A Mapping Study on Privacy Attacks in Big Data and IoT

Raisa Islam
Computer Science and Engineering
New Mexico Tech
Socorro, USA
raisa.islam@student.nmt.edu

Mohammad Sahinur Hossen
Computer Science and Engineering
New Mexico Tech
Socorro, USA
mohammad.hossen@student.nmt.edu

Dongwan Shin
Computer Science and Engineering
New Mexico Tech
Socorro, USA
dongwan.shin@nmt.edu

2022 13th International Conference on Information and Communication Technology Convergence (ICTC) | 978-1-6654-9939-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICTC55196.2022.9952824

***Abstract*—Application domains like big data and IoT require a lot of user data collected and analyzed to extract useful information, and those data might include user's sensitive and personal information. Hence, it is strongly required to ensure the privacy of user data before releasing them in the public space. Since the fields of IoT and big data are constantly evolving with new types of privacy attacks and prevention mechanisms, there is an urgent need for new research and surveys to develop an overview of the state-of-art. We conducted a systematic mapping study on selected papers related to user privacy in IoT and big data, published between 2010 to 2021. This study focuses on identifying the main privacy objectives, attacks and measures taken to prevent the attacks in the two application domains. Additionally, a visualized classification of the existing attacks is presented along with privacy metrics to draw similarities and dissimilarities among different attacks.**

***Index Terms*—Big Data, IoT, privacy, attack**

I. Introduction

The concept of smart environment introduced a novel paradigm, *Internet of Things* (IoT), which is a rapidly growing network of interconnected physical objects (the *things*). According to the report of Statistica, the worldwide number of the total installed base of connected IoT devices will be approximately 75.44 billions by 2025 [1]. All these interconnected devices collect, store, process and exchange an enormous amount of information with each other. Big data and its analysis, associated with IoT, have drawn a lot of attention in recent years, and the characteristics of big data are defined by the three V's: *volume*, *velocity* and *variety*. Volume indicates the amount of data which is received in real-time or in batches. On the other hand, velocity is the rate of data received. Since data is collected from various sources, the heterogeneous nature of data is referred as variety. Benefited from those characteristics, the IoT data collection and analysis can be used for many purposes such as autonomous suggestions and business analytics. Since the collected data may contain sensitive user data, releasing these kinds of data should be done with the utmost caution for privacy reasons.

Identifying privacy goals and the types of attacks that affect them can help us to better understand how sharing private data hinders privacy. Several surveys have been conducted on IoT and big data in the past since both fields are very popular [2, 3, 4]. Nevertheless, both fields are still evolving with new types of attacks and prevention mechanisms, so new research and surveys are urgently required to develop a comprehensive overview. In this paper, we performed a systematic mapping study to identify the main privacy objectives, attacks and measures taken to prevent the attacks in IoT and big data domains. Specifically, we developed the following key research questions to study different topics and methods for our mapping study:

1) What are the user privacy objectives in big data and IoT?
2) What type of attacks were there to breach each privacy objective?
3) Is there any categorical approach to classify the attacks?
4) What kind of measures are taken to prevent these attacks?

In our mapping study, we aim to understand what are the privacy objectives with respect to IoT and Big Data, the types of privacy attack that violate those objectives and the existing solutions to prevent these attacks.

This paper is organised as follows : In Section II we discuss background and related work. Section III describes our research methodology. Results and analysis of our study are presented in Section IV. In Section V we discuss different types of privacy attack. Finally we conclude our paper in Section VI.

II. Background

Privacy can be described as the right of an entity to be secure from unauthorized disclosure of sensitive information from any unwanted party. A privacy attack leverages seemingly benign data to deduce personal information. By combining anonymized data from one dataset with non-anonymized data from another dataset, an attacker can violate user privacy. Since a huge amount of user data is gathered for various purposes in IoT and big data, it is easier to connect the dots and comprehend present behaviors of a user, predict subsequent behaviours, and create deep and complete profiles of the user's lives and interests.

