A Secure Mechanism for Big Data Collection in Large Scale Internet of Vehicle

A Secure Mechanism for Big Data Collection in Large Scale Internet of Vehicle

Longhua Guo, Mianxiong Dong, Kaoru Ota, Qiang Li, Tianpeng Ye, Jun Wu and Jianhua Li

***Abstract*—As an extension for Internet of Things (IoT), Internet of Vehicles (IoV) achieves unified management in smart transportation area. With the development of IoV, an increasing number of vehicles are connected to the network. Large scale IoV collects data from different places and various attributes, which conform with heterogeneous nature of big data in size, volume, and dimensionality. Big data collection between vehicle and application platform becomes more and more frequent through various communication technologies, which causes evolving security attack. However, the existing protocols in IoT cannot be directly applied in big data collection in large scale IoV. The dynamic network structure and growing amount of vehicle nodes increases the complexity and necessary of the secure mechanism. In this paper, a secure mechanism for big data collection in large scale IoV is proposed for improved security performance and efficiency. To begin with, vehicles need to register in the big data center to connect into the network. Afterwards, vehicles associate with big data center via mutual authentication and single sign-on algorithm. Two different secure protocols are proposed for business data and confidential data collection. The collected big data is stored securely using distributed storage. The discussion and performance evaluation result shows the security and efficiency of the proposed secure mechanism.**

***Index Terms*—Internet of Vehicle (IoV), big data, large scale, secure mechanism.**

I. Introduction

With the rapid development of communication and computation technologies, a growing number of vehicles are connected to the Internet of Things (IoT) [1]. As a huge interactive network, Internet of Vehicles (IoV) has become an important issue of mobile Internet [2]. Information such as vehicles' location, speed and driven route are collected to central processing system using particular sensors and devices [3]. Huge research value and commercial interest will be promised after computing and analyzing vehicles' information [4]. Large scale IoV achieves unified management as an extension for IoT in smart transportation area [5].

In IoV, the vehicles' trajectory is subject to the road distribution in a wide range of physical area. A large number of traffic information is shared through IoV which contributes to the smart management and road optimization [6]. With the development of society, the increasing number of vehicles and roads lead to extended scale of IoV which covers a wide range of physical area. Deployed on the vehicles, different kinds of sensors provide a large amount of data about vehicles' attribute information, driving state information and traffic information [7]. The data is spatio-temporal in nature for its dependence upon time and location. The increasing number of vehicles collect data from different places and various attributes, which converges big data of heterogeneous nature with variation in size, volume, and dimensionality [8].

With the spread and development of IoV, the collected contents involve not only personal privacy for example vehicle's real-time location, but also some important data including vehicle running parameter which is closely related to traffic safety [9]. However, the fraudulent messages may be sent by malicious vehicle nodes to jeopardize the traffic system or purse their own profit [10]. Hence, it is significant to design a mechanism to ensure that the transmission of vehicle data resource is trusted and not tampered with. As the intelligent transportation system is continuously developing and big data applied in the IoV [11], big data collection between vehicle and application platform becomes more and more frequent through various communication technologies, which causes evolving security attack. How to secure the big data collection in large scale IoV is meaningful and deserves researching.

Nowadays, there existed some related works which focus on security of big data and IoV. Khaleel *et al.* in [12] proposed a security scheme of data messages exchanged between users and RSUs, but the scalability of IoV is still a remained problem to solve. Wu *et al.* in [10] proposed an efficient system for balancing public safety and vehicle privacy that guarantees message trustworthiness. Wang *et al.* in [13] proposed a secure mechanism for privacy-preserving communication with available cryptographic primitives in vehicle-to-grid (V2G) networks. The authors in [14-15] work at the big data area and developed the security and privacy mechanisms. As an important technology in big data area, the security of Hadoop is also addressed in [16-17]. Liu *et al.* proposed a key exchange scheme for secure scheduling of big data applications in [18]. The authors in [19-21] proposed security models to solve authentication, privacy issues in related areas. However, the existing protocols in the related area cannot be directly applied in big data collection in large scale IoV. As a result, the security and efficiency issue for big data collection still deserves

L. Guo, Q. Li, J. Wu, T. Ye and J. Li are with School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai Key Laboratory of Integrated Administration Technologies for Information Security, Shanghai 200240, China.

M. Dong and K. Ota are with Department of Information and Electronic Engineering, Muroran Institute of Technology, Japan. Correspondence should be addressed to Mianxiong Dong (e-mail: mx.dong@csse.muoran-it.ac.jp).

research.

In this paper, a secure information collection scheme for big data in large scale IoV is proposed. To begin with, vehicles need to register in the big data center to connect in the network. After the initialization phase, vehicles associate with big data center via authentication towards both sides using single sign-on algorithm. The collected information is transferred under security protection with improved efficiency. The collected big data will be secure stored using distributed storage architecture to achieve the unified management. The remainder of the paper is organized as follows. Section II describes the background related to this paper and presents the security requirements for big data collection in large scale IoV. Section III presents the proposed system model. In section IV, the details of proposed security mechanism are given, followed by the discussion in section V as well as performance evaluation in section VI respectively. Finally, we draw our conclusion and give the future work in section VII.

II. BACKGROUND

A. IoV

According to particular communication protocols and data interaction standards, IoV is an integrated network based on in-vehicle network, Vehicular Ad Hoc Network and vehicle-mounted mobile Internet. It is an extended application of Internet of Things which achieves intelligent traffic management control, vehicles intellectualization control and intelligent dynamic information service [22-23].

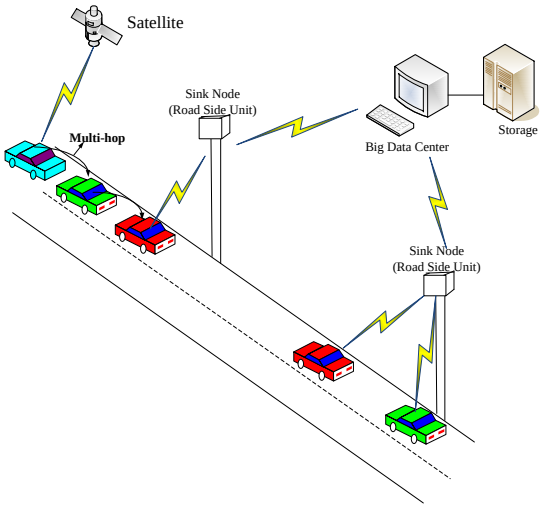


Fig.1: Basic architecture of IoV.

