ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 3, March 2018

Big Data Based Security Analytics for Protecting Virtualized Infrastructures in Cloud Computing

[1] J.Yamuna
BEE (PG scholar), Computer Science and Engineering
Francis Xavier Engineering College, Tirunelveli.

Abstract: **- In this Paper Virtualized infrastructure in cloud computing has become an attractive target for cyber attackers to launch advanced attacks. This paper proposes a novel big data-based security analytics approach to detecting advanced attacks on virtualized infrastructures. Network logs, as well as user application logs collected periodically from the guest virtual machines (VMs), are stored in the Hadoop Distributed File System (HDFS). Then, extraction of attack features is performed through graph-based event correlation and Map Reduce parser based identification of potential attack paths. Next, determination of attack presence is performed through two-step machine learning, namely, logistic regression is applied to calculate attack's conditional probabilities with respect to the attributes, and belief propagation is applied to calculate the belief in the existence of an attack based on them. Experiments are conducted to evaluate the proposed approach using well-known malware as well as in comparison with existing security techniques for virtualized infrastructure.**

I. INTRODUCTION

Imagine a world without data storage; a place where every detail about a person or organization, every transaction performed, or every aspect which can be documented is lost directly after use. Organizations would thus lose the ability to extract valuable information and knowledge, perform detailed analyses, as well as provide new opportunities and advantages. Anything ranging from customer names and addresses, to products available, to purchases made, to employees hired, etc. has become essential for day-to-day continuity. Data is the building block upon which any organization thrives. Now think of the extent of details and the surge of data and information provided nowadays through the advancements in technologies and the internet. With the increase in storage capabilities and methods of data collection, huge amounts of data have become easily available. Every second, more and more data is being created and needs to be stored and analyzed in order to extract value. Furthermore, data has become cheaper to store, so organizations need to get as much value as possible from the huge amounts of stored data. The size, variety, and rapid change of such data require a new type of big data analytics, as well as different storage and analysis methods. Such sheer amounts of big data need to be properly analyzed, and pertaining information should be extracted. The contribution of this paper is to provide an analysis of the available literature on big data analytics. Accordingly, some of the various big data tools, methods, and technologies which can be applied are discussed, and their applications and opportunities provided in several decision domains are portrayed.

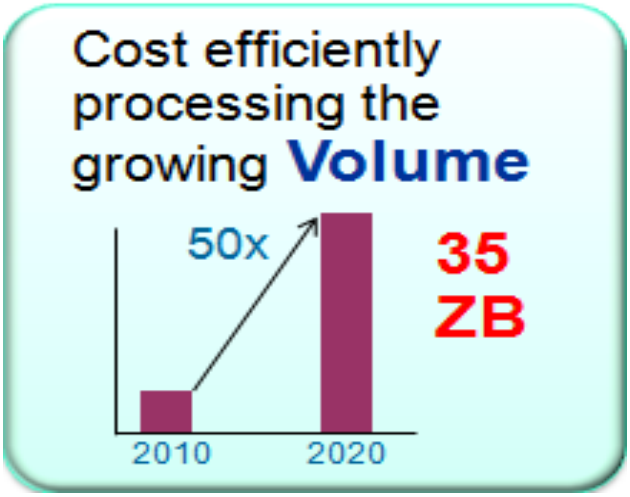
BIG DATA ANALYTIC

The term “Big Data” has recently been applied to datasets that grow so large that they become awkward to work with using traditional database management systems. They are data sets whose size is beyond the ability of commonly used software tools and storage systems to capture, store, manage, as well as process the data within a tolerable elapsed time [12]. Big data sizes are constantly increasing, currently ranging from a few dozen terabytes (TB) to many petabytes (PB) of data in a single data set. Consequently, some of the difficulties related to big data include capture, storage, search, sharing, analytics, and visualizing. Today, enterprises are exploring large volumes of highly detailed data so as to discover facts they didn’t know before [17]. Hence, big data analytics is where advanced analytic techniques are applied on big data sets. Analytics based on large data samples reveals and leverages business change. However, the larger the set of data, the more difficult it becomes to manage [17]. In this section, we will start by discussing the characteristics of big data, as well as its importance. Naturally, business benefit can commonly be derived from analyzing larger and more complex data sets that require real time or near-real time capabilities; however, this leads to a need for new data architectures, analytical methods, and tools. Therefore the successive section will elaborate the big data analytics tools and methods, in particular, starting with the big data storage and management, then moving on to the big data analytic processing. It then concludes with some of the various big data analyses which have grown in usage with big data

