International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 08, 2020
ISSN: 1475-7192

Cloud Threat Detection: Big Data Based Security Analytics Approach

[1]Krishnavandan Padhye, [2]Shreya Dwivedi, [3]S. Sharanya

ABSTRACT *— The cloud infrastructure consists of VMs (Virtual Machines) and hosting hardware which creates multiple instances of resources defined by the software. The VMs manage, support and monitor the software defined multi-instance framework. The potential of system to perform real time resource scaling has led to widespread implementation of virtualized infrastructure for cloud computing. This has caused virtual infrastructure to become a tempting* target *for cyber attackers to launch attacks and to gain illegal access. This paper analyses the threats posed by cyber criminals and the development of threat detection methods over time this gives us insight on what are the current requirements of security researchers to make the system more efficient and accurate to detect threats in real time. The paper also discusses about leveraging technologies like big data and machine learning which can handle huge amount unstructured data and are designed to run on a distributed network.*

Keywords—*Big Data, Cloud Threats, Security Analysis, Machine learning, Map Reduce, Graph Based Event Correlation, Cloud infrastructure, Random Forest*

I. INTRODUCTION

As the entire world is moving forward in the age of computing all the systems are getting more connected and interlinked to form a huge network. Every organization is taking advantage of computing services to expand their reach to remote locations. Organizations need servers and computing resources which are very costly and difficult to build and maintain.

Cloud service providers (CSP's) provide and maintain computing resources for such organizations. Organizations don't have to care about maintaining and investing they can rely on CSP's for their enterprise operations, infrastructure and network services. The exponential rise in the use of cloud services has made it an attractive target for cyber attackers.

Over the years, cyber attackers have posed many threats to cloud which includes breach of data, loss of data, DoS (Denial of Service) attacks etc. also they are devising newer methods to bypass the security systems. It has become very important for the CSP's to remain updated and secure against attacks [1]. The security system must be quick to respond and must be capable of handling the huge amount of data being generated by the network. The CSP networks are generating huge amount of semi structured or unstructured data which is very difficult to manage thus demanding the use of new age technology like big data analytics for attack detection.

[1] *Department of CSE,padhyevandan@gmail.com, +91-8827423371, SRM Institute of Science and Technology,Chennai*

[2] *Department of CSE,shreyadwivedi15@gmail.com, +91-8754510998, SRM Institute of Science and Technology,Chennai*

[3] *Department of CSE,sharanya.se@ktr.srmuniv.ac.in, +91-9003444631, SRM Institute of Science and Technology,Chennai*

Received: 21 Dec 2019 | Revised: 18 Jan 2020 | Accepted: 05 Feb 2020

Open source frameworks like Hadoop, hive etc. are used for implementing scalable threat detection systems with big data analytics which are capable of handling high network bandwidth.

Existing security approaches generally come under two categories malware threat detection and security analytics [2]. The first category of security approach relies on signature and rule based detection techniques in which a threat signature database is maintained and is updated on a regular basis, this database is used as a reference to compare network logs for identifying attack presence. These detection methods are at risk from newer attacks which are not yet been detected by security researchers, known as zero-day attacks and have new attack signatures. Big data based security analytics overcomes this limitation by removing the need for signature database. HDFS is used to store network logs on distributed network and MapReduce parser is used find potential attack paths from the attack features extracted through graph based event correlation. Machine learning can be leveraged to confirm attack presence.

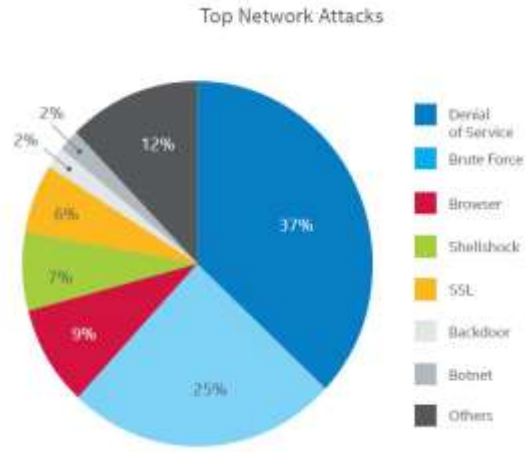

Fig. 1: Network attacks

From the above figure it is clearly visible that the attackers use a lot of different attacks against the cloud infrastructure out of these attacks DDoS attack is the most common type of attack. It is a very common type of attack and comparatively easy to perform.

