and Application demonstration of C-V2X

In order to support the development of the Intelligent Connected Vehicle (ICV) industry in China, on November 13, 2018, Radio Administration of MIIT issued "Regulations for the use of the 5905-5925MHz frequency band in direct-connected communications on the internet of vehicles (ICV) (provisional)", which plan 20MHz bandwidth in 5905-5925MHz band for using the LTE-V2X direct connection communication technology of ICV. The managements of related frequency, station, equipment and interference coordination is also regulated.

In the aspect of LTE-V2X communication chip and terminal product development, Datang has officially released LTE-V2X commercial PC5 Mode 4 mode communication module DMD31, RSU (DTVL3000-RSU) and OBU (DTVL3000-OBU) products; Huawei released Balong 765, a multimode 4.5G LTE modulation and demodulation chip including LTE-V2X, during MWC 2018, and released the first commercial C-V2X RSU during MWC in Shanghai in June 2018. In addition, Qualcomm, an international chipmaker, expects to make commercial 9150 c-v2x chipsets in the second half of 2018. Meanwhile, based on these commercial modules and chips, some equipment manufacturers have been able to provide OBU and RSU such as Huawei, Datang, Nebula link, Wanji, Jinyi, Savari and China Mobile.

In the aspect of LTE-V2X application demonstration, MIIT signed the demonstration cooperation framework of "Intelligent vehicle and intelligent traffic application demonstration based on broadband mobile Internet" with Beijing, Chongqing, Zhejiang, Jilin and Hubei, carried out the "national intelligent transportation comprehensive test base co-construction and cooperation" project in Wuxi with the ministry of public security and the government of Jiangsu province, supported the Shanghai international automobile city to establish the "national intelligent network vehicle (Shanghai) pilot demonstration project", promoted the formation of the "5+2" vehicle networking demonstration area pattern, and supported the establishment of the demonstration base in Sichuan Province through the Sino-German project of pilot demonstration about ICV, Internet of Vehicle (IOV) Standard, Testing and Verification. The Wuxi LTE-V2X city-level demonstration application project was officially launched in May 2018. Which Marks the world's first city-level vehicle-road coordination platform has entered the full implementation stage. In September 2018, Open road demonstration area for 240 intersections and 5 viaducts of LTE-V2X covering Wuxi Old City, Taihu New City, high-speed rail station, airport, Xuelang test ground were completed.

3-B: C-V2X Test and Evaluation

LTE-V2X test and validation is implemented in three phases of "LTE-V2X Communications Product and solution Development-- End-to-End Verification of typical V2X Application --Large-scale Test and Verification." Up to now, the first two phases of testing and validation had been completed basically.

In the aspect of laboratory testing, by simulating various traffic scenarios and V2V wireless communication environment, the basic functions of LTE-V2X, end-to-end communication delay, packet deliver rate and other performance indicators were tested. And the interoperation between devices of different manufacturers based on 3GPP R14 international standard is realized. In addition, China Institute of Information and Communication Technology (CAICT), together with Neusoft and other partners, has built a conformance testing platform for LTE-V2X communication protocols to support automated conformance testing of protocols at the network and application levels, provided complete test interface control (TCI) specification and technical documentation for the vendor under test. In September-October 2018, IMT-2020 (5G) Promotion Group (PG) C-V2X working Group (WG) organized and completed the network layer and application layer interoperability and protocol conformance testing with nine equipment manufacturers of Huawei, Datang, Wanji, Jinyi, Nebula Link, Neusoft, Savari, Qianfang, and Huali-tec in the Laboratory of CAICT, which is according to standards of "Cooperative intelligent transportation system – Dedicated short range communications – Part 3: Technical requirements for the network layer and application layer", "Test Specification for Network layer Conformance of LTE-V2X Terminal (Labs)" and "Test Specification for Application layer Conformance of LTE-V2X Terminal (Labs)". And in mid-October, Datang and Huawei jointly completed the 3GPP R14-based chip-level direct communication (PC5) interoperability test in the laboratory of CAICT.

In the aspect of outfield testing, the LTE-V2X function, performance test of typical V2X application verification about safety warning have been carried out in the closed test field of Chongqing and Shanghai demonstration areas, and the inter-operation testing between LTE-V2X devices of different manufacturers under the real vehicle condition is also realized. Laboratory and field tests verify that the realization of LTE-V2X communication, coverage distance, end-to-end delay, packet deliver reliability and other performance indicators can meet the V2V communication requirements of typical scenarios.

In addition, during the annual meeting of China-SAE in November 2018, CAICV, C-V2X WG of IMT-2020 (5G) PG and Shanghai International Automobile City jointly held the LTE-V2X "Three-Layers" Interoperability V2X Application Demonstration based on Chinese standards. Which achieves the world first cross-communication module (chip), cross-terminal providers, cross-vehicle manufacturers interoperability demonstration.

4. C-V2X Working Group

Although C-V2X technology is developing rapidly in recent years, due to its outstanding cross-industry characteristics, it is still facing some challenges in its commercial promotion. Neither C-V2X business models nor network deployment plans are clear now. Due to its industry chain, C-V2X business models is different from the traditional IOV business models. There is not yet a strong leader arising from many manufacturers to unify C-V2X network deployment plan. To solve these problems, promote cross-sector collaboration and form the core cohesiveness of the entire industry, the C-V2X WG was established by IMT-2020 (5G) PG.

IMT-2020 (5G) PG jointly established by three ministries in China (the Ministry of Industry and Information Technology, the National Development and Reform Commission and the Ministry of Science and Technology) in February 2013, based on the original IMT-Advanced Promotion Group. It is the major platform to promote the research of 5G in China. Its members include the leading operators, vendors, universities, and research institutes in the field of mobile communications.

The C-V2X WG was established in June 2017 to study V2X key technologies, carry out V2X trial, and conduct industrial and application promotion. Now it has 106 members including OMEs, supplier, chips companies, research institutes etc. Now C-V2X WG organizes to started research on spectrum, testing architecture, demonstration, security, Multi-access Edge Computing (MEC), business model and high precision positioning based on C-V2X. It also promote C-V2X international cooperation with 5GAA, ETSI etc.