A better understanding of privacy and privacy metrics can help the data owner to be more aware of these attacks. However, there are no general rules to decide on privacy metrics. In [5, 6, 7], researchers tried to identify privacy metrics, but in our research perspective those metrics are

unhelpful. In this paper, the term privacy metric will be used as a standard of privacy criterion in whole/partial. Since one single metric can not cover all the aspects of privacy thus we will use a combination of multiple metrics. For rest of our paper, we will use the followings as privacy metrics:

- **Anonymity** - Within a system, it is impossible to identify an item of interest (IoI). For example, the anonymity of a sender requires that the other users are unable to determine the identity of the sender within a set of potential senders.
- **Unlinkability** - Within a system, it is impossible to identify if two or more entities are related or not. If two events occur within a set, i.e., sending messages, unlinkability refers that an attacker will not be able to differentiate if the messages were sent by the same sender or different senders.
- **Undetectability** - Within a system, the attacker cannot adequately recognize whether an IoI exits or not. If we consider messages as IoIs, then undetectability means that these messages are indiscernible from random noises.
- **Unobservability** - Within a system, the same type of IoIs are indistinguishable. From the perspective of sending/receiving messages, unobservability refers that it is sufficiently undetectable whether any sender has sent any messages to a set of could-be recipients.

To preserve sensitive information, several countermeasures have been proposed in the past. We can classify these techniques into following categories:

- **Statistical** techniques that apply perturbation based on probability distribution. Differential privacy, k-anonymity are examples of few popular heuristic-based approach. Most of the recent aggregated data protection mechanisms use this approach.
- **Cryptographic** techniques that apply different encryption algorithms such as blockchain, multi-party computation e.t.c. This approach is mainly designed for distributed scenario.

III. Research Methodology

In order to conduct our mapping study, we have defined the research questions, planned search strategy, screened relevant sources, and finally analyzed the results. In this section, we will discuss designing research questions, planning the search strategy and filtering sources. Result will be presented in the next section.

A. Research Questions

To determine the state of the literature, we have identified the following research questions showed in Table I.

B. Search Strategy

We have designed our search string (shown in Table II) in such manner so that the resulting papers can answer our research questions. Since the primary focus of this study is user privacy in the field of IoT and big data, the papers must contain the phrase "privacy" as author keywords and either "iot" or "big data" must be found within the abstract of the paper. Other words are used to make sure that the resulted paper answers at least one of the research questions. Additionally, as IoT and big data is a vast research field, we have applied filter to trim down any paper that dated back than 2010. The main digital libraries that were used to search for primary studies were: IEEEXplore, ACM Digital Library etc. Unfortunately, the search facility provided by those database differs in many ways, hence the search string had to be transformed according to the native syntax of each database.

RQ#	Question	Rationale
RQ1	What are the user privacy objectives in big data and IoT?	The answer delivers an overview of main goals in our filed of interest
RQ2	What type of attacks are performed to breach each privacy objective?	This question answers the ways user privacy is compromised
RQ3	Is there any existing categorical approach to classify attacks that hamper those objectives?	This answer will help to gain an overview of classification of existing attacks and their similarity with each other.
RQ4	What kind of privacy measures are taken to prevent these attacks?	This answers the most important question about the available approaches to protect user privacy

TABLE I
RESEARCH QUESTIONS OF OUR STUDY

Search String
("Author Keywords":"privacy") AND ("Abstract":"iot" OR "Abstract":"big data") AND ("Abstract":"attack" OR "Abstract":"threat" OR "Abstract":"challenge") AND ("Abstract":"solution" OR "Abstract":"method" OR "Abstract":"technique" OR "Abstract":"mechanism")

TABLE II
SEARCH STRING

C. Screening

After applying our search string on digital databases, we skimmed through the resulted papers to further filter out irrelevant sources. We excluded the sources that were:

- either mapping study or survey paper
- not relevant to information technology
- full-text not available or incomplete paper
- written in languages other than English

Moreover, we screened the bibliography of resulting papers to identify other possible relevant sources to our research.

