paper did not use our format,

Big Data Analytics in Cyber Security and Threat Research

Tousif Ahmed
Indiana University Bloomington
107 S Indiana Ave
Bloomington, Indiana 47408
touahmed@indiana.edu

ABSTRACT

The introduction of big data poses new security threats for the organizations and the consequences can cause catastrophic damages for organizations. Newer threats are sophisticated enough that the existing security mechanisms might be ineffective to thwart the attacks. However, big data analytics and tools can help the organizations to detect the threats and take protective measures. Moreover, big data analytics and machine learning together can detect newer threats which was not possible before. In this paper, we discuss some new cybersecurity threats and challenges that has been bolstered by big data and then we discuss some new big data related security mechanisms which can help the organizations to protect their resources.

KEYWORDS

HID 237,E534, Big Data, Cyber Securty, Threat Intelligence.

1 INTRODUCTION

With the rapid growth of consumer based services, artificial intelligence, and social medias, the number security and privacy threats are also increasing. Emerging cyberattacks are not only a threat for industries, end-users are at more risk than ever. A recent stat published by MalwareBytes suggests that there are a rapid proliferation of number of security threats recently and more than 1 billion malwares were detected in their collected data of six month time span [1]. The consequence of the cyberattacks can be catastrophic to the Organization and the security of the consumers. Therefore, proper defensive mechanism and security controls need to be placed to impede cyber threats. However, newer threats have become so sophisticated that now it's extremely difficult detect and mitigate newer threats. Existing security mechanisms are ineffective due to the large volume of traffic on the internet and lack of resources.

To accommodate and protect from new line cyber attacks, industries now rely on big data analytics. These new big data analytics shows promises for these new sophisticated cyberattacks and a new area of research has established which studies newer threats and explore defensive mechanisms. Advanced analytics and a combination of machine learning can improve the cyber security and provide a new defensive mechanisms. For example, with the large number of traffic on websites big data analytics enables the security administrators to detect intruders effectively and prevent them from compromising the system. Moreover, the big data analytics can also provide newer lines of defense against consumer's security threat. For example, financial industries can analyze the user's financial behavior and detect anomalies which may help users to mitigate the consequence of stolen credit card or identity theft. The introduction big data analytics poses interesting conundrum for security and privacy, on one hand it creates new security and privacy risks and on the other hand big data analytics provides newer tools for mitigating security threats.

In this paper, we first discuss the cyber security and privacy threats of big data analytics on organizations and consumers, then we discuss some potential applications that can be used to mitigate cyber security risks.

2 CYBER SECURITY THREATS AND CHALLENGES

In this section, we briefly discuss various security threats that is posed by the introduction of big data analytics:

2.1 Increased potential for security breaches

Now, organizations are collecting more data which increased the motivation of the attackers to exploit the organization's vulnerabilities and breach their security. The main objective of the attackers are accessing and downloading consumer's data and that data can be sold to other companies or those information can be used to infiltrate more sensitive data [3]. For example, if the user's email and birth date can be revealed from one system, that data can be used to infer other sensitive information like banking information. The availability of the data motivates the attacker to attack a system and gain access. Recent series of security breaches on high profile companies like Yahoo [9] and Equifax [12] provides examples of increased potential, and the number of affected users provides an example of the consequences.

2.2 Threats to consumers privacy

With more data and available tools, consumer's privacy is at more risk than ever. Users share bits of their personal information on various websites and the combined information from various websites poses new security and privacy threats to the consumers. Consumer's privacy threat is a big risk for an organization as the security of the user's data is correlated with the reputation of the organization. Therefore, thwarting the security attacks have become an important issue for the organizations.

2.3 Sophisticated vulnerabilities

With sophisticated big data tools, attackers are now implementing more intelligent spams, malware, and website threats [5].By using machine learning tools, newer generation of spams have been proliferated which is now hard to detect. The rise of chatbots and automated text generating tools have enabled creating spams extremely easy. Moreover, social medias have made spam distribution

[1] https://press.malwarebytes.com/2017/01/31/malwarebytes-releases-global-state-of-malware-report-finds-2016-as-year-threat-reality-catches-up-to-threat-hype/

extremely easy. Besides spams and malwares, increasing number of device usages (smartphone, IoT devices) have increased the number of BotNets. Using analytics tools, now botnet distribution and management has become pretty easy [4, 10]. The emergence of deep learning has also motivated newer types of security threats.