VOLUME

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 3, March 2018



The size of available data has been growing at an increasing rate. The volume of data is growing. Experts predict that the volume of data in the world will grow to 25 Zettabytes in 2020. That same phenomenon affects every business – their data is growing at the same exponential rate too. This applies to companies and to individuals. A text file is a few kilo bytes, a sound file is a few mega bytes while a full length movie is a few giga bytes. More sources of data are added on continuous basis. For companies, in the old days, all data was generated internally by employees. Currently, the data is generated by employees, partners and customers. For a group of companies, the data is also generated by machines. For example, Hundreds of millions of smart phones send a variety of information to the network infrastructure. This data did not existfiveyearsago. More sources of data with a larger size of data combine to increase the volume of data that has to be analyzed. This is a major issue for those looking to put that data to use instead of letting it just disappear. Peta byte data sets are common these days and Exa byte is not far away.

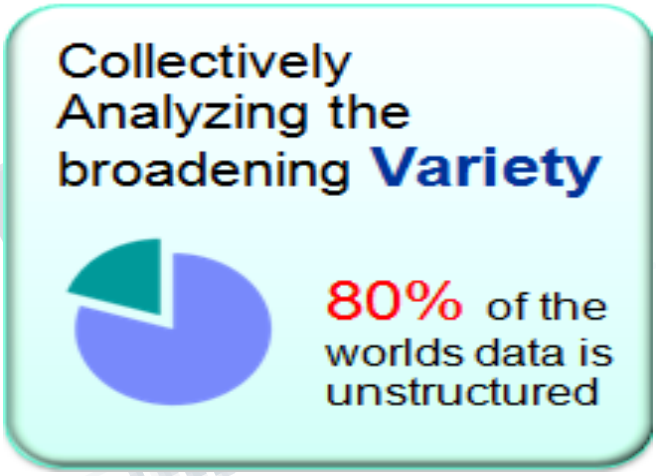
VELOCITY

Data is increasingly accelerating the velocity at which it is created and at which it is integrated. We have moved from batch to a real-time business. Initially, companies analyzed data using a batch process. One takes a chunk of data, submits a job to the server and waits for delivery of the result. That scheme works when the incoming data rate is slower than the batch-processing rate and when the result is useful despite the delay. With the new sources of data such as social and mobile applications, the batch process breaks down. The data is now streaming into the server in real time, in a continuous fashion and the result is only useful if the delay is very short.

Data comes at you at a record or a byte level, not always in bulk. And the demands of the business have increased as well – from an answer next week to an answer in a minute. In addition, the world is becoming more instrumented and interconnected. The volume of data streaming off those instruments is exponentially larger than it was even 2 years ago.

VARIETY

Variety presents an equally difficult challenge. The growth in data sources has fuelled the growth in data types. In fact, 80% of the world's data is unstructured. Yet most traditional methods apply analytics only to structured information.

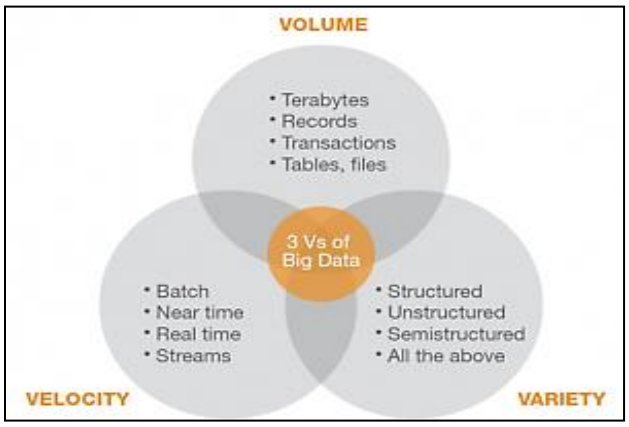


From excel tables and databases, data structure has changed to loose its structure and to add hundreds of formats. Pure text, photo, audio, video, web, GPS data, sensor data, relational data bases, documents, SMS, pdf, flash, etc. One no longer has control over the input data format. Structure can no longer be imposed like in the past in order to keep control over the analysis. As new applications are introduced new data formats come to life. The variety of data sources continues to increase. It includes