As shown in Fig.1, vehicle nodes, sink nodes and big data center constitute the basic architecture of the Internet of Vehicle. The big data center managements and processes data which are collected by vehicle nodes and transferred by sink nodes. In the onboard unit of vehicle node, vehicle gateway collects the information from orientation module, vehicle station parameter collection module and so on. As sink nodes, rode side units and users' communication devices help transfer the information.

In contrast to other ad hoc networks, IoV have some different features. As the vehicle nodes may change its locations at a high speed, node topological structure is dynamic and changing. It is hard to build accurate neighborhood. What's more, the process of information exchange in IoV has serious Doppler Effect and attenuation which has bad influence to the efficiency of information collection. With the data increasing, it causes worse effect to large scale IoV in the big data environment.

B. Big Data

Big Data is a system that let digitize large amount of information and combine it with existing databases. Big data is defined based on three primary characteristics, also known as the 3Vs: volume, variety, and velocity [24]. The increasing number of vehicles collect data from different places and various attributes, which converge big data of heterogeneous nature with variation in size, volume, and dimensionality. The integration of big data and IoV has been a trend with the development of new information technologies [25-26]. Big data collection can improve decision making, especially path planning in IoV. As for the government, the collected big data helps analyses and solve the traffic problems. As for the company like real-time transportation company, it helps optimize the vehicle resource. As a result, the government and companies demand and start building the big data platform for the large scale IoV.

C. Security Requirements for IoV

According to the features of IoV, the secure information collection scheme has to meet the requirements to ensure the data collection security. The security requirements with operational functions and management functions include [27-29]:

1) Authentication to identify the vehicle node, sink node and big data center;

2) Integrity to protect messages against modification or destruction;

3) Confidentiality to protect the information sent to appropriate entity. The business data like temperature parameters can be transferred in plain text form while the confidential data like location data need to be transferred in cipher text form;

4) Nonrepudiation to prevent deny afterward;

5) Authorization to ensure that only authorized nodes access to the resource. As for the high dynamic topological structure, single sign-on mechanism for nodes is necessary.

III. SYSTEM MODEL

To address the security requirements in large scale IoV, a secure data collection scheme for big data is proposed as shown in Fig.2. The increasing number of large scale vehicle nodes generate various attributes data from different places. These data will be collected by big data center with secure protection and stored in distributed storage system using Hadoop architecture. In the initialization phase, association with authentication towards all new adding vehicle nodes forms the first security line of defense against illegal nodes. These nodes will register in the system and exchange necessary information with the data center. After the initialization phase, the proposed secure single sign-on algorithm improves the efficiency of the

logon protocol. Besides, the collected information is transferred under security protection until logout

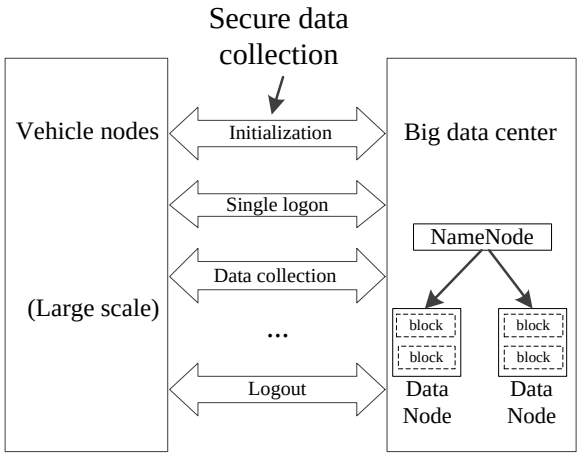


Fig.2: Secure data collection model.

IV. THE PROPOSED SCHEME

To address the security problems in wide area IoV, a secure information collection scheme for big data is proposed [30]. To begin with, vehicles need to register in the big data center to connect in the network. After the initialization phase, the vehicles associate with the big data center via authentication towards both sides using single sign-on algorithm. The collected information is transferred under security protection with improved efficiency. Big data of vehicle information is collected as designed format using distributed storage.

Table I

Symbols used by secure information collection scheme

Symbol	Explanation
V	Vehicle node
S	Sink node
C	Big data center
Cert_veh	Certification of the vehicle nodes
Cert_sink	Certification of the sink nodes
Cert_cen	Certification of big data center
pk_veh	Public key of the vehicle nodes
pk_sink	Public key of the sink nodes
pk_cen	Public key of big data center
pk_veh	Public key of the vehicle nodes
pk_sink	Public key of the sink nodes
pk_cen	Public key of big data center
sign_veh	Signature of the vehicle nodes
sign_sink	Signature of the sink nodes
sign_cen	Signature of big data center
key_vs	Session key between vehicle node and sink node
key_vc	Session key between vehicle node and big data center
key_sc	Session key between sink node and big data center
T_k	A random key
nonsense	A random number to fight against replay attack
Ts	Timestamp of the message
Period	The period of validity for the sign-on
H(m)	The hash value for the message *m*
HMAC(m)	Calculate *MAC* for message *m* based hash function
m_1, m_2, m_3, m_4	Parameters involved in the scheme
M_1, M_2	The business data and confidential data

A. Initialization

To support different kinds of big data platform, we assume that each vehicle is equipped with a certificate issued by outside Certification Authority (CA). In the initialization phase, vehicles need to register in the big data center to connect in the network. Vehicle nodes and big data center generate public key and private key of themselves respectively. As shown in Fig.3, certification, with their corresponding public keys as a pat, is exchanged between vehicle nodes and big data center. If the certificates pass the inspection, the corresponding ID will be registered as a valid account. Sink nodes are responsible for message forwarding. What's more, sink nodes are also necessary to register in this phase.

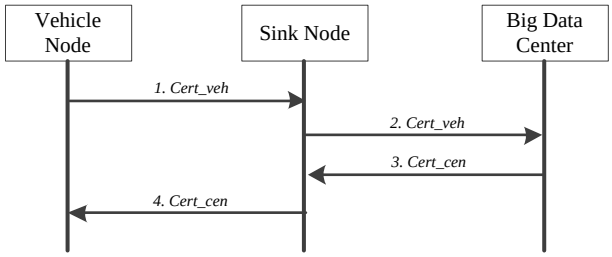


Fig.3: Message exchange in initialization phase.

B. Logon for the First-time

With the development of IoV, an increasing number of vehicles are connected to the network. Vehicles may run at a high speed and connect to different sink nodes. With the high dynamic topological structure, IoV requires single sign-on for nodes to achieve authorization ensuring that only authorized nodes access to the resource. The secure information collection scheme proposed single sign-on algorithm which improves the efficiency of the logon protocol. The expandability is enhanced utilizing the proposed scheme. After initialization phase, sink nodes and vehicle nodes connect to the big data center using different protocol as shown in Fig.4 and Fig.5.