The world is growing exponentially dependent on the digital services which has led to increase in the frequency of such attacks. Attackers are also providing these services to other people in exchange for money. There are many websites and tools available on the internet today which specializes in providing on demand DDoS attack such as LOIC (Low Orbit Ion Cannon), HULK (HTTP Unbearable Load King), GoldenEye etc.

II. LITERATURE SURVEY

Attack on the cloud infrastructure results in huge loss to the companies and causes inconvenience to the customers. Attackers use these attacks to kill services, take down websites, steal unauthorized data and extort money. Dan, Jeremy, Evan, Zev, and Dulani proposes a model in their paper on Cloud Trust [3] that estimates an effective security metric to measure a level of integrity and confidentiality given by cloud computing systems which can be used to identify the level of security for IaaS architectures installed with alternative cloud security

controls. The major threat to the reliability of cloud services is denial of service attacks (DoS) which cyber attackers use to take down entire network. [4]

In DoS attack the attacker tries to make a machine or network resource unavailable to the intended users by disrupting the services of host which is done by flooding the targeted host with requests in an attempt to overload systems and prevent it from processing legitimate requests. To make the attack more effective the attackers use an advanced network of distributed denial of service which uses multiple sources for flooding the host.

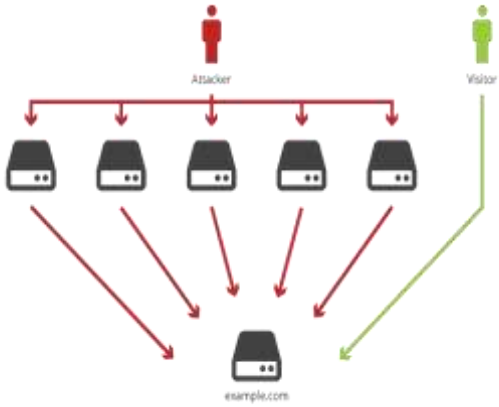

Fig. 2:DoS attack illustration

To tackle DDoS attacks first step is identifying what is actually an attack which is discussed in the paper that tries to differentiate between an actual attack and congestion of network due to actual requests by Shevtekar, Amey and Ansari, Nirvan.[5] The methods proposed in the paper are based on adaptive management of queue and aggregate congestion control. The author demonstrates the inability of representative defense mechanism to detect the attacks and introduce a newer attack model called quiet model. The method was able to reduce the ftp flow by almost 70%. The cyber attackers are evolving with time and using newer attacks like the flash crowd attacks.

Flash crowd attacks are discussed in the paper by Shui, Wanlei Zhou, Yong Xiang, and Tang Feilong.[6] The authors discussed the about detection of botnet DDoS attacks that try to mimic patterns of a flash crowd to disguise themselves as the real network traffic.[7] The proposed flow correlation method highly relies on the size of botnet organizations being known to the CSP which is very difficult to predict reliably as the organizations are evolving continuously. Botnet size determination[8] is a very tough task for the security providers they use honey pots to trick cyber attackers to attack them so that they can determine their attack signature and size of attack which is studied by Bill McCarty in the paper Botnets: Big and Better.[9]

The main focus of security researchers is to make the system detect threats and that too in real time. To achieve this goal the authors in their paper about security analytics approach to protect cloud infrastructure [10] have used the open source frameworks for big data such as Hadoop and hive[11] to create a scalable and real time threat detection system. The approach proposed by the author is to use the graph based event correlation [12] to perform attack feature extraction, then use map reduce parser to find the paths that might represent the potential attacks then to confirm the presence of attack they are using logistic regression in step one and then applying belief propagation in step two. [13]

Received: 21 Dec 2019 | Revised: 18 Jan 2020 | Accepted: 05 Feb 2020

Use of machine learning to confirm attack presence allows the system to become more accurate over time to detect attacks. Cyber attackers are using more devices to increase the scale of their attacks. [14] Everything in the current world is getting digitized and connected to internet from a coffee machine to cars but manufacturers are not very focused on providing proper security by updating security patches and firewall. This leads to a serious concern of unauthorized access and cyber attackers are using this opportunity to exploit the consumer IoT devices to act as an attack source which provides them with a huge army of attacking sources which are distributed all over the globe.[15] In last one year the average DDoS attack volume has grew by 194%. [16] In Q4 of 2017 the average attack volume was 1.7 Gbps whereas in Q4 of 2018 the average attack volume was 5 Gbps. [17] the attacks are getting more complex and are using multiple vectors at the same time. The most complex attack in Q4 of 2018 used nine different attack vectors.

