JES
Journal of Engineering Sciences
Vol 13, Issue 01, Jan /2022
ISSN NO:0377-9254

SPCSS: Social Network Based Privacy-Preserving Criminal Suspects Sensing

[1]Podila Poojasree, [2]Mrs.P. Chaitnaya
[1]M.Tech [CSE], Eluru college of Engineering and Technology, Duggirala - 534004, Andhra Pradesh
[2]Professor, Eluru college of Engineering and Technology, Duggirala - 534004, Andhra Pradesh

Abstract— With development of online social networks, many criminal suspects use social network to communicate with each other. In order to obtain valuable criminal clues, considerable research works have been done to analyze criminal suspects' social data. However, most of them did not pay much attention on privacy-preserving problems, which may leak some sensitive data in the analysis process. To solve this problem, we propose a novel analysis approach of criminal suspects by exploiting social data and crime data that are collected by social network and police information systems. We enable the social cloud server and public security cloud server to exchange social information of criminal suspects and user's public information in a privacy- preserving way. Specifically, we propose a privacy-preserving data retrieving method based on oblivious transfer to guarantee that only the authorized entities can perform queries on suspects' social data, while the social cloud server cannot infer anything during the query. Moreover, several building blocks, such as encrypted data comparing, secure classification and regression tree (CART) model are also proposed. Based on these building blocks, we designed a privacy-preserving criminal suspects sensing scheme. Finally, we demonstrate a performance evaluation which shows that our scheme can enhance analysis of criminal suspects without privacy leakage, while with low overhead.

Index Terms— Classifier, criminal suspects analysis, decision tree, privacy-preserving, social network.

I. INTRODUCTION

WITH the continuous development of the Internet, online social networks have emerged rapidly, such as WeChat, Facebook, and Twitter, which has greatly changed the way people communicate, expanded people's social circle, and abstracted people's concern on social network analysis and mining. At the same time, criminal behavior is also emerging towards gang and organizational development. From a psychological and sociological point of view, people with strong social relations and similar spatial trajectories (such as, frequent access in the same internet cafe) are possible to be of the same group. One traditional approach of gang criminal suspects' investig ation is to determine the specific target of several suspects in advance, and manually monitor and collect information of specific suspects to discover other related criminal suspects or criminal gangs that are closely related to. In such a scenario, the police needs to equip enough human and material resources, which undoubtedly increases labor costs, material and financial expenses, and even causes anxiety or panic of the society.

To resolve such problem, a cloud server associated with crime analysis was established by the police to continuously collect information related with public security, i.e., location, criminal records, and credibility in image and text format. The server uses these data to analyze the potential connections among the suspects and provide clues for excavating criminal gangs, and excluding undiscovered

suspects [1]. Moreover, it helps to analyze whether the user is a suspect. However, it is the lack of sufficient social information to infer whether there are any potential suspects in their personal social circle [2]. Considerable applications in social networks were proposed to analyze the user's social data during their social interaction [3]. For example, the flow of funds from banks and the purchase records of e-commerce can help alert crimes; face recognition technology can help locate suspects through online photo identification. The combination of these social data and monitored personal data can strengthen the analysis of criminal suspects. Suppose Eve is a specific suspect arrested by the police, and grants the police the access authorization, and if police finds that Alice frequently contacts with Eve, who has several criminal records before, thus, Alice has high possibility to be in a potential suspected crime. Personal data, i.e., criminal records, location, credibility, and social data, i.e., contact duration, contact frequency, are usually collected and stored by different service providers, such as police's cloud server and social network service providers (Twitter). To protect data privacy, data sharing among these parties becomes very important for the analysis of potential criminal suspects [4], [5].