- Internet data (i.e., click stream, social media, social networking links)
- Primary research (i.e., surveys, experiments, observations)
- Secondary research (i.e., competitive and marketplace data, industry reports, consumer data, business data)
- Location data (i.e., mobile device data, geospatial data)
- Image data (i.e., video, satellite image, surveillance)
- Supply chain data (i.e., EDI, vendor catalogs and pricing, quality information)
- Device data (i.e., sensors, PLCs, RF devices, LIMs, telemetry)

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 3, March 2018



Big Data Analytic Application:

Big Data Analytics Tools and Methods

With the evolution of technology and the increased multitudes of data flowing in and out of organizations daily, there has become a need for faster and more efficient ways of analyzing such data. Having piles of data on hand is no longer enough to make efficient decisions at the right time. Such data sets can no longer be easily analyzed with traditional data management and analysis techniques and infrastructures. Therefore, there arises a need for new tools and methods specialized for big data analytics, as well as the required architectures for storing and managing such data. Accordingly, the emergence of big data has an effect on everything from the data itself and its collection, to the processing, to the final extracted decisions. Consequently, [8] proposed the Big – Data, Analytics, and Decisions (B-DAD) framework which incorporates the big data analytics tools and methods into the decision making process [8]. The framework maps the different big data storage, management, and processing tools, analytics tools and methods, and visualization and evaluation tools to the different phases of the decision making process. Hence, the changes associated with big data analytics are reflected in three main areas: big data storage and architecture, data and analytics processing, and, finally, the big data analyses which can be applied for knowledge discovery and informed decision making. Each area will be further discussed in this section. However, since big data is still evolving as an important field of research, and new findings and tools are constantly developing, this section is not exhaustive of all the possibilities, and focuses on providing a general idea, rather than a list of all potential opportunities and technologies

Customer Intelligence

Big data analytics holds much potential for customer intelligence, and can highly benefit industries such as retail, banking, and telecommunications. Big data can create transparency, and make relevant data more easily accessible to stakeholders in a timely manner [14]. Big data analytics can provide organizations with the ability to profile and segment customers based on different socioeconomic characteristics, as well as increase levels of customer satisfaction and retention [4]. This can allow them to make more informed marketing decisions, and market to different segments based on their preferences along with the recognition of sales and marketing opportunities.

II. CONCLUSION

In this research, we have examined the innovative topic of big data, which has recently gained lots of interest due to its perceived unprecedented opportunities and benefits. In the information era we are currently living in, voluminous varieties of high velocity data are being produced daily, and within them lay intrinsic details and patterns of hidden knowledge which should be extracted and utilized. Hence, big data analytics can be applied to leverage business change and enhance decision making, by applying advanced analytic techniques on big data, and revealing hidden insights and valuable knowledge. Accordingly, the literature was reviewed in order to provide an analysis of the big data analytics concepts which are being researched, as well as their importance to decision making. Consequently, big data was discussed, as well as its characteristics and importance. Moreover, some of the big data analytics tools and methods in particular were examined. Thus, big data storage and management, as well as big data analytics processing were detailed. In addition, some of the different advanced data analytics techniques

REFERENCES

1. Adams, M.N.: Perspectives on Data Mining. International Journal of Market Research 52(1), 11–19 (2010)

2. Asur, S., Huberman, B.A.: Predicting the Future with Social Media. In: ACM International Conference on Web Intelligence and Intelligent Agent Technology, vol. 1, pp. 492–499 (2010)

3. Bakshi, K.: Considerations for Big Data: Architecture and Approaches. In: Proceedings of the IEEE Aerospace Conference, pp. 1–7 (2012)

4. Cebr: Data equity, Unlocking the value of big data. in: SAS Reports, pp. 1–44 (2012)

5. Cohen, J., Dolan, B., Dunlap, M., Hellerstein, J.M., Welton, C.: MAD Skills: New Analysis Practices for Big Data. Proceedings of the ACM VLDB Endowment 2(2), 1481–1492(2009)