Date of current version January 13, 2022.

Digital Object Identifier 10.1109/ACCESS.2020.3036101

EDITORIAL

IEEE ACCESS SPECIAL SECTION: PRIVACY PRESERVATION FOR LARGE-SCALE USER DATA IN SOCIAL NETWORKS

Social networks have become one of the most popular platforms for people to communicate and interact with their friends and share personal information and experiences (e.g., Facebook owns over 1.23 billion monthly active users). The increasing popularity of social networks has generated extremely large-scale user data (e.g., Twitter generates 500 million tweets per day and around 200 billion tweets per year). These data can help improve people's quality of life as well as benefit various interest groups such as advertisers, application developers, and so on. However, privacy may be compromised if learning algorithms are used to infer unpublished privacy information from published data. Hence, user data privacy preservation has become one of the most urgent research issues in social networks.

A great deal of effort has been devoted to protecting user data privacy, such as work on cryptography, security protocols, and some industry standards. However, these still lack techniques that can effectively prevent the indirect disclosure of privacy in social networks. Therefore, effective privacy protection techniques and tools are actively sought to prevent malicious inference of private information.

We have accepted 24 excellent articles in this Special Section after a rigorous review, covering all aspects of data privacy protection for large-scale users in social networks. In the following, we will introduce these articles and highlight their main contributions.

Phishing is a popular information-stealing technique for attackers to obtain victim users' sensitive information, such as username with passwords, social security numbers, credit card numbers, and so on. In the article "Phishing-alarm: Robust and efficient phishing detection via page component similarity," Mao *et al.* proposed a robust phishing detection approach, called Phishing-Alarm, which is based on CSS features of web pages, to detect phishing attacks using features that are hard to evade by attackers. Developing techniques to identify effective CSS features, and algorithms to efficiently evaluate page similarity, Phishing-Alarm is prototyped as an extension to the Google Chrome browser. The effectiveness of this new approach is demonstrated in the evaluation using real-world phishing samples.

In the article "Perturbation-based private profile matching in social networks," by Li *et al.*, a novel perturbation-based private profile matching mechanism is proposed, which preserves privacy by mixing the private data with random noise. The authors consider the case where the profiles are fine-grained, meaning that each attribute is associated with a user-specific numerical value to indicate the level of interest. By carefully tuning the amount of information owned by each party, privacy is effectively preserved while the matching result of users' profiles can be cooperatively obtained. They improve the basic scheme by considering collusion resistance and verifiability, which is effective in privacy protection and efficient in computational cost.

In the article "A new structure-hole-based algorithm for influence maximization in large online social networks," Zhu *et al.* developed a novel algorithm to solve the problem of influence maximization based on structure hole theory. The algorithm utilizes structure hole information to significantly decrease the number of candidates of seed nodes. To evaluate the structural importance of nodes in networks, a structure hole value calculate algorithm is proposed to calculate the structural hole value of nodes. Proving the structure-hole-based influence maximization algorithm is NP-hard; therefore, a structure-based greedy algorithm is developed to select seeds with wide influence spread and high structural hole value. Experiments are conducted on real data sets to verify the algorithm's time efficiency and accuracy, and the results show that the algorithms are much more efficient and scalable than the existing algorithms.

In the article "Achieving effective k-anonymity for query privacy in location-based services," Wang *et al.* investigated effective k-anonymity-based solutions for query privacy in location-based services. The authors formulated a probabilistic framework PkA, under which k-anonymity-based mechanisms can be initiated and analyzed a recent proposed algorithm DLS as an instance of PkA. To obtain more effective query privacy in general cases, two algorithms, MEE and MER, are proposed, which optimize two individual privacy metrics, and denoted expected entropy and expected max–min ratio. The algorithms satisfy both No More Leakage and k-Effectiveness, two practical properties for effective query privacy. Evaluations are conducted based on real-life data sets and synthetic distributions of query interests, and the results demonstrate that the proposed algorithms produce significantly improved query privacy.