2.4 Complex security management and monitoring

With high volume of data, now security management is extremely complex. It is very difficult to correctly assess the risks of a system and monitor the networks. With millions of users accessing websites, now it is almost impossible to scrutinize the network traffics. The high volume of data creates additional security risks for organizations. Existing signature based intrusion detection has become irrelevant with higher number of traffics and unpredictable nature of the users.

3 BIG DATA ANALYTICS FOR CYBER DEFENCE AND THREAT RESEARCH

In response to newer security threats, big data analytics have been used to provide newer set of tools to the security administrators. Based on the existing researches and news, in this section we discuss some newer techniques for cyber defence:

3.1 Scalable Anomaly detection

The most widely used big data tools for cyber defense is the anomaly detection. Now, with the help of numerous data it has become extremely easy to detect anomalies. Anomaly or abnormal behaviour detection is pretty easy to detect with large volume of data, as most user's exhibits common behavior or patterns. Illegal or bad actors act differently while accessing a system and using clusters it has become very easy to detect anomalies. Nowadays, anomaly detection systems have been incorporated to detect scammers, credit card theives, hackers, and potential intruders. Network monitoring schemes have become extremely scalable and efficient, so that it can easily raise an alert once an abnormal behavior exhibits [7, 8].

3.2 Effective Malware Analysis

The existing ways to detect malwares are highly inefficient as it highly relies on the previously seen malwares and signatures. Once a software behaves in an inappropriate way (e.g., accessing files that the software does not supposed to, creating multiple copies, logging keys), then the antiviruses generates an alert and then the software is matched with the virus database. With the new attack mechanisms, the malware analysis and reverse engineering of the softwares are highly time consuming and inefficient. Moreover, they do not always help to prevent a security breach. With the help of big data analytics tool, now it's become extremely easy to analyze high volume of software behaviors, network traffics, file-system modification. Therefore, big data analytics shows promises of a more intelligent antiviruses with more effective malware analysis [6].

3.3 Fake user detection and prevention at scale

With the growing number of services, one problem that the organizations regularly face is that the number of fake users. Often fake users create profiles in various platforms and websites. These fake users often create problems on the platforms ecology and exhibits abusive behaviors towards legitimate users. Identifying fake users often extremely difficult with the large number of legitimate users. Now, big data analytics provides various tools to analyze networks effectively which allowed the platforms to detect fake users by analyzing their behaviors. Often these fake user's creates a large networks and by clustering algorithms it has become pretty easy to isolate the group of fake users.

3.4 Spam fighting and detect Botnets at scale

Spam and Botnet are one of the major security problem for organizations. They cost useful resources and the underground economy of spam suggests that spam accounts have high benifits [11]. Everyday hundreds of users falls into phishing attack which cost the users moentary loss. The proliferation of social medias and crowd sourced systems have increased the spam distribution. However, now organizations are extremely effective on detecting spams and botnets. Various data anayltics are helping organizations to prevent spammers and protect naive users [10].

3.5 Automated security management

As mentioned earlier, security management has become extremely complex with higher volumes of traffic and data. However, big data management tools provide better security management and new data analytics and visualization tools provide automated approach of security management. Now, it's not necessary to manually investigate the behaviors and generate rules. Using the tools and machine learning, now it is possible to predict threats and automate the security and risk management.

3.6 Better surveillance and cyber safety

Since 9/11, we have seen an increasing usage of communication technology by terrorists or malicious users. However, with the variety of platforms it is has become hard to identify these malicious actors. Big data analytics tools provide a better tool for government surveillance. Although such massive surveillance compromises public privacy, still such surveillance has become effective to thwart dangerous national attacks and so far more than 50 terror plots have been thwarted [13]. Such massive surveillance have become possible due to the big data tools and analytics. They have been successful detecting anomalous behaviors and identifying the bad actors. Similar to the terrorists, these new tools are helpful for detecting social menaces like pedophiles online and keep people safe online.

4 CONCLUSION

Although big data tools and analytics has created cyber threats, it is also helping defending the threats and shown promises on successful defending in the future. According to recent stats, security breaches are declining with the help of big data analytics [2]. However, with the apparent benefits still companies have not widely adopted big data tools for security, only one in five companies are

using big data security at this moment [1]. The main reason for not adopting big data analytics is the high cost and lack of human resources. However, it is expected that the cost will be reduced and more people will be interested on big data related tools which may influence widespread use of big data analytics. With the increasing usage and better tools, threats will be more sophisticated and defensive mechanisms need to be advanced on parallel.

ACKNOWLEDGMENTS

The authors would like to thank Professor Gregor von Laszewski for helping us with the instruction and resources that was required to complete this paper. We would also to like to thank the associate instructors for being available on the course website all the time and helping us with their answers.

