frontiers in Big Data

SPECIALTY GRAND CHALLENGE
published: 14 February 2019
doi: 10.3389/fdata.2019.00001

Research Challenges at the Intersection of Big Data, Security and Privacy

Murat Kantarcioglu[1]* *and Elena Ferrari*[2]

[1] Department of Computer Science, University of Texas at Dallas, Richardson, TX, United States, [2] Department of Theoretical and Applied Science, University of Insubria, Varese, Italy

Keywords: big data, security, privacy, cybersecurity, sharing, machine learning

1. OVERVIEW

As reports from McKinsey Global Institute (Mckinsey et al., 2011) and the World Economic Forum (Schwab, 2016) suggest, capturing, storing and mining "big data" may create significant value in many industries ranging from health care to government services. For example, McKinsey estimates that capturing the value of big data can create $300 billion dollar annual value in the US health care sector and $600 billion dollar annual consumer surplus globally (Mckinsey et al., 2011). Still, several important issues need to be addressed to capture the full potential of big data. As shown by the recent Cambridge Analytica scandal (Cadwalladr and Graham-Harrison, 2018) where millions of users profile information were misused, security and privacy issues become a critical concern. As big data becomes the new oil for the digital economy, realizing the benefits that big data can bring requires considering many different security and privacy issues. This in return implies that the entire big data pipeline needs to be revisited with security and privacy in mind. For example, while the big data is stored and recorded, appropriate privacy-aware access control policies need to be enforced so that the big data is only used for legitimate purposes. On the other hand, while linking and sharing data across organizations, privacy/security issues need to be considered. Below, we provide an overview of novel research challenges that are at the intersection of cybersecurity, privacy and big data.

2. STORING AND QUERYING BIG DATA

One of the ways to securely store big data is using encryption. Once data is encrypted, if the encryption keys are safe, then it is infeasible to retrieve the original data from the encrypted data alone. At the same time, encrypted data must be queried efficiently. Encrypted storage and querying of big data have received significant attention in the literature (e.g., Song et al., 2000; Hacigumus et al., 2002; Golle et al., 2004; Ballard et al., 2005; Chang and Mitzenmacher, 2005; Kantarcıoğlu and Clifton, 2005; Canim and Kantarcioglu, 2007; Shi et al., 2007; Shaon and Kantarcioglu, 2016). Many techniques ranging from simple encrypted keyword searches to fully homomorphic encryption have been developed (e.g., Gentry, 2009). Although there have been major progress in this line of research, breakthroughs are still needed to scale encryption techniques for big data workloads in a cost effect manner. In addition, more practical systems need to be developed for end users. Recent developments that leverage advances in trusted execution environments (TEEs) (e.g., Ohrimenko et al., 2016; Chandra et al., 2017; Shaon et al., 2017; Zheng et al., 2017) offer much more efficient solutions for processing encrypted big data under the assumption that hardware provides

OPEN ACCESS

Edited and reviewed by:
Jorge Lobo,
Catalan Institution for Research and Advanced Studies, Spain

***Correspondence:**
Murat Kantarcioglu
muratk@utdallas.edu

Specialty section:
This article was submitted to Cybersecurity and Privacy, a section of the journal Frontiers in Big Data

Received: *24 July 2018*
Accepted: *10 January 2019*
Published: *14 February 2019*

Citation:
Kantarcioglu M and Ferrari E (2019) Research Challenges at the Intersection of Big Data, Security and Privacy. Front. Big Data 2:1. doi: 10.3389/fdata.2019.00001

some security functionality. Still, the risks of using encrypted data processing (e.g., access pattern disclosure Islam et al., 2012) and TEEs need to be further understood to provide scalability for the big data while minimizing realistic security and privacy risks.

