A Survey of Security and Privacy in Big Data

Haina Ye, Xinzhou Cheng, Mingqiang Yuan, Lexi Xu, Jie Gao, and Chen Cheng
Department of Network Optimization and Management,
China Unicom Network Technology Research Institute, Beijing, P.R.China
Email: yehn3@chinaunicom.cn

***Abstract*—Big data has been arising a growing interest in both scientific and industrial fields for its potential value. However, before employing big data technology into massive applications, a basic but also principle topic should be investigated: security and privacy. In this paper, the recent research and development on security and privacy in big data is surveyed. First, the effects of characteristics of big data on information security and privacy are described. Then, topics and issues on security are discussed and reviewed. Further, privacy-preserving trajectory data publishing is studied due to its future utilization, especially in telecom operation.**

I. Introduction

Big data has emerged to a new paradigm for data applications. Due to significant benefits, big data arises a growing interest in many industry fields, such as telecom operation [1], [2], healthcare [3], [4] and so on. Many efforts on big data have been working on the data storage, data mining, and data application. However, the widespread usage of big data relies not only on the promising solutions and mechanisms of data analysing, but also on security protection and privacy preserving.

Information security can be improved by big data technology, which is beneficial from security tools such as network monitoring, security information, and event management [5], [6]. However, on the down-side, there are additional security challenges brought by the big data technology, including cryptography algorithms, data provenance, secure data storage, access control, real time monitoring and so on [7]. Identifying and analysing the security issues will bring a better usage of big data. Thus, in this paper, we will first survey existing research on security and privacy. Then, we will focus on an essential type of data: trajectory.

Trajectory data represents the mobility of moving objects, such as people, vehicles, and so on. Spatio-temporal trajectories provide significant and valuable information, and foster a broad range of applications, such as intelligent transportation system, commercial site planning and so on. Therefore, trajectory data mining has become an increasingly interesting research topic, attracting attentions from numerous fields, especially in telecom operation. As a trustful data owner, telecom operators are authorized to have large amounts of location data of mobile phone customers. Making a good use of trajectory data can help telecom operators optimize the network and promote social services as well. However, location and trajectory data can be sensitive for individuals. Attackers may infer individuals' privacy such as personal habits or personal details from trajectories [8]. In order to preserve privacy, techniques and algorithms should be applied in the case that trajectories are released to third party for data analysis or data mining results of trajectories are published.

The main objective of this paper is to survey the literature related to security and privacy in big data to provide a comprehensive reference of the challenges and risks to which a big data application chain is facing. Then, we focus on the research and industry approaches to trajectory preserving issues in big data.

In this work, we provide a context to the work by introducing the security and privacy challenges triggered by characteristics of big data in Section II. In Section III, we present big data system security and privacy analyses. Research work on trajectory publishing is then highlighted in Section IV. In Section V, we conclude this paper.

II. Security and Privacy Challenges Triggered by 5Vs

In this section, the discussion begins with understanding the impacts of big data characteristics on security and privacy. According to the definition and principle of big data [9], the characteristics of big data are summarized as "5Vs", i.e., Volume, Variety, Velocity, Value, and Veracity. Existing literatures have given detail descriptions and explanations about the "5Vs" [10]. In order to explicate why and how security and privacy issues are magnified by big data, we specify the challenges and risks of security issues that are triggered by the "5Vs" characteristics of big data.

1) "Volume" points to the size of data. There is a huge amount of data generated by organizations, individuals and sensors every second in every fields. It is nearly impossible for data providers to supervise or control all the data they "actively" or "passively" provide to others. By using these data, people's identifications or behaviors can be predicted, which may further infer to individuals' privacy. Accordingly, the big volume of data increases the risk of information leakage. Besides, existing infrastructure strategies, such as regular tracking, monitoring, auditing or security scanning technology are not sufficient any more. Because it is complicated and costly to implement those methods on the large-scale big data scenario [11].
2) "Variety" indicates the diversity of data formats and sources. The data format includes structured, semi-structured, and unstructured ones, while the filetype consists of texts, figures, and videos. Large-scaled cloud

978-1-5090-4099-5/16/$31.00 ©2016 IEEE

infrastructure, as an optional method to manage and store date, makes the traditional storage and management measures invalid [13]. Consequently, not only the infrastructure security facing massive data up to PB (Petabyte) level, but also data management methods addressing data provenance, needs to be considered.