REFERENCES

[1] Bi-survey.com. 2016. Big Data Security Analytics: A Weapon Against Rising Cyber Security Attacks? . https://bi-survey.com/big-data-security-analytics. (2016). Online; accessed Sept 30, 2017.

[2] CSO Online. 2016. How Big Data is Improving Cyber Security. https://www.csoonline.com/article/3139923/security/how-big-data-is-improving-cyber-security.html. (2016). Online; accessed Sept 29, 2017.

[3] Hervais Simo Fhom. 2015. Big Data: Opportunities and Privacy Challenges. *CoRR* abs/1502.00823 (2015). http://arxiv.org/abs/1502.00823

[4] Y. Gahi, M. Guennoun, and H. T. Mouftah. 2016. Big Data Analytics: Security and privacy challenges. In *2016 IEEE Symposium on Computers and Communication (ISCC)*. IEEE, 501 Hoes Lane, 3rd Floor, Piscataway, NJ 08855., 952–957. https://doi.org/10.1109/ISCC.2016.7543859

[5] T. Mahmood and U. Afzal. 2013. Security Analytics: Big Data Analytics for cybersecurity: A review of trends, techniques and tools. In *2013 2nd National Conference on Information Assurance (NCIA)*. 129–134. https://doi.org/10.1109/NCIA.2013.6725337

[6] Rahul Dasgupta. 2015. Big data analytics leads the way for next-gen malware protection. http://techspective.net/2015/04/27/big-data-analytics-leads-the-way-for-next-gen-malware-protection/. (2015). Online; accessed Sept 27, 2017.

[7] A. Razaq, H. Tianfield, and P. Barrie. 2016. A Big Data Analytics Based Approach to Anomaly Detection. In *2016 IEEE/ACM 3rd International Conference on Big Data Computing Applications and Technologies (BDCAT)*. 187–193.

[8] L. Rettig, M. Khayati, P. Cudr-Mauroux, and M. Pirkowski. 2015. Online anomaly detection over Big Data streams. In *2015 IEEE International Conference on Big Data (Big Data)*. 1113–1122. https://doi.org/10.1109/BigData.2015.7363865

[9] Selena Larson.CNN. 2017. Every single Yahoo account was hacked - 3 billion in all. http://money.cnn.com/2017/10/03/technology/business/yahoo-breach-3-billion-accounts/index.html. (2017). Online; accessed Sept 27, 2017.

[10] Kamaldeep Singh, Sharath Chandra Guntuku, Abhishek Thakur, and Chittaranjan Hota. 2014. Big Data Analytics framework for Peer-to-Peer Botnet detection using Random Forests. *Information Sciences* 278, Supplement C (2014), 488 – 497. https://doi.org/10.1016/j.ins.2014.03.066

[11] Brett Stone-Gross, Thorsten Holz, Gianluca Stringhini, and Giovanni Vigna. 2011. The Underground Economy of Spam: A Botmaster's Perspective of Coordinating Large-scale Spam Campaigns. In *Proceedings of the 4th USENIX Conference on Large-scale Exploits and Emergent Threats (LEET'11)*. USENIX Association, Berkeley, CA, USA, 4–4. http://dl.acm.org/citation.cfm?id=1972441.1972447

[12] Tara Bernard and Tiffany Hsu and Nicole Perlroth and Ron Lieber. 2017. Equifax Says Cyberattack May Have Affected 143 Million in the U.S. http://money.cnn.com/2017/10/03/technology/business/yahoo-breach-3-billion-accounts/index.html. (2017). Online; accessed Sept 27, 2017.

[13] The Washington Post. 2013. NSA head: Surveillance helped thwart more than 50 terror plots. https://www.washingtonpost.com/news/post-politics/wp/2013/06/18/nsa-head-surveillance-helped-thwart-more-than-50-terror-attempts/?utm_term=.784e59848c4f. (2013). Online; accessed Sept 29, 2017.

5 BIBTEX ISSUES

Warning–page numbers missing in both pages and numpages fields in Fhom15

Warning–empty publisher in Gahi:2016

DONE:

Fixed Publisher

Warning–empty address in Gahi:2016

DONE:

Fixed Address

Warning–empty publisher in Mahmood:2013

Warning–empty address in Mahmood:2013

Warning–empty publisher in Razar:2016

Warning–empty address in Razar:2016

Warning–empty publisher in Rettig

Warning–empty address in Rettig

Warning–numpages field, but no articleno or eid field, in Stone-Gross:2011

(There were 10 warnings)

6 ISSUES

A format issue cause another one more page