The article titled "Authentication with block-chain algorithm and text encryption protocol in calculation of social network," by Yu *et al.*, proposed a protocol and authentication with a block-chain algorithm to protect user privacy information in social community detection. During the expansion of communities on the base of mining seed, in order to prevent others from malicious users, users' identities need to be verified after they send a request. In addition, to prevent honest but curious users from illegal access to other users' information, plaintext is not sent directly after the authentication, and attributes are hashed using mixed hash encryption to make sure that users can only calculate the matching degree rather than know specific information of other users. Experiments show that the protocol could serve well against different types of attacks.

In the article "A novel cooperative jamming scheme for wireless social networks without known CSI," by Huang *et al.*, a cooperative jamming scheme is presented based on the space power synthesis with unknown channel state information (CSI) of eavesdroppers. The authors analyze the superimposed effects of jammers with different locations in a fixed area and present corresponding jamming schemes to minimize synthetic jamming power at a legitimate receiver but satisfy basic interference in other locations. At the same time, power allocation schemes are provided to maximize the worst-case secrecy rate of a legitimate receiver. Theoretical analysis and numerical simulation results demonstrate the effectiveness of the proposed schemes.

In the article "Attribute couplet attacks and privacy preservation in social networks," Yin *et al.* raised a new privacy attack risk termed as attribute couplet attack which utilizes the couplets of attributes to infer the identities in anonymized social networks. To achieve privacy-preservation under attribute couplet attacks, a new anonymity concept known as k-couplet anonymity is proposed. Then, the authors design and implement two heuristic algorithms to promote the k-couplet anonymity. An approximate algorithm for multiple-attribute social networks is designed to realize the k-couplet anonymity. According to the evaluations on multiple public data sets, the proposed algorithms can preserve the privacy and utility of the social network data set effectively under the attribute couplet attacks.

The article "ClickLeak: Keystroke leaks through multimodal sensors in cyber-physical social networks," by Li *et al.*, explored a novel and practical multi-modal side-channel keystroke recognition system, named ClickLeak, which can infer the PIN code/password entered on a numeric keypad by using the commodity Wi-Fi devices. ClickLeak is built on the observation that each key input makes a unique pattern of hand and finger movements, and this generates unique distortions to multi-path Wi-Fi signals. Acceleration and microphone sensors of smartphones determine the starting and ending time of keystrokes, while the time series of channel state information is analyzed to determine the keystrokes. The evaluation results have shown that with large-scale data collections from public social settings, the key recognition accuracy can reach higher than 83%.

With the development of mobile crowdsourcing systems, location privacy protection has become the research focus. "The novel location privacy-preserving CKD for mobile crowdsourcing systems," by Chi *et al.*, proposed a location privacy-preserving mechanism CKD through combining k-anonymity and differential privacy-preserving to prevent mobile user's location privacy from being leaked. Furthermore, the tradeoff between privacy protection and service quality is solved based on a Stackelberg game. The effectiveness of the proposed CKD and the adaptability of the proposed optimal strategy have also been verified through comparison experiments.

JavaScript applications are widely used in a range of scenarios, including Web applications, mobile applications, and server-side applications. However, the flexibility of the JavaScript language introduces new security challenges in these platforms. In the article "Detecting malicious behaviors in JavaScript applications," Mao *et al.* proposed a detection technique to identify malicious behaviors in JavaScript applications. They prototyped the solution on the popular JavaScript engine V8 and used it to detect attacks on the android system. The evaluation shows the effectiveness of the approach in detecting injection attacks to JavaScript applications.