3) "Velocity" shows the continuousness and high frequencies of data. This feature makes information security and privacy issues even more severe. Fast growing and iterating data requires non-relational databases, thus distributed programming frameworks should have been developing with security and privacy in mind [14]. Besides, the hacker can launch advanced persistent threats (APT) more easily, while it is hard to be detected by the traditional protection strategy.
4) "Value" refers to the outputs that gains from huge data sets. The highly potential value and intensely integrated data attracts hackers [12]. Hackers who successfully attack the database would obtain a larger number of data and more sensitive information, and thus the cost of the attack is decreased. This may lead to a higher probability of cyber-attack. In addition, the purpose of data mining is to analyse data and extract useful information from data sets, which helps to predict the future and make decision. Individuals, corporations and organizations would gain benefits from big data predictive analysis, but on the other hand, they will be easily identified and treated worse at the same time. Accordingly, the tradeoff between the privacy preserving and the benefits brought from data utilization should be seriously considered.
5) "Veracity" refers to the trustworthiness, applicability, noise, bias, abnormality and other quality properties of the data [9]. Usually the results of data mining are used for commercial or public decision making, thus, the main concerns is whether the mining results are credible. This feature involves in the whole big data chain, from the authenticity of the original data, through the integrity of the mined data, to the credibility of the published data.

III. Security and Privacy Issues in Big Data

Given the big data characteristics and the impacts triggered by the characteristics on security and privacy in Section II, existing security and privacy topics and issues are discussed and surveyed in this section.

The authors of [7], [14] have proposed some conceptual and operational taxonomies of security and privacy to introduce vulnerabilities of big data system. However, in order to provide a causal relationship between the big data characteristics and vulnerabilities, considering the topics and issues in research fields, we refine and propose the following category, as shown in Fig.1:

- Infrastructure Security
- Data Privacy
- Data Management

A number of previous surveys study the big data security and privacy in various perspectives. As illustrated in Fig.2,

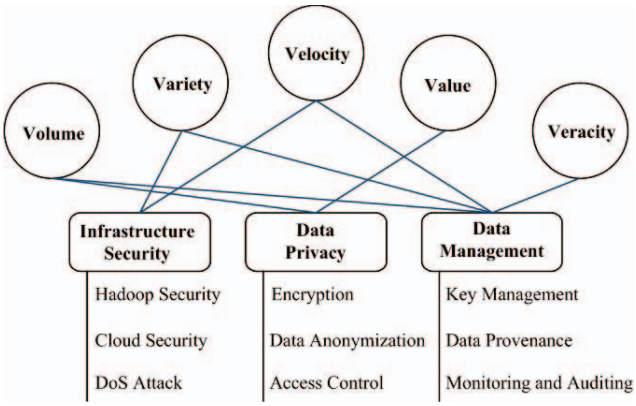

Fig. 1. Category of Security Challenges in Big Data.

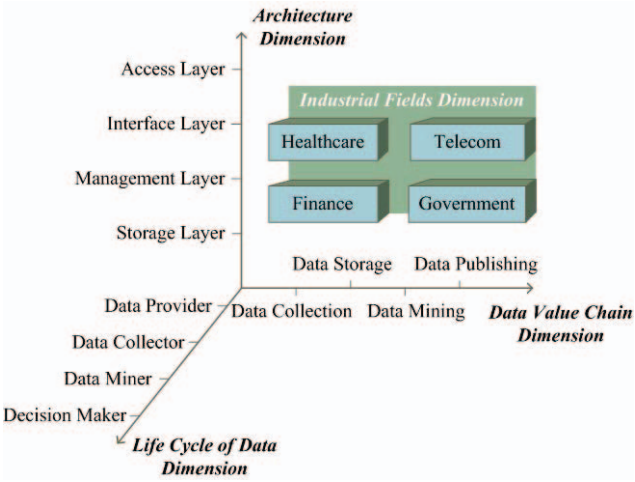

Fig. 2. Perspectives of Big Data Security and Privacy Analytics.

different perspectives can help having a better understanding of the security and privacy problem.