Even if the data is stored in an encrypted format, legitimate users need to access the data. This implies that we need to have effective access control techniques that allow users to access the right data. Although the research community has developed a plethora of access control techniques for almost all of the important big data management systems (e.g., Relational databases Oracle, 2015, NoSql databases Ulusoy et al., 2015a; Colombo and Ferrari, 2018, social network data Carminati et al., 2009) with important capabilities, whether the existing techniques and tools could easily support the new regulatory requirements such as the ones introduced by European Union General Data Protection Directive GDPR (Voigt and Bussche, 2017) is an important question. For example, to address new regulations such as right-to-be-forgotten where users may require the deletion of data that belongs to them, we may need to better understand how the data linked and shared among multiple users in a big data system. For example, multiple users that are tagged in the same picture may have legitimate privacy claims about the picture. This implies that access control systems need to support policies based on the relationships among users and data items (e.g., Pasarella and Lobo, 2017). These observations indicate that understanding how to provide scalable, secure and privacy-aware access control mechanisms for the future big data applications ranging from personalized medicine to Internet of Things systems while satisfying new regulatory requirements would be an important research direction.

3. LINKING AND SHARING BIG DATA

In many cases, data that belongs to different sources need to be integrated while satisfying many privacy requirements. For example, a patient may visit multiple health care providers and his/her complete health records may not be available in one organization. As another example, passenger data coming from airlines may need to be linked to governmental terrorist watch lists to detect suspicious activity. To protect individual privacy, only the records belonging to government watch lists may be shared. Clearly, these types of use cases require linking potentially sensitive data belonging to the different data controllers. Over the years, private record linkage research has addressed many issues ranging from handling errors (e.g., Kuzu et al., 2013) to efficient approximate schemes that leverage cryptographic solutions (e.g., Inan et al., 2008). Still, the scalability of these techniques for multiple data sources with different privacy and security requirements have not been explored. More research is needed to make these recent developments to be deployed in practice by addressing these scalability issues.

Once data is collected and potentially linked/cleaned, it may be shared across organizations to enable novel applications and unlock potential value. For example, location data collected from mobile devices can be shared with city planners to better optimize transportations networks. Unfortunately, privacy and security issues may prevent such data sharing. Even worse, in some cases such data may be distributed among multiple parties with potentially conflicting interests. For example, different organizations may not want to share their cybersecurity incident data because of the potential concerns where a competitor may use this information for their benefit. Therefore, many issues ranging from security to privacy to incentives for sharing big data need to be considered.

From a privacy point of view, novel privacy-preserving data sharing techniques, based on a theoretically sound privacy definition named differential privacy, have been developed (e.g., Dwork, 2006). These techniques usually work by adding noise to shared data and may not be suitable in some application domains where noise free data need to be shared (e.g., health care domain). In addition, in some cases, these techniques require adding significant amount of noise to protect privacy. This in return may significantly reduce the data utility. On the other hand, some practical risk–aware data sharing tools have been developed (e.g., Prasser et al., 2017). Unfortunately, these practical risk-aware data sharing techniques do not provide the theoretical guarantees offered by differential privacy. Therefore, better understanding of the limits of privacy-preserving data sharing techniques that balance privacy risks vs. data utility need to be developed.

In many cases, misaligned incentives among the data collectors and/or processors may prevent data sharing. For example, instead of getting lab tests conducted by another health care provider, for a hospital, it may be more profitable to redo the tests. To address this type of incentive issues, secure distributed data sharing protocols that incentivize honest sharing of data have been developed (e.g., Buragohain et al., 2003). These protocols usually leverage ideas from economics and game theory to incentivize truthful sharing of big data where security concerns prevent direct auditing (e.g., Kantarcioglu and Nix, 2010; Kantarcioglu and Jiang, 2012). Still addressing incentive issues ranging from compensating individuals for sharing their data (e.g., data market places [1]) to payment systems for data sharing among industry players need to be addressed. More research that integrates ideas from economics, and psychology with computer science techniques is needed to address the incentive issues in sharing big data without sacrificing security and/or privacy.

4. ANALYZING BIG DATA

Another important research direction is to address the privacy and the security issues in analyzing big data. Especially, recent developments in machine learning techniques have created important novel applications in many fields ranging from health care to social networking while creating important privacy challenges.