Meanwhile, both personal data and social data, such as criminal records and contact information, are sensitive [4], [6]. For a specific criminal suspect ui, the police can obtain the ui's social data from service providers. The analysis service provider (ASP) hosts a learned model, and provides suspects analysis service for the police to use such a model remotely. In such a scenario, the personal and social data are private to the suspects which should be protected against the service providers, while the model is a valuable asset to classifier owner, which should not be disclosed to untrusted party, and analysis data and classification results are also private to the police. To solve such a problem, personal and social data are encrypted and stored in service providers, and through data sharing, police can securely obtain the plaintext of personal and social data. Moreover, the analysis data should also be in ciphertext form when the police submits it to the ASP for analysis. However, such method may limit the data processing ability of the ASPs [7]. Therefore, it is a serious challenge to complete the data analysis while protecting privacy of potential criminal suspects. In addition, the query target and results are valuable assets to the police, which may contain some sensitive information about specific suspects and unknown suspects, such as identity, which should also be protected against service providers. Therefore, access pattern protection is also a hard task when using social data to strengthen the analysis of potential suspects. In this article, we propose a privacy-preserving criminal suspects sensing (SPCSS) scheme considering social data associated with personal data to perform criminal suspects analysis. This scheme employs a privacy-preserving data retrieving (PPDR) method based on oblivious transfer to enable access pattern protection, and several building blocks to construct SPCSS to enable the cloud servers to infer criminal suspects status, and preserve data privacy using classification and regression tree (CART) model. The main contributions of this article are as follows.

1) First, we analyze the organization structure and personnel affinity of gang crime through the existing personal data of gang crime, and social data among members. According to the analysis results, several key factors that can reflect the characteristics and intimacy of the criminal suspects are extracted, such as the criminal records, the contact duration, and the location similarity.

2) Second, we put forward a PPDR method (PPDR) based on oblivious transfer for

authenticated entity (police) to query the social cloud server for social data of specific suspects, while the social cloud server is unable to know neither the target nor results of the query. Meanwhile, the unauthorized entities cannot access the social cloud server. Moreover, we combine this method with the proxy reencryption technique to enable data sharing between police and social cloud server, while preventing man-in-the-middle attacks.

3) Third, for SPCSS, we present a new system model which includes classifier owner, cloud server, ASP and police. The classifier owner owns the tree model and outsource it to ASP to provide criminal suspects analysis service. Personal and social data of suspects are owned by the public security cloud server, and social cloud server. On investigation, the police can obtain the authorization of suspects. Using authorization, the police can query social data from social cloud server through PPDR. The police combined the personal data and social data as a query, and launched the query for the ASP and wait for results in return. The main ciphertext computation work was done by the ASP, while preventing the tree model, the query data, and classification results from reveling to an untrusted party. In addition, through data simulation and experiments, we demonstrate that SPCSS can effectively analyze potential suspects over encrypted social data using classifier, and have lower computational overhead even if the query data, and model are confidential.

II. LITERATURE SURVEY

Social data analysis has attracted extensive attention in academia and industry, such as infection analysis [8], emotion analysis [9], and especially plays an important role in the analysis of gang criminal behavior [10], [11]. There exist considerable applications about potential crime analysis based on machine learning in social networks. Rigopoulos and Karadimas [12] developed a model for the assignment based on the application of NexClass methodology, and decision support system in order to assign crime types into a number of categories according to predefined criteria. Ingilevich and Ivanov [13] used linear regression, logistic regression and gradient enhancement methods to predict the number of crimes in different regions of city based on gang robbery crime data of the Russian Federation. Prathap and Ramesha [14] proposed a novel approach to analyze the Twitter sentiments of the users about a particular crime event tweets posted by the active users, thus find out public option changes, and emotion distribution on different types of crimes. In scenarios of most applications, we can see social network analysis (SNA) is now a common tool in criminal investigations; however, evidence collection, and analysis are often limited by data privacy laws. In recent years, many research works consider data availability [5], and privacy protection in data analysis. Sun et al. [15] proposed an improved fully homomorphic encryption (FHE) scheme based on HElib [28] by reducing the ciphertext size, the modulus, and decryption noise. Based on these, they implemented a private decision tree classifier, and the result showed that it has a better performance. Nevertheless, it has low efficiency in practical application. Later, Abadi et al. [16] presented a differential privacy based deep learning scheme. The security measure used in this scheme is to add noise to the original data of the data owner before training in order to resist the inverse attack of extracting the data set directly from the training model, and this privacy preserving scheme is noncloud aided. Olimenco et al. [17] proposed a novel multiparty machine learning method, where a trusted SGX (SoftwareGuard Extensions) processor was used to training oblivious data in cloud. Either non cloud or cloud- assisted privacy preserving scheme focus on privacy issues in the data