In the architecture dimension, the research objective is considered as a big data management platform. There are four layers, to which physical or logical entities correspond: storage layer, including secure storing, distributed equipment, monitoring and control etc.; management layer, including Hadoop distributed file system(HDFS), data encryption, content distribution etc; interface layer, including identity authentication, access management, public application programming interface(API) etc; access layer, including the cyber security of users. In the life cycle of data dimension, there are four types of users involving in data mining life cycle. For each type of user, the concerns and the methods adopted for each user role can be diverse [15]. In the data value chain dimension, it considers the stages in the process to obtain data value. For each stage, the risks, requirements as well as methods are different [16]. In the industrial fields dimension, [17] analyses the most concerning security issues in different application area. For example, in big data for healthcare, preserving customers' sensitive information is one of the top targets. While in big data for financial, such as banking system, the

TABLE I
COMPARISON OF SURVEYS ON SECURITY ISSUES IN BIG DATA

Research	Topics			OF	Big Data Process Layer/Interface			
	Infrastructure Security	Data Privacy	Data Management		Data Collecting	Data Storing	Data Mining	Data Publishing
[11], [15]		✓	✓		✓	✓	✓	✓
[10], [12]	✓	✓	✓		✓	✓	✓	✓
[18]	✓	✓			✓	✓		
[19]		✓	✓		✓	✓		
[16], [20]		✓					✓	
[21], [22]		✓						✓
[23]		✓					✓	✓

credibility of the data should be mostly taken into consider.

At the beginning of the research, surveys focus on analyse vulnerabilities and risks, which are brought or increased in big data era [10]–[12], [15]. However, the more recent work also analyse more specific technologies or mechanisms to enhance the security [18]–[20], [22]. The security and privacy analyses are comparatively presented in Table I in terms of their consideration of topics and big data process layer/interfaces discussed.

The challenges of operating the distributed and scalable environments, along with the increasing attention on privacy makes security and privacy in big data become a hot topic. In [15], authors first illustrate the privacy-preserving issues of four types of users in data life cycle from their unique perspectives. Then, they mainly focus on how privacy-preserving data publishing(PPDP) is realized in two emerging applications, i.e., social networks and location-based services. Rather than portraying the whole application chain, some researches focus on depicting typical issues. In [12], authors discuss big data security management platform, information security system and relevant laws and regulations. In [10], authors present related research work on five subjective: Hadoop security, cloud security, monitoring and auditing, key management and anoymization.

Data collecting and storing are essential phases in big data applications. The vulnerabilities and enhancements in secure collection and storage have been presented in [18], [19]. [18] discusses the security issue on NoSQL databases; several database products of four types of database, such as key-value database, column-oriented database, document based database and group database, are studied and their merits and weakness are revealed. Access control models are compared as well. [19] not only considers the privacy and data confidentiality in big data, but data provenance and data trustworthiness are also taken into account.

Data mining is the principal process of discovering knowledge. However, since data mining enables efficiently discover valuable, non-obvious information from large volumes of data, it may result in an extraction of sensitive information. [20] and [16] provide reviews on privacy-preserving data mining techniques and analyse those methods.The transformation methods on the original data in order to preserve privacy is classified as randomization methods, anoymization and distributed methods in [20]. The analytics and comparison of privacy preserving in clustering and association rule mining are given as well. In [16], privacy-preserving techniques in data mining, including privacy-preserving aggregation, operations over encrypted data, and de-identification, are reviewed. Besides, a cosine similarity computing protocol was proposed to transform original data.

