# Chapter 1
# Big Data and Privacy: Challenges and Opportunities


**Amin Azmoodeh and Ali Dehghantanha**

## 1 Introduction

The contemporary decade is distinguished for the explosion of information that is generating, transferring, and storing over vast and complex networks [1]. Technological advancements in information technology are creating a sea change in today's life. A majority of public and private sectors [2] beyond different industries are utilizing digital devices and procedures to provide their clients with high quality and reliable services. This widespread usage ranging from healthcare [3] and transport systems [4] to smart grids [5] and military services [6] has resulted in an inconceivable volume of data being generated and processed. The importance and sensitivity of such big data have turned it into an invaluable target for cybercriminals.

The privacy of big data has acquired new urgency due to the different issues linked to it [7, 8]. Regulating the pace of data growth with confidentiality, integrity, and availability of data processing technique is a challenging issue [9] which should be addressed. Moreover, investigation of big data on cloud-based platforms to identify and recover traces of criminal activities for forensic investigations is a time-consuming process [10] that demands novel approaches to overcome this challenge. Besides, big data storage, processing, sharing and management are crucial procedures [11] that should be carefully tuned because it may increase attack surface for malicious activities and data leakage.

On the other hand, and in terms of big data advantages, big data provides exemplary opportunities to leverage the high volume of data. It is projected that increasing data will lead to in-depth knowledge about the domain of data. Consequently, extracting in-depth knowledge from big data paves the way for proposing robust and

A. Azmoodeh (✉) · A. Dehghantanha
Cyber Science Lab, University of Guelph, Guelph, ON, Canada
e-mail: amin@cybersciencelab.org; ali@cybersciencelab.org



K.-K. R. Choo, A. Dehghantanha (eds.), *Handbook of Big Data Privacy*,
https://doi.org/10.1007/978-3-030-38557-6_1

outstanding mechanisms for protecting data and securing information technology networks [12]. Besides, storing big data and recovery mechanisms designed for it provides forensic investigators with more pieces of evidence that lead them to quick and accurate decision making [13].

## 2 Book Outline

This handbook presents existing state-of-the-art advances from both academia and industry, in big data and privacy. The remainder of the book is structured as follows. The second chapter [14] reviews security challenges and concerns related to critical infrastructure and methods that utilize artificial intelligence to protect these infrastructures. In the third chapter [15], authors survey new concepts, methodologies, and applications to achieve full autonomy in industry 4.0. In the fourth chapter [16], Moghadam et al. propose a privacy protection key agreement protocol for smart grid based on energy consumption controllers (ECC).

The fifth chapter [17] reviews the application of machine learning for the Internet of Things and discuss about their challenges and issues. In the subsequent chapter [18] (sixth chapter), Peters et al. apply different machine learning methods on the Internet of Things malware dataset and compare their performance and discuss the results. Singh et al. [19](seventh chapter) survey about the latest artificial intelligence based researches and methodologies undertaken for measuring and managing industrial cyber threats risks and security metrics that have been identified as a barrier to implementing these methodologies.

Eighth chapter [20] gives information about traditional machine learning based threat detection techniques for network security that are incapable of facing with huge amount of data so as to obtain more efficient knowledge to design and choose such techniques. In the next chapter, Sharma et al. [21] propose a multi-level network security and privacy evaluation scheme to evaluate and assess the security of cyber physical systems. Chapter 10 [22] is dedicated to machine learning approaches for cyber physical system anomaly detection. Then, through a case study, authors demonstrate the effectiveness of machine learning techniques for classifying False Data Injection attacks. The next chapter (Chapter 11) [23] briefly introduces renewable energy resources as well as different aspects and relations of security and big data for power systems using such resources. In the subsequent chapter, Cabello et al. [24] describe the importance of using cyber-physical systems and big data in healthcare sector. Chapter 13 [25] proposes a deep learning approach for abnormality detection while preserve privacy for a medical images dataset.

In order to provide a clear insight about researches related to security of smart farming, Nakhodchi et al. [26] in the fourteenth chapter propose a bibliometric analysis to comprehensively assess security and privacy of smart agriculture systems and related literature. In the next chapter, Amrollahi et al. [27] highlight the impact of big data and privacy in financial systems and survey the work related to FinTech banking cyber security concerns and detection methods. Chapter 16 [28] proposes

a hybrid deep generative metric learning approach for intrusion detection and protect critical infrastructures. Nassiri et al. [29] present a method that combines the static and dynamic machine learning based malware detection methods. They experimentally demonstrate the performance of their proposed method. In the subsequent chapter [30], BehradFar et al. introduce a machine learning algorithm that applies a two-layer feature selection to obtain the optimum set of features for Remote access Trojan (RAT) detection and achieve high performance for RAT detection. In the last chapter [31], Azmoodeh et al. propose an active spectral clustering method to tackle problem of massive data in botnet detection research sphere that consumes the minimum number of similarity between network nodes to identify botnets.

