ISSN 2394-739X

International Journal for Research in Science Engineering and Technology

BIG DATA SECURITY CHALLENGES

[1] **K. Jayabharathi,**

[1] Assistant. professor.

[1] Dept of Computer Science,

[1] Guru Nanak College,

[1] Chennai.

Abstract:-

Big Data consists of huge modules, difficult, growing data sets with numerous and , independent sources. With the fast development of networking, storage of data, and the data gathering capacity, Big Data are now quickly increasing in all science and engineering domains, as well as animal, genetic and biomedical sciences. This paper elaborates a HACE theorem that states the characteristics of the Big Data revolution, and proposes a Big Data processing model from the data mining view. This data-oriented model contains demand-driven aggregation of data sources, mining and study, user knowledge modeling, and security and privacy issues. We examine the difficult issues in the data-oriented model and also in the Big Data revolution.

Keywords: - Big Data, Data Sources, Big Data Revolution

1. INTRODUCTION

Securing big data comes with its own unique challenges beyond being a high-value target. It’s not that big data security is fundamentally different from traditional data security. Big data security challenges arise because of incremental differences, not fundamental ones. The differences between big data environments and traditional data environments include:

- The data collected, aggregated, and analyzed for big data analysis.
- The infrastructure used to store and house big data.
- The technologies applied to analyze structured and unstructured big data.

Use of Big Data to Manage Security Threats Because of the scale of the Internet and the fact that the world’s population is steadily coming online, protecting users from cybercrime can be viewed as a numbers game. The same forces that are driving big data are driving threats concurrently. New methods of addressing cyber threats are needed to process the enormous amount of data emerging from the world and to stay ahead of a sophisticated, aggressive, and ever-evolving threat landscape. No off-the-shelf solution can address a problem of this magnitude. The traditional rules of engagement no longer apply. Scaling up to manage the changes in the threat landscape is necessary, but it must be done intelligently.

Best Practices in Achieving End Users Results

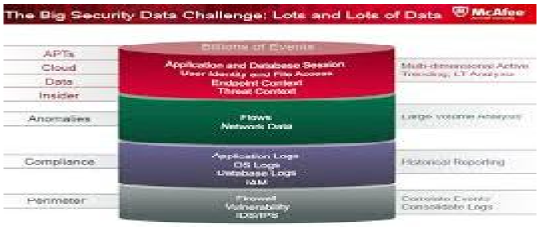
Best Practices in Achieving End User Results Addressing today’s threat landscape requires a synergistic relationship with customers and other third parties that are constantly exposed to ever-evolving malicious content. A licensing agreement that allows customers to anonymously donate suspicious data for analysis and reverse engineering can provide valuable access to real data on real machines operating in the real world. Based on data gathered from this community network, specialized search algorithms, machine learning, and analytics can then be brought to bear on this data to identify abnormal patterns that can signal a threat.

For example, many computer users follow a typical daily pattern. That pattern may consist of visiting a news site, encountering several ad servers, and logging on to Facebook. If that pattern suddenly changes, perhaps moving the user to a domain never previously visited, this incident can be immediately prioritized for further analysis. These types of complex correlations can be identified only by a system that can perform a very large number of database searches per second.

A feedback loop for process improvement is another critical component. Keen observation and curation of key data that is fed back into the process allows for continual process improvement. Over time, the process can predict malicious behavior long before it occurs.

While big data in security is a numbers game, human experts need to play the most important role. Trained analysts need to constantly evolve the combination of methodologies, apply human intuition to complex problems, and identify trends that computers miss. Using the right approach when an attack slips through the cracks is also crucial.

A savvy security software company works directly with the ISP involved in an attack to drive a better end result. This often involves working closely with law enforcement agencies. Ultimately, relationships are formed with ISPs that drive a symbiotic relationship with a common threatprotection goal.The end user result is safer Internet.



2. BIG DATA

Big Data is a comprehensive term for any collection of data sets so large and multifarious that it becomes difficult to process them using conventional data processing applications. The challenges include analysis, capture, search, sharing, storage, transfer, revelation, and privacy violations. The tendency to larger data sets is due to the additional information derivable from analysis of a single large set of related data, as compared to separate smaller sets with the same total amount of data, allowing correlations to be found to "spot business trends, prevent diseases, combat crime and so on. There are two types of Big Data: structured and unstructured.