After data mining and analyzing, basic operations include showing the interesting mining results in proper ways. Linkage of private information is the one of the top security risks in this stage. Thus, PPDP is required to publish useful information while preserving data privacy. Fung et al. systematically summarize and evaluate different approaches in their frequently cited survey [21]. Rashid et al. study the PPDM and PPDP, and present the differences and requirements between PPDP and other related problems [23].

IV. PRIVACY-PRESERVING TRAJECTORY PUBLISHING

In the Section III, the surveys and achievements of security and privacy in big data are reviewed. In this section, we focus on privacy-preserving trajectory publishing techniques in big data.

A. What is Trajectory?

Advancement of wireless communication enables a large number of location based applications and services, along with a massive collection of location information. It is obvious that sharing location information can help improve users' quality of lives, while on the other hand, it may reveal sensitive and private information about individuals.

Compared with single location, a trajectory is an entire set of discrete location samples. There are two major scenarios that we need to protect a user's trajectory data from the privacy leak. One is in on-line location-based services. In this scenario, a user may not want to exactly disclose his or her current location when using a service [24]. The other is the off-line historical trajectories. Based on a collection of trajectories, an adversary may discover an individual's most frequent places, and therefore identify the individual, or even infer sensitive personal information like health condition, religious and sexual preferences [25]. In this paper, we concentrate on the latter.

Trajectory is usually samples of a mobile object's true movements. For the purpose of data analysis, approaches of reducing the uncertainty of a trajectory are studied [26]. On the other hand, to protect a user from the privacy leak caused by the disclosure of the user's trajectories, a trajectory should

be even more uncertain. Thus, privacy-preserving trajectory publishing requires mechanisms to blur a user's trajectory, while ensuring the utility of data at the same time.

B. Mechanisms in Privacy-Preserving Trajectory Publishing

Anonymization technique is an efficient method to realize privacy-preserving, and it can be also utilized for trajectory data set. However, spatio-temporal data, different from relational data, have some unique features, including time dependence, location dependence and high dimensionality [15]. Thus, tailored privacy-preserving methods should be considered. In this subsection, three common mechanisms in privacy-preserving trajectory publishing are presented, and existing research under each class are reviewed.

Generalization and Suppression

Generalization and suppression are the most common anonymity operations used to implement k-anonymity. Generalization means replacing one or multiple specific values with a more general one. Suppression involves deleting values or records of data. A number of solutions provide anonymization protection based on generalization and suppression.

In [27], Terrovitis et al. proposed an anonymization algorithm that iteratively suppresses selected locations from the original trajectories, taking into consider of the benefit in terms of privacy and the deviation from the main direction of the trajectory.

The authors of [31] consider the trajectories as a collection of points, each point represented by intervals on the three dimensions. Then, k-anonymity model is built.

Yarovoy studies the case that each user has a different set of quasi-identifier(QID)(location,time) pairs for which he or she requires protection [32]. Based on the graph theory, the authors build the k-anonymity model.

Generalization and suppression operations are feasible and easy. However, the main negative side is that replacing or deleting real values leads to a high possibility of information loss.

Perturbation

While data semantics are retained at a record level by the generalization and suppression mechanisms, the perturbation techniques retain data semantics at an aggregate level [28]. Perturbation techniques usually are based on randomization. Adding noise and swapping data are common means of perturbation [37], [38]. Some studies based on perturbation have been developed.

Abul et al. consider the problem of publishing a complete sequence of individuals' trajectory [29], [30]. In [29], a (k, θ)-anonymity by space translation is proposed to preserve the individuals' privacy. Then, there are k different trajectories coexisting in a cylinder of the radius θ. It has been later improved by removing some constraints about the input datasets and scales to large datasets at the cost of higher computational requirements [30].

The main idea of [34] is also adding noise, however, instead of spatial distortion, the scheme proposed is built on time distortion. Promesse is designed to smooth the users' speed from original data to a constant speed, and then blur endpoints at the same time. Hence, the users' interests spots and endpoint are preserved.