training phase. For privacy issues at the classification stage, Bost et al. [18] proposed several building blocks, such as secure comparison, secure dot product, and secure argmax. Based on these building blocks, they constructed several types of machine learning classifiers, i.e., secure hyperplane decision-based classifier, secure naïve Bayesian classifier, and secure decision tree classifier. Hassani et al. [19] proposed a privacy-preserving social network analysis solution based on differential privacy. Wu et al. [20] proposed the secure evaluation based on homomorphic encryption for decision tree and random forests. Tai et al. [21] proposed a privacy- preserving decision tree evaluation for semi-honest, and one- side secure model. They replace the polynomial evaluation step in [20] via linear functions. This leads to computational complexity reduce, and better performance for sparse decision trees. However, the model owner must maintain online to provide interactive classification services for users, and equipped adequate storage and computing power which is hard for the owner. Joye and Salehi [22] modified the DGK (Damgard, Geisler and Krøigaard) comparison protocol in [23]. This leads to better performance for encrypted data comparison, which can also go against timing attacks. They borrowed from [20] the astute idea of using a random permutation to hide the indexes of the comparison nodes at each level of the tree, and reduced the comparison numbers from decision nodes m to tree level d. However, it has to continually exchange the decision tree structure in the classification phase, which is a costly operation.

However, most of the existing studies are concentrated on a single cloud platform. Because gang crime is a geospatial phenomenon, it has geospatial, topical, and temporal relevance. As a member of gang crime, its published data on Facebook, LinkedIn, and other public sources can help determine the leadership role of the organization. Therefore, when conducting analysis of unknown potential criminal suspects based on existing criminal suspects' information, it is necessary to analyze the suspects by considering all kinds of user information, such as personal data, and social data. At the same time, data storage by different independent cloud servers poses a huge challenge to the data management, and collaboration of traditional police crime analysis platforms.

III. SYSTEM MODEL AND DESIGNED GOALS

In this section, we proposed the system model of SPCSS, and pointed out the adversary model, and design goals.

A. System Model

The proposed SPCSS consists of seven entities: key generation center (KGC), classifier owner (CO) (i.e., trusted scientific

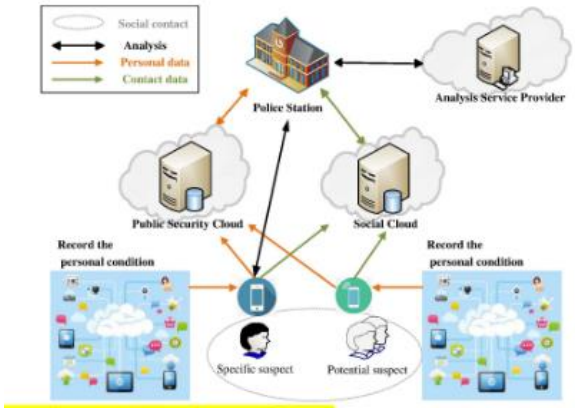


Fig. 1. Criminal suspects sensing system

considering social network and personal data research center), users (i.e., suspects), police station (PS), public security cloud (PSC), social

cloud (SC) and ASP as shown in Fig. 1. The system is divided into preparation domain and analysis domain according to different goals of processed data. Suspects CO, SC, PSC and PS have operations in the preparation domain, while suspects, CO, ASP and PS (as a query requestor) are involved in the analysis domain. The details of each entity in SPCSS are represented as follows:

1) Key Generation Center (KGC): The trusted KGC is responsible for parties' registration, key's generation, and management for legal party in our system. After- ward, KGC is not involved in network and parties' interactions.

2) Users (U): U first enrolls to the KGC and generates public and private keys in the preparation domain. In this article, we mainly concern the users who are considered as suspects or have contacts with suspects. Their personal data are recorded via P's cloud server, and constantly send personal data to PSC. The contact information of ui and uj, such as identity, duration, and social ties, are recorded by smartphones, and have periodically restored to the SC.

3) Classifier Owner (CO): CO is a third party trusted by PS in the preparation domain. We suppose that the tree model has been trained from training data sets on CO, and has already been encrypted and stored in ASP to provide criminal suspects analysis. CO is semihonest in the analysis domain. When PS launches query request to ASP, CO can assist the ASP does the classification work, but the main operations are done by ASP.

4) Public Security Cloud (PSC): PSC is a cloud server with a strong storage capacity, mainly responsible for collecting and storing encrypted personal data of users.

5) Social Cloud (SC): SC is a cloud server for encrypted social data collection, and storage (such as, sender, receiver, duration, contact frequency), which does the similar work to PSC. SC only provides social data for PS. Later, it can remain offline.