In the phase of sink nodes' sign-on, *ID*, *nonsense* and *Ts* are sent to big data center with sink nodes' signature as shown in Fig.4. According to received message, big data center checks the signature and *ID* of sink nodes. What's more, *Ts* guarantees the time-efficiency while *nonsense* resists replay attack. If the messages are legal from valid account, the big data center generates the unique *key_sc*. The nonsense and *key_sc* will be encrypted using *pk_cen* and sent to sink node afterwards. Sink node will acquire the *key_sc* after decrypting the ciphertext using *sk_sink*.

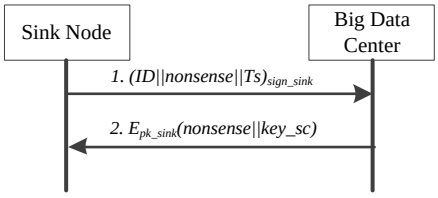


Fig.4: sink nodes' logon for the first time.

In the phase of vehicle nodes' sign-on, a similar process is designed as shown in Fig.5. Big data center receives and checks m_1 with vehicle and sink node's signature. After checking the signatures, m_2 is calculated with big data

center's signature. The "ticket" m_2 is stored in vehicle node and works as an important parameter for vehicle node's single sign-on afterwards. What's more, *pk_veh* encrypts *key_vc* as m_3 which is used for protecting the messages between vehicle node and big data center. *key_sc* are utilized for ensuring the security between sink node and big data center. Sink node generates *key_vs* to encrypt m_2 and m_3. Apart from $E_{pk\_veh}(key\_vs)$ and $E_{key\_vs}(m_2,m_3)$, *cert_sink* is required to ensure the legal identity for the sink node.

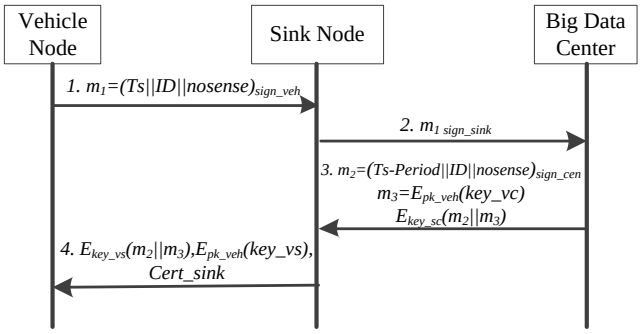


Fig.5: vehicle nodes' logon for the first time.

C. Logon once again

When the vehicle nodes leave the region of its first logon sink node, it has to access to the new arriving sink node with another logon. As for this kind of vehicle nodes, the proposed scheme simply the logon process afterwards. As shown in Fig.6, interaction between vehicle node and sink node can improve the efficiency of the logon process and update the session key. Besides the stored "ticket" m_2, certificate and *Ts* are sent to sink node with vehicle node's signature. Signature of big data center in m_2 proves that the ticket was awarded by big data center. If the ID in certificate matches with that in m_2 and the timestamp doesn't exceed the period limit, the vehicle will be regarded as legal node and log in the system. The certificate of sink node and *key_vs* which is encrypted by *pk_sink* will be sent to vehicle node afterwards.

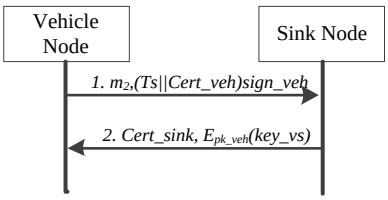


Fig.6: vehicle nodes' logon later.

D. Secure Data Collection

The above subsections set the secure premise for data collection in the large scale Internet of Vehicle. In condition that the vehicle nodes have succeed to log in the system, the business data and confidential data will be collected using following algorithms as shown in Fig.7 and Fig.8 respectively.

M_1 and M_2 represent the business data which is the main object for interaction. The business data like temperature parameters can be transferred in plain text form. m_4 is calculated by the concatenation of vehicle node's ID and M_1. To improve the calculation efficiency, hash value of m_4 is utilized for calculating the *HMAC*. As *key_vc* and *key_vs* has been shared in advance, *HMAC* helps prevent tampering with data and guarantee the identity of data sender. Sink nodes verify *HMAC (key_vs, H(m_4))* and transmit *HMAC(key_vc, H(m_4))*. When big data center publishes M_2, the same algorithm is proposed in step 3 and 4.

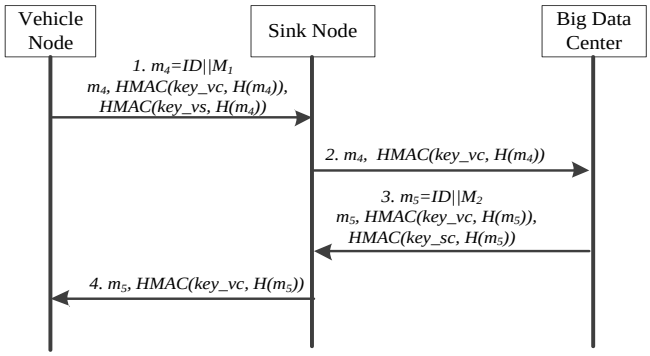


Fig.7: Message exchange for business data collection.

Different from business data collection, confidential data has to be transferred in ciphertext form. As a random key, T_k is utilized for encrypting. To share T_k with sink node and big data center, *key_vc* and *key_vs* helps achieve the confidence of T_k. As T_k is shorter than m_4, the utilization of T_k decrease the calculation complexity.

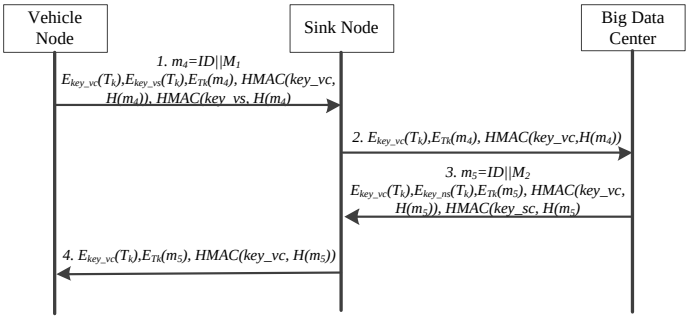


Fig.8: Message exchange for confidential data collection.