IV. Result

A total of 580 papers resulted after applying the search string over several digital libraries. However, after screening process, the number of useful papers scaled downed to a smaller number. Table III shows the overall number of papers returned from each digital library, as well as the number of papers that were relevant following screening. We have applied filter to trim any publication that dated back than 2010 to perceive a most recent overview of the literature.

A. Research Question 1

Our first research question deals with user privacy objectives in big data and IoT. Individual privacy can be protected by anonymizing personal information, shifting attribute values

Source	I. Number	F. Number
IEEE Xplore	87	41
ACM Digital Library	12	11
Elsevier ScienceDirect	271	0
Semantic Scholar	23	3
arXiv	174	5
AIS eLibrary	13	0

TABLE III
SEARCH RESULTS FROM DIGITAL REPOSITORIES

across people, or introducing random noise. When disclosing information, the definition of privacy must specify what information is sensitive and needs to be safeguarded, how disclosure occurs, and when privacy disclosure occurs [8]. In this mapping study, we categorized the privacy goals stated in the resulted papers. Depending on the findings, we observed that there are three privacy objectives described as followings

- **Data Privacy** - Data privacy refers to keeping the private and sensitive data safe from any adversary. Nowadays, a large amount of data is available in the public domain for research and other purposes, which can be misused by malicious parties.
- **Location Privacy** - With the rapidly growth of location based services, many people share their location in the public domains. These information could be used by malicious parties to harm the user.
- **Identity Privacy** - An individual's identity privacy refers to protecting all their attributes from adversaries, like their name, address, health card number, etc.

Table IV presents the reference of the papers that discussed the above privacy goals.

Privacy Goal	Count	Reference
Data privacy	35	[8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25] [26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42]
Identity privacy	7	[43, 44, 45, 46, 47, 48, 49]
Location privacy	8	[50, 51, 52, 53, 54, 55, 56, 57]

TABLE IV
PRIVACY GOALS COVERED IN SOURCES

B. Research Question 2

Our second research question deals with the type of attacks performed to breach user privacy. There exists a variety of attacks which aim to compromise user privacy. In this section, organized different types of attacks mentioned in the resulted sources. Table V shows the attacks along with reference source. From the table it is visible that Inference attack, Membership inference attack (MIA), Linkage attack and Collusion attacks are the most popular privacy attacks. All the attacks are discussed in Section V.

C. Research Question 3

Following the answer to the second research question, one can ask whether there is any existing approach to categorizing privacy attacks. A proper categorization of privacy attacks will provide a better understanding of similar attacks and how they violate user privacy. However, there is no existing categorization of attacks based on privacy in our study resources. According to the types of data, we classified privacy attacks into two major branches: aggregated data attacks and single stream data attacks. Furthermore, aggregated data attacks can be divided into two sub-branches; sufficient public information and compulsatory auxiliary information. Figure 1 shows our approach to categorize privacy attacks including the privacy metrics which were violated by those attacks.

Attack	Count	Reference
Linkage Attack	6	[8, 10, 11, 43, 46, 47]
Database linking Attack	1	[13]
Graph Matching Attack	2	[12, 48]
Inference Attack	12	[8, 14, 15, 16, 17, 18, 19, 20, 41, 50, 52, 54]
Differencing Attack	1	[8]
Membership Inference Attack	7	[24, 25, 26, 27, 57, 56, 58]
Reconstruction Attack	4	[8, 9, 21, 22]
Location Attack	4	[34, 37, 55, 56]
Correlation Attack	1	[8]
NILM Attack	2	[38, 59]
Collusion Attack	10	[28, 29, 30, 31, 32, 33, 35, 39, 40, 41]

TABLE V
ATTACKS COVERED IN SOURCES

D. Research Question 4

Our last research question answers the most important question about the available privacy measures to protect user privacy. Table VI represents the solutions along with count and reference sources.