Figure 2: Members of C-V2X WG

The C-V2X WG mainly carried out seven tasks in 2018:

- 1) supporting the radio administration of MIIT to complete the planning and allocation of dedicated frequency for IOV in China which we mentioned above;
- 2) promoting the establishment of C-V2X test, certification and evaluation architecture, which carry out specific research on test environment and test standard in three aspects of communication module, parts and vehicle, and writing research report of test, certification and evaluation;
- 3) promoting the LTE-V2X large-scale test and application demonstration, supporting C-V2X application scenarios testing, exploring the V2X testing solutions, and providing network and application layer conformance testing services;
- 4) studying and verifying the LTE-V2X security certification mechanism, carrying out research on security risks and requirements, sorting out the security mechanisms, and writing "White Paper on V2X Security Technology based on LTE";
- 5) carrying out the research on the fusion of C-V2X and MEC, studying the application scenarios, key technology and networking scheme of the integration of MEC and vehicle networking, and working out the research report on "Integration of C-V2X and MEC";
- 6) strengthening the C-V2X application promotion and business model research, combing out the 5G-V2X applications and different types of service, and forming the business model white paper;
- 7) carrying out the research of high-precision positioning enhancement technology, studying the positioning requirements of C-V2X scenarios, positioning technology, solutions of C-V2X high-precision positioning technology.

5. Conclusion

This paper explains the concepts of V2X and C-V2X in IOV, introduces the progress of C-V2X standardization, the

development of C-V2X and its application in China, and the progress of C-V2X testing and evaluation. Finally, the tasks of C-V2X Working Group are also introduced. V2X allows the vehicle not only to obtain more information than the unlinked vehicles, which stimulates the innovation and application of autonomous driving technology, but also to develop ITS which fosters new modes and formats of automotive and transportation services. This is significant for improving traffic efficiency and management, conserving resources, reducing pollution and reducing accident incidence. C-V2X-based IOV sets a model for cross-industry and cross-sector integration and forms an important development area for automotive, communications and transportation industries. However, there are still many problems to be solved in the current industrial situation, including technologies, regulations, policies, and business models. Therefore, all shareholders should work together to promote industrial synergy and accelerate C-V2X development and landing.

Acknowledgement

This work was subsidized by the project of Research on Key Technologies and Social Legal Problems of Autonomous Electric Vehicles, project ID: 2018YFB0105202.

References

- [1] IMT-2020 (5G) Promotion Group, (Aug. 2018). C-V2X White Paper. [Online]. Available: <http://www.imt-2020.org.cn/zh/documents/1>
- [2] Chen, Shanzhi, et al. "Vehicle-to-everything (v2x) services supported by LTE-based systems and 5G." IEEE Communications Standards Magazine 1.2 (2017): 70-76.
- [3] 3GPP TS 23.303 V15.0.0, Technical Specification Group Services and System Aspects; Proximity-based services (ProSe); Stage 2 (Release 15), June 2017.
- [4] 3GPP TS 36.300 V15.0.0, Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2 (Release 15), December. 2017.
- [5] Uhlemann, Elisabeth. "Initial steps toward a cellular vehicle-to-everything standard [connected vehicles]." IEEE Vehicular Technology Magazine 12.1 (2017): 14-19.
- [6] Chen, Shanzhi, et al. "LTE-V: A TD-LTE-based V2X solution for future vehicular network." IEEE Internet of Things journal 3.6 (2016): 997-1005.
- [7] Campolo, Claudia, et al. "5G network slicing for vehicle-to-everything services." IEEE Wireless Communications 24.6 (2017): 38-45.
- [8] 3GPP TR 22.185 V14.3.0, Technical Specification Group Services and System Aspects; Service requirements for V2X services; Stage 1 (Release 14), March 2017.
- [9] 3GPP TR 23.285 V14.3.0, Technical Specification Group Services and System Aspects; Architecture enhancements for V2X services (Release 14), June 2017.
- [10] 3GPP TR 22.886 V15.2.0, Technical Specification Group Services and System Aspects; Study on enhancement of 3GPP Support to 5G V2X Services (Release 15), June 2018.
- [11] 3GPP TR 22.186 V16.0.0, Technical Specification Group Services and System Aspects; Enhancement of 3GPP support for V2X scenarios; Stage 1 (Release 16), September 2018.



Yuming Ge received his Ph.D. degree in Institute of Computing Technology, Chinese Academy in 2014. He is interested in topics related to vehicular networks. He is currently senior engineer at China Academy of Information and Communications Technology. mainly focuses on standardization, innovation research and international cooperation in the area of Connected and Automated Vehicles and Industrial Internet. He is the Leader C-V2X WG of IMT-2020 (5G) PG, the Co-Chair of Automotive TG of IIC, the Chairman of International and LWG of AII, the leader of C-V2X WG of CCSA TC10, and the committee member of SAC/TC114/SC34.



Rundong Yu received his master's degree in Electrical and Computer Engineering from Worcester Polytechnic Institute in 2015. He is interested in topics related to vehicular networks. Currently he is engineer in China Academy of Information and Communications Technology.



Cheng Li received his MS. degree in Computer Science from Beijing University of Posts and Telecommunications in 2013. He is interested in topics related to security of vehicular networks. Currently he is a senior engineer in China Academy of Information and Communications Technology.



Lu Li received her bachelor's degree in information security in 2007. Currently she is an engineer in China Academy of Information and Communications Technology.

Intelligent Network Slicing in 5G for V2X Services

Jie Mei¹, Xianbin Wang², Kan Zheng¹

¹Wireless Signal Processing and Network (WSPN) Lab, Key Laboratory of Universal Wireless Communication, Ministry of Education, Beijing University of Posts and Telecommunications (BUPT)

²Department of Electrical and Computer Engineering, University of Western Ontario
meijie.wspn@bupt.edu.cn, xianbin.wang@uwo.ca, zkan@bupt.edu.cn

1. Introduction

The dramatic development of intelligent driving has brought numerous vehicle-to-everything (V2X) applications, including both safety related service and entertainment services, which directly generate a wide spectrum of quality of service (QoS) requirements in vehicular networks [1]. Provisioning of V2X services require diverse data rate, reliability and latency from the vehicular networks [2], [3]. For instance, autonomous driving requires a communication latency below a few milliseconds, and a reliability of close to 100%. In comparison, entertainment services mainly focus on high data and thus can tolerate much longer latency and low reliability. Although the Long Term Evolution (LTE) technology has incorporated V2X communication and its enhanced version (eV2X) [4], current LTE networks still cannot support diverse QoS requirements of V2X services effectively with its “one size fits all” architecture.