E. Secure Data Storage

In the above subsection, a secure information collection scheme for big data is proposed. Apart from business data and confidential data, some necessary security data of vehicle and sink nodes need to be stored in big data center. Data structure for vehicle and sink nodes stored in big data center is designed as shown in Table II. ID and certificate represent identification of the nodes. It is worth mentioning that if the vehicle node is successfully registered in the system, the statue changes from "off" to "on". Once abnormal actions are detected through analysis in big data center, the statue will be changed from "on" to "off". Another initialization process is necessary to fight against illegal nodes. If the timestamp is beyond the valid period, the node has to logon as the new adding node. Session key and public key are significant to achieve confidentiality and protect the information sent to appropriate entity.

Table II

Data structure for vehicle and sink nodes stored in big data center

Items	ID	Certificate	Statue	Valid period	Encrypted Session key
Vehicle node	*ID_veh*	*Cert_veh*	*on/off*	*Ts-Period*	*key_vc*
Sink node	*ID_sink*	*Cert_sink*	*on/off*	*Ts-Period*	*key_sc*

The business data like temperature parameters can be stored in plain text form while confidential data has to be stored in ciphertext form. As for the confidential data, *key_vc* is utilized to encrypt it for each vehicle node. Only when vehicle node interacts with big data center, its confidential data can be accessed and decrypted using *key_vc* while other vehicle nodes wouldn't be able to acquire the confidential data. In the data storage, the proposed algorithm helps achieve authorization to ensure that only authorized nodes access to the resource.

However, with the growing number of vehicles nodes, big data center processes and collects larger amount of data. The data size is much larger than that one single disk can load. Distributed file system is badly demand for the big data in large scale IoV. To address this issue, Hadoop Distributed File System (HDFS) enjoys great popularity in big data systems. In proposed scheme, the collected big data is stored using secure distributed storage algorithm in the basis of HDFS. When the data is requested, localized image file will be called using stream process. It doesn't matter where the file is located and which format is stored. One hdfs cluster contains one NameNode and several DataNodes. NameNode is master node in charge of managing file system including namespace and block. DataNode is utilized for data file storage. HDFS cuts one file into several blocks which are stored in DataNode. In addition to the original HDFS, enhanced security scheme ensure that only authorized nodes can access to the resource. To read the stored files, client interacts with NameNode to acquire the access token of the corresponding block where the target file is stored. These token has been allocated in advance in NameNode. If the block access token matches with the allocated token, the stored file can be access to the client.

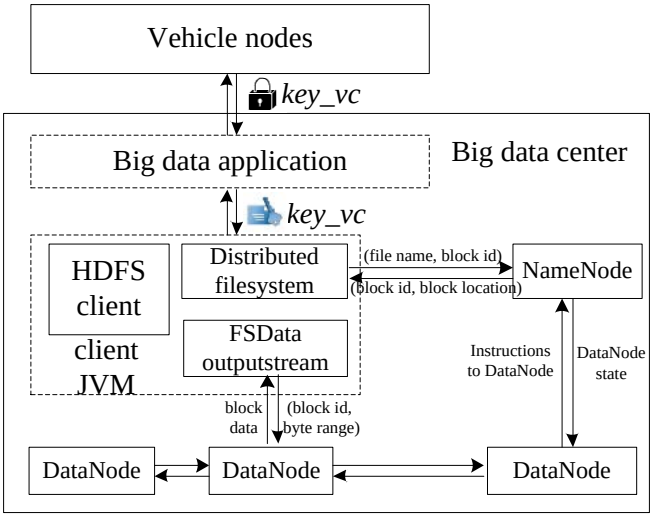


Fig.9: Secure distributed storage for collected big data.

As shown in Fig.9, *key_vc* encrypts the messages transferred between vehicle nodes and big data application. In big data center, *key_vc* works as certificate for controlling the big data application's access to the confidential data of the corresponding vehicle node. If available, the client JVM request file name and block id to NameNode through Distributed filesystem while block id and location will be sent as response. Afterwards, FSData outputstream sends block id and byte range to DataNode to acquire block data. NameNode controls the state of DataNode while the DataNode will send the instructions for searching the stored data. As the business data is stored in plain text form, it can be used directly while the ciphertext form of confidential data has to be decrypted using *key_vc*.

V. Discussion

A. Security Analysis

With the development of advanced information technologies, large scale IoV has occupied huge research value and commercial interest. The security of big data collection is of great significance. To ensure the data collection security, the proposed mechanism meets the requirements including authentication, integrity, confidentiality, nonrepudiation and authorization. In addition to the security requirements, security attacks such as Man-in-the-Middle (MITM) attack, replay attack, masquerade attack and message manipulation attack are also prevented by the proposed mechanism.

Nodes in the system are authenticated using certificates which are issued by CA. Compared with the traditional username/password token scheme, the certificates in proposed scheme fight against brute force and can't be forged which is more reliable for authentication. In the message exchange from initialization phase to data collection phase, signature is utilized for ensuring the integrity against modification or destruction. Public key is exchanged in the initialization phase and private key helps encrypt symmetric key which is utilized for encrypting confidential data. Combining with public/private key, symmetric key protects the information to be sent to appropriate entity. To meet the requirement of confidentiality, the confidential data is transferred in cipher text form. Stored by itself, the private key is utilized to calculate signature for nonrepudiation to prevent deny afterward. In the data storage, the session key "*key_vc*" works as certificate for controlling the big data application's access to the confidential data of the corresponding vehicle node. Only authorized nodes that have the certificate can access to the resource.

In our scheme, m_2 is utilized as the "ticket" for single sign-on. However, an attacker cannot success to sign-on even if it makes a copy of the previous m_2, certificate and *Ts* are also required for checking the identity of the vehicle nodes. So the malicious nodes fail to conduct replay attack to the big data center according to the proposed mechanism. MITM attacker intercepts the exchanging massages and tampers with the data. However, the messages transferred in the proposed mechanism are encrypted using session key and signature also helps fight against MITM attack. In masquerade attack [31], the attackers send wrong messages through pretending to be valid nodes which have bad influence to the security of information system. In the proposed scheme, all the nodes in the large scale network are authenticated using certificates and signatures. Therefore, the masquerade attackers cannot send wrong messages between the valid nodes because they cannot pass the authentication and pretend to be the valid nodes. In message manipulation attack, the exchanged messages may be dropped, modify or even forged to interrupt the data collection by attacker. In our mechanism, an attacker is too hard to forge the packet or path. Thus, the message manipulation attack is not effective to conduct with this mechanism.

B. *Efficiency Analysis*

To improve the efficiency of the secure big data collection process, single sign-on algorithm, message digest and random key (T_k) are designed and utilized in the proposed mechanism.

As for the high dynamic topological structure, single sign-on algorithm contributes to the simplification of logon process. For the vehicle nodes that connect to the new sink node, they will just interact with the new sink node while a tripartite interaction is required. The stored "ticket" m_2 will certificate the valid identity for the vehicle node while the certification of sink node will be sent back afterwards. The simplification of logon process helps improve the efficiency of the mechanism. The expandability is also enhanced utilizing the proposed scheme.