Solution	Count	References
Differential Privacy	13	[11, 13, 15, 17, 18, 21, 22, 25, 26, 27, 31, 49, 51]
Blockchain and Smart contract	2	[30, 46]
k-anonymity variation	6	[10, 12, 13, 34, 52, 53]
Sanitization and Randomization	5	[9, 20, 23, 37, 43]
Trust model	2	[29, 33]
Cryptographic techniques	4	[35, 39, 40, 41]
Others	8	[14, 16, 28, 36, 38, 42, 47, 54]

TABLE VI
PRIVACY SOLUTIONS COVERED IN SOURCES

Differential privacy (DP) is one of the most popular privacy preserving methods that allows to share aggregated data by adding noise with the promise that no one will be able to distinguish if a dataset contains data of a specific user [60]. Multiple authors proposed personalized DP solutions to prevent attacks, i.e., inference attack, database linking attack, collusion attack etc. k-anonymity is another popular privacy-preserving method. The basic idea of k-anonymity is that each released record should be indistinguishable from at least (K-1) others in a block. To provide better privacy, multiple variants has been proposed by extending the concept of k-anonymity such as l-diversity, t-closeness etc. k-anonymity based solutions can be used to prevent background knowledge attack, inference attack, database linking attack etc. Other than these two, different cryptographic techniques, data sanitization and randomization, etc., are also adopted by the researchers to prevent privacy attacks.

V. PRIVACY ATTACKS

In this section, we briefly discuss some of the important privacy attacks identified in the previous section.

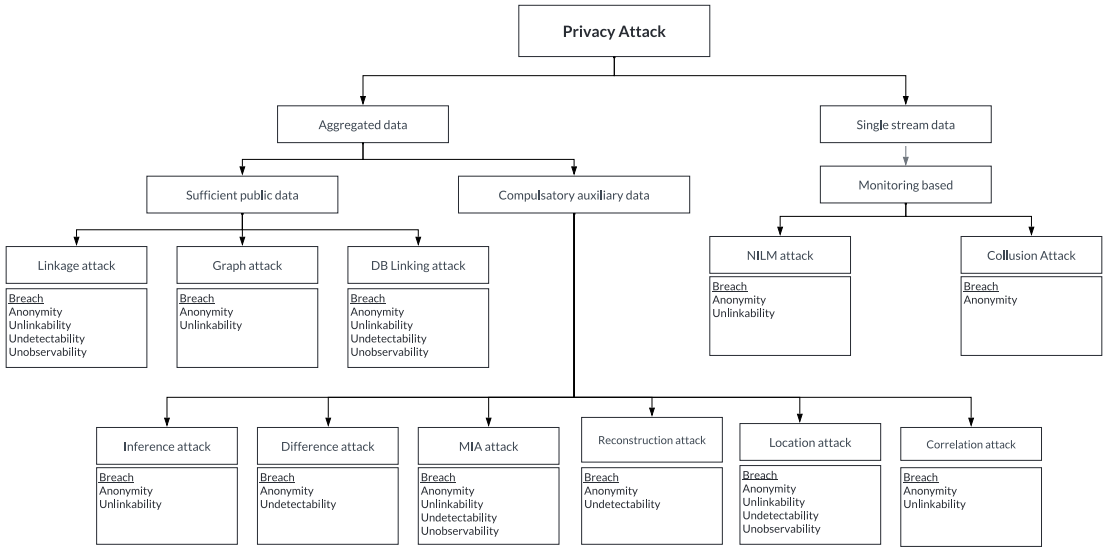

Fig. 1. Privacy attack categorization tree

A. Aggregated data attack

Aggregated data is a collection of data that come from multiple sources. Based on the requirement of auxiliary information, we can sub-divide aggregated data attacks into two categories.

1) Sufficient public information: Publicly available information is sufficient to reveal individual information.