In [35], a (K, t)-privacy metric based on the idea of swapping data, is proposed. The algorithm is designed to exchange multiple users' pseudonyms only when they meet the same location, so as to eliminate the linkability of their pseudonyms before and after the exchange. This algorithm can be used in the scenario that many people move through hub locations, such as a train station.

However, schemes based on the partition-based privacy model, including generalization, suppression and perturbation, have been found to be vulnerable to many types of privacy attacks, such as composition attack, deFinetti attack, and foreground knowledge attack [39].

Differential Privacy

Differential privacy is recently introduced to privacy preserving data publishing. The privacy preserving model is designed to ensure an equal probability of any released data among all nearly identical input data sets, and due to this reason, it guarantees that all outputs are insensitive to individuals. Adding random noise to the true output of the function is a common method. The Laplace mechanism [40] and the Exponential mechanism [41] are two major techniques. For real outputs, the Laplace mechanism is used that the noises are generated based on Laplace distribution. When outputs are not real, the Exponential mechanism assigns exponentially greater probability to a output with a higher score, with which it is more likely to be selected. As a consequence, the final output would be close to the optimum with respect to utility function. Some research have been working on differential private publication of trajectory data.

Chen et al. first introduce differential privacy to trajectory publishing [42]. A non-interactive data-dependent sanitization algorithm is proposed. The efficiency of the approach is guaranteed by narrowing down the output domain by constructing a noisy prefix tree under Laplace mechanism. Then Chen et al. develop techniques making use of the inherent Markov assumption in the variable-length n-gram model in order to improve the utility [43].

Considering that there are not many common prefixes or n-grams of raw trajectories, Hua et al. propose location generalization algorithm based on the exponential mechanism for preparation; then design authors design a release algorithm which leverages a noise counting scheme based on Laplace mechanism [39].

The differential privacy is proved to have a better private utility for trajectory data, however, the future research should pay more attention to adapt the huge volume of data in real-time scenario.

V. Conclusion

Big data has become one of the most promising and prevailing technology to predict future trends. In these circumstances, security and privacy should be taken into consideration for applications. In this paper, we have analysed the effects of big data characteristics on security and privacy, which requires conceptual and operational study on infrastructure security, data privacy and data management. Then, having surveyed the research on big data security and privacy in big data, a set of topics and issues have been illustrated and compared. Finally, we have focused on privacy-preserving trajectory data publishing mechanisms, due to sensitivity and widely-usage of trajectory data in telecom operation. Three common mechanisms of privacy-preserving trajectory publishing: generalization and suppression, perturbation and differential privacy have been presented, and relative research under each class have been also reviewed.

References

[1] X. Cheng, L. Xu, et al., *A Novel Big Data Based Telecom Operation Architecture*[C], in Proc. 2015 International Conference on Signal and Information Processing(ICSINC), Beijing, China, Oct. 2015.

[2] L. Xu, Y. Luan, et al., *WCDMA Data based LTE Site Selection Scheme in LTE Deployment*[C], in Proc. 2015 International Conference on Signal and Information Processing(ICSINC), Beijing, China, Oct. 2015.

[3] M. Viceconti, P. Hunter, et al., *Big Data, Big Knowledge: Big Data for Personalized Healthcare*[J], IEEE Journal of Biomedical and Health Informatics, vol. 19, no. 4, pp. 1209-1215, July 2015.

[4] M. Herland, T. M. Khoshgoftaar, et al., *Survey of Clinical Data Mining Applications on Big Data in Health Informatics*[C], in Proc. 2013 12th International Conference on Machine Learning and Applications (ICMLA), Miami, FL, 2013, pp. 465-472.

[5] Deng Y, Wang L, et al, *Artificial-Noise Aided Secure Transmission in Large Scale Spectrum Sharing Networks*[J], IEEE Trans. on Communications,vol. 64, no. 5, pp. 2116-2129, May 2016.