In the large scale IoV, the increasing number of vehicle nods generates growing big data in size, volume, and dimensionality. Message digest is utilized for decreasing the length of exchanged message in the business data and confidential data collection. m_4 is calculated by the concatenation of vehicle node's ID and M_1 which is far longer than hash value of m_4. The utilization of message digest improves the calculation efficiency for big data collection.

In big data collection, confidential data are transferred in cipher text form. The vehicle nodes are required to encrypt the confidential data to sink node and big data center. If the confidential data is encrypted using session keys including *key_vc* and *key_vs* directly, the calculation time will be effected by the length of the data which is often large in big data. In the proposed scheme, random key (T_k) encrypts the confidential data while the session keys ensure the length of T_k which is far shorter than the confidential data. The utilization of random ley also contributes to improve the efficiency for encrypting the big data in large scale IoV.

VI. Performance Evaluation

In the large scale IoV, an increasing number of vehicles are connected to the network which generates growing big data. The security schemes of big data collection are required to take the efficiency into consideration. In this section, we present the simulation results for evaluating the performance of our proposed mechanism. An overall time sequence will be given to shown the data flow of the proposed scheme. The simulation result of single sign-on algorithm, message digest and random key (T_k) shows the efficiency in logon process and data collection phase. With the increasing amount of data, the computing time and transmitting time for data collection directly reflect the performance of the security mechanism. As a result, we will compare our mechanism with others in these two aspects.

A. *Overall Time Sequence*

To evaluate performance of the proposed secure mechanism, the simulation of the entire data collection process is conducted using the network simulator software Opnet. As shown in Fig.10, overall time sequence for three kinds of nodes is given. A complete secure big data collection between vehicle node, sink node and big data center is presented to show how the work flow goes. As defined in the IEEE 802.11 about telecommunications and information exchange in vehicle-to – infrastructure (V2I), we assume that the transfer rate is up to 12 megabytes per second [32]. The abscissa axis records the time of interactions while vertical axis shows the transfer rate of each node using the security protocols. Interactions including initialization phase, sink node's logon, vehicle node's logon and data collection are simulated continuously. For each node, the time period when the blue line is over zero represents the transmitting time for each action. For two interacting objects, the difference time value of between two following actions represents the calculation time. The results show that the proposed mechanism is available in the environment of large scale IoV.

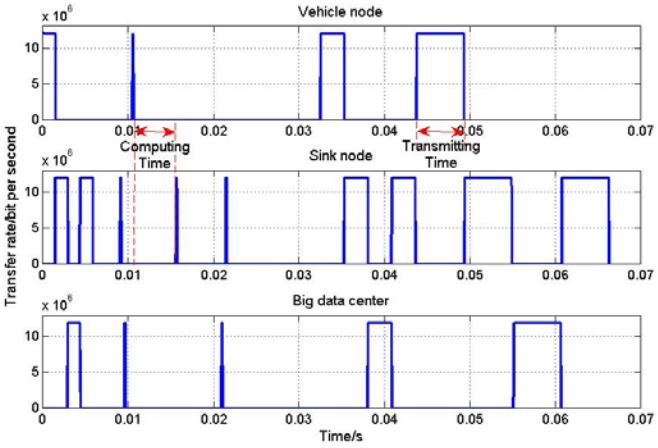


Fig.10: Overall time sequence of the proposed scheme.

B. *Single Sign-on Comparison*

As the vehicle nodes may change its location at a high speed, node topological structure is dynamic and changing. The vehicle nodes may connect to different sink nodes with the changing of network topological. The asymmetric encryption utilizes RSA algorithm in which the length of public/private key are both 1024 bites while AES-128 algorithm is utilized for symmetrical encryption. The asymmetric encryption costs much more than the symmetrical encryption does. As a result, the sink node's logon time cost and vehicle node's logon time cost for the first time are far more than that of other phases as shown in Fig.11. So the big data collection in large scale IoV deserves a secure single sign-on algorithm.

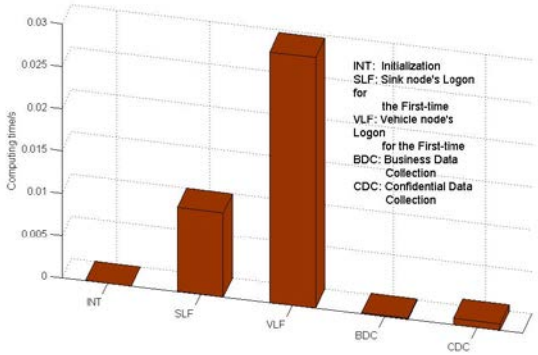


Fig.11: Calculation time for each phase.

Compared with Mutual Authentication (MA), the proposed mechanism just designs an interaction with the new sink node for vehicle node while a tripartite interaction is required in the traditional sign-on. As shown in Fig.12, the computing time of vehicle node, sink node and big data center using SSO and MA

are presented. All the three kinds of nodes using SSO cost less time than that using MA. It shows that SSO decreases the computing cost for logon which is more significant for the logon with the dynamic network structure in the environment of large scale IoV.

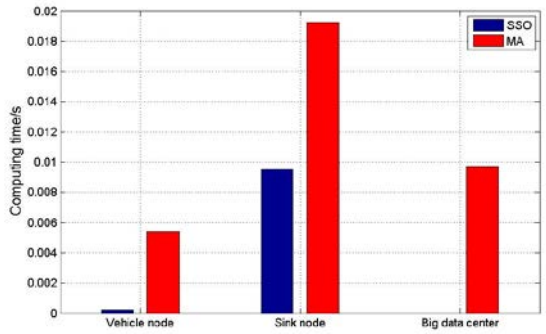


Fig.12: The calculation time of each node using SSO and MA.

As shown in Fig.13, the computing time and transmitting time for SSO and MA are presented. The computing time for SSO is shorter than MA. As the proposed SSO algorithm is designed to transfer certificate twice, the transmition of certificate costs more time. The huge computation time decrease at the price of little increase of transmitting time. With the enhanced security requirements, the length of the key for asymmetric encryption grows which may lead to the increase of the encryption/decryption time. The proposed SSO algorithm shows better efficiency in the large scale IoV.