The forthcoming fifth generation (5G) wireless networks is expected to meet various QoS requirements for V2X services. To achieve this goal, 5G wireless network is intended to adopt many emerging technologies. Apart from new radio access technologies (e.g., non-orthogonal multiple access (NOMA), massive multiple-input-multiple-output (MIMO) [5]), network softwarization and intelligentization are the most distinctive and critical aspects of the 5G network for V2X service provisioning.

Compared with existing rigid network architecture, network softwarization aims at providing highly flexible and programmable end-to-end communication. Through software-defined networking (SDN) and network functions virtualization (NFV), network functions and resources are not restricted to the dedicated physical network infrastructures. By using network softwarization, mobile network operator can allow service providers customize their own logical (virtual) networks, which can better guarantee the demand QoS of services, that is, “everything as a service” or “X as a service (XaaS)” [6], [7]. These logical networks are referred to as network slices, which can be defined as a collection of core network (CN) and radio access network (RAN) functions whose settings are configured to meet the specific requirements of services [8]-[10]. Therefore, with network slicing, operator can flexibly compose slices for guaranteeing the diverse requirements of different V2X services. Furthermore, it is also beneficial to implement the network slicing in the vehicular networks in achieving improved network agility and re-configurability.

Due to the complexity and diverse operational conditions of future networks, precise modelling of network environment and timely decision-making constitute the two most critical challenges of 5G [11]. To cope with these challenges, network intelligentization is proposed due to recent success of advanced Artificial Intelligence (AI) / machine learning (ML) technologies (e.g., deep learning applied in network operation). Compared to traditional mathematical-oriented system/network design methods, AI/ML techniques can automatically extract dynamics of network and make decisions based on historical observations, without the need of making unrealistic modelling of system/network (e.g., data following Gaussian distribution) [12]. Furthermore, the advanced ML/AI can be applied for different domains, ranging from network design, mobile data analysis, and network management. Therefore, it is necessary to utilize the advanced ML/AI to facilitate the deployment and management of wireless network, in order to realize automation and self-evolution of 5G network.

Motivated by these two trends in 5G networks, we believe it is essential to utilize the network intelligentization in realizing the automated network slicing for V2X service provisioning. By exploiting the advantage of the advanced AI/ML, we can intelligently extract high-level patterns of network slice that has complex structure and inner correlations. However, this is difficult to achieve through the traditional prediction-optimization approaches. Therefore, the scope of this short paper is thus to bridge the gap between network softwarization and network intelligentization, by presenting the preliminary study on the AI/ML based 5G network slice architecture for V2X services in the vehicular network.

2. Diverse Requirements of V2X Services

To improve transportation efficiency, safety, and comfort on the road, V2X services have various and complex application scenarios, ranging from enhancing real-time safety and efficiency on road, to autonomous driving, to HD

video streaming played for passengers in the vehicle. Based on the diverse service requirements, V2X services can be categorized into following three main types as shown in Figure 1. To support V2X services with versatile communication type, V2X communication is proposed and classified into four patterns, i.e., vehicle-to-pedestrian (V2P), vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I) and vehicle-to-network (V2N) communication.

- **Traffic safety and efficiency service:** It aims at reducing the possibility of traffic accidents and improvement of traffic efficiency. For instance, in order to avoid the accidents generated by blind spot, vehicles broadcast messages to surrounding vehicles via V2V communications. The periodic and event-driven messages carry kinematics parameters of the transmitting vehicle to allow other vehicles and pedestrians to sense the surrounding environment. Low latency and high reliability are considered as highly demanding requirements for this kind of services.
- **Vehicular internet and infotainment service:** The main objective here is to enable a more comfortable infotainment experience via V2N communication. Web browsing, social media access, and HD video streaming can provide on-demand entertainment information to drivers and passengers. This kind of services would become even more ordinary with the more widespread of automobile driving, where the driver needs media consumption.
- **Autonomous driving related service:** In principle, automated driving is possible without V2X communication. However, the sensing ability of a single autonomous driving vehicle (ADV) is limited. With the aid of cooperative awareness messages between other vehicles or infrastructures, the driving reliability of ADV can be significantly improved. For instance, ADV can also better adapt to the traffic situation by forming vehicle platoon and ADV can also download high-resolution digital maps by V2N communication. Compared to safety related service, ADV related service requires ultra-low latency, ultra-high reliability and high data rate requirements.