Detection of such attacks is another problem which can be solved with the help of machine learning as discussed in the paper by Dashi Rohan, Apthrope Noah and Feamster Nick about leveraging machine learning to prevent the use of consumer IoT devices in DDoS attacks. It helps in differentiating between attacking IoT device and benign IoT device.

III. EXPERIMENTAL APPROACH

Idea that is being proposed here is to detect threats to the network infrastructure posed by any malware of network attack in real time. It leverages the big data solutions to deal with the huge volume of network and VM logs. It also helps in utilizing distributed computing which increases the efficiency of system. The first part of the system deals with the extraction of attack features through graph-based event correlation. The parser focuses on the features like source IP, source port, destination IP, destination port, how long the connection lasted, protocol used etc. All this data is taken into consideration in the form of a graph. From this graph, map-reduce parser finds the potential attack paths out of all paths.

Once a path is detected as a potential attack path then it is passed to the second part of the system that uses probabilistic machine learning algorithm to confirm the presence of attack. It enables the system to detect zero-day attacks and over time accuracy of the system increases. It is a novel approach for protecting cloud infrastructure with the use of big data and machine learning.

IV. ANALYSIS AND DISCUSSION

Security researchers are developing newer and advance methods of protecting cloud infrastructure. In table 1 we have discussed some of the currently available security techniques which are used to detect and mitigate attacks. [18]

Table 1: Security Approaches

Approach	**Characteristic**	**Strength**	**Limitation**

Signature based malware detection	Regularly updated signature table	Identify attacks accurately	Cannot detect new unknown attacks
Bot Cloud	Use of page rank algorithm	Identifying botnet	Limited security
Large Scale Botnet Detection	Clustering based correlation	Detect huge well known botnet attacks	Vulnerable to zero day attacks
BDSA	Event correlation and logistic regression	Real time threat detection	Occasional latency issue

The dependence of entire world on computing is increasing every day and that has made cloud an attractive target for cyber attackers. Protection of these services in real time [19] is the key focus of security researchers. Technologies like big data and machine learning can help us to make our security systems more efficient and accurate at detecting threats. The current analytics methods can be replaced with newer methods that make the security system more scalable and easily deployable to meet the exponential expansion rate of cloud and can detect threats in real time.

V. CONCLUSION

In this process we have analyzed the previous and current security threat detection techniques, their method of working and limitations and how the researchers are devising newer methods to combat the cyber-attacks. The result has showed that the malware detection techniques that use rule based and signature based detection system are vulnerable to zero day attack [20] and the security analytics can be made more scalable to make it easily deployable over large CSP networks. [21] Big data based security analytics can help CSP's to handle the huge amount of network data being generated and to detect threats in real time.

REFERENCES

1. W. Jansen and T. Grance, "Guidelines on security and privacy in public cloud computing," NIST Spec. Publ. 800–144, National Institute of Standards and Technology, Gaithersburg, MD 20899, Dec. 2011.
2. T. Mahmood and U. Afzal, "Security analytics: Big data analytics for cybersecurity: A review of trends, techniques and tools," in Proc. 2nd Nat. Conf. Inf. Assurance, 2013, pp. 129–134.

