Peer-to-Peer Networking and Applications
https://doi.org/10.1007/s12083-020-00922-5

Security and privacy of machine learning assisted P2P networks

Hongwei Li[1] · Rongxing Lu[2] · Mohamed M. E. A. Mahmoud[3]

Advances in machine learning (ML) in recent years have enabled a dizzying array of applications in diverse areas of networks and communications. Specifically, the development of Peer-to-Peer (P2P) networks is promoted by either traditional or most advanced ML techniques in terms of efficiency, functionality as well as the scalability. Examples of such promotions are easy to find. As typical functions in P2P networks, the performance of neighbor selection, traffic classification and resource management could be significantly improved with the help of ML. Furthermore, with the rapid growing of network nodes and the network traffic, the diversity and the vast volume of the data also poses great challenges to the management and QoS of network. Therefore, ML cloud be one of the most promising novel technologies to resolve or mitigate such challenges.

Although remarkable and promising benefits are brought through leveraging ML techniques to solve the problems in P2P networks, the security and privacy issues are becoming one of the most troublesome barriers. On one hand, personal data leakage, neighbor selection attacks, system penetration, Botnet attacks, and numerous traditional security and privacy threatens continuously hinder the development of the P2P network. Fortunately, ML is already proved to be one of the practical techniques to resist such attacks. On the other hand, the ML-assisted P2P network may introduce new security and privacy issues. Numerous previously conducted researches have shown that the newly trained models may suffer from intensive black box and white box attacks. From the distribution of the training data set to the model parameters, numerous information could be the target of internal and external attackers. Therefore, such problems still exist in the ML-assisted P2P networks. Moreover, some new and unexplored security and privacy problems also deserve further study. Hence, it is urgent to conduct researches to address security and privacy issues in ML-assisted P2P networks.

This special issue has gained overwhelming attention and received 37 submissions from researchers and practitioners working on security and privacy of ML-assisted P2P networks. Each paper is selected into the regular rigorous review process and each submission has been reviewed by at least three reviewers. After 2–3 round reviews, eventually nine quality papers are recommended to be included into this special issue, which are summarized as below.

The paper titled "CSNN: Password Guessing Method Based on Chinese Syllables and Neural Network" has proposed s a password guessing method CSNN for Chinese password sets. This method treats Chinese Syllables as integral elements and uses them to parse and process passwords. The processed passwords are trained in Long Short-Term Memory Neural Network, and the trained model is used to generate password dictionaries (guessing sets). In order to evaluate the effectiveness of CSNN, the hit rates of guessing sets generated by CSNN on target password sets (test sets) are compared with Probability Context-Free Grammar (PCFG) and 5th-order Markov Chain Model. In hit rate experiment, the results show that the comprehensive performance of guessing sets generated by CSNN is better than PCFG and 5th-order Markov Chain Model. Compared with PCFG, different scales of CSNN guessing sets can improve up to 9% in hit rate on some test sets; compared with 5th-order Markov Chain Model, the highest increase is 14.6%.

In "Achieving Reliable Timestamp in the Bitcoin Platform", the author have given a thorough discussion on the topic of Blockchain, which is the underlying technology of the Bitcoin cryptocurrency, is an innovation of information

Guest Editors: Hongwei Li, Rongxing Lu and Mohamed Mahmoud

✉ Hongwei Li
hongweili@uestc.edu.cn

Rongxing Lu
RLU1@unb.ca

Mohamed M. E. A. Mahmoud
mmahmoud@tntech.edu

[1] University of Electronic Science and Technology of China, Chengdu, China

[2] University of New Brunswick, Fredericton, NB, Canada

[3] Tennessee Technological University, Cookeville, TN, USA

Published online: 15 June 2020

technology. The blockchain technology has been widely applied in the evidence storage scenarios to prove that an event occurred at a certain time due to its publicity and immutability. However, the timestamp of a block in the blockchain is introduced by the blockchain node and can be manipulated in hours. This will either lead the failure of the evidence storage system built on top of the blockchain platform or increase the risk of double spending of the blockchain platform itself. The authors have introduced an optimized blockchain timestamp mechanism. The authors also narrow the range of the timestamp in a block to an average of 10 min by leveraging an outside trust timestamp service to the blockchain consensus. Finally, the authors present a security analysis of the proposed scheme.