Again differential privacy ideas have been applied to address privacy issues for the scenarios where all the needed data is controlled by one organization (e.g., McSherry, 2009). These techniques usually require adding noise to the results. Still, it

[1] https://datum.org

is shown that given large amount of data, these techniques can provide useful machine learning models. To address the scenarios where machine learning models need to be built by combining data that belong to different organization, many different privacy-preserving distributed machine learning protocols have been developed (e.g., Clifton et al., 2003; Kantarcıoğlu and Clifton, 2004; Vaidya and Clifton, 2005). Using cryptographic techniques, these algorithms usually provide security/privacy proofs that show nothing other than the final machine learning models are revealed. Furthermore, these results suggest that most of the privacy-preserving distributed machine learning tasks could be securely implemented by using few basic "secure building blocks" such as secure matrix operations, secure comparison, etc. (Clifton et al., 2003). Still many challenges remain in both settings. In the case of differential private techniques, for complex machine learning tasks such as deep neural networks, the privacy parameters need to adjusted properly to get the desired utility (e.g., classifier accuracy Abadi et al., 2016). The practical implications of setting such privacy parameters need to be explored further. In the case of privacy-preserving distributed machine learning techniques, except few exceptions, these techniques are not efficient enough for big data. Although leveraging trusted execution environments showed some promising results, potential leaks due to side channels need to be considered (Schuster et al., 2015; Costan and Devadas, 2016; Shaon et al., 2017). Therefore, more research is needed to scale these techniques without sacrificing security guarantees.

Unfortunately, securely building machine learning models by itself may not preserve privacy directly. It has been shown that machine learning results may be used to infer sensitive information such as sexual orientation, political affiliation (e.g., Heatherly et al., 2013), intelligence (e.g., Kosinski et al., 2013) etc. Although differential privacy techniques have shown some promise to prevent such attacks, recent results have shown that it may not be effective against many attack while providing acceptable data utility (Fredrikson et al., 2014). These results indicate the need to do more research on understanding privacy impact of machine learning models and whether the models should be built in the first place (e.g., machine learning model that tries to predict intelligence).

5. ACCOUNTABILITY ISSUES IN BIG DATA

As machine learning algorithms affect more and more aspects of our lives, it becomes crucial to understand how these algorithms change the way decisions are made in today's data-driven society. The lack of transparency in data-driven decision-making algorithms can easily conceal fallacies and risks codified in the underlying mathematical models, and nurture inequality, bias, and further division between the privileged and the under-privileged (Sweeney, 2013). Although the recent research tries to address these transparency challenges (Baeza-Yates, 2018), more research is needed to ensure fairness, and accountability in usage of machine learning models and big data driven decision algorithms. Understanding the data provenance (e.g., Bertino and Kantarcioglu, 2017) (i.e., how the data is created, who touched it etc.) have shown to improve trust in decisions and the quality of data used for decision making.

In addition to increasing accountability in decision making, more work is needed to make organizations accountable in using privacy sensitive data. With the recent regulations such as GDPR (Voigt and Bussche, 2017), using data only for the purposes consented by the individuals become critical, since personal data can be stored, analyzed and shared as long as the owner of the data consent the data usage purposes. At the same time, it is not clear whether the organizations who collect the privacy sensitive data always process the data according to user consent. An example of this problem is reflected in the recent Cambridge Analytica scandal (Cadwalladr and Graham-Harrison, 2018). In this case, it turns out that the data collected by Facebook is shared for purposes that are not explicitly consented by the individuals which the data belong. As more and more data collected, making organizations accountable for data misuse becomes more critical. It is not clear whether purely technical solutions can solve this problem, even though some research try to formalize purpose based access control and data sharing for big data (e.g., Byun and Li, 2008; Ulusoy et al., 2015b). Legal and economic solutions (e.g., rewarding insiders that report data misuse) need to be combined with technical solutions. Research that addresses this interdisciplinary area emerges as a critical need.