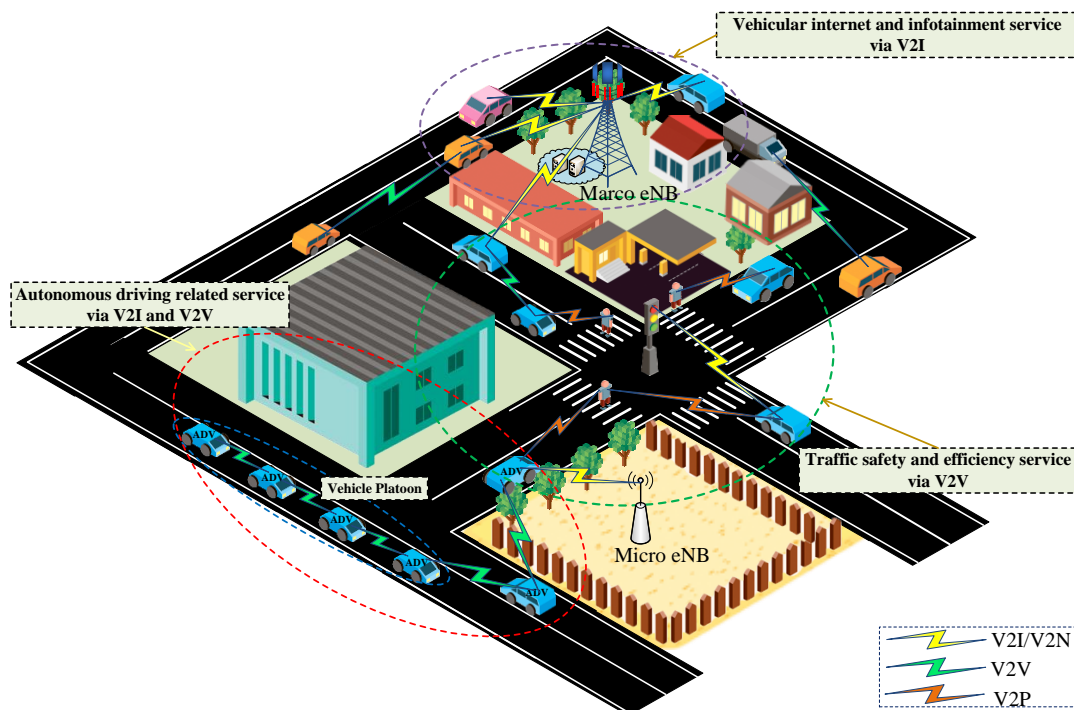


Figure 1: Illustration of three typical V2X services in the vehicular network.

As mentioned before, network slicing can effectively cope with V2X services with diverse demands provided over the 5G wireless network. However, network slicing will introduce more complexity into already too complicated network, which make traditional prediction-optimization based network management schemes ineffective. Therefore, in the next section, we will elaborate the opportunity of using advanced AI/ML technologies for autonomous and effective of network slicing.

3. Intelligent Network Slicing Architecture for V2X Services

In this part, we discuss how to design an intelligent network slicing architecture with high flexibility. Firstly, to provide a centralized controlling environment for network slicing, we present the cloud-based framework with the SDN technology for the vehicular network. Thanks to the centralization of this cloud-based framework, the intelligent

network slicing architecture can conveniently be implemented for V2X services.

3-A: Cloud-based Framework for Vehicular Network

In order to perform a centralized control of network slicing, a cloud-based framework with SDN technology is presented by integrating various network infrastructures, in which important functions of the control plane are provided by the core cloud and user plane functions are shifted to the edge cloud and remote cloud.

Core Cloud: To reduce the burden of backhaul, control signaling and delays of data transmissions, the core network in the traditional centralized network has evolved into the core cloud, in which the control plane and the user plane are split and decoupled at the gateway.

Edge Cloud: The edge cloud, which hosts service entities in close physical proximity for vehicular user equipments (VUEs), can ensure the low end-to-end response time of V2X services. The service area of the edge cloud is defined as a basic geographical unit, which ranges from a single macrocell to a cluster of macrocells. The edge cloud physically consists of the remote radio heads (RRHs), computation infrastructures and storage infrastructures. Then, all the network functions of RAN can be run by the computation infrastructures.

Remote Cloud: For a certain edge cloud, there are surplus radio, computing and storage resources in its neighboring edge clouds. These resources can be defined as the remote cloud. When the capability of this edge cloud cannot meet the QoS requirements of V2X services, offloading the service requests of VUEs in its service area to the remote cloud becomes a good alternative. However, to access the remote cloud, it needs not only wireless links but also wired links, which will increase the latency of V2X services.

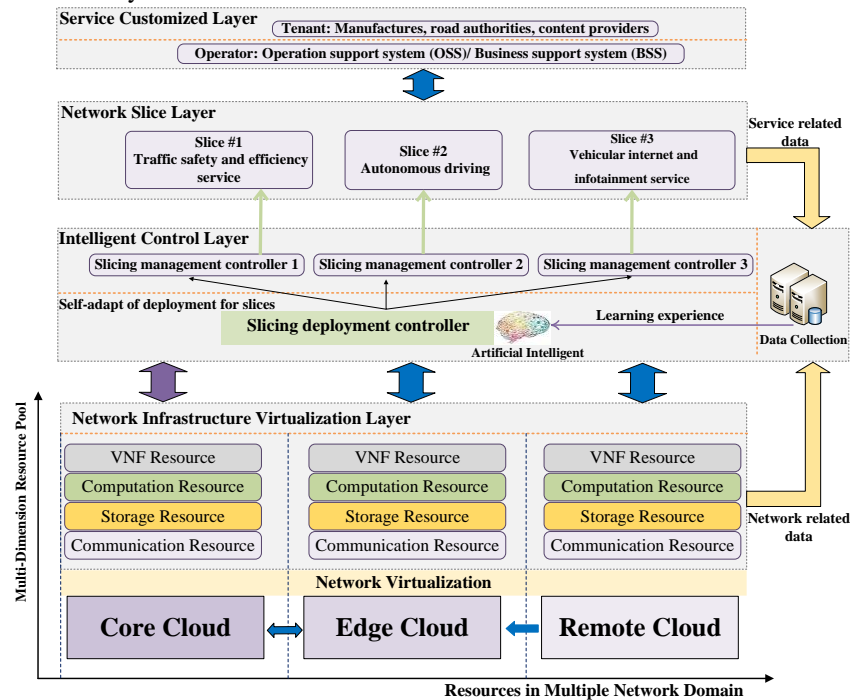


Figure 2: Architecture of intelligent network slicing for V2X services

3-B: Design of Intelligent Network Slicing Architecture

The centralized cloud-based framework provides the feasibility to apply the network slicing and AI/ML based intelligent control for diverse V2X services. Then, as shown in Figure 2, the intelligent network slicing architecture is proposed and divided as four layers: the network infrastructure layer, intelligent control layer, network slice layer and service layer. The network infrastructure layer decouples kinds of resources from dedicated hardware of network without interoperability and compatibility issues, which makes network slicing more flexible and programmable [8]. To deal with the increased complexity in network slicing, intelligent control layer with the AI/ML technologies can provide robust capabilities of orchestrating and managing of network slices.