6) Analysis Service Provider: ASP is a semihonest party which has powerful computational and storage capability to perform the complicated and time-consuming operations on PS's query data. ASP stored encrypted model, and provide analysis service for PS. When received query from PS, ASP would do suspect analysis using encrypted model, and return encrypted result to PS.

7) Police Station (PS): PS is the users of SPCSS in this system. The goal of PS is to find the suspect status of ui's contacted. Note that ui was considered as a criminal suspect of gang, PS first queried ui's social data from SC. Then combined with the personal data of ui's contacted, thus ASP and CO can perform criminal suspect analysis, and return encrypted result to PS. After decrypting, PS can analyse whether ui's contacted users are criminal suspects or not, and send back the analysis conclusion to police officer who is responsible for criminal suspects.

B. Adversary Model

In this article, we consider two kinds of adversaries in the semihonest model, namely, external attacker, and internal attacker. An external attacker is an untrusted entity such that it may steal some of the communication requests or responses from the common channel for later analysis to obtain private information. It acts as an external eavesdropper. An internal attacker is a semihonest passive entity, it will honestly follow the execution of the model, but it can retain the information obtained during the interaction for the inference of private information of PSC, SC, PS or ASP. We divided

the internal adversary into three types: type-I adversary, PSC, and

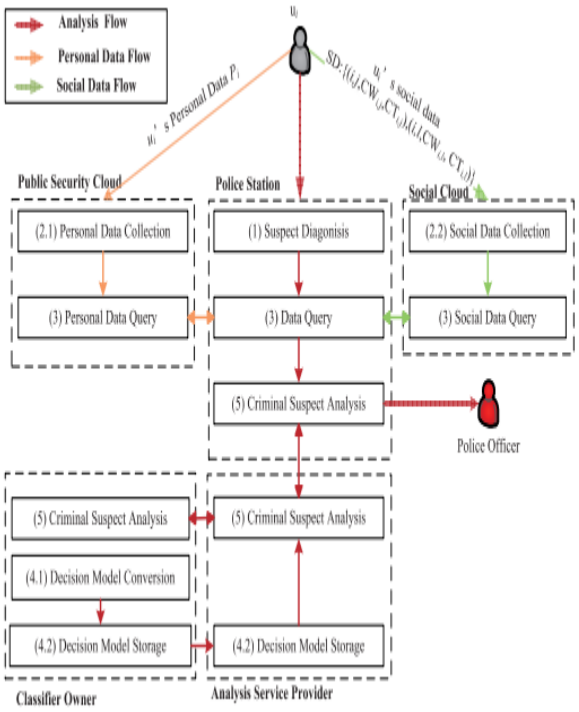


Fig. 2. Illustration of privacy-preserving criminal suspects analysis scheme.

SC may retrieve the queried target (potential suspect uj) of PS through access pattern; PS is the type-II adversary, which may extract the structure and value of model W; type-III adversary includes CO and ASP, they are enabled to revel the query instance x, and corresponding classification result v. Moreover, ASP may learn the classification model W.

C. Design Goals

In this article, the design goals we need to achieve are shown as follows:

1) Model Privacy: W is learned and encrypted by CO and stored in ASP to provide classification services for PS. Hence, the privacy of model should be protected against the untrusted entities such as semihonest ASP, and PS avoid revealing of plaintext of model W.

2) Instance Security and Privacy: A classified instance vector x consist of two types: personal data from PSC, and social data from SC, which are combined by PS. In our scheme, the personal data and social data contains sensitive information of suspects which should be revealed to neither PSC or SC in preparation domain. Moreover, the access pattern also should be protected. In addition, the x is the property of PS, which cannot be disclosed to CO or ASP.

3) Result Confidentiality and Accuracy: It means that the classification result v of x can only be known by the owner of x (i.e., PS). Moreover, it is really important that the classification result accuracy must be guaranteed when applying the privacy-preserving strategy. Hence, our scheme should keep the same accuracy with plain text data classification.

V. PERFORMANCE EVALUATION

In this section, we evaluate the performance of the SPCSS based on simulated data, and computational test.

A. Implementation Details

We implemented our scheme on several computers running on Ubuntu 18.04.2 64-Bit Version with Inter(R) Core(TM) i5-3230M CPU (2.60 GHz), four core and 2 GB of RAM memory on VMware Workstation in the LAN in C++ language. Two of them acts as the PS and CO, the others act as the CS, PSC, and ASP, respectively.