- *Linkage attack* - Linkage attack aims to re-identify people by linking data from one dataset with the data of another one [8]. The 'linking' employs indirect identifiers, also known as quasi-identifiers, such as zip or postcode, gender, salary etc [10, 11, 43].
- *Graph attack* - Graph attack extracts a collection of characteristics from the encoded and plain-text databases by creating two similarity graphs. This is done by identifying groups of comparable feature vectors and building a bipartite graph connecting nodes based on their similar feature vectors using locality sensitive hashing [12], and an one-to-one matching uncovers sensitive information.
- *Database linking attack* - Database linking attack allows an attacker to join two datasets and obtain additional information for a specific row without knowing a person's identity [13].

2) Compulsatory auxiliary information: Adversaries require additional information i.e., background information or location information to reveal individual information.

- *Inference attack* - Different data mining techniques such as bayesian reasoning and association rules are used to launch attack in order to obtain private information from publicly available information added with some other auxiliary information.
- *Difference attack* - Differencing attack leverages the differences between multiple statistical queries or aggregate queries across datasets to identify people [8].
- *Membership inference attack(MIA)* - MIA leverages the fact that machine learning models perform differently on training and test data [27]. Based on the output of the target model for the supplied record, MIA predicts if a particular record was utilized in training a target model [24, 25, 26].
- *Reconstruction attack* - Adversaries usually have limited access to private datasets, but can access aggregate statistics. If the statistics are not sufficiently skewed, it is possible to reconstruct an accurate probabilistic version of the original dataset using multiple queries [8].
- *Location attack* - In location attack, an adversary uses location information to predict a user's private information, such as social affiliations, personal activities, or future physical location [34, 37, 55, 56].
- *Correlation attack* - An adversary with background knowledge or auxiliary information on correlation i.e., interactive links in social networking data and spatio-temporal correlations in position trajectories, has a significant possibility of getting private information and thus infringing on privacy [8].

B. Single stream data attack

A collection of data from a single source is referred as single stream data. Single stream data attacks are mostly monitoring based. A monitoring attack involves tracking down a single data source, leading to classification and identification of sensitive information.

- *Non intrusive load monitoring(NILM) attack* - A NILM attack is a load prediction approach that employs energy disaggregation on smart meters via false data injection [59, 38].
- *Collusion attack* - Two or more adversaries with a shared interest can initiate collusion attacks. Each attacker has a piece of raw data with varied injected noise from the

same source. As a result, launching a collusion assault in a fog computing situation is optimal [31, 32, 33].

VI. Conclusion

In this systematic mapping study, we studied different types of user privacy attacks in IoT and big data and classified them to better understand the nature of those attacks, thereby promoting further research in this important and interesting field of information security. Additionally, this research made a contribution to the development of privacy preserving techniques by providing a reference for their threat model development. Our immediate future work will focus on the development of a formal computing model to evaluate or compare various privacy preserving techniques used in the IoT and big data environments.

Acknowledgment

This work was supported at the Secure Computing Laboratory at New Mexico Tech by the grant from National Science Foundation (NSF-OIA-1757207).

References

[1] Internet of things (iot) and non-iot active device connections worldwide from 2010 to 2025, 2021.

[2] S. Yu. Big privacy: challenges and opportunities of privacy study in the age of big data. *IEEE Access*, 4, 2016.

[3] P. Goswami and S. Madan. Privacy preserving data publishing and data anonymization approaches: a review. In *2017 International Conference on Computing, Communication and Automation (ICCCA)*, 2017.

[4] S. Imtiaz, R. Sadre, and V. Vlassov. On the case of privacy in the iot ecosystem: a survey. In *2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, 2019.

[5] I. Wagner and D. Eckhoff. Technical privacy metrics: a systematic survey. *ACM Comput. Surv.*, 51(3), June 2018.

[6] E. Bertino, D. Lin, and W. Jiang. *A survey of quantification of privacy preserving data mining algorithms*. In Oct. 2008.

[7] I. Wagner and I. Yevseyeva. Designing strong privacy metrics suites using evolutionary optimization. *ACM Trans. Priv. Secur.*, 24(2), Jan. 2021.

[8] T. Wang, Z. Zheng, M. H. Rehmani, S. Yao, and Z. Huo. Privacy preservation in big data from the communication perspective—a survey. *IEEE Communications Surveys Tutorials*, 21(1), 2019.