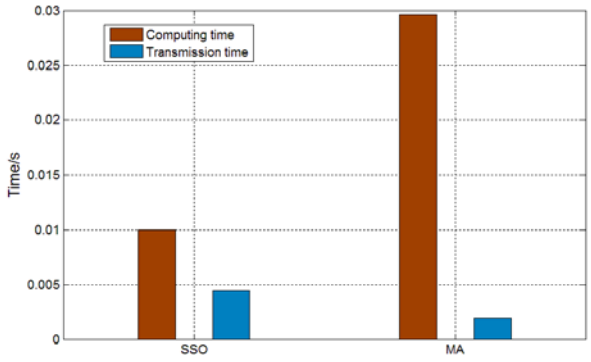


Fig.13: The computing and transmission time using SSO and MA.

C. *Message Digest and Random Key Comparison*

In big data collection, confidential data are transferred in cipher text form which is required to encrypt the confidential data to sink node and big data center. As shown in Fig. 14, the transmitting time of data collection increases with the growing of collected data size. Other phases set a security foundation for data collection and the size of collected data size doesn't affect the efficiency of these phases. As a result, how to improve the efficiency of the big data collection deserves researching in the large scale IoV.

In our scheme, message digest and random key (T_k) are utilized for improving the efficiency. Different from our scheme in Fig.14 and Fig.15, the message is directly processed using HMAC algorithm without calculating message digest in advance in Scheme 1 and 3 while the confidential data is encrypted using session keys including *key_vc* and *key_vs* directly in Scheme 2 and 3. As shown in Fig.15, the computing time using our scheme for vehicle nodes and big data center is much shorter than that using other schemes. The computing time using our scheme for sink node is almost equal to that using other schemes.

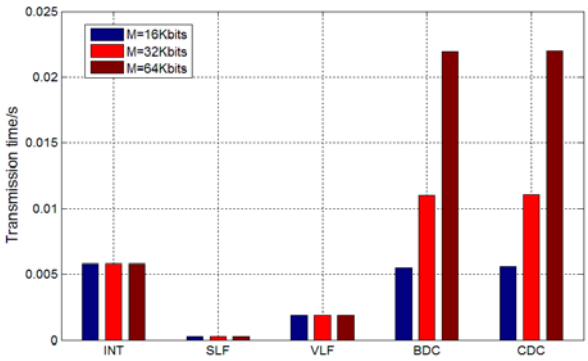


Fig.14: Transmission time of collected data's size for each phase.

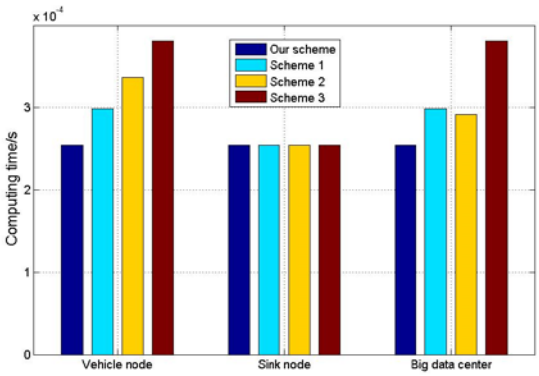


Fig.15: Computing time for each node using our scheme and other schemes.

As the utilization of random key, the exchange of encrypted T_k costs extra transmitting time for data exchange. As shown in Fig.16, our proposed scheme performs much better than Scheme 1 and Scheme 3 in the interaction between vehicle node to sink node as well as big data center to sink node. However, our proposed scheme performs little worse than Scheme 1 and Scheme 3 in the interaction between sink node to big data center as well as sink node to vehicle node. The calculation cost decreases at the price of little increase of transmitting time in some interactions. In terms of overall mechanism, the big data secure collection improves the efficiency using our proposed scheme.

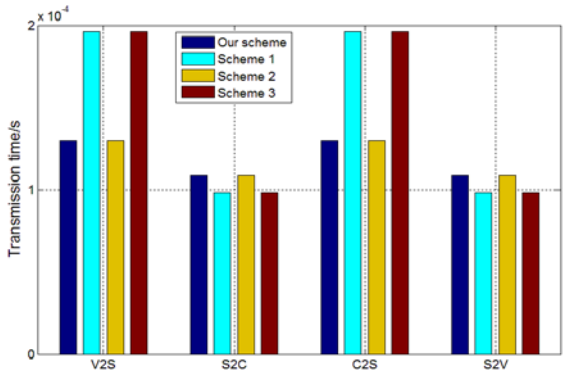


Fig.16: Transmission time for each node using our scheme and other schemes.

VII. CONCLUSION

In this paper, a secure information collection scheme for big data in large scale IoV is proposed. Single sign-on algorithm for authentication are utilized with improved efficiency. The proposed secure data exchange algorithm using message digest and random key contributes to overhead reduction. The business data is transferred in plain text form while the confidential data is transferred in cipher text form. The collected big data will be processed using hadoop architecture to achieve the unified management. The evaluation result and discussion show the proposed secure information collection scheme achieves high efficiency and security for big data in large scale IoV.

In the future, our work will consider developments in the following three aspects. Firstly, a demonstration experiment is necessary to verify our proposed scheme's efficiency and security. Secondly, with the increasing amount of vehicles in the IoV, we could do some further research about the routing protocol of IoV to optimize our security scheme. Thirdly, with the development of the new communication technology, such as 5G, we would pay attention on the security scheme to fit with these changes.

ACKNOWLEDGMENT

This work is supported by National Natural Science Foundation of China (Grant No. 61401273, 61571300, 61562004 and 61431008). Additionally, this work was supported in part by the JSPS KAKENHI Grant No. JP15K15976, JP16K00117, 26730056 and the JSPS A3 Foresight Program.

REFERENCES

[1] J. A. Guerrero-ibanez, S. Zeadally, J. C. Castillo, "Integration challenges of intelligent transportation systems with connected vehicle, cloud computing, and internet of things technologies", IEEE Wireless Communications, vol. 22, no. 6, pp. 122-128, Dec. 2015.

[2] M. Jin, X. Zhou, E. Luo, and X. Qing, "Industrial-QoS-Oriented Remote Wireless Communication Protocol for the Internet of Construction Vehicles", IEEE Transactions on Industrial Electronics, vol. 62, no. 11, pp 7103-7113, Nov. 2015.

[3] N. Kumar, J. J. P. C. Rodriguesand and N. Chilamkurti, "Bayesian Coalition Game as-a-Service for Content Distribution in Internet of Vehicles", IEEE Internet of Things Journal, vol. 1, no. 6, pp. 554-555, Dec.2014.

[4] J. Fu, Z. Chen, R. Sun and B. Yang, "Reservation Based Optimal Parking Lot Recommendation Model in Internet of Vehicle Environment", China Communications, pp 38-48, vol.11, no.6, Oct. 2014.