[6] A. A. Cardenas, P. K. Manadhata, et al., *Big Data Analytics for Security*[J], IEEE Security and Privacy, vol. 11, no. 6, pp. 74-76, Nov.-Dec. 2013.

[7] Cloud Security Alliance Big Data Working Group, *Expanded Top Ten Big Data Security and Privacy Chanllenges*[R], Apr. 2013.

[8] Y. Deng, L. Wang, M. Elkashlan, et al., *Physical Layer Security in Three-Tier Wireless Sensor Networks: A Stochastic Geometry Approach*[J], in IEEE Trans. on Information Forensics and Security, vol. 11, no. 6, pp. 1128-1138, Jun. 2016.

[9] ISO/IEC JTC 1, *Big Data* [R], Preliminary Report, 2014.

[10] D. S. Terzi, R. Terzi, et al., *A Survey on Security and Privacy Issues in Big Data*[C], in Proc. 2015 IEEE International Conference on Internet Technology and Secured Transactions(ICITST' 2015), 2015.

[11] B. Matturdi, X. Zhou, et al., *Big Data Security and Privacy: A Review*[J], China Communications Magazine, 2014.

[12] M.Yang, X.Zhou, et al., *Challenges and Solutions of Information Security Issues in the Age of Big Data*[J]. China Communications Magazine, Mar. 2016.

[13] D. Mittal, D. Kaur, et al., *Secure Data Mining in Cloud Using Homomorphic Encryption*[C], in Proc. 2014 IEEE International Conference on Cloud Computing in Emerging Markets(CCEM' 2014), Oct. 2014.

[14] NIST, *NIST Big Data Interoperability Framework: Volume 4, Security and Privacy*[R], National Institute for Standards and Technology, 2015, http://dx.doi.org/10.6028/NIST.SP.1500-4.

[15] K. Hu, D. Liu, et al., *Research on Security Connotation and Response Strategies for Big Data*[J], Telecommunications Science, vol.2, pp.112-117, Feb. 2014.

[16] R. Lu, H. Zhu, et al., *Toward Efficient and Privacy-Preserving Computing in Big Data Era*[J], IEEE Network, Aug. 2014.

[17] L. Xu, C. Jiang, et al., *Information Security in Big Data: Privacy and Data Mining*[J]. IEEE Access, vol.2, pp. 1149-1176, Oct. 2014.

[18] E. Sahafizadeh, and M.A.Nematbakhsh, *A Survey on Security Issues in Big Data and NoSQL*[J]. Advances in Computer Science: an International Journal(ACSIJ), vol.4, no.16, pp.68-72, Jul. 2015.

[19] E. Bertino, *Big Data-Security and Privacy*[C]. in Proc. 2015 IEEE International Congress on Big Data, Jun. 2015.

[20] K. Saranya, K. Premalatha, et al., *A Survey on Privacy Preserving Data Mining*[C], in Proc. 2015 IEEE Sponsored 2nd International Conference on Electronics and Communication System(ICECS'15), 2015.

[21] B. C. Fung, K. Wang, et al, *Privacy-Preserving Data Publishing: A Survey of Recent developments*[J], ACM Comput.Surv., vol. 42, no.4, Jun. 2010.

[22] J. Abawajy, M. I. Ninggal, et al., *Privacy Preserving Social Network Data Publication*[J], IEEE Comm. Surveys Tutorials, vol. pp, no.x, Mar. 2016.

[23] A. H. Rashid, and N. Yasin *Privacy Preserving Data Publishing: Review*[J], International Journal of Physical Sciences, vol.10 pp. 239-247, Apr. 2015.

[24] A.G.Divanis, P. Kalnis, et al., *Providing K-Anonymity in Location Based Services*[J]. SIGKDD Explorations, vol.12, no.1, pp.3-10, 2010.

[25] F. Bonchi, L. V. S. Lakshmanan, et al.,*Trajectory Anonymity in Publishing Personal Mobility Data*[J], SIGKDD Explorations, vol.13, no.1, pp.30-42, 2011.