[9] K. Kasichainula, H. Mansourifar, and W. Shi. Privacy preserving proxy for machine learning as a service. In *2020 IEEE International Conference on Big Data (Big Data)*, 2020.

[10] N. Elanshekhar and R. Shedge. An effective anonymization technique of big data using suppression slicing method. In *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, 2017.

[11] K. Jung, J. Kim, Y. Kim, and S. Park. Drako: differentially private algorithm to meet k-anonymity for online portal service. In *2018 IEEE International Conference on Big Data (Big Data)*, 2018.

[12] W. Wang, Q. Mu, Y. Pu, D. Man, W. Yang, and X. Du. Sensitive labels matching privacy protection in multi-social networks. In *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, 2020.

[13] Y. Park, J. Ghosh, and M. Shankar. Perturbed gibbs samplers for generating large-scale privacy-safe synthetic health data. In *2013 IEEE International Conference on Healthcare Informatics*, 2013.

[14] G. Yang, S. Luo, H. Zhu, Y. Xin, K. Xiao, Y. Chen, M. Li, and Y. Wang. A mechanism to improve effectiveness and privacy preservation for review publication in lbs. *IEEE Access*, 7, 2019.

[15] T. Jung, K. Jung, S. Park, and S. Park. A noise parameter configuration technique to mitigate detour inference attack on differential privacy. In *2017 IEEE International Conference on Big Data and Smart Computing (BigComp)*, 2017.

[16] K. Li, G. Luo, Y. Ye, W. Li, S. Ji, and Z. Cai. Adversarial privacy-preserving graph embedding against inference attack. *IEEE Internet of Things Journal*, 8(8), 2021.

[17] J. Liu, C. Zhang, and Y. Fang. Epic: a differential privacy framework to defend smart homes against internet traffic analysis. *IEEE Internet of Things Journal*, 5(2), 2018.

[18] Z. He and J. Zhou. Inference attacks on genomic data based on probabilistic graphical models. *Big Data Mining and Analytics*, 3(3), 2020.

[19] H. Shang and Z. He. Kin genomic data inference attacks through factor graph. In *2019 IEEE 16th International Conference on Mobile Ad Hoc and Sensor Systems Workshops (MASSW)*, 2019.

[20] Z. He, Z. Cai, and J. Yu. Latent-data privacy preserving with customized data utility for social network data. *IEEE Transactions on Vehicular Technology*, 67(1), 2018.

[21] W. Xue, Y. Shen, C. Luo, W. Hu, and A. Seneviratne. Acies: a privacy-preserving system for edge-based classification. In *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*, 2018.

[22] Q. Zhang, J. Ma, Y. Xiao, J. Lou, and L. Xiong. Broadening differential privacy for deep learning against model inversion attacks. In *2020 IEEE International Conference on Big Data (Big Data)*, 2020.

[23] L. Yao, D. Liu, X. Wang, and G. Wu. Preserving the relationship privacy of the published social-network data based on compressive sensing. In *2017 IEEE/ACM 25th International Symposium on Quality of Service (IWQoS)*, 2017.

[24] S. Truex, L. Liu, M. E. Gursoy, L. Yu, and W. Wei. Demystifying membership inference attacks in machine learning as a service. *IEEE Transactions on Services Computing*, 2019.

[25] B. Zhang, R. Yu, H. Sun, Y. Li, J. Xu, and W. Wang. Privacy for all: demystify vulnerability disparity of differential privacy against membership inference attack. *ArXiv*, abs/2001.08855, 2020.

[26] J. Chen, W. Wang, and X. Shi. Differential privacy protection against membership inference attack on machine learning for genomic data. In *PSB*, 2021.

[27] M. Rahman, T. Rahman, R. Laganière, and N. Mohammed. Membership inference attack against differentially private deep learning model. *Trans. Data Priv.*, 11, 2018.