[5] J. Cheng, J. Cheng, Me. Zhou, F. Liu, S. Gao and C. Liu, "Routing in Internet of Vehicles A Review", IEEE Transactions on Intelligent Transportation Systems, vol.16, No. 5, pp 2339-2351, Oct. 2015.

[6] A. Dua, N. Kumar, and S. Bawa, "A systematic review on routing protocols for Vehicular Ad Hoc Networks," Vehicular Communications, 1, vol. 1,pp. 33-52, 1// 2014.

[7] B. Li, C. Zhao, H. Zhang, X. Sun, "Characterization on Clustered Propagations of UWB Sensors in Vehicle Cabin: Measurement, Modeling and Evaluation," IEEE Sensors Journal, vol.13, no.4, pp. 1288-1300, Apr. 2013.

[8] N. Kumar, S. Misra, J. Rodrigues, M. S. Obaidat. "Coalition Games for Spatio-Temporal Big Data in Internet of Vehicles Environment: A Comparative Analysis", IEEE Internet of Things Journal, vol.2 no.4, pp. 310-320, Aug. 2015.

[9] Y. Zhou, S. Chen, Y. Zhou, M. Chen. "Privacy-Preserving Multi-Point Traffic Volume Measurement Through Vehicle-to-Infrastructure Communications", IEEE Transactions on Vehicular Technology, vol. 64, no.12, pp. 5619-5630, Dec. 2015.

[10] Q. Wu, J. D. Ferrer, Ú. G. Nicolas. "Balanced Trustworthiness, Safety, and Privacy in Vehicle-to-Vehicle Communications", IEEE Transactions on Vehicular Technology, vol. 59, no. 2, pp. 559 – 573, Feb. 2010.

[11] J. Soares, N. Borges, B. Canizes, Z. Vale. "Probabilistic estimation of the state of Electric Vehicles for smart grid applications in big data context", 2015 IEEE Power & Energy Society General Meeting, Denver, CO, July 2015, pp. 1-5.

[12] K. Mershad, H. Artail, "A Framework for Secure and Efficient Data Acquisition in Vehicular Ad Hoc Networks", IEEE Transactions on Vehicular Technology, vol. 62, no. 2, pp.536 – 551, 2013.

[13] H Wang, B Qin, Q. Wu, L. Xu, J. D. Ferrer, "TPP: Traceable Privacy-Preserving Communication and Precise Reward for Vehicle-to-Grid Networks in Smart Grids", IEEE Transactions on Information Forensics and Security, vol. 10, no. 11, pp. 2340-2351, Nov. 2015.

[14] A. A. Cárdenas, P. K. Manadhata and S. P. Rajan, "Big Data Analytics for Security", IEEE Security & Privacy, Vol.11, no. 6, pp 74-76, Dec. 2013.

[15] L.Xu, C. Jiang, J. Wang, J. Yuan, And Y. Ren, "Information Security in Big Data Privacy and Data Mining", IEEE Access, vol.2, pp 1149-1176, Oct. 2014.

[16] M. Rezaei Jam, L. M. Khanli, M. K. Akbari and M. S. Javan, "A Survey on Security of Hadoop", in Proc. 4th International Conference on Computer and Knowledge Engineering (ICCKE), Mashhad , Oct. 2014, pp 716-721.

[17] P. Adluru, S. S. Datla and X. Zhang, "Hadoop Eco System for Big Data Security and Privacy", in Proc. 2015 IEEE Long Island Systems, Applications and Technology Conference (LISAT), Farmingdale, NY May. 2015, pp.1-6.

[18] C. Liu, X. Zhang, C. Liu, Y. Yang, R. Ranjan, D. Georgakopoulos and J. Chen, "An Iterative Hierarchical Key Exchange Scheme for Secure Scheduling of Big Data Applications in Cloud Computing", in Proc. 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications, Melbourne, VIC, Jul. 2013, pp10-16.

[19] H. Li, Y. Yang, T. H. Luan, X. Liang, L. Zhou, and X. (Sherman) Shen, "Enabling Fine-grained Multi-keyword Search Supporting Classified Sub-dictionaries over Encrypted Cloud Data," IEEE Transactions on Dependable and Secure Computing, 2015.

[20] H. Li, X. Lin, H. Yang, X. Liang, R. Lu, and X. (Sherman) Shen, "EPPDR: An Efficient Privacy-Preserving Demand Response Scheme with Adaptive Key Evolution in Smart Grid," IEEE Transactions on Parallel and Distributed Systems, 2014, vol. 25, no.8, pp. 2053 - 2064.

[21] H. Li, R. Lu, L. Zhou, B. Yang, and X. (Sherman) Shen, "An Efficient Merkle Tree Based Authentication Scheme for Smart Grid," IEEE SYSTEMS Journal, 2014, vol. 8, no.2, pp. 655 – 663.

[22] M. A. Salahuddin, A. Al-Fuqaha and M. Guizani, "Software-Defined Networking for RSU Clouds in Support of the Internet of Vehicles", IEEE Internet of Things Journal, vol. 2, no. 2, pp 133-144, Apr. 2015.

[23] K. M. Aalm, M. Saini and A. E. Saddik, "Toward Social Internet of Vehicles Concept, Architecture, and Applications", IEEE Access, vol.3, pp 343-357, Mar. 2015.

[24] Z. Su, Q. Xu, Q. Qi, "Big data in mobile social networks: a QoE-oriented framework", IEEE Network, vol.30, no.1, pp.52-57, 2016.

[25] D. Tracey, C. Sreenan, "A Holistic Architecture for the Internet of Things, Sensing Services and Big Data", in Proc. 13th IEEE/ACM International Symposium on Cluster, Cloud, and Grid Computing, Delft , May. 2013, pp.546-553.

[26] C. Cecchinel, M. Jimenez, S. Mosser and M. Riveill, "An Architecture to Support the Collection of Big Data in the Internet of Things", in Proc. IEEE 10th World Congress on Services, Anchorage, AK, Jun. 2014, pp.442-449.

[27] L. Guo, J. Wu, Z. Xia, J. Li, "Proposed Security Mechanism for XMPP-Based Communications of ISO/IEC/IEEE 21451 Sensor Networks", IEEE Sensors Journal, vol.15, no.5, pp. 2577-2586, May 2015.

[28] L. Guo, M. Dong, K. Ota, W. Jun, J. Li. "Event-Oriented Dynamic Security Service for Demand Response in Smart Grid employing Mobile Networks", China Communications, vol.12, no.12, pp. 63-75, Dec. 2015.