We conducted data simulations based on Nursery(Simular data 1), and ECG(Simular dara 2) data sets which come from UCI Machine Learning Repository.1 We use the five features of the data sets to represent CWi,j , CTi,j , LSi,j , CRi,j , Lui , which can be the parameters of personal and social data. In the simulation, we aim to show the accuracy and efficiency of

criminal suspects analysis on encrypted data. We implemented our homomorphic encryptions with GMP,2 NTL,3 HELib4 library. Moreover, we constructed several building blocks, such as encrypted data comparing, and PPDR. Based on these building blocks, we constructed our SPCSS scheme. To implement the scheme more securely, we improved the modulus n of the Paillier to 1024 bits, and comparison bit length to l = 64. Since the homomorphic encryption only supports integer operations, we used the IEEE 754 double precision floating points to represent the real number, and the accuracy is 52 bits, later by multiplication, a very large real integer was transformed to integer according to the method mentioned in Bost et al. [18].

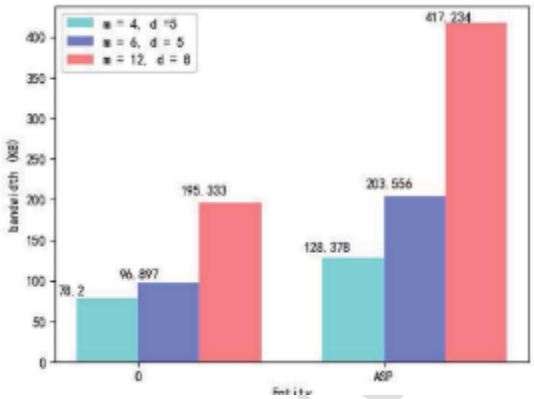
B. Performance

According to the description of Section V-A, SC, and PSC only respond to the social data and personal data query requests launched by PS in privacy-preserving way, in which time cost only impacts the preparation time cost, however does little influence on classification that we were mainly concerned. Therefore, we mainly focus on the performance on the side of ASP, CO, and PS in the classification phase.

TABLE IV

COMPUTATION TIME (d: DEPTH OF TREE, m: NO. OF DECISION NO

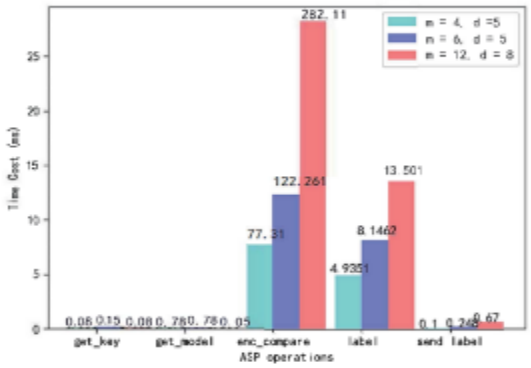
Query Data	m	d	Method	Client bench time (ms)	Server ben time(ms
Simular data 1	4	5	Bost [19]	135.508	202.095
			Our	78.2	128.378
Simular data 2	6	5	Bost [19]	196.095	344.408
			Our	96.897	203.556



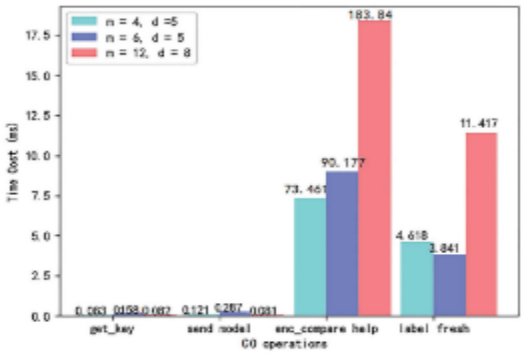
Bandwidth of CART decision tree classifier-based SPCSS. In our settings, preparation works have been completed in the start of the experiments, where the tree model W has been trained, converted, encrypted, and stored in ASP to provide criminal suspects analysis service for P S remotely. Moreover, PS has queried ui and uj's personal and social data from PSC and SC securely. To evaluate the performance and feasibility of secure decision-tree model, the benchmark computation time and communication bandwidth, the experiments were mainly examined in the classification phase.

