■ 751

Editorial

Fortifying Big Data infrastructures to Face Security and Privacy Issues

Tole Sutikno*[1], Deris Stiawan[2], Imam Much Ibnu Subroto*[3]
[1]Department of Electrical Engineering, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
[2]Department of Computer System Engineering, Universitas Sriwijaya, Palembang, Indonesia
[3]Department of Informatics Engineering, Universitas Islam Sultan Agung, Semarang, Indonesia
*Corresponding author, email: tole@ee.uad.ac.id[1], deris.stiawan@gmail.com[2], imam.utm@gmail.com[3]

The explosion of data available on the internet is very increasing in recent years. Human beings create more than 4 quintillion bytes of data every day in 2013, which come from individual archives, sensors embedded in smartphones, social networks, internet of things, enterprise and internet in all scales and formats [1],[2]. One of the most challenging issues is how to effectively manage such a large amount of data and identify new ways to analyze large amounts of data and unlock information [3]. The issue is well-known as Big Data, which has been emerging as a hot topic in current information and communication technologies research because of facing many challenges, such as its efficient encryption and decryption algorithms, encrypted information retrieval, attribute based encryption, attacks on availability, reliability and integrity [4],[5].

Big data comes in many forms. It can come with big differences. It is properly be tagged as the four HVs: high-volume, high-variety, high-velocity, and high-veracity [6]. Big Data is a term defining data that has four main characteristics [7]: 1) It involves a great volume of data, 2) the data cannot be structured into regular database tables; 3) the data is produced with great velocity and must be captured and processed rapidly, and 4) sometimes there is a very big volume of data to process before finding valuable needed information. Google, Microsoft, Yahoo, YouTube, Twitter, and Facebook are early innovators in big data infrastructure. Although the companies have been developing big data infrastructure since their inception, only more recently have big data workloads been running in the public cloud [8]. Facebook reports about 6 billion new photos every month and 72 hours of video are uploaded to YouTube every minute [9]. So, it's a big challenge. Hence, organizations must find a way to manage their data in accordance with all relevant privacy regulations without making the data inaccessible and unusable.

Cloud Security Alliance (CSA) has released that the top 10 challenges, which are as follows [10]: 1) secure computations in distributed programming frameworks, 2) security best practices for non-relational data stores, 3) secure data storage and transactions logs, 4) end-point input validation/filtering, 5) real-time security monitoring, 6) scalable and composable privacy-preserving data mining and analytics, 7) cryptographically enforced data centric security, 8) granular access control, 9) granular audits, 10) data Provenance. The challenges themselves can be organized into four distinct aspects of the Big Data ecosystem [10].

1. Infrastructure Security (included: 1st and 2nd challenges)

The distributed computations and data stores must be secured in order to secure the infrastructure of Big Data systems [10]. A way of supporting the security to detect anomalies which constitute threats to the system is currently placed in critical infrastructures by using behavioural observation and big data analysis techniques [11].

2. Data Privacy (included: 6th, 7th and 8th challenges)

For securing the data itself, information dissemination must be privacy-preserving and cryptography and granular access control must be used to protect sensitive data [10]. Cloud

computing provides promising scalable IT infrastructure to support various processing of a variety of big data applications, but brings about privacy concerns potentially if the information is released or shared to third-parties in cloud [12].

3. Data Management (included: 3rd, 9th and 10th challenges)

Big Data is a colossal amount of data that cannot be handled by the traditional data management system [13]-[15]. A modern data management system is very needed for or storage and retrieval of the big data.

4. Integrity and Reactive Security (included: 4th and 5th challenges)

For integrity purpose, the streaming data emerging from diverse end-points must be checked; and for reactive security purpose the streaming data can be used to perform real-time analytics in order to ensure the infrastructure health [10].

This journal is sponsoring an international conference entitled "the 2015 International Conference on Science in Information Technology" on October 27-28, 2015 which will be conducted under the theme: "The Role of Business Intelligence in Big Data Management". This conference is hoped to be a forum for dialogues to look at various issues, aiming at finding the right solutions to initial round of big data management technologies are inherently well-suited for real-time operations according to the top 10 challenges above.

References

[1] Veiga Neves, M., et al. *Pythia: Faster Big Data in Motion through Predictive Software-Defined Network Optimization at Runtime*. in Parallel and Distributed Processing Symposium, 2014 IEEE 28th International. 2014.
[2] Yangqing, Z., Z. Jun. *Probe into setting up big data processing specialty in Chinese universities*. in Computer Science & Education (ICCSE), 2014 9th International Conference on. 2014.
[3] Malik, P., Governing Big Data: Principles and practices. *IBM Journal of Research and Development*, 2013; 57(3/4): 1-13.
[4] Takaishi, D., et al. *Towards Energy Efficient Big Data Gathering in Densely Distributed Sensor Networks.* Emerging Topics in Computing, IEEE Transactions on, 2014; **PP**(99): 1-1.
[5] Juan, Z., Z. Yanqin, J. Fangyuan. *Practical and Secure Outsourcing of Linear Algebra in the Cloud*. in Advanced Cloud and Big Data (CBD), 2013 International Conference on. 2013.
[6] Courtney, M. Puzzling out big data [Information Technology Analytics]. *Engineering & Technology*. 2013; **7**(12): 56-60.
[7] Garlasu, D., et al. *A big data implementation based on Grid computing*. in Roedunet International Conference (RoEduNet), 2013 11th. 2013.
[8] Collins, E. Big Data in the Public Cloud. *Cloud Computing, IEEE*. 2014; 1(2): 13-15.
[9] Bari, N., D. Liao, S. Berkovich. *Organization of Meta-knowledge in the Form of 23-Bit Templates for Big Data Processing*. in Computing for Geospatial Research and Application (COM.Geo), 2014 Fifth International Conference on. 2014.
[10] *Expanded Top Ten Big Data Security and Privacy Challenges*. 2013.
[11] Hurst, W., M. Merabti, P. Fergus. *Big Data Analysis Techniques for Cyber-threat Detection in Critical Infrastructures*. in Advanced Information Networking and Applications Workshops (WAINA), 2014 28th International Conference on. 2014.
[12] Zhang, X., et al. *Proximity-Aware Local-Recoding Anonymization with MapReduce for Scalable Big Data Privacy Preservation in Cloud.* Computers, IEEE Transactions on, 2014. **PP**(99): p. 1-1.
[13] Padhy, R., U. Berhampur. Big Data Processing with HadoopMapReduce in Cloud Systems. *International Journal of Cloud Computing and Services Science (IJ-CLOSER)*. 2013; 2(1): 16-27.
[14] Han, H., et al. A Privacy DataOriented Hierarchical MapReduce Programming Model. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2013; 11(8): 4587-4593.
[15] Liu, Y., et al. Checkpoint and Replication Oriented Fault Tolerant Mechanism for MapReduce Framework. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2014; 12(2): 1029-1036.