Structured data are numbers and words that can be easily categorized and analyzed. These data are generated by things like network sensors embedded in electronic devices, smart phones, and global positioning system (GPS) devices. Structured data also include things like sales figures, account balances, and transaction data.

- **Unstructured data**include more multifarious information, such as customer reviews from feasible websites, photos and other multimedia, and comments on social networking sites. These data can not be separated into categorized or analyzed numerically.

3. THE MASSIVE SCOPE OF BIG DATA SECURITY

To establish comprehensive big data security, executives and administrators have to address the following areas:

3.1 Data sources

To most fully exploit the advantages of big data, organizations leverage various forms of data, including both structured data in a range of heterogeneous applications and databases and unstructured data that comes in a number of file types. Organizations may leverage data from enterprise resource planning systems, customer relationship management platforms, video files, spreadsheets, social media feeds, and many other sources. Further, more data sources are added all the time. Today, you don't know where new data sources may come from tomorrow, but you can have some certainty that there will be more to contend with and more diversity to accommodate. These big data sources can include personally identifiable information, payment card data, intellectual property, health records, and much more. Consequently, the data sources being compiled need to be secured in order to address security policies and compliance mandates.

4. SAFEGUARDING BIG DATA ANALYTICS

Big data output comes in many forms, including on-demand dashboards, automated reports, and ad hoc queries. Very often, these outputs contain intellectual property that is very valuable to an organization—and a potential target of attack. To provide big data analytics security for these confidential assets, security teams can use the following solutions:

- **Vormetric Transparent Encryption.** This encryption product can easily be deployed on servers, where it can encrypt big data outputs and control and monitor who accesses them.

- **Vormetric Application Encryption.** You can use this encryption product to secure specific fields that may be created in analytics applications.

- **Big data frameworks.** Within the big data environment itself—whether it's powered by Hadoop, MongoDB, NoSQL, Teradata, or another system—massive amounts of sensitive data may be managed at any given time. Sensitive assets don't just reside on big data nodes, but they can come in the form of system logs, configuration files, error logs, and more.

- **Analytics.** The ultimate fruit of a big data initiative is the output, the analytics that help the business optimize and innovate. This information can be presented in dashboards and reports, and accessed via on-demand queries. In some businesses, big data analytics represent the most sensitive asset of all, intelligence that provides a critical competitive differentiator—and a huge competitive exposure if it falls into the wrong .

5. THE TECHNOLOGY

An additional big data security challenge is that big data programming tools, including Hadoop and NoSQL databases, were not originally designed with security in mind.

For example, Hadoop originally didn't authenticate services or users, and didn't encrypt data that's transmitted between nodes in the environment. This creates vulnerabilities for authentication and network security. NoSQL databases lack some of the security features provided by traditional databases, such as role-based access control.

The advantage of NoSQL is that it allows for the flexibility to include new data types on the fly, but defining security policies for this new data is not straightforward with these technologies.

CONCLUSION

Trend Micro blocks 200 million threats per day within their network of customers. Effectively managing and prioritizing the volume, variety, and velocity of data requires human insight, a multi-pronged approach, and multiple layers of defense. Using big data tools to analyze the massive amount of threat data received daily, and correlating the different components of an attack, allows a security vendor to continuously update their global threat intelligence and equates to improved threat knowledge and insight. Customers benefit through improved, faster, and broader threat protection.

By reducing risk, they avoid potential recovery costs, adverse brand impacts, and legal implications. Smarter Protection Through Global Intelligence The Trend Micro™ Smart Protection Network™ cloud security infrastructure rapidly and accurately identifies new threats, delivering global threat intelligence to all our products and services. Ongoing advances in the depth and breadth of the Smart Protection Network allow Trend Micro to monitor more extensively for threat data, and respond to new threats more effectively, to secure data wherever it resides.

Watch for future white papers to discuss more specific sets of best practices that are incorporated into Trend Micro's approaches and its Smart Protection Network infrastructure.

REFERENCES

[1] Bilge, L. & T. Dumitras. (2012, October) Before We Knew It: An empirical study of zero-day attacks in the real world. Paper presented at the ACM Conference on Computer and Communications Security (CCS), Raleigh, NC.

[2] Bryant, R., R. Katz & E. Lazowska. (2008). Big-Data Computing: Creating revolutionary breakthroughs in commerce, science and society. Washington, DC: Computing Community Consortium.