6. BLOCKCHAINS, BIG DATA SECURITY AND PRIVACY

The recent rise of the blockchain technologies have enabled organizations to leverage a secure distributed public ledger where important information could be stored for various purposes including increasing in transparency of the underlying economic transactions. The first application of Blockchain has been the Bitcoin (Nakamoto, 2008) cryptocurrency. Bitcoin's success has resulted in more than 1000 Blockchain based cryptocurrencies, known as alt-coins.

It turns out that blockchains may have important implications for big data security and privacy. On the one hand, combined with other cryptographic primitives, blockchain based tools (e.g., Androulaki et al., 2018) may enable more secure financial transactions (e.g., Cheng et al., 2018), data sharing (e.g., Kosba et al., 2016) and provenance storage (e.g., Ramachandran and Kantarcioglu, 2018) . On the other hand, the data stored on blockchains (e.g., financial transactions stored on Bitcoin blockchain) may be analyzed to provide novel insights about emerging data security issues. For example, it seems that cryptocurrencies are used in payments for human trafficking (Portnoff et al., 2017), ransomware (Huang et al., 2018), personal blackmails (Phetsouvanh and Oggier, 2018), and money laundering (Moser and Breuker, 2013), among many others. Blockchain Data Analytics tools (Akcora et al., 2017) and big data analysis algorithms can be used by law agencies to detect such misuse (for Law Enforcement Cooperation, 2017).

7. ADVERSARIAL ML AND ML FOR CYBERSECURITY

Like many application domains, more and more data are collected for cyber security. Examples of these collected data include system logs, network packet traces, account login formation, etc. Since the amount of data collected is ever increasing, it became impossible to analyze all the collected data manually to detect and prevent attacks. Therefore, data analytics are being applied to large volumes of security monitoring data to detect cyber security incidents (see discussion in Kantarcioglu and Xi, 2016). For example, a report from Gartner claims (MacDonald, 2012) that "Information security is becoming a big data analytics problem, where massive amounts of data will be correlated, analyzed and mined for meaningful patterns." There are many companies that already offer data analytics solutions for this important problem. Of course, data analytics is a means to an end where the ultimate goal is to provide cyber security analysts with prioritized actionable insights derived from big data.

Still, direct application of data analytics techniques to the cyber security domain may be misguided. Unlike most other application domains, cyber security applications often face adversaries who actively modify their strategies to launch new and unexpected attacks. The existence of such adversaries in cyber security creates unique challenges compared to other domains where data analytics tools are applied. First, the attack instances are frequently being modified to avoid detection. Hence a future dataset will no longer share the same properties as the current datasets. For example, attackers may change the spam e-mails written by adding some words that are typically associated with legitimate e-mails. Therefore, the spam e-mail characteristics may be changed significantly by the spammers as often as they want. Secondly, when a previously unknown attack appears, data analytics techniques need to respond to the new attack quickly and cheaply. For example, when a new type of ransomware appears in the wild, we may need to update existing data analytics techniques quickly to detect such attacks. Thirdly, adversaries can be well-funded and make big investments to camouflage the attack instances. For example, a sophisticated group of cyber attackers may create malware that can evade all the existing signature-based malware detection tools using zero day exploits (i.e., software bugs that were previously unknown). Therefore, there is an urgent need to protect machine learning models against potential attacks. Although there is an active research directions for addressing adversarial attacks in machine learning (e.g., Zhou et al., 2012; Szegedy et al., 2013; Goodfellow et al., 2014; Papernot et al., 2016; Zhou and Kantarcioglu, 2016), more research that also leverages human capabilities may be needed to counter such attacks.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

MK research was supported in part by NIH award 1R01HG006844, NSF awards CNS-1111529, CICI- 1547324, and IIS-1633331 and ARO award W911NF-17-1-0356.