We evaluated the secure decision-tree model in once classification. The results are shown in Table IV. From Table IV, we can see that there exists two types of secure decision tree model: Bost [18], and ours. Either both Bost, or ours, the server always takes on more calculation works, however, ours gives better performance. Moreover, for once classification, the time cost in client and server is less than 300 ms which is acceptable in practice. Therefore, we choose our decision-tree model in SPCSS to provide service for PS. Without consideration of network communications, the computation time spends on CO and ASP in once classification is less than 1000 ms, and the total bandwidth are less than 500 KB in one side as seen in Figs. 5 and 6, where CO acted as the client, and ASP acted as the server. In these tree

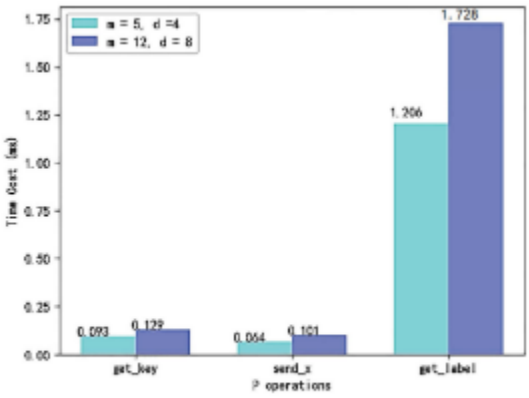
models, m denotes the number of decision tree nodes, and d denotes the depth of tree. In once-time comparison protocol, compared with existed works [20], [21], we just need one bit data join the interaction between CO and ASP, thus decreases the iteration counts between two parties. The classification works runs among PS, ASP, and CO. In the classification, for party PS, there are three resources of the time-consuming overheads, which are get public key, send instance, and get_label. From Fig. 6(c), we can see that get_label operation is the most time-consuming operation. For party PS and ASP, there are three main operations that



(a) Time consuming of ASP.



(b) Time consuming of CO.



(c) Time consuming of PS.

Fig. 6. Time cost of CART decision tree classifier-based SPCSS. (a) Time consuming of ASP. (b) Time consuming of CO. (c) Time consuming of PS.

VI. CONCLUSION

In this article, we have proposed a criminal suspects analysis approach by utilizing social data and crime data to enhance crime analysis without privacy leakage. In our scheme, nothing of personal and social data is leaked to either of the service providers. Moreover, the access pattern is protected and CART model has been trained, encrypted, and outsourced to the ASP to provide criminal suspects analysis. During the analysis phase, any untrusted party can deduce nothing from the classification model, the police station's inputs, and analysis results. Besides, in our scheme, the police station does not need to take part in the analysis, i.e., they just send a query and wait for the results. The experiments evaluation results show that our approach can achieve good analysis results with the acceptable overhead. For the future work, we plan to extend our work to support CO offline.

REFERENCES

[1] H. Arshad, A. Jantan, and E. Omolara, "Evidence collection and forensics on social networks: Research challenges and directions," Digit. Invest., vol. 28, pp. 126–138, Mar. 2019.

[2] S. Seo et al., "Partially generative neural networks for gang crime classification with partial information," in Proc. AAAI/ACM Conf. AI, Ethics, Soc., New York, NY, USA, 2018, pp. 257–263, doi: 10.1145/3278721.3278758.

[3] D. Ramalingam, V. Chinnaiah, and A. Jeyagobi, "Privacy preserving schemes for secure interactions in online social networks," in Proc. Int. Conf. Soft Comput. Syst., vol. 837, 2018, pp. 548–557.

[4] S. Jiang, M. Duan, and L. Wang, "Toward privacy-preserving symp- toms matching in SDN-based mobile healthcare social networks," IEEE Internet Things J., vol. 5, no. 3, pp. 1379–1388, Jun. 2018, doi: 10.1109/JIOT.2018.2799209.

[5] K. Zhang, J. Ni, K. Yang, X. Liang, J. Ren, and X. S. Shen, "Security and privacy in smart city applications: Challenges and solutions," IEEE Commun. Mag., vol. 55, no. 1, pp. 122–129, Jan. 2017, doi: 10.1109/MCOM.2017.1600267CM.

[6] R. Yu, J. Kang, X. Huang, S. Xie, Y. Zhang, and S. Gjessing, "MixGroup: Accumulative pseudonym exchanging for location privacy enhancement in vehicular social networks," IEEE Trans. Depend. Secure Comput., vol. 13, no. 1, pp. 93–105, Jan./Feb. 2016.