1). **Network Infrastructure Virtualization Layer:** This layer is located at the bottom of this architecture, which aims at virtualizing different kinds of resources in multiple network domains (i.e., core, edge and remote cloud). As shown in the bottom of Figure 2, the resources in each network domain are mapped to the multi-dimension resource pool. According to its capability and ability, in this pool, the resource is further divided into four dimensions: communication, computing, storage and VNF resource.

- 2). **Intelligent Control layer:** This layer collects mobile data, save it in database, and updated as well. Based on these data, the intelligent control layer exhibits control behaviors by allocating, managing, and releasing the resources in multiple network domains for network slices. To perform the deployment and management of network slices fast and efficiently, we propose a hierarchical control layer, which consists of the slicing deployment controller and the slicing management controller. The former is in charge of network slicing deployment based on a global view and historical observations of the vehicular network, while the latter one directly manage multi-dimension resources in each slice.
- 3). **Network slice layer:** After network virtualization and intelligentization of vehicular network, network slice can be defined as an appropriate collection of multi-dimension resources in the multiple network domains. Then, based on the QoS requirements of the typical V2X services stated in Table 1, we propose the following set of slices.
- 4). **Service customized layer:** The service customized layer is the top layer in this architecture, which captures the Qos requirements of a given service level agreement (SLA) as agreed by V2X service provider. Thus, mobile operator can support on-demand network slices for the V2X services.

4. Conclusion

Intelligent network slicing is a promising direction in meeting the diverse QoS requirements of V2X services. Multi-dimension resources can be virtualized and pooled, which can be exploited by network slicing to meet the stringent requirements of V2X services. In this short paper, we have presented the design of intelligent network slicing architecture for V2X service provisioning. A new intelligent control layer is proposed to achieve automated deployment of network slicing based on historical observation of the vehicular network. Although we have presented a preliminary study, many more future studies on intelligent network slicing are needed in order to fulfill the critical role of V2X in the anticipated autonomous driving era.

References

- [12] S. Chen et al., "Vehicle-to-Everything (V2X) Services Supported by LTE-Based Systems and 5G," in *IEEE Communications Standards Magazine*, vol. 1, no. 2, pp. 70-76, 2017.
- [13] 3GPP, "Study on LTE Support for Vehicle to Everything (V2X) Services," document TR 22.885 V14.0.0, TSG service area, 3GPP, 2015.
- [14] K. Zheng, Q. Zheng, P. Chatzimisios, W. Xiang and Y. Zhou, "Heterogeneous Vehicular Networking: A Survey on Architecture, Challenges, and Solutions," in *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 2377-2396, Fourth quarter 2015.
- [15] 3GPP, "Technical Specification Group Services and System Aspects; Study on Architecture Enhancements for LTE Support of V2X Services (Release 14)," TR 23.785 v. 1.1.0, July 2016.
- [16] B. Di, L. Song, Y. Li and Z. Han, "V2X Meets NOMA: Non-Orthogonal Multiple Access for 5G-Enabled Vehicular Networks," in *IEEE Wireless Communications*, vol. 24, no. 6, pp. 14-21, Dec. 2017.
- [17] NGMN-Alliance, "Description of network slicing concept," NGMN 5G, vol. 1, 2016.
- [18] K. Zheng, L. Hou, H. Meng, Q. Zheng, N. Lu and L. Lei, "Soft-defined heterogeneous vehicular network: architecture and challenges," in *IEEE Network*, vol. 30, no. 4, pp. 72-80, July-August 2016.
- [19] C. Campolo, A. Molinaro, A. Iera and F. Menichella, "5G Network Slicing for V2X Services," in *IEEE Wireless Communications*, vol. 24, no. 6, pp. 38-45, Dec. 2017.
- [20] K. Katsalis, N. Nikaein, E. Schiller, A. Ksentini and T. Braun, "Network Slices toward 5G Communications: Slicing the LTE Network," in *IEEE Communications Magazine*, vol. 55, no. 8, pp. 146-154, 2017.
- [21] J. Ordonez-Lucena, P. Ameigeiras, D. Lopez, J. J. Ramos-Munoz, J. Lorca and J. Folgueira, "Network Slicing for 5G with SDN/NFV: Concepts, Architectures, and Challenges," in *IEEE Communications Magazine*, vol. 55, no. 5, pp. 80-87, May 2017.
- [22] R. Li et al., "Intelligent 5G: When Cellular Networks Meet Artificial Intelligence," in *IEEE Wireless Communications*, vol. 24, no. 5, pp. 175-183, October 2017.
- [23] C. Zhang, P. Paul, and H. Hamed. "Deep Learning in Mobile and Wireless Networking: A Survey." arXiv preprint arXiv:1803.04311, Mar. 2018.

IEEE COMSOC MMTC Communications - Frontiers



Jie Mei received his B.S. degree from Nanjing University of Posts and Telecommunications (NJUPT), China, in 2013. Since then, he has been working toward a Ph.D. degree at Beijing University of Posts and Telecommunications (BUPT). His research interests include radio resource management in wireless network and Vehicle-to-Everything (V2X) communication.



Xianbin Wang (S'98-M'99-SM'06-F'17) is a Professor and Tier-I Canada Research Chair at Western University, Canada. He received his Ph.D. degree in electrical and computer engineering from National University of Singapore in 2001.

Prior to joining Western, he was with Communications Research Centre Canada (CRC) as a Research Scientist/Senior Research Scientist between July 2002 and Dec. 2007. From Jan. 2001 to July 2002, he was a system designer at STMicroelectronics, where he was responsible for the system design of DSL and Gigabit Ethernet chipsets. His current research interests include 5G technologies, Internet-of-Things, communications security, machine learning and locationing technologies. Dr. Wang has over 300 peer-reviewed journal and conference papers, in addition to 26 granted and pending patents and several standard contributions.