REFERENCES

Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., et al. (2016). "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security* (New York, NY: ACM), 308–318. doi: 10.1145/2976749.2978318

Akcora, C. G., Gel, Y. R., and Kantarcioglu, M. (2017). Blockchain: A Graph Primer. *arXiv preprint arXiv:1708.08749* , 1–17

Androulaki, E., Barger, A., Bortnikov, V., Cachin, C., Christidis, K., De Caro, A., et al. (2018). "Hyperledger fabric: a distributed operating system for permissioned blockchains," in *Proceedings of the Thirteenth EuroSys Conference* (New York, NY: ACM), 30.

Baeza-Yates, R. (2018). Bias on the web. *Commun. ACM* 61, 54–61. doi: 10.1145/3209581

Ballard, L., Kamara, S., and Monrose, F. (2005). "Achieving efficient conjunctive keyword searches over encrypted data," in *Seventh International Conference on Information and Communication Security (ICICS 2005)* (Heidelberg: Springer), 414–426.

Bertino, E., and Kantarcioglu, M. (2017). "A cyber-provenance infrastructure for sensor-based data-intensive applications," in *2017 IEEE International Conference on Information Reuse and Integration, IRI 2017* (San Diego, CA), 108–114. doi: 10.1109/IRI.2017.91

Buragohain, C., Agrawal, D., and Suri, S. (2003). "A game theoretic framework for incentives in p2p systems," in *P2P '03: Proceedings of the 3rd International Conference on Peer-to-Peer Computing* (Washington, DC: IEEE Computer Society) 48.

Byun, J.-W., and Li, N. (2008). Purpose based access control for privacy protection in relational database systems. *VLDB J.* 17, 603–619. doi: 10.1007/s00778-006-0023-0

Cadwalladr, C. and Graham-Harrison, E. (2018). *Revealed: 50 million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach.* Available online at: https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election (Accessed on 12/21/2018).

Canim, M., and Kantarcioglu, M. (2007). "Design and analysis of querying encrypted data in relational databases," in *The 21th Annual IFIP WG 11.3 Working Conference on Data and Applications Security* (Berlin, Heidelberg: Springer-Verlag), 177–194.

Carminati, B., Ferrari, E., Heatherly, R., Kantarcioglu, M., and Thuraisingham, B. M. (2009). "A semantic web based framework for social network access control," in *SACMAT*, eds B. Carminati and J. Joshi (New York, NY: ACM), 177–186.

Chandra, S., Karande, V., Lin, Z., Khan, L., Kantarcioglu, M., and Thuraisingham, B. (2017). "Securing data analytics on sgx with randomization," in *Proceedings of the 22nd European Symposium on Research in Computer Security* (Oslo).

Chang, Y., and Mitzenmacher, M. (2005). "Privacy preserving keyword searches on remote encrypted data," in *Proceedings of ACNS'05* (New York, NY), 442–455.

Cheng, R., Zhang, F., Kos, J., He, W., Hynes, N., Johnson, N. M., et al. (2018). Ekiden: a platform for confidentiality-preserving, trustworthy, and performant smart contract execution. *CoRR* abs/1804.05141

Clifton, C., Kantarcıoğlu, M., Lin, X., Vaidya, J., and Zhu, M. (2003). Tools for privacy preserving distributed data mining. *SIGKDD Explorat.* 4, 28–34. doi: 10.1145/772862.772867

Colombo, P., and Ferrari, E. (2018). "Access control enforcement within mqtt-based internet of things ecosystems," in *Proceedings of the 23nd ACM on Symposium on Access Control Models and Technologies, SACMAT 2018* (Indianapolis, IN), 223–234. doi: 10.1145/3205977.3205986

Costan, V., and Devadas, S. (2016). *Intel sgx Explained*. Techinical Report, Cryptology ePrint Archive, Report 2016/086, 20 16. Available online at: http://eprint.iacr.org

Dwork, C. (2006). "Differential privacy," in *33rd International Colloquium on Automata, Languages and Programming- ICALP 2006* (Venice: Springer-Verlag), 1–12.

for Law Enforcement Cooperation, E. U. A. (2017). *Internet Organised Crime Threat Assessment (iocta)*. Available online at: https://www.europol.europa.eu/activities-services/main-reports/internet-organised-crime-threat-assessment-iocta-2017

Fredrikson, M., Lantz, E., Jha, S., Lin, S., Page, D., and Ristenpart, T. (2014). "Privacy in pharmacogenetics: An end-to-end case study of personalized warfarin dosing," in *23rd USENIX Security Symposium (USENIX Security 14)* (San Diego, CA: USENIX Association), 17–32.