While enjoying the convenience of location-based services (LBSs) in everyday life, wireless device users could also put their location privacy at risk. The article "Cognitive approach for location privacy protection," by Han *et al.*, developed a new cognitive approach that enables near-complete privacy protection for LBS users by leveraging existing social network resources. A heterogeneous multi-server architecture that cuts off the direct connection between the LBS queries and the query issuers is designed, while the authors introduce an auction-based incentive mechanism for guaranteed user participation, which is critical for the success of the proposed architecture. A simulation system and a smartphone application were developed, and evaluation results show that the proposed method can not only achieve near-total privacy protection for LBS users but also significantly improves the quality of the services.

In the article "A blockchain-based privacy-preserving incentive mechanism in crowdsensing applications," by Wang *et al.*, a privacy-preserving blockchain incentive mechanism in crowdsensing applications is proposed. In the distributed crowdsensing system, the sensing data qualities are evaluated via the EM algorithm and contributions are quantified via mutual information by miners. A signcryption method is used to prevent miners and other adversaries from violating users' privacy, which saves computing costs compared to operating sequentially of the signature and encryption. In addition, the node cooperation-based privacy protection mechanism is developed, which will keep the user's privacy hidden within a group, to deal with impersonation attacks in the open and transparent blockchain. Theoretical

analysis and simulation experiments demonstrate the feasibility and security of the incentive mechanism.

To mitigate investments, stock price forecasting has attracted more attention in recent years. Aiming at the discreteness, non-normality, and high-noise in high-frequency data, a support vector machine regression (SVR) algorithm is introduced in "An adaptive SVR for high-frequency stock price forecasting," by Guo *et al*. However, SVR with fixed parameters is difficult to satisfy with the constantly changing data flow. To tackle this problem, an adaptive SVR is proposed for stock data at three different time scales, including daily data, 30-min data, and 5-min data. Experiments show that the improved SVR with dynamic optimization of learning parameters by particle swarm optimization can get a better result than compared methods including SVR and back-propagation neural network.

With the exponential growth of data, how to efficiently utilize the data becomes a critical issue. The article "A survey on big data market: Pricing, trading, and protection," by Liang *et al.*, addressed the issue of big data trading. The authors conducted a comprehensive survey on the lifecycle of data and data trading, then focused on the design of data trading platforms and schemes, supporting efficient, secure, and privacy-preserving data trading. Furthermore, digital copyright protection mechanisms and challenges in data protection in the data trading lifecycle were outlined.

Electronic medical records (EMRs) play a significant role in healthcare networks. In "Privacy preservation for outsourced medical data with flexible access control," Zhou *et al.*, proposed two anonymous Role-based access control (RBAC) schemes for the EMR system. The first scheme achieves moderate security, where adversaries choose attack targets before obtaining information from the EMR system, while the second scheme achieves full security, where adversaries adaptively choose attack targets after interaction with the EMR system. Rigorous proof demonstrates the security and anonymity of the schemes. In addition, the authors developed an approach in which EMR owners can search for their EMRs in an anonymous system.

The article "Enhanced instant message security and privacy protection scheme for mobile social network systems," by Wang *et al.*, proposed an enhanced secure Instant Messaging (IM) system that is based on the Elliptic Curve Cryptosystem to overcome the data security and privacy protection problems of mobile social networks. The proposed scheme supports an offline key agreement between users and data privacy protection on a device. In addition, the scheme utilizes timestamps to deny replaying attacks and utilizes the elliptic curve digital signature algorithm to sign and verify the messages that are transferred in the IM system. The comparison results of the proposed scheme with others and the results of an experiment show that it is a comprehensive, secure scheme with high security and good practicability.

The article "Towards understanding community interests with topic modeling" by Wang *et al*., presented a new methodology using topic modeling to verify structure-based communities based on whether their members share strong common interests or not. A Latent Dirichlet Allocation topic model was trained to capture the topics in the aggregated tweets of each user in a community. Moreover, new distance metrics were proposed to quantify the topic similarity of individual users, cliques, and communities. By building a Twitter topic modeling system to interpret the communities identified by two community detection algorithms in a large-scale Twitter topology, it is evident that Twitter users in a community show common interests, in general, was discovered.