Dr. Wang is a Fellow of Canadian Academy of Engineering, a Fellow of IEEE and an IEEE Distinguished Lecturer. He has received many awards and recognitions, including Canada Research Chair, CRC President's Excellence Award, Canadian Federal Government Public Service Award, Ontario Early Researcher Award and five IEEE Best Paper Awards. He currently serves as an Editor/Associate Editor for IEEE Transactions on Communications, IEEE Transactions on Broadcasting, and IEEE Transactions on Vehicular Technology and He was also an Associate Editor for IEEE Transactions on Wireless Communications between 2007 and 2011, and IEEE Wireless Communications Letters between 2011 and 2016. Dr. Wang was involved in many IEEE conferences including GLOBECOM, ICC, VTC, PIMRC, WCNC and CWIT, in different roles such as symposium chair, tutorial instructor, track chair, session chair and TPC co-chair.



Kan Zheng (S'02-M'06-SM'09) received the B.S., M.S., and Ph.D. degrees from Beijing University of Posts and Telecommunications (BUPT), China, in 1996, 2000, and 2005, respectively, where he is currently a Professor. He has rich experiences on the research and standardization of the new emerging technologies. He is the author of more than 200 journal articles and conference papers in the field of wireless networks, IoT, vehicular communication, and so on. He holds editorial board positions for several journals and has organized several special issues, including IEEE COMMUNICATIONS SURVEYS & TUTORIALS, IEEE Communication Magazine, and IEEE SYSTEM JOURNAL.

Machine Learning Enabled Data Preprocessing in Cyber Security Applications

Shengjie Xu and Yi Qian

Department of Electrical and Computer Engineering, University of Nebraska-Lincoln

shengjie.xu@huskers.unl.edu, yqian@unl.edu

1. Introduction

The growing popularity and development of big data applications are bringing serious threats to the availability of critical online systems [1]. Among the massive big data applications that have been adopted and processed all over the world, threats based on anomalous behaviors have never been more outrageous. An Intrusion Detection System (IDS) is a complicated computing system tool that provides monitoring on malicious behaviors and detects violations on a specific operation. Most recently, network-based IDSs start to deal with big data, which also exhibit unique characteristics as compared to traditional data. Big data is unstructured with various types, and retains a need of real-time analysis [2]. This development calls for new IDS protocols and architectures for data acquisition, real-time data learning and large-scale data processing mechanisms.

A general flow of data learning process in cyber security applications is illustrated in Fig. 1. One simple and classical approach to speed up data learning process is to apply principle component analysis (PCA) algorithm [3], which conducts a dimensionality reduction process on major high-dimensional data sets. When it comes to big data applications, PCA is also a great approach to perform for the sake of less memory consumption, less disk space occupancy, and faster processing on statistical learning algorithms. However, as pointed out by [4], classical PCA algorithm suffers from a list of limitations. One of which states that PCA may be sensitive to the presence of outliers remained in the data set.

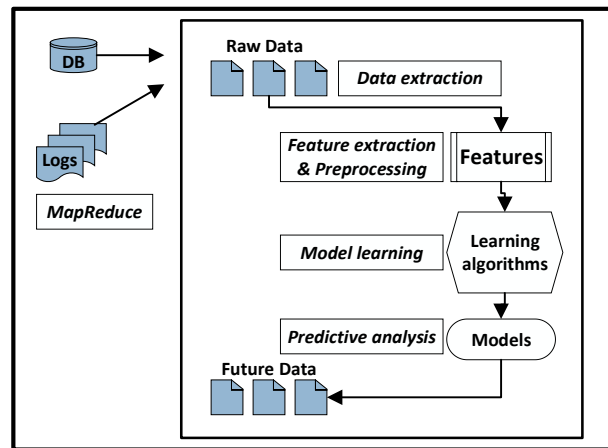


Figure 1 Flow of data learning process in cyber security applications

To fill in such gap, a new approach to detect anomalies based on a robust preprocessing scheme for cyber security applications is proposed. In this work, we aim to introduce a machine learning enabled data preprocessing scheme. Specifically, we make extensive use of robust processing to build robust principal components (PCs), and utilize real-time data learning to speed up the process.

The remainder of this article is organized as follows. Section II covers the preliminaries and related work of our proposed scheme. Section III describes the proposed data processing scheme, with the focus on the methodologies of building robust PCs, reducing dimensionality and outliers removal. Section IV presents the discussions of the proposed scheme. Section V presents a case study by applying the proposed scheme. Section VI concludes with our current work.

2. Preliminaries and related work

- A. PCA: PCA is mainly adopted as a dimensionality reduction technique. In real-world applications, human ability to visualize data is usually limited to three or less than three dimensions. With the use of PCA, the generated lower dimension data could be used to visualize the process. Some of the related research findings have presented on detecting anomalous behaviors and building robust PCs. In [5], the authors introduced the masking effect that illustrates the reason that why traditional computations on mean and covariance matrix are not usually able to detect multivariate-based outliers. In [4], the authors presented robust feature selection and robust PCA for the detection of Internet traffic anomaly. In [6], the authors presented a simulation study of detecting outliers by robust principal component analysis.
- B. Median and Median absolute deviation (MAD): As introduced by [6, 7], MAD is a robust estimator and an alternative of empirical standard deviation. For a data $X = (x_1, \dots, x_n)^t$, the MAD of this data can be

represented as $MAD(X) = c \cdot \text{med}_i |x_i - \text{med}(X)|$, where constant $c = 1.4826$ for normal distributions and $\text{med}()$ represents the median of this data. Authors from [6, 7] have adopted robust statistical concepts such as median and MAD to perform outlier detection.

- C. Multivariate statistical analysis: The generalized Euclidean distance can be applied to develop basic approaches for outlier detection. Specifically, given a data point $x_i \in \mathbb{R}^p$ with multivariate information, the square the generalized Euclidean distance can be described as $D_M^2(x_i) = (x_i - \mu)^T \Sigma^{-1} (x_i - \mu)$, where μ is the mean vector of the original data X and Σ^{-1} needs to be a robust estimate of the covariance matrix. This equation is known as the Mahalanobis distances (MD).

3. Proposed machine learning enabled data preprocessing scheme

The objective is to exploit the properties of PCs to identify outliers in the transformed space, which would not only produce robust PCs, but also lead to significant computational advantages for large scale data. A detailed flow of the proposed scheme is shown in Fig. 2.