The paper titled "Efficient distributed privacy-preserving collaborative outlier detection", the authors have presented an efficient protocol for privacy preserving collaborative outlier detection from arbitrarily partitioned data using Local Distance-based Outlier Factor (LDOF). The proposed protocols fall in the two-server model where data owners distribute their private data among two non-colluding servers who detect outlier on the joint data by secure two-party computation. In particular, we perform arithmetic operations which takes place inside LDOF on arithmetic circuits instead of Boolean circuits, and perform sorting operations on Boolean circuits. Such a design enables standard operations are performed with suitable circuits, and thus our scheme is more efficient. In addition, to further improve protocol efficiency, local sensitive hash (LSH) is utilized to filter out data which do not need secure computation to reduce the amount of shared data. The system is implemented using C++ on real data. And the results have demonstrated the superiority of the proposed scheme.

The *k*-means clustering has been well-studied by a significant amount of works, most of the existing schemes are not designed for peer-to-peer (P2P) networks. P2P networks impose several efficiency and security challenges for performing clustering over distributed data. In "Privacy-Preserving *k*-Means Clustering with local synchronization in Peer-to-Peer Networks", the authors have proposed a novel privacy-preserving k-means clustering scheme over distributed data in P2P networks, which achieves local synchronization and privacy protection. Specifically, the authors have designed a secure aggregation protocol and a secure division protocol based on homomorphic encryption to securely compute clusters without revealing the privacy of individual peer. Moreover, the authors also proposed a novel massage encoding method to improve the performance of the proposed aggregation protocol. We formally prove that the proposed scheme is secure under the semi-honest model and demonstrate the performance of our proposed scheme.

The paper titled "An Optimal Uplink Traffic Offloading Algorithm via Opportunistic Communications Based on Machine Learning", the author focusses on maximizing the probability of offloading data to Wi-Fi APs by fragmenting the data and assigning the fragments to different direct or indirect paths generated by opportunistic contacts. Firstly, they propose two methods based on mobility prediction, which is realized by machine learning, to separately calculate the probability of offloading data to Wi-Fi APs by the direct offloading path considering multiple opportunistic contacts and contact duration, and the probability of indirectly offloading data to Wi-Fi APs by the indirect offloading path. Then, based on the probability calculation methods the offloading probability maximization is formulated as a nonlinear integer programming problem, and they also propose a distributed heuristic algorithm to solve it considering complexity of the probability calculation and limited computation capacities of devices. Simulation results prove the data offloading probability of the proposed algorithm outperforms other algorithms under different simulation environment.

In "A Trust-based Minimum Cost and Quality Aware Data Collection Scheme in P2P Network", the authors utilize the idea of machine learning to select trusted data reporter to collect data. The data collection optimization is translated into how to maximize data coverage and minimize the cost under given budget in malicious network. Then a Trust-based Minimum Cost Quality Aware (TMCQA) data collection scheme is proposed to perform data collection optimization. The data collection in TMCQA scheme has the following innovations. (1) A trust evaluation mechanism utilizing the idea of machine learning is established to evaluate the trust of the data reporter. (2) An optimized data reporter selection strategy is proposed to select optimized reporters based on the three key evaluation indices to improve data collection performance. Finally, the authors also verify the validity of the TMCQA scheme proposed in this paper through various experiments. Comparing to Contribution-based with Trust Value Scheme (CNTVS), Random Data Reporter Selection Scheme (RDRSS), and No Time Decay Scheme (TNTDS), with the same budget, the QoS can be improved by 32.21%, 49.39%, 23.68% respectively. The performance of the TMCQA scheme in a malicious P2P network is significantly better than previous strategies.