"Modeling privacy leakage risks in large-scale social networks," by Du *et al.*, proposed a privacy disclosure attack-defense tree to describe a series of attack steps launched by the attackers to achieve their ultimate goals and the corresponding countermeasures that can be adopted by the social network security defenders. To further illustrate a dynamic attacking process, the authors extended a Markov chain-based approach to model a temporal-aware attack-defense tree. At the same time, an attack-defense game was introduced to analyze the potential strategies performed by the attacker and the defender. Experimental evaluations on three real-world data sets illuminated the privacy risk management of contemporary social network service providers.

The article "A blockchain based truthful incentive mechanism for distributed P2P applications," by He *et al.*, proposed a blockchain-based truthful incentive mechanism for distributed P2P applications which applies a cryptocurrency such as Bitcoin to incentivize users for cooperation. In the incentive mechanism, intermediate nodes who contribute to a successful delivery can obtain rewards from Blockchain transactions. A secure validation method and a pricing strategy were presented and integrated into the incentive mechanism, as users and miners in the Blockchain P2P system may exhibit selfish actions or collude with each other. The effectiveness and security strength of the incentive mechanism were demonstrated through a game theoretical analysis and evaluation study.

The article "An empirical study on the privacy preservation of online social networks," by Siddula *et al.*, presented a survey of research on a wide range of privacy-enhancing methods in social networks. The authors investigated various privacy-preserving models and methods including naive anonymization, perturbation, or building a complete alternative
network. The work done by multiple researchers in the past was reviewed, where social networks are stated as network graphs with users represented as nodes and friendship between users represented as links between the nodes. Other systems proposed, along with all the available databases, were also presented in the article for future researchers in this area.

Ensuring the privacy security of network user data in data mining is an important and challenging problem. In the article "DP-MCDBSCAN: Differential privacy preserving multi-core DBSCAN clustering for network user

data,'' Ni *et al.* focused on the privacy preservation in clustering analysis of network user data. They put forward a differential privacy preservation multiple cores DBSCAN clustering schema based on the powerful differential privacy and DBSCAN algorithm, which enhanced data clustering efficaciously by adding Laplace noise. Extensive theoretical analysis and simulation results demonstrated that the schema shows better efficiency, accuracy, and privacy preservation effect than previous schemas.

In the article ''Physical layer security in 5G based large scale social networks: Opportunities and challenges,'' Gao *et al.* discussed the physical layer security in large-scale social networks. The opportunities in large-scale social networks are summarized in physical, link, and upper layer using the cross-layer optimization; traditional physical layer security would meet new problems not yet tackled. Furthermore, the challenges due to some ideal considerations are also summarized: the detection of wire-tap users, the utilization of high dynamic range, and the information exchange in cross-layer design.

With the inclusion of mobile devices and ubiquitous connectivity of smart devices in Internet of Things, secure key management is mandatory to ensure privacy for information exchange. In the article ''Distributed multi-party key management for efficient authentication in the Internet of Things,'' by Mahmood *et al.*, a novel distributed key management scheme is presented by utilizing Chebyshev polynomials and chaotic maps for cryptographic operations. The authors established the session key between the group heads and server and presented the intra- and inter-group secret key establishment schemes between the GH and smart devices. Moreover, a testbed is set up for group head to server-level authentication and key establishment. Results proved the supremacy of the scheme as compared with preliminaries in terms of computation cost, communication cost, and resilience.

The article ''Efficient compressed ciphertext length scheme using multi-authority CP-ABE for hierarchical attributes,'' by Zhang *et al.*, proposed a hierarchical multi-authority, attribute-based encryption on prime order groups to tackle the problem of low calculation efficiency and ciphertext information redundancy. The encryption technique has a polycentric attribute authorization system based on an AND gate access structure, with a unified attribute index established by each attribute authority throughout the system, to form a binary tree. The state value of the parent node can be determined by the state of its child node in an attribute access tree. This representation can effectively decrease encryption and decryption computations and ciphertext length for the access control structure with many *Don't Care* element values.