Gentry, C. (2009). *A Fully Homomorphic Encryption Scheme*. Ph.D. thesis, Stanford University. Available online at: crypto.stanford.edu/craig

Golle, P., Staddon, J., and Waters, B. (2004). "Secure conjunctive keyword search over encrypted data," in *Applied Cryptography and Network Security (ACNS 2004)* M. Jakobsson, M. Yung, and J. Zhou (Berlin, Heidelberg: Springer), 31–45. doi: 10.1007/978-3-540-24852-1_3

Goodfellow, I. J., Shlens, J., and Szegedy, C. (2014). Explaining and harnessing adversarial examples. *arXiv[Preprint]. arXiv:1412.6572*.

Hacigumus, H., Iyer, B. R., Li, C., and Mehrotra, S. (2002). Executing SQL over encrypted data in the database-service-provider model. in *Proceedings of the 2002 ACM SIGMOD International Conference on Management of Data* (Madison, WI), 216–227.

Heatherly, R., Kantarcioglu, M., and Thuraisingham, B. M. (2013). Preventing private information inference attacks on social networks. *IEEE Trans. Knowl. Data Eng.* 25, 1849–1862. doi: 10.1109/TKDE.2012.120

Huang, D. Y., McCoy, D., Aliapoulios, M. M., Li, V. G., Invernizzi, L., Bursztein, E., et al. (2018). "Tracking ransomware end-to-end," in *Tracking Ransomware End-to-end* (San Francisco, CA: IEEE), 1–12.

Inan, A., Kantarcioglu, M., Bertino, E., and Scannapieco, M. (2008). A hybrid approach to private record linkage. *IEEE 24th International Conference on Data Engineering, 2008. ICDE 2008* 496–505. doi: 10.1109/ICDE.2008.4497458

Islam, M. S., Kuzu, M., and Kantarcioglu, M. (2012). "Access pattern disclosure on searchable encryption: Ramification, attack and mitigation," in *19th Annual Network and Distributed System Security Symposium, NDSS 2012* (San Diego, CA).

Kantarcıoğlu, M., and Clifton, C. (2004). Privacy-preserving distributed mining of association rules on horizontally partitioned data. *IEEE TKDE* 16, 1026–1037. doi: 10.1109/TKDE.2004.45

Kantarcıoğlu, M., and Clifton, C. (2005). "Security issues in querying encrypted data," in *The 19th Annual IFIP WG 11.3 Working Conference on Data and Applications Security* (Storrs, CT).

Kantarcioglu, M., and Jiang, W. (2012). Incentive compatible privacy preserving data analysis. *IEEE Transactions on Knowledge and Data Engineering* (IEEE), 1323–1335. doi: 10.1109/TKDE.2012.61

Kantarcioglu, M., and Nix, R. (2010). "Incentive compatible distributed data mining," in *Proceedings of the 2010 IEEE Second International Conference on Social Computing, SocialCom/IEEE International Conference on Privacy, Security, Risk and Trust, PASSAT 2010, Minneapolis, Minnesota, USA, August 20-22, 2010*, eds A. K. Elmagarmid and D. Agrawal (Minneapolis, MN: IEEE Computer Society), 735–742. doi: 10.1109/SocialCom.2010.114

Kantarcioglu, M., and Xi, B. (2016). "Adversarial data mining: Big data meets cyber security," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security* (Vienna), 1866–1867. doi: 10.1145/2976749.2976753