Received: 21 Dec 2019 | Revised: 18 Jan 2020 | Accepted: 05 Feb 2020

3. Dan Gonzales, Jeremy M. Kaplan, Evan Saltzman, Zev Winkelman and Dulani Woods, "Cloud Trust – a Security Assessment Model for Infrastructure as a Service (IaaS) Clouds", IEEE Transactions on Cloud Computing (Volume: 5 , Issue: 3 , July-Sept. 1 2017)
4. G. Somani, M.S. Gaur, D. Sanghi, and M. Conti, "DDoS Attacks in Cloud Computing: Collateral Damage to Non-targets," Computer Networks, vol. 109, no. 2, 2016, pp. 157–171.
5. Amey Shevtekar, and Nirvan Ansari, "Is it congestion or a DDoS attack?", IEEE Communications Letters (Volume: 13 , Issue: 7 , July 2009)
6. Shui Yu, Wanlei Zhou, Weijia Jia, Song Guo, Yong Xiang and Feilong Tang, "Discriminating DDoS Attacks from Flash Crowds Using Flow Correlation Coefficient", IEEE Transactions on Parallel and Distributed Systems (Volume: 23 , Issue: 6 , June 2012)
7. S. Yu et al., "Discriminating DDoS attacks from flash crowds using flow correlation coefficient," IEEETrans. Parallel Distrib. Syst., vol. 23, no. 6, pp. 1073–1080, Jun. 2012.
8. Nazrul Hoque, Dhruba K. Bhattacharyya, and Jugal K. Kalita, "Botnet in DDoS Attacks: Trends and Challenges", IEEE Communication Surveys & Tutorials, Vol. 17, No. 4, Fourth Quarter 2015
9. Bill McCarty, "Botnets: Big and Bigger", IEEE Security & Privacy, vol. 1, no. , pp. 87-90, 2003. doi:10.1109/MSECP.2003.1219079
10. Thu Yein Win and Quentin Mair, "Big Data Based Security Analytics Approach For Protecting Virtualized Infrastructure in Cloud Computing", IEEE Transactions On Big Data, Vol. 4, No. 1, January-March 2018
11. K. Singh, S. C. Guntuku, A. Thakur, and C. Hota, "Big data analytics framework for peer-to-peer botnet detection using random forests," Inf. Sci., vol. 278, pp. 488–497, 2014.
12. Tao Qin , Yuli Gao, Lingyan Wei, Zhaoli Liu and Chenxu Wang, "Potential threats mining methods based on correlation analysis of multi-type logs", IET Networks (Volume: 7 , Issue: 5 , 9 2018)
13. Rohan Doshi, Noah Apthorpe and Nick Feamster, "Machine Learning DDoS Detection for Consumer Internet of Things Devices", 2018 IEEE Symposium on Security and Privacy Workshops
14. Natalija Vlajic and Daiwei Zhou, "IoT as a Land of Opportunity for DDoS Hackers", IEEE Computer Society, Computer (Volume: 51 , Issue: 7 , July 2018)
15. Rémi Cogranne, Guillaume Doyen, Nisrine Ghadban, and Badis Hammi, "Detecting Botclouds at Large Scale: A Decentralized and Robust Detection Method for Multi-Tenant Virtualized Environments", IEEE Transactions on Network and Service Management (Volume: 15 , Issue: 1 , March 2018)
16. Andrew Ross, "Cybercrime on the rise: DDoS attack volumes have trebled in past year, says study", https://www.information-age.com/cyber-crime-on-the-rise-ddos-attack-123478977/
17. DDoS Report, Link 11, https://www.link11.com/en/ddos-report/
18. Muyowa Mutemwa and Francois Mouton, "Cyber security threats and mitigation techniques for multifunctional devices", 2018 Conference on Information Communications Technology and Society (ICTAS)
19. Vasco Samuel Carvalho, Maria João Polidoro and João Paulo Magalhães, "OwlSight: Platform for Real-Time Detection and Visualization of Cyber Threats", 2016 IEEE 2nd International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS)

20. Ashwini Mujumdar, Gayatri Masiwal and Dr. B. B. Meshram, “Analysis of Signature-Based and Behavior-Based Anti-Malware Approaches”, International Journal of Advanced Research in Computer Engineering and Technology (IJARCET) Volume 2, Issue 6, June 2013
21. A. Lioy, G. Gardikis, B. Gaston, L. Jacquin, M. De Benedictis, Y. Angelopoulos and C. Xylouris, “NFV-based network protection: the SHIELD approach”, 2017 IEEE Conference on Network Function Virtualisation and Software Defined Networks (NFV-SDN)

Received: 21 Dec 2019 | Revised: 18 Jan 2020 | Accepted: 05 Feb 2020