- A. Log-transformation: In the first step of the proposed scheme, we apply the log transformation on the original numerical data X with n observations and p features. Applying log transformation can make highly skewed distributions less skewed. This is also significant for both making patterns in the data more interpretable and for helping to meet the assumptions of inferential statistics [9].
- B. Robust estimator selection: Once log transformation is performed, the second step is to identify the distribution of a given data. If the given data follows or appropriately follows the normal distribution, then robust statistical concepts, such as median and MAD, can be applied to process the data. If not, the idea of robust estimator S_n is adopted to process the transformed data. S_n , denoted as $S_n = c \cdot \text{med}_i |x_i - x_j|$, was proposed by [8] as an alternatives to the MAD. Not only has it performed well on the normal data, but also on the non-normal data. S_n attains a Gaussian efficiency of 58%, whereas MAD attains 37%.
- C. Data scaling: In the third step, we scale to the log-transformed data by the idea of coordinate-wise median and robust estimator. We robustly scale the given data in two ways. If the given log-transformed data possesses a density of normal distribution, then we scale the given data by $(X_{\text{med}_i}(x))/MAD(X)$. If the given log-transformed data possess a density of non-normal distribution, then we scale the given data by $(X_{\text{med}_i}(x))/S_n$. By this fashion, we replace the mean operation by the process of median, and similarly the empirical standard deviation by the MAD and S_n . This provides a newly and robustly generated data X' .
- D. Applying PCA: The goal of PCA is to minimize the information loss, which is equivalent to minimize the projection error. We now minimize the square project error by $\|X' - X'vv^T\|^2$, where v is the unit vector of the first PC. This error can be represented by the trace of two product, denoted as $\text{tr}((X' - X'vv^T)(X' - X'vv^T)^T)$. By calculation, this produces $c \cdot (1 - v^T \Sigma v)$, where c is a constant value. This form is then equivalent to maximize the $v^T \Sigma v$, which is the covariance matrix. We apply singular value decomposition (SVD) on the covariance matrix Σ to obtain the robust PCs.
- E. Robust PC selection: We now choose the number of robust PCs that need to be determined, in this fifth step. Ideally, the most percentage of the total variance should be retained by as much PCs as possible. However, in practical, this may not be the case since different applications possess different requirements. Assume that at least an $\alpha\%$ of total variance is required to retain, we denote k as the number of PCs retained to satisfy this requirement. The project data after PCA is denoted as Z with n observations and k features.
- F. Boundary setting: We conclude the scheme by computing the squared MD of data Z . A quantile of χ_k^2 distribution [4] is adopted for the distance metric as a separation boundary for outliers.

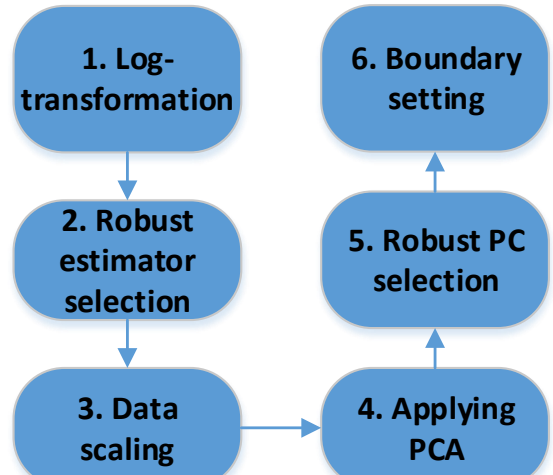


Figure 2 Flow of the proposed data preprocessing scheme

4. Discussions

The proposed scheme mainly adopts distance-based method to perform outlier identification and to seek for robust principle components. The use of the MD shows robustness and removes several of the limitations of the Euclidean metric. It automatically accounts for the scaling of the coordinate axes, corrects for correlation between the different features, and provides curved as well as linear decision boundaries.

However, there exist some potential drawbacks when Mahalanobis distance (MD) was adopted. One of them is that MD may not perform well when the number of features is larger than the number of observations. This is because when a data has more numbers of dimensions than observations, then the covariance matrix will be singular and a robust MD might not be computed properly. Additionally, high-dimensional data suffers from computational overhead rapidly with p than with n . The inverse of covariance matrix performed in MD is a polynomial based operation.

In summary, there is a price to pay for these advantages. The covariance matrices can be hard to determine accurately, and the memory and time requirements grow in a quadratic way rather than linearly with the number of features. These problems may be insignificant when only a few features are needed, but they can become quite serious when the number of features becomes large.

5. Case study

In this section, we present a case study to perform major data mining based classification method on a data set modified by KDD'99 [10]. We consider applying logistic regression (LR) [3] as a statistical method to find out the mean error rate. We use both common statistical and computational metrics to evaluate the proposed scheme.

A. Data preprocessing

Feature selection on the data is performed manually, and two zero-based features are removed. Since the proposed scheme aims to conduct preprocessing on numerical data, the categorical features that contain in the original training set are then removed. There are four categorical features in the original training set, including the response variable. We remove the first three categorical predictors, reserve the response variable for the subsequent detection, and form a new training set with numerical predictors. After performing log transformation, we identify that the data does not exactly follow a normal distribution, therefore we apply the robust estimator to shape the data and obtain X' .

PCA algorithm is applied on the data before any model is run. We choose three robust PCs to explain the variance of X' . We choose the first 12, 16 and 18 PCs, to represent the cumulative proportion of approximately 80%, 90% and 95% variance of X' , as shown in Table 1.

We then apply the idea of MD to the projected data Z . When the 0.975 quantile is adopted as the threshold to identify the outliers, 12345 observations are found. Once the preprocessing steps are completed, we start to apply LR to detect the anomalies and compare the predicted result with the actual values from the response variable.