[7] B. Desmet and V. Hoste, "Online suicide prevention through optimized text classification," Inf. Sci., vol. 439, pp. 61–78, May 2018.

[8] K. Zhang, X. Liang, J. Ni, K. Yang, and X. Shen, "Exploiting social network to enhance human-to-human infection analysis without privacy leakage," IEEE Trans. Depend. Sec. Comput., vol. 15, no. 4, pp. 607–620, Jul./Aug. 2018, doi: 10.1109/TDSC.2016.2626288.

[9] B. Desmet and V. Hoste, "Online suicide prevention through optimized text classification," Inf. Sci., vols. 439–440, pp. 61–78, May 2018, doi: 10.1016/j.ins.2018.02.014.

[10] Z. Yu, F. Yi, Q. Lv, and B. Guo, "Identifying on-site users for social events: Mobility, content, and social relationship," IEEE Trans. Mobile Comput., vol. 17, no. 9, pp. 2055–2068, Sep. 2018, doi: 10.1109/TMC.2018.2794981.

[11] A. Tundis, A. Jain, G. Bhatia, and M. Muhlhauser, "Similarity analysis of criminals on social networks: An example on Twitter," in Proc. 28th Int. Conf. Comput. Commun. Netw. (ICCCN), Valencia, Spain, Jul./Aug. 2019, pp. 1–9, doi: 10.1109/ICCCN.2019.8847028.

[12] G. Rigopoulos and N. V. Karadimas, "Military student assignment using NexClass decision support system," in Proc. 3rd Int. Conf. Math. Comput. Sci. Ind. (MCSI), Chania, Greece, Aug. 2016, pp. 213–218, doi: 10.1109/MCSI.2016.047.

[13] V. Ingilevich and S. Ivanov, "Crime rate prediction in the urban environ- ment using social factors," Procedia Comput. Sci., vol. 136, pp. 472–478, Jan. 2018.

[14] B. R. Prathap and K. Ramesha, "Twitter sentiment for analyzing different types of crimes," in Proc. Int. Conf. Commun., Com- put. Internet Things, Chennai, India, Feb. 2018, pp. 483–488, doi: 10.1109/IC3IoT.2018.8668140.

[15] X. Sun, P. Zhang, J. K. Liu, J. Yu, and W. Xie, "Private machine learning classification based on fully homomorphic encryption," IEEE Trans. Emerg. Topics Comput., to be published, doi: 10.1109/TETC.2018.2794611.

[16] M. Abadi et al., "Deep learning with differential privacy," in Proc. ACM SIGSAC Conf. Comput. Commun. Secur., New York,

NY, USA, 2016, pp. 308–318, doi: 10.1145/2976749.2978318.

[17] O. Ohrimenko et al., "Oblivious multi-party machine learning on trusted processors," in Proc. USENIX Secur., vol. 16, 2016, pp. 619–636. [18] R. Bost, R. A. Popa, S. Tu, and S. Goldwasser, "Machine learning classification over encrypted data," in Proc. NDSS, 2015.

[19] H. Hassani, X. Huang, M. Ghodsi, and E. S. Silva, "A review of data mining applications in crime," Stat. Anal. Data Mining, ASA Data Sci. J., vol. 9, no. 3, pp. 139–154, Apr. 2016, doi: 10.1002/sam.11312.

[20] D. J. Wu, T. Feng, M. Naehrig, and K. Lauter, "Privately evaluating decision trees and random forests," in Proc. Privacy Enhancing Technol., vol. 4, pp. 335–355, 2016.

[21] R. K. H. Tai, J. P. K. Ma, Y. J. Zhao, and S. S. M. Chow, "Privacy- preserving decision trees evaluation via linear functions," in Proc. Eur. Symp. Res. Comput. Secur. (Lecture Notes in Computer Science), vol. 10493. Berlin, Germany: Springer, 2017, pp. 494–512.

[22] M. Joye and F. Salehi, "Private yet efficient decision tree evaluation," in Data and Applications Security and Privacy XXXII. Berlin, Germany: Springer, 2018, pp. 243–259, doi: 10.1007/978-3-319-95729-6_16.

[23] T. Veugen, "Improving the DGK comparison protocol," in Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS), Tenerife,