Kosba, A., Miller, A., Shi, E., Wen, Z., and Papamanthou, C. (2016). "Hawk: the blockchain model of cryptography and privacy-preserving smart contracts," in *2016 IEEE Symposium on Security and Privacy (SP)* (San Jose, CA: IEEE), 839–858. doi: 10.1109/SP.2016.55

Kosinski, M., Stillwell, D., and Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proc. Natl. Acad. Sci. U.S.A.* 110, 5802–5805. doi: 10.1073/pnas.1218772110

Kuzu, M., Kantarcioglu, M., Durham, E. A., Tóth, C., and Malin, B. (2013). A practical approach to achieve private medical record linkage in light of public resources. *JAMIA* 20, 285–292. doi: 10.1136/amiajnl-2012-000917

MacDonald, N. (2012). *Information Security is Becoming a Big Data Analytics Problem.* Available online at: https://www.gartner.com/doc/1960615/information-security-big-data-analytics (Accessed Jul 15, 2018).

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., et al. (2011). *The Big Data: The Next Frontier for Innovation, Competition, and Productivity*. Mckinsey & Company.

McSherry, F. D. (2009). "Privacy integrated queries: an extensible platform for privacy-preserving data analysis," in *SIGMOD*. (New York, NY), 19–30.

Moser, M., Bohme, R., and Breuker, D. (2013). "An inquiry into money laundering tools in the bitcoin ecosystem," in *eCrime Researchers Summit*, 1–14. doi: 10.1109/eCRS.2013.6805780

Nakamoto, S. (2008). *Bitcoin: A Peer-to-Peer Electronic Cash System.* Available online at: https://bitcoin.org/bitcoin.pdf

Ohrimenko, O., Schuster, F., Fournet, C., Mehta, A., Nowozin, S., Vaswani, K., et al. (2016). "Oblivious multi-party machine learning on trusted processors," in *25th USENIX Security Symposium (USENIX Security 16)* (Austin, TX: USENIX Association), 619–636.

Oracle (2015). *Access Control in Oracle*. Available online at:http://goo.gl/cnwQVv

Papernot, N., McDaniel, P. D., Jha, S., Fredrikson, M., Celik, Z. B., and Swami, A. (2016). "The limitations of deep learning in adversarial settings," in *IEEE European Symposium on Security and Privacy, EuroS&P 2016* (Saarbrücken), 372–387. doi: 10.1109/EuroSP.2016.36

Pasarella, E., and Lobo, J. (2017). "A datalog framework for modeling relationship-based access control policies," in *Proceedings of the 22nd ACM on Symposium on Access Control Models and Technologies, SACMAT 2017* (Indianapolis), 91–102. doi: 10.1145/3078861.3078871

Phetsouvanh, A. D. S., and Oggier, F. (2018). "Egret: extortion graph exploration techniques in the bitcoin network," in *IEEE ICDM Workshop on Data Mining in Networks (DaMNet)*.

Portnoff, R. S., Huang, D. Y., Doerfler, P., Afroz, S., and McCoy, D. (2017). "Backpage and bitcoin: Uncovering human traffickers," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Halifax, NS), 1595–1604. doi: 10.1145/3097983.3098082

Prasser, F., Gaupp, J., Wan, Z., Xia, W., Vorobeychik, Y., Kantarcioglu, M., et al. (2017). "An open source tool for game theoretic health data de-identification," in *AMIA 2017, American Medical Informatics Association Annual Symposium* (Washington, DC).