B. Results

In this study, a 5-fold cross validation is performed to obtain the mean error rate. An experiment is run by logistic regression using R, by a machine with a CPU capability of 6 Intel Xeon X5660 CPUs \times 2.8 GHz and a RAM of 23.987 GB. The results are presented as follows: Once we conduct the proposed preprocessing scheme on the data, the mean error rate has been decreased by 0.6%, while the computational cost has been saved for approximately 30 seconds, as shown in Table 2. In Table 3, we present the comparison among robustly dimensionality-reduced and original data sets in terms of some common metrics. The first three reduced data sets retained a variance proportion approximately 80%, 90% and 95% of the total data. As the dimensionality is been reduced, the computational cost has been declined correspondingly, as well as the storage consumption. In an example, the reduced data with 18 PCs only consumes 18% of the total storage occupancy, while it has already preserve 95% of the information. All three reduced data sets retain a small amount of data storage consumption. The mean error rate of each data is also decreasing in general.

Table 1 3 PCs with their cumulative proportion of variances when α being 80%, 90% and 95%

	12th PC	16th PC	18th PC
Value of variance	0.92587	00.77461	0.70664
Proportion of variance	0.02572	0.02152	0.01963
Cumulative proportion	0.81768	0.91060	0.95062

Table 2 Results based on original data and preprocessed data (with all 36 PCs).

	Mean error	Computational cost (seconds)
LR (Original Data)	0.007044237	228.060
LR (Preprocessed Data)	0.001087868	197.556

Table 3 Metrics comparison among robustly reduced data sets and the original data

	12 PCs	16 PCs	18 PCs	Original data
Proportion of variance explained (%)	81.768	91.060	95.062	100
Computational cost (Seconds)	164.236	212.896	227.824	228.060
Storage consumption (Mb)	47.8	62.5	69.8	374.4
Mean error rate by LR (%)	0.1042	0.0490	0.04890	0.7044

6. Conclusion

In this article, a machine learning enabled data preprocessing scheme in cyber security applications is presented. The proposed scheme not only preserves the critical property of dimensionality reduction for high-dimensional data, but also introduces a robust detection boundary to the presence of outliers as well. In the end, an evaluation of the proposed scheme is presented by a case study, in terms of proportion of variance retained, computational overhead, storage consumption, and mean error rate.

References

- [1] S. Xu, Y. Qian, and R. Q. Hu, "A data-driven preprocessing scheme on anomaly detection in big data applications," *Proceedings of 2017 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2017.
- [2] H. Hu, Y. Wen, T. S. Chua, and X. Li, "Toward Scalable Systems for Big Data Analytics: A Technology Tutorial," *IEEE Access*, vol. 2, pp. 652-687, 2014.
- [3] G. James, D. Witten, T. Hastie, and R. Tibshirani, "An Introduction to Statistical Learning," *Springer*, vol. 6, 2013.
- [4] C. Pascoal, M. R. de Oliveira, R. Valadas, P. Filzmoser, P. Salvador, and A. Pacheco, "Robust feature selection and robust PCA for internet traffic anomaly detection," *Proceedings of 2012 IEEE INFOCOM*, 2012.
- [5] C. Becker and U. Gather, "The masking breakdown point of multivariate outlier identification rules," *Journal of the American Statistical Association*, vol. 94, no. 447, pp. 947-955, 1999.
- [6] C. Pascoal, M. R. Oliveira, A. Pacheco, and R. Valadas, "Detection of outliers using robust principal component analysis: A simulation study," *Combining Soft Computing and Statistical Methods in Data Analysis*, pp. 499-507, 2010.
- [7] P. Filzmoser, R. Maronna, and M. Werner, "Outlier identification in high dimensions," *Computational Statistics & Data Analysis*, vol. 52, no. 3, pp. 1694-1711, 2008.
- [8] P. J. Rousseeuw and C. Croux, "Alternatives to the median absolute deviation," *Journal of the American Statistical Association*, vol. 88, no. 424, pp. 1273-1283, 1993.
- [9] D. M. Lane, "Online Statistics Education: A Multimedia Course of Study," 2017. [Online]. Available: <http://onlinestatbook.com/>. [Accessed: 20- November- 2018]
- [10] M. Lichman, "UCI Machine Learning Repository," 2013. [Online]. Available: <http://archive.ics.uci.edu/ml>. [Accessed: 20- November- 2018]

MMTC OFFICERS (Term 2018 — 2020)

CHAIR

Honggang Wang
UMass Dartmouth
USA

STEERING COMMITTEE CHAIR

Sanjeev Mehrotra
Microsoft
USA

VICE CHAIRS

Pradeep K Atrey (North America)
Univ. at Albany, State Univ. of New York
USA

Wanqing Li (Asia)
University of Wollongong
Australia

Lingfen Sun (Europe)
University of Plymouth
UK

Jun Wu (Letters&Member Communications)
Tongji University
China

SECRETARY

Shaoen Wu
Ball State University
USA

STANDARDS LIAISON

Guosen Yue
Huawei
USA

MMTC Communication-Frontier BOARD MEMBERS (Term 2016—2018)

Dalei Wu	Director	University of Tennessee at Chattanooga	USA
Danda Rawat	Co-Director	Howard University	USA
Melike Erol-Kantarci	Co-Director	University of Ottawa	Canada
Kan Zheng	Co-Director	Beijing University of Posts & Telecommunications	China
Rui Wang	Co-Director	Tongji University	China
Lei Chen	Editor	Georgia Southern University	USA
Tasos Dagiuklas	Editor	London South Bank University	UK
ShuaiShuai Guo	Editor	King Abdullah University of Science and Technology	Saudi Arabia
Kejie Lu	Editor	University of Puerto Rico at Mayagüez	Puerto Rico
Nathalie Mitton	Editor	Inria Lille-Nord Europe	France
Zheng Chang	Editor	University of Jyväskylä	Finland
Dapeng Wu	Editor	Chongqing University of Posts & Telecommunications	China
Luca Foschini	Editor	University of Bologna	Italy
Mohamed Faten Zhani	Editor	l'École de Technologie Supérieure (ÉTS)	Canada

IEEE COMSOC MMTC Communications - Frontiers

Armir Bujari	Editor	University of Padua	Italy
Kuan Zhang	Editor	University of Nebraska-Lincoln	USA