k-nearest neighbor (*k*-NN) query is widely applied to various networks, such as mobile Internet, peer-to-peer (P2P) network, urban road networks, and so on. The location-based service in the outsourced environment has become a research hotspot with the rise of cloud computing. Meanwhile, various privacy issues have been increasingly prominent. In "Efficient and Secure k-Nearest Neighbor Query on Outsourced Data", the authors have proposed an efficient privacy-preserving query protocol to accomplish the *k*-nearest neighbor (*k*-NN) query processing on outsourced data. The authors adopt the Moore curve to transform the spatial data into one-dimensional sequence and utilize the AES to encrypt the original data. According to the cryptographic transformation, the proposed protocol can minimize the communication overhead to achieve efficient *k*-NN query while protecting the spatial data and location privacy. Furthermore, the proposed efficient scheme offers considerable performance with privacy preservation. Experiments show that the proposed scheme achieve high accuracy and efficiency while preserving the data and location privacy when compared with the existing related approach.

The paper titled "Functional Encryption with Application to Machine Learning: Simple Conversions from Generic Functions to Quadratic Functions", the authors have studied the theoretical problem of functional encryption for machine learning. They present a simple transformation from functional encryption that supports generic functions to one that supports quadratic functions (that include inner product function). In the proposed scheme, the ciphertexts in the proposed schemes have the improved size of length $O((n + m) \log q)$ rather than $O (nm \log q)$ which is the case in a previously proposed scheme where an FE scheme is proposed for inner product functions (this scheme can implement the inner product functionality if taking the key matrix as the identity matrix. The proposed schemes remove the bilinear maps that is used to implement the quadratic functionality. This schemes also allow for instantiations under other standard assumptions such as the decisional diffie-hellman (DDH), RSA, learning with errors (LWE), and learning parity with noise (LPN).

In "A Dynamic and Verifiable Multi-keyword Ranked Search Scheme in the P2P Networking Environment", the authors have proposed a dynamic verifiable multi-keyword ranked search scheme under the background of P2P network and cloud storage service (CSS). On the basis of using secure *kNN* algorithm to encrypt index and traditional inner product algorithm to obtain ranked results, the scheme in this paper realizes forward and backward security by changing the structure of file vector and using modular residual computation. Meanwhile, the integrity and freshness of search results are verified by combining time-stamp chain and Merkle tree. Finally, the security of this scheme under two threat models is analyzed, and the performance evaluation experiment is carried out on the document set.

Finally, we would like to appreciate all authors who submitted manuscripts for consideration, and many anonymous dedicated reviewers for their criticism and time to help us making final decisions. Without their valuable and strong supports, we cannot make this special issue successful. Our sincere gratitude will also go to the PPNA EiC, Prof. Xuemin (Sherman) Shen, and the Springer Journal Editorial Office for helping us to presenting this special issue to readers.

Hongwei Li is currently the Head and a Professor at Department of Information Security, School of Computer Science and Engineering, University of Electronic Science and Technology of China. He received the Ph.D. degree from University of Electronic Science and Technology of China in June 2008. He worked as a Postdoctoral Fellow at the University of Waterloo from October 2011 to October 2012 under the supervision of Prof. Sherman Shen. His research interests include network security and applied cryptography. His research is supported by National Science Foundation of China, Ministry of Science and Technology of China, Ministry of Industry and Information Technology, and China Unicom. Dr. Li has published more than 80 technical papers. He is the sole author of a book, Enabling Secure and Privacy Preserving Communications in Smart Grids (Springer, 2014). Dr. Li serves as the Associate Editor of IEEE Internet of Things Journal, and Peer-to-Peer Networking and Applications, the Guest Editor of IEEE Network and IEEE Internet of Things Journal. He also serves/served the technical symposium co-chair of ACM TUR-C 2019, IEEE ICCC 2016, IEEE GLOBECOM 2015 and IEEE BigDataService 2015, and many technical program committees for international conferences, such as IEEE INFOCOM, IEEE ICC, IEEE GLOBECOM, IEEE WCNC, IEEE SmartGridComm, BODYNETS and IEEE DASC. He won the Best Paper Award from IEEE MASS 2018 and IEEE HELTHCOM 2015. He is the Senior Member of IEEE, the Distinguished Lecturer of IEEE Vehicular Technology Society.