Ramachandran, A., and Kantarcioglu, M. (2018). "Smartprovenance: a distributed, blockchain based dataprovenance system," in *Proceedings of the Eighth ACM Conference on Data and Application Security and Privacy, CODASPY 2018* (Tempe, AZ), 35–42. doi: 10.1145/3176258.3176333

Schuster, F., Costa, M., Fournet, C., Gkantsidis, C., Peinado, M., Mainar-Ruiz, G., et al. (2015). "Vc3: trustworthy data analytics in the cloud using sgx," in *2015 IEEE Symposium on Security and Privacy (SP)* (San Jose, CA: IEEE), 38–54. doi: 10.1109/SP.2015.10

Schwab, K. (2016). *The Fourth Industrial Revolution*. Available online at: http://www3.weforum.org/docs/Media/KSC_4IR.pdf. (Accessed on 10/17/2016)

Shaon, F., and Kantarcioglu, M. (2016). "A practical framework for executing complex queries over encrypted multimedia data," in *Proceedings on 30th Annual IFIP WG 11.3 Conference on Data and Applications Security and Privacy XXX DBSec 2016* (Trento), 179–195. doi: 10.1007/978-3-319-41483-6_14

Shaon, F., Kantarcioglu, M., Lin, Z., and Khan, L. (2017). "Sgx-bigmatrix: a practical encrypted data analytic framework with trusted processors," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, CCS 2017* (Dallas, TX), 1211–1228. doi: 10.1145/3133956.3134095

Shi, E., Bethencourt, J., Chan, T.-H. H., Song, D., and Perrig, A. (2007). "Multi-dimensional range query over encrypted data," in *SP '07: Proceedings of the 2007 IEEE Symposium on Security and Privacy* (Washington, DC: IEEE Computer Society), 350–364.

Song, D. X., Wagner, D., and Perrig, A. (2000). "Practical techniques for searches on encrypted data," in *IEEE SP* (Washington, DC), 44–55.

Sweeney, L. (2013). Discrimination in online ad delivery. *Commun. ACM* 56, 44–54. doi: 10.1145/2447976.2447990

Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I. J., et al. (2013). Intriguing properties of neural networks. *arXiv[Preprint]. arXiv:1312.6199*.

Ulusoy, H., Colombo, P., Ferrari, E., Kantarcioglu, M., and Pattuk, E. (2015a). "Guardmr: fine-grained security policy enforcement for mapreduce systems," in *Proceedings of the 10th ACM Symposium on Information, Computer and Communications Security, ASIA CCS* (Singapore), 285–296. doi: 10.1145/2714576.2714624

Ulusoy, H., Kantarcioglu, M., Pattuk, E., and Kagal, L. (2015b). "Accountablemr: toward accountable mapreduce systems," in *2015 IEEE International Conference on Big Data, Big Data 2015* (Santa Clara, CA), 451–460. doi: 10.1109/BigData.2015.7363786

Vaidya, J., and Clifton, C. (2005). "Privacy-preserving decision trees over vertically partitioned data," in *The 19th Annual IFIP WG 11.3 Working Conference on Data and Applications Security* (Storrs, CT: Springer).

Voigt, P., and Bussche, A. V. D. (2017). *The EU General Data Protection Regulation (GDPR): A Practical Guide*. Springer Publishing Company, Incorporated.

Zheng, W., Dave, A., Beekman, J., Popa, R. A., Gonzalez, J., and Stoica, I. (2017). "Opaque: a data analytics platform with strong security," in *14th USENIX Symposium on Networked Systems Design and Implementation (NSDI 17)* (Boston, MA: USENIX Association).

Zhou, Y., and Kantarcioglu, M. (2016). "Modeling adversarial learning as nested Stackelberg games," in *Advances in Knowledge Discovery and Data Mining - 20th Pacific-Asia Conference, PAKDD 2016, Proceedings, Part II*, vol. 9652 of *Lecture Notes in Computer Science* eds J. Bailey, L. Khan, T. Washio, G. Dobbie, J. Z. Huang, and R. Wang (Auckland: Springer), 350–362.

Zhou, Y., Kantarcioglu, M., Thuraisingham, B., and Xi, B. (2012). "Adversarial support vector machine learning," in *Proceedings of the 18th ACM SIGKDD international Conference on Knowledge Discovery and Data Mining, KDD '12* (New York, NY: ACM), 1059–1067. doi: 10.1145/2339530.2339697

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.