Security Threats for Big Data: An Empirical Study

Muhammad Nazrul Islam, Military Institute of Science and Technology (MIST), Bangladesh
Tarannum Zaki, Military Institute of Science and Technology (MIST), Bangladesh
Md. Sami Uddin, Military Institute of Science and Technology (MIST), Bangladesh
Md. Mahedi Hasan, Military Institute of Science and Technology (MIST), Bangladesh

ABSTRACT

With the advancement of modern science and technology, data emerging from different fields are escalating gradually. Recently, with this huge amount of data, Big Data has become a source of immense opportunities for large scale organizations related to various business sectors as well as to information technology (IT) professionals. Hence, one of the biggest challenges of this context is the security of this big set of data in different ICT based organizations. The fundamental objective of this article is to explore how big data may create security challenges in email communication. As an outcome, this article first shows that big data analysis helps to understand the behavior or interest of email users, which in turn can help phishers to create the phishing sites or emails that result in IT security threat; second, the article finds that phishing e-mail generation based on the (email) users’ behavior can break an organization’s IT security; and finally, a framework was proposed that would help to enhance the security of email communication.

KEYWORDS

Big data, Empirical Study, Enron, Information Security, Phishing, Security

1. INTRODUCTION

A fundamental and concerning attribute in the field of information and communication technology is security. Security is a very important factor in every part of life whether it is personal, social or commercial. People need to protect data that are subtle and often accessible to different applications such as online banking, shopping, tourism, interpersonal communication, and so forth (Mindfulsecurity.com, 2009). Likewise, different organizations need to secure their clients’ and workers’ data related to a company’s business strategy, monetary results and many more (The Barking Seal, 2008). The results of security breach are extremely intense on both personal and corporate level. Such outcomes may bring about budgetary misfortune, notoriety, composition of protection and many more serious cases. As a result, conservation of data security, understanding the related dangers and results of security break are some incredible concerns related to the protection of data from unapproved and felonious activities.

Big data suggests a huge quantity of dataset that can reveal interesting and useful patterns related to user interactivity through computational investigation (Oxford Dictionaries, n.d.). The idea of big data is regularly embraced by vast associations where a huge amount of information is kept up. Presently, various types of data analysis method can be utilized to reveal a lot of valuable patterns or
trends related to human conduct and communication. For example, for a certain period, a significant change is observed in a person’s social activity on internet (e.g. Facebook) that does not go with his regular conduct and hence it is a common notion that his account has been hacked. Hence, it has become an absolute necessity to take over concrete and compelling safety efforts in order to anticipate such immense mass of information leakage.

These days, the most broadly utilized methods for correspondence is email because of its ease of use, lessened time zone barriers, quick delivery, a decrease in topographical troubles, inexpensiveness and effectiveness (Acevedo, 2018). Due to these inborn points of interest and utilization of email correspondence, creating defensive measures for email security is an essential issue to guarantee IT security. There are three essential characteristics of email security – Confidentiality, Integrity, and Availability (Cocca, 2004). Abuse of one of the characteristics hugely affects email security. The personal and corporate messages that are sent over the external and untrusted networks through email communication are huge in quantity, therefore it is an excellent prey for cybercriminals to damage the email security. Malware attack, email spoofing, spam and phishing, entities with malicious intent are some instances that can hamper the security of email communication by unlawfully accessing into private data which can eventually destroy an entire organization (Cocca, 2004; CDW.com, n.d.).

Therefore, this paper has aimed to demonstrate how big data may create security threats to corporative email users by generating phishing sites based on the (email) users’ behavior or interest. In order to attain this aim, the first study that the authors have conducted is a case study on Enron Email Dataset (Edrm.net n.d.). The second study was conducted on an engineering institute in Bangladesh to observe how the phishing email generation based on users’ (students) behavior or interest may break the IT security of that institute. It is worthy to mention here that the authors have conducted two studies on two organizations because, on one hand, the phishing emails (generated based on the Enron’s users’ behavior or interest) were not possible to send to Enron email users since the Enron email addresses are not still functioning. On the other hand, the authors of this article are working in an engineering institute where it was easy to collect a relevant dataset in order to understand the students’ behavior or interest and also, they had access to their email addresses.

The overall research question thus is formulated as ‘How does big data create security challenges in email communication?’ To provide an effective answer, the research project will show firstly, how big data analysis helps to depict the (email) users’ behavior or interest, and secondly, how phishing email generation based on users’ intention or behavior may break the security system, and finally propose a framework to improve the email security system for an institute or organization.

The paper is organized as follows. In Section 2, the related work of this research is discussed briefly to show the contribution of big data in IT as well as to highlight the research gaps in the field of big data and IT security for future research. In Section 3, a detail study methodology of Study I is discussed. The findings of Study I are briefly presented in Section 4. The study methodology and outcomes of Study II are discussed in Section 5. The proposed framework to protect email from phishing threats are presented in Section 6. The limitations, future work, and a brief concluding remark are presented in the final section.

2. RELATED WORK

This section briefly presents the works relevant to the respective research of this paper.

In (Khoni et al., 2013), a literature survey was conducted to explore the phishing detection and the anti-phishing software techniques. In this work, they showed a high-level overview of phishing mitigation techniques, that include, for example, offensive defense, detection, prevention, and correction. They found that blacklists, rule-based heuristics, visual similarity, and machine learning-based classifiers are mostly used software detection solutions that exist in the literature. They also found that the use of Machine Learning techniques is promising to compare with the rule-based heuristic techniques. In another study (Sahu & Dubey, 2014), a survey was carried out on phishing
New Perspectives on Adoption of RFID Technology for Agrifood Traceability
www.igi-global.com/chapter/new-perspectives-adoption-rfid-technology/52434?camid=4v1a

Economics & Practical Applications for Applied Trauma Theory: Sustainable Energy & Rural Tourism
www.igi-global.com/article/economics-practical-applications-applied-trauma/77645?camid=4v1a