[26] Y. Zheng, *Trajectory Data Mining: An Overview*[J]. ACM Transactions on Intelligent Systems and Technology, vol.6, no.3, Article 29, May 2015.

[27] M. Terrovitis and N. Mamoulis, *Privacy Preservation in the Publication of Trajectories*[C], in Proc 9th International Conference on Mobile Data Management(MDM 2008), pp.65-72, Beijing, Apr. 2008.

[28] Junqiang Liu, *Privacy-Preserving Data Publishing: Current Status and New Directions*[J], Information Technology Journal, vol. 11, no. 1, pp.1-9, 2012.

[29] O. Abul, F. Bonchi, et al., *Never Walk Alone: Uncertainty for Anonymity in Moving Objects Databases*[C], in Proc. 2008 IEEE International Conference on Data Engineering(ICDE'08), 2008.

[30] O. Abul, F. Bonchi, et al., *Anonymization of moving objects databases by clustering and perturbation*[J], Information Systems, vol.35, no. 8, pp. 884-910, Dec. 2010.

[31] M. E. Nergiz, M. Atzori, et al., *Towards Trajectory Anonymization: a Generalization-Based Approach*[C], in Proc. ACM GIS Workshop on Security and Privacy in GIS and LBS, 2008.

[32] R. Yarovoy, F. Bonchi, et al,*Anonymizing moving objects (how to hide a MOB in a crowd?*[J], Extending Database Technology, 2009.

[33] Z. Cai, H. Yang, et al ,*A Clustering-Based Privacy-Preserving Method for Uncertain Trajectory Data*[C]. in Proc. 2014 IEEE 13th International Conference on Trust, Security and Privacy in Computing and Communications(TrustCom), Beijing, Sep. 2014.

[34] V. Primault, S. B. Mokhtar, et al. *Time Distortion Anonymization for the Publication of Mobility Data with High Utility*[C], in Proc. 2015 IEEE Trustcom/BigDataSe/ISPA, Helsinki, Aug. 2015.

[35] K. Mano, K. Minami, et al.,, *Pseudonym Exchange for Privacy-Preserving Publishing of Trajectory Data Set*[C]. in Proc. 2014 IEEE 3rd Global Conference on Consumer Electronics (GCCE), Tokyo, 2014.

[36] G. Poulis, S. Skiadopoulos, et al., *Distance-Based k^m-Anonymization of Trajectory Data*[C], in Proc. 2013 IEEE 14th International Conference on Mobile Data Management(MDM), Milan, Jun. 2013.

[37] Y. Xu, T. Ma, et al., *A Survey of Privacy Preserving Data Publishing using Generalization and Suppression*[J]. Appl. Math. Inf. Sci. vol. 8, no. 3, pp. 1103-1116, 2014.

[38] A. Al-Talabani, Y. Deng, et al., *Enhancing Secrecy Rate in Cognitive Radio Networks via Multilevel Stackelberg Game*[J], IEEE Commun. Letters, vol. 20, no. 6, pp. 1112-1115, Jun. 2016.

[39] J. Hua, Y. Gao, et al.,*Differentially Private Publication of General Time-Serial Trajectory Data*[C], in Proc. 2015 IEEE Conference on Computer Communications(INFOCOM), pp.549-557, Kowloon, Apr. 2015.

[40] C. Dwork, F. Mcsherry, et al., *Calibrating Noise to Sensitivity in Private Data Analysis*[C], in Proc. Theroy of Cryptography Conference, Jan. 2006.

[41] F. Mcsherry, and K. Talwar, *Mechanism Design via Differential Privacy*[C], in Proc. Foundations of Computer Science, 2007.

[42] R. Chen, B. C. M Fung, et al., *Differentially Private Trajectory Data Publication*[J]. Arxiv e-prints, Dec. 2011.

[43] R. Chen, G. Acs, et al., *Differentially Private Sequential Data Publication via Variable-Length N-Grams*[C], in Proc. ACM Computer and Communication Security (CCS), Oct 2012, Raleigh, United States. 2012.