## References

1. A. Azmoodeh, A. Dehghantanha, K.-K.R. Choo, *Big Data and Internet of Things Security and Forensics: Challenges and Opportunities* (Springer International Publishing, Cham, 2019), pp. 1–4
2. V. Ho, A. Dehghantanha, K. Shanmugam, A guideline to enforce data protection and privacy digital laws in Malaysia, in *2010 Second International Conference on Computer Research and Development* (IEEE, Piscataway, 2010), pp. 3–6
3. S. Walker-Roberts, M. Hammoudeh, A. Dehghantanha, A systematic review of the availability and efficacy of countermeasures to internal threats in healthcare critical infrastructure. IEEE Access **6**, 25167–25177 (2018)
4. G. Epiphaniou, P. Karadimas, D. Kbaier Ben Ismail, H. Al-Khateeb, A. Dehghantanha, K.R. Choo, Nonreciprocity compensation combined with turbo codes for secret key generation in vehicular Ad Hoc social IoT networks. IEEE Internet Things J. **5**(4), 2496–2505 (2018)
5. H. Karimipour, A. Dehghantanha, R.M. Parizi, K.R. Choo, H. Leung, A deep and scalable unsupervised machine learning system for cyber-attack detection in large-scale smart grids. IEEE Access **7**, 80778–80788 (2019)
6. A. Azmoodeh, A. Dehghantanha, K.R. Choo, Robust malware detection for internet of (battlefield) things devices using deep eigenspace learning. IEEE Trans. Sustain. Comput. **4**(1), 88–95 (2019)
7. R. Bao, Z. Chen, M.S. Obaidat, Challenges and techniques in big data security and privacy: a review. Secur. Priv. **1**(4), e13 (2018)
8. S. Yu, Big privacy: challenges and opportunities of privacy study in the age of big data. IEEE Access **4**, 2751–2763 (2016)
9. S. Nepal, R. Ranjan, K.R. Choo, Trustworthy processing of healthcare big data in hybrid clouds. IEEE Cloud Comput. **2**(2), 78–84 (2015)
10. Y. Teing, A. Dehghantanha, K.R. Choo, Z. Muda, M.T. Abdullah, Greening cloud-enabled big data storage forensics: syncany as a case study. IEEE Trans. Sustain. Comput. **4**(2), 204–216 (2019)
11. C. Yang, Q. Huang, Z. Li, K. Liu, F. Hu, Big data and cloud computing: innovation opportunities and challenges. Int. J. Digital Earth **10**(1), 13–53 (2017)
12. P.J. Taylor, T. Dargahi, A. Dehghantanha, *Analysis of APT Actors Targeting IoT and Big Data Systems: Shell_Crew, NetTraveler, ProjectSauron, CopyKittens, Volatile Cedar and Transparent Tribe as a Case Study* (Springer International Publishing, Cham, 2019), pp. 257–272
13. Y.-Y. Teing, A. Dehghantanha, K.-K.R. Choo, CloudMe forensics: a case of big data forensic investigation. Concurr. Comput. Pract. Exp. **30**(5), e4277 (2017)