In conclusion, we would like to acknowledge our appreciation of all the authors for their excellent contributions. We also thank all reviewers for their hard work and valuable comments and suggestions for enlightening the quality of the articles. Finally, we appreciate the advice and support of the journal editors.

YUAN GAO, *Associate Editor*
Tsinghua University
Beijing 100084, China

YI LI, *Guest Editor*
The High School Affiliated to Renmin University
Beijing 100080, China

YUNCHUAN SUN, *Guest Editor*
Beijing Normal University
Beijing 100875, China

ZHIPENG CAI, *Guest Editor*
Georgia State University
Atlanta, GA 30302, USA

LIRAN MA, *Guest Editor*
Texas Christian University
Fort Worth, TX 76129, USA

MATEVŽ PUSTIŠEK, *Guest Editor*
University of Ljubljana
1000 Ljubljana, Slovenia

SU HU, *Guest Editor*
University of Electronic Science and Technology
Chengdu 610054, China

YUAN GAO (Member, IEEE) is currently an Associate Research Fellow with Tsinghua University. He has published more than 80 academic papers in peer-reviewed international journals and conferences. His research interests include wireless communication systems, satellite communication systems, network control theory, and big data. He is a member of ACM. He is an Associate Editor of several international journals. He is also a Guest Editor of several special issues. He also served as a Guest Reviewer and a TPC Member of several journals and international conferences, including IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS (JSAC), IEEE TRANSACTIONS ON WIRELESS COMMUNICATION, IEEE TRANSACTIONS ON COMMUNICATION, IEEE COMMUNICATIONS LETTER, ICC, and WCNC.

YI LI (Member, IEEE) received the dual degrees in economics from Peking University and the Ph.D. degree in electronic engineering from Tsinghua University, Beijing, China, in 2014. She was a Visiting Scholar with the Department of Electrical Engineering, Columbia University, in 2012. In 2014, she joined the High School Affiliated to Renmin University as a mathematics, physics, and computer science teacher. At the same time, she is currently leading an optional course called Mathematical Modeling and Students Inquiry Study in Information and Communication Engineering. She is a member of CCF and CSIAM. She is also a Reviewer of IEEE ACCESS and an Invited Editor of Special Issue Privacy Preservation for Large-Scale User Data in Social Networks. She has published more than 20 papers in the top class of conferences and journals. She also won outstanding awards in National Teaching Competitions. Students under her supervision have made great achievements in international mathematical modeling competitions and STEM research. Five of her students' research articles have been published in international conference (EI index). She also devotes herself to educational research, especially in interdisciplinary education, STEM education, and big data in educational assessment.

YUNCHUAN SUN (Senior Member, IEEE) received the Ph.D. degree from the Institute of Computing Technology, Chinese Academy of Science, Beijing, China, in 2009. He is currently a Professor and the Director of the International Institute of Big Data in Finance, Beijing Normal University, Beijing. He has published more than 80 papers at international conferences and in journals. His research interests include FinTech, digital economy, the Internet of Things, event-linked networks, knowledge engineering, and information security. He is a member of the China Computer Federation (CCF) and the Big Data Committee of CCF. He has been an Associate Editor of *Personal and Ubiquitous Computing* since 2012. He has served as the Vice Chair and the Secretary for the IEEE Communications Society Technical Subcommittee for the Internet of Things and the Active Chair for Emergent Technologies Technical Committee (ETTC) Task Force on Smart World at IEEE Computational Intelligent Society. He is also the Referee of the *International Journal of Electronic Commerce*. As one of the founders and program co-chairs, he has successfully organized the international IIKI series events IIKI2012–IIKI2019. He has also organized more than 20 special issues on relevant topics in several international journals. He is involved in several research projects, including NSFC 973 and 863 and Program of China projects.