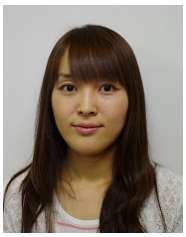
[29] H. Li, D. Liu, Y. Dai and T. H. Luan, "Engineering Searchable Encryption of Mobile Cloud Networks: When QoE Meets QoP," IEEE Wireless Communications, 2015, vol.22, no.4, pp. 74 -80.

[30] K. Yu, M. Arifuzzaman, W. Zheng, D. Zhang, T. Sato. "A Key

Management Scheme for Secure Communications of Information Centric Advanced Metering Infrastructure in Smart Grid", IEEE Transactions on Instrumentation and Measurement, vol. 64, no. 8, pp. 2072-2085, Aug. 2015.
[31] M. Rahman, K. El-Khatib, "Secure Time Synchronization for Wireless Sensor Networks Based on Bilinear Pairing Functions", IEEE Transactions on Parallel and Distributed Systems, vol. pp, no.99, pp. 1-15, May, 2010.
[32] IEEE 802.11, Telecommunications and information exchange between systems. Local and metropolitan area networks. Specific requirements. Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, IEEE Standards 802.11-2012.

Longhua Guo was born in Shanxi, China, in 1991. He received the B.S. degree in electronic information engineering from Tianjin University, Tianjin, China, in 2013 and is currently pursuing the Ph.D. degree in Shanghai Jiao Tong University, Shanghai, China. He participates in many national projects, such as National Natural Science Foundation of China, National "973" Planning of the Ministry of Science and Technology Program, China, etc. His research interests include sensor network security, social network analysis, etc.

Mianxiong Dong (M'13) received B.S., M.S. and Ph.D. in Computer Science and Engineering from The University of Aizu, Japan. He is currently an Assistant Professor with Department of Information and Electronic Engineering, Muroran Institute of Technology, Japan. He was a Researcher with National Institute of Information and Communications Technology (NICT), Japan. He was a JSPS Research Fellow with School of Computer Science and Engineering, The University of Aizu, Japan and was a visiting scholar with BBCR group at University of Waterloo, Canada supported by JSPS Excellent Young Researcher Overseas Visit Program from April 2010to August 2011. From January 2007 to March 2007, he was a visiting scholar of West Virginia University, USA. Dr. Dong was selected as Foreigner Research Fellow (A total of 3 recipients all over Japan) by NEC C&C Foundation in 2011. He is the Best Paper Award winner of IEEE HPCC2008 and IEEE ICESS 2008. Dr. Dong is currently a research scientist with A3Foresight Program (2011-2014) funded by Japan Society for the Promotion of Sciences (JSPS), NSFC of China, and NRF of Korea. His research interests include wireless sensor networks, vehicular ad-hoc networks and wireless security.

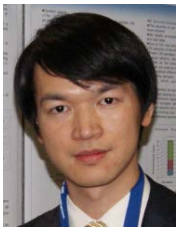


Kaoru Ota (M'12) received M.S. degree in Computer Science from Oklahoma State University, USA in 2008 and Ph.D. degree in Computer Science and Engineering from The University of Aizu, Japan in 2012. She is currently an Assistant Professor with Department of Information and Electronic Engineering, Muroran Institute of Technology, Japan. From March 2010 to March 2011, she was a visiting scholar with BBCR group at University of Waterloo, Canada. Also she was a prestigious Japan Society of the Promotion of Science (JSPS) research fellow with Graduate School of Information Sciences at Tohoku University, Japan from April 2012 to April 2013. Dr. Ota is a recipient of JSPS Grant-in-Aid for Research Activity Start-up in 2013. She has joined JSPS A3 foresight program as one of primary researchers since 2011 which is supported by Japanese, Chinese and Korean government. She serves as a Guest Editor of IEICE Transactions on Information and Systems, Special Section on Frontiers of Internet of Things 2014, Editor of Peer-to-Peer Networking and Applications (Springer), Journal of Cyber-Physical Systems, and International Journal of Embedded Systems. Dr. Ota has actively involved in international conferences in present and past. Currently she is the Publicity Co-chair of the 2014 ICC workshop on Secure Networking and Forensic Computing and 2014 IEEE ICCC Workshop on Internet of Things. Her research interests include wireless sensor networks, vehicular networks, and ubiquitous computing.

Qiang Li is an assistant professor in the School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China. His research interests are in the fields of information security, including secret sharing, multi-party computation and key management.

TianPeng Ye was born in Jiangsu, China, in 1993. He is a graduate student of Information Security Engineering in Shanghai Jiao Tong University, Shanghai, China. He received the B.S. Degree in Communication Engineering from Southwest Jiaotong University, Chengdu, China, in 2015. His research interests include software defined network (SDN), network function virtualization (NFV) and Internet of Things (IoT) security.

Jun Wu is an Associate Professor of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, China. He obtained his PH.D. Degree in Information and Telecommunication Studies at Waseda University, Japan. He was a postdoctoral researcher for the Research Institute for Secure Systems (RISEC), National Institute of Advanced Industrial Science and Technology (AIST), Japan, from 2011 to 2012. He worked as a researcher for the Global Information and Telecommunication Institute (GITI), Waseda University, Japan, from 2011 to 2013. His research interests include the advanced computation and communications techniques of smart sensors, wireless communication systems, industrial control systems, wireless sensor networks, smart grids, and more. He has hosted and participated in several research projects for the National Natural Science Foundation of China, National 863 Plan and 973 Plan, Japan Society of the Promotion of Science (JSPS) projects, etc. He has been a Guest Editor for the IEEE Sensors Journal and a TPC Member of several international conferences including WINCON 2011, GLOBECOM 2015, etc. He is a member of IEEE.



Jianhua Li is a professor/Ph.D. supervisor and the vice dean of School of Information Security Engineering, Shanghai Jiao Tong University, Shanghai, China. He got his BS, MS and Ph.D. degrees from Shanghai Jiao Tong University, in 1986, 1991 and 1998, respectively. He is the director expert of information security committee of National High Technology Research and Development Program of China (863 Program) of China. He is the member of the committee of information security area of the state 10th five-year plan of China. Also, he is a committee expert of China State Secrecy Bureau and Shanghai Secrecy Bureau. He is also a committee expert of Information Technique Standardization Committee of Shanghai, China. He was the leader of more than 30 state/province projects of China, and published more than 200 papers. He published 6 books and has about 20 patents. He made 3 standards and has 5 software copyrights. He got the Second Prize of National Technology Progress Award of China in 2005. He got the First Prize of National Technology Progress Award of Shanghai in 2003 and 2004, and he got two First Prize of National Technology Progress Awards of Shanghai in 2004. His research interests include information security, signal process, computer network communication, etc.