14. J. Sakhnini, H. Karimipour, A. Dehghantanha, R.M. Parizi, AI and security of critical infrastructure, in *Big Data and Privacy*, ed. by K.-K.R. Choo, A. Dehghantanha (Springer, Cham, 2020). https://doi.org/10.1007/978-3-030-38557-6_2
15. A. Al-Abassi, H. Karimipour, H.H. Pajouh, A. Dehghantanha, R.M. Parizi, Industrial big data analytics: challenges and opportunities, in *Big Data and Privacy*, ed. by K.-K.R. Choo, A. Dehghantanha (Springer, Cham, 2020). https://doi.org/10.1007/978-3-030-38557-6_3
16. M.F. Moghadam, A. Mohajerzdeh, H. Karimipour, H. Chitsaz, R. Karimi, B. Molavi, A privacy protection key agreement protocol based on ECC for smart grid, in *Big Data and Privacy*, ed. by K.-K.R. Choo, A. Dehghantanha (Springer, Cham, 2020). https://doi.org/10.1007/978-3-030-38557-6_4
17. S. Yousefi, F. Derakhshan, H. Karimipour, Applications of big data analytics and machine learning in the internet of things, in *Big Data and Privacy*, ed. by K.-K.R. Choo, A. Dehghantanha (Springer, Cham, 2020). https://doi.org/10.1007/978-3-030-38557-6_5
18. W. Peters, A. Dehghantanha, R.M. Parizi, G. Srivastava, A comparison of various machine learning models for opcode based internet of things malware detection, in *Big Data and Privacy*, ed. by K.-K.R. Choo, A. Dehghantanha (Springer, Cham, 2020). https://doi.org/10.1007/978-3-030-38557-6_6
19. S. Singh, H. Karimipour, H. HaddadPajouh, A. Dehghantanha, Artificial intelligence and security of industrial control systems, in *Big Data and Privacy*, K.-K. R. Choo, A. Dehghantanha (Springer, Cham, 2020). https://doi.org/10.1007/978-3-030-38557-6_7
20. M. Amrollahi, S. Hadayeghparast, H. Karimipour, F. Derakhshan, G. Srivastava, Enhancing network security via machine learning: opportunities and challenges, in *Big Data and Privacy*, ed. by K.-K.R. Choo, A. Dehghantanha (Springer, Cham, 2020). https://doi.org/10.1007/978-3-030-38557-6_8
21. M. Sharma, H. Elmiligi, F. Gebali, Network security and privacy evaluation scheme for cyber physical systems (CPS), in *Big Data and Privacy*, ed. by K.-K.R. Choo, A. Dehghantanha (Springer, Cham, 2020). https://doi.org/10.1007/978-3-030-38557-6_9
22. H.M. Rouzbahani, H. Karimipour, A. Rahimnejad, A. Dehghantanha, G. Srivastava, Anomaly attack detection in cyber-physical systems using machine learning, in *Big Data and Privacy*, ed. by K.-K.R. Choo, A. Dehghantanha (Springer, Cham, 2020). https://doi.org/10.1007/978-3-030-38557-6_10
23. H.M. Rouzbahani, H. Karimipour, G. Srivastava, Big data application for renewable energy resource security, in *Big Data and Privacy*, ed. by K.-K.R. Choo, A. Dehghantanha (Springer, Cham, 2020). https://doi.org/10.1007/978-3-030-38557-6_11
24. J.C. Cabello, A.N. Jahromi, H. Karimipour, A. Dehghantanha, R.M. Parizi, Big-data and cyber-physical systems in healthcare: challenges and opportunities, in *Big Data and Privacy*, ed. by K.-K. R. Choo, A. Dehghantanha (Springer, Cham, 2020). https://doi.org/10.1007/978-3-030-38557-6_12
25. W. Han, A. Azmoodeh, H. Karimipour, S. Yang, Privacy preserving abnormality detection: a deep learning approach, in *Big Data and Privacy*, ed. by K.-K.R. Choo, A. Dehghantanha (Springer, Cham, 2020). https://doi.org/10.1007/978-3-030-38557-6_13
26. S. Nakhodchi, A. Dehghantanha, H. Karimipour, Privacy and security in smart and precision farming: a bibliometric analysis," in *Big Data and Privacy*, ed. by K.-K.R. Choo, A. Dehghantanha (Springer, Cham, 2020). https://doi.org/10.1007/978-3-030-38557-6_14
27. M. Amrollahi, A. Dehghantanha, R.M. Parizi, A survey on application of big data in fin tech banking security and privacy, in *Big Data and Privacy*, ed. by K.-K.R. Choo, A. Dehghantanha, (Springer, Cham, 2020). https://doi.org/10.1007/978-3-030-38557-6_15
28. M. Saharkhizan, A. Azmoodeh, H. HaddadPajouh, A. Dehghantanha, R.M. Parizi, G. Srivastava, A hybrid deep generative local metric learning method for intrusion detection, in *Big Data and Privacy*, ed. by K.-K.R. Choo, A. Dehghantanha (Springer, Cham, 2020). https://doi.org/10.1007/978-3-030-38557-6_16

29. M. Nassiri, H. HaddadPajouh, A. Dehghantanha, H. Karimipour, R.M. Parizi, G. Srivastava, Malware elimination impact on dynamic analysis: an experimental analysis on machine learning approach, in *Big Data and Privacy*, ed. by K.-K.R. Choo, A. Dehghantanha (Springer, Cham, 2020). https://doi.org/10.1007/978-3-030-38557-6_17
30. M.M. BehradFar, H. HaddadPajouh, A. Dehghantanha, A. Azmoodeh, H. Karimipour, R.M. Parizi, G. Srivastava, Rat hunter: building robust models for detecting rats based on optimum hybrid features, in *Big Data and Privacy*, K.-K.R. Choo, A. Dehghantanha (Springer, Cham, 2020). https://doi.org/10.1007/978-3-030-38557-6_18
31. A. Azmoodeh, A. Dehghantanha, R.M. Parizi, S. Hashemi, B. Gharabaghi, G. Srivastava, Active spectral botnet detection based on eigenvalue weighting, in *Big Data and Privacy*, ed. by K.-K.R. Choo, A. Dehghantanha (Springer, Cham, 2020). https://doi.org/10.1007/978-3-030-38557-6_19