ZHIPENG CAI (Senior Member, IEEE) received the B.S. degree from the Beijing Institute of Technology and the M.S. and Ph.D. degrees from the Department of Computing Science, University of Alberta.

He is currently an Assistant Professor with the Department of Computer Science, Georgia State University. He has published more than 60 SCI journal articles, including more than 20 IEEE/ACM Transactions articles, such as IEEE/ACM TRANSACTIONS ON NETWORKING, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, *ACM Transactions on Sensor Networks*, IEEE TRANSACTIONS ON CLOUD COMPUTING, IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT, *Theoretical Computer Science*, IEEE TRANSACTIONS ON MOBILE COMPUTING, IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, and IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. He also published more than 50 conference papers, including articles published in INFOCOM, MOBIHOC, ICDCS, SECON, DASFAA, and IPSN. His research interests include privacy, social networks, the Internet of Things, and big data.

Dr. Cai was a recipient of the NSF CAREER Award, which is the most competitive and prestigious award from National Science Foundation to junior faculty members in science and engineering fields. He served as the General Chairs / Program Chairs for COCOA 2018, ISBRA 2017, IPCCC 2016, SOCIALCOM 2016, WASA 2014, COCOON 2014, IPCCC 2013, and ISBRA. He is currently a Steering Committee Co-Chair of the International Conference on Wireless Algorithms, Systems, and Applications (WASA). He also serves as Steering Committee Member for the International Computing and Combinatorics Conference (COCOON) and the IEEE International Performance Computing and Communications Conference (IPCCC). He has also served on the technical program committee for numerous international conferences, including INFOCOM, MobiHoc, ICDCS, ICCCN, IPCCC, MASS, DASFAA, WASA, ICC, GLOBECOM, ICPADS, SEDE, MSN, ISBRA, CIKM, COCOON,

COCOA, and ISAAC. He is an Editor/Guest Editor of *Journal of Network and Computer Applications*, *International Journal of Distributed Sensor Networks*, *Sensor*, *Tsinghua Science and Technology*, *Algorithmica*, *Theoretical Computer Science*, *Journal of Combinatorial Optimization*, IEEE/ACM TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS, and *International Journal of Sensor Networks*.

LIRAN MA (Member, IEEE) received the D.Sc. degree in computer science from The George Washington University in 2008. He is currently an Associate Professor with the Department of Computer Science, Texas Christian University. His current research interests include wireless networking, systems security, mobile health and safety, data privacy, and cloud computing. He has published more than 40 papers in renown journals and international conferences. He is also serving as a Technical Program Committee (TPC) Co-Chair for the 12th International Conference on Wireless Algorithms, Systems, and Applications (WASA 2017). In addition, he has served as a TPC Member for many conferences, including IEEE International Conference on Computer Communications (INFOCOM) and a Reviewer for a number of journals, including IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS (JSAC), IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS (TWC), and IEEE TRANSACTIONS ON MOBILE COMPUTING (TMC).

MATEVŽ PUSTIŠEK received the Ph.D. degree in telecommunications from the University of Ljubljana in 2009. He is currently a Senior Lecturer with the Faculty of Electrical Engineering as well as a Researcher with the Laboratory for Telecommunications. He is also a Visiting Lecturer with the Bonch-Bruevich Saint-Petersburg State University of Telecommunications. He has been working in the field of ICT since 1994. He first specialized in network traffic analysis and network simulation. Also at the center of his work are services and applications of the Internet of Things (IoT). His interest is in particular in user- and usability-related aspects as well as in security in IoT.

SU HU received the M.S. and Ph.D. degrees from the National Key Laboratory on Communications, University of Electronic Science and Technology of China (UESTC), in 2007 and 2010, respectively. From February 2011 to August 2012, he was a Research Fellow with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. He is currently a Professor with the UESTC. His research interests include sequence design with good correlation properties and physical-layer design for wireless communication systems, such as filter bank multicarrier systems and cognitive radio networks.