[28] L. Dong, X. Chen, J. Zhu, H. Chen, K. Wang, and C. Li. A secure collusion-aware and probability-aware range query processing in tiered sensor networks. In *2015 IEEE 34th Symposium on Reliable Distributed Systems (SRDS)*, 2015.

[29] V. B. Reddy, A. Negi, S. Venkataraman, and V. R. Venkataraman. A similarity based trust model to mitigate badmouthing attacks in internet of things (iot). In *2019 IEEE 5th World Forum on Internet of Things (WF-IoT)*, 2019.

[30] N. S. Patel, P. Bhattacharya, S. B. Patel, S. Tanwar, N. Kumar, and H. Song. Blockchain-envisioned trusted random oracles for iot-enabled probabilistic smart contracts. *IEEE Internet of Things Journal*, 2021.

[31] X. Wang, B. Gu, Y. Qu, Y. Ren, Y. Xiang, and L. Gao. Reliable customized privacy-preserving in fog computing. In *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, 2020.

[32] Z. Zhu, Z. Jiang, and R. Jiang. The attack on mona: secure multi-owner data sharing for dynamic groups in the cloud. In *2013 International Conference on Information Science and Cloud Computing Companion*, 2013.

[33] R. Dwarakanath, B. Koldehofe, Y. Bharadwaj, T. A. B. Nguyen, D. Eyers, and R. Steinmetz. Trustcep: adopting a trust-based approach for distributed complex event processing. In *2017 18th IEEE International Conference on Mobile Data Management (MDM)*, 2017.

[34] J. Xiong, M. Zhao, M. Z. A. Bhuiyan, L. Chen, and Y. Tian. An ai-enabled three-party game framework for guaranteed data privacy in mobile edge crowdsensing of iot. *IEEE Transactions on Industrial Informatics*, 17(2), 2021.

[35] S. Badsha, I. Khalil, X. Yi, and M. Atiquzzaman. Designing privacy-preserving protocols for content sharing and aggregation in content centric networking. *IEEE Access*, 6, 2018.

[36] A. Alshehri, J. Granley, and C. Yue. Attacking and protecting tunneled traffic of smart home devices. In *Proceedings of the Tenth ACM Conference on Data and Application Security and Privacy*, CODASPY '20, New Orleans, LA, USA. Association for Computing Machinery, 2020.

[37] A. Ukil, S. Bandyopadhyay, J. Joseph, V. Banahatti, and S. Lodha. Negotiation-based privacy preservation scheme in internet of things platform. In *Proceedings of the First International Conference on Security of Internet of Things*, SecurIT '12, Kollam, India. Association for Computing Machinery, 2012.

[38] R. Dong, L. J. Ratliff, A. A. Cárdenas, H. Ohlsson, and S. S. Sastry. Quantifying the utility–privacy tradeoff in the internet of things. *ACM Trans. Cyber-Phys. Syst.*, 2(2), May 2018.

[39] X. Yang, R. Lu, J. Shao, X. Tang, and H. Yang. An efficient and privacy-preserving disease risk prediction scheme for e-healthcare. *IEEE Internet of Things Journal*, 6(2), 2019.

[40] X. Guo, H. Lin, C. Xu, and W. Lin. A data clustering strategy for enhancing mutual privacy in healthcare system of iot. In *2019 IEEE International Conferences on Ubiquitous Computing Communications (IUCC) and Data Science and Computational Intelligence (DSCI) and Smart Computing, Networking and Services (SmartCNS)*, 2019.

[41] H. Pang and B. Wang. Privacy-preserving association rule mining using homomorphic encryption in a multikey environment. *IEEE Systems Journal*, 15(2), 2021.

[42] K. Alanezi and S. Mishra. A privacy negotiation mechanism for the internet of things. *2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech)*, 2018.

[43] A. Pandya, P. Kostakos, H. Mehmood, M. Cortes, E. Gilman, M. Oussalah, and S. Pirttikangas. Privacy preserving sentiment analysis on multiple edge data streams with apache nifi. In *2019 European Intelligence and Security Informatics Conference (EISIC)*, 2019.

