

Data Security on Big Data

Research Scholar : SARAFUDHEEN. M. T,

PhD in Computer Science (Pursuing) Bharathiar University

Abstract

When we talk about Big data we immediately realize about any important and secure data availability. Big Data is an immensely popular talking point, From a security perspective, there are two distinct issues: securing the organisation and its customers