[44] A. Acar, H. Fereidooni, T. Abera, A. K. Sikder, M. Miettinen, H. Aksu, M. Conti, A.-R. Sadeghi, and S. Uluagac. Peek-a-boo: i see your smart home activities, even encrypted! In *Proceedings of the 13th ACM Conference on Security and Privacy in Wireless and Mobile Networks*, WiSec '20, Linz, Austria. Association for Computing Machinery, 2020.

[45] O. Tomashchuk. Threat and risk management framework for ehealth iot applications. In *Proceedings of the 24th ACM International Systems and Software Product Line Conference - Volume B*, SPLC '20, Montreal, QC, Canada. Association for Computing Machinery, 2020.

[46] A. Dorri, S. S. Kanhere, and R. Jurdak. Towards an optimized blockchain for iot. In *Proceedings of the Second International Conference on Internet-of-Things Design and Implementation*, IoTDI '17, Pittsburgh, PA, USA. Association for Computing Machinery, 2017.

[47] F. C. Sangogboye, R. Jia, T. Hong, C. Spanos, and M. B. Kjærgaard. A framework for privacy-preserving data publishing with enhanced utility for cyber-physical systems. *ACM Trans. Sen. Netw.*, 14(3–4), Nov. 2018.

[48] A. Vidanage, P. Christen, T. Ranbaduge, and R. Schnell. A graph matching attack on privacy-preserving record linkage. *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 2020.

[49] J. Ren, Z. Yao, J. Xiong, Y. Zhang, and A. Ye. A secure data deduplication scheme based on differential privacy. In *2016 IEEE 22nd International Conference on Parallel and Distributed Systems (ICPADS)*, 2016.

[50] B. Li, H. Zhu, and M. Xie. Lisc: location inference attack enhanced by spatial-temporal-social correlations. In *2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, 2019.

[51] L. Ou, Z. Qin, Y. Liu, H. Yin, Y. Hu, and H. Chen. Multi-user location correlation protection with differential privacy. In *2016 IEEE 22nd International Conference on Parallel and Distributed Systems (ICPADS)*, 2016.

[52] L. Zhang, C. Fang, Y. Li, H. Zhu, and M. Dong. Optimal strategies for defending location inference attack in database-driven crns. In *2015 IEEE International Conference on Communications (ICC)*, 2015.

[53] M. Ma. Enhancing privacy using location semantics in location based services. In *2018 IEEE 3rd International Conference on Big Data Analysis (ICBDA)*, 2018.

[54] H. Xu, Y. Zheng, J. Zeng, and C. Xu. Location-semantic aware privacy protection algorithms for location-based services. In *2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, 2018.

[55] H. B. Alharbi, N. Abdulrazak Baghanim, and A. Munshi. Cyber risk in internet of things world. In *2020 3rd International Conference on Computer Applications Information Security (ICCAIS)*, 2020.

[56] G. Zhang, A. Zhang, and P. Zhao. Locmia: membership inference attacks against aggregated location data. *IEEE Internet of Things Journal*, 7(12), 2020.

[57] A. Pyrgelis, C. Troncoso, and E. D. Cristofaro. Knock knock, who's there? membership inference on aggregate location data. *ArXiv*, abs/1708.06145, 2018.

[58] L. Van De Kamp, C. Ugwuoke, and Z. Erkin. Economy: ensemble collaborative learning using masking. In *2019 IEEE 30th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC Workshops)*, 2019.

[59] N. Iliaee, S. Liu, and W. Shi. Non-intrusive load monitoring based demand prediction for smart meter attack detection. In *2021 International Conference on Control, Automation and Information Sciences (ICCAIS)*, 2021.

[60] J. Marks, B. Montano, J. Chong, M. Raavi, R. Islam, T. Cerny, and D. Shin. Differential privacy applied to smart meters: a mapping study. In *Proceedings of the 36th Annual ACM Symposium on Applied Computing*, SAC '21, New York, NY, USA. Association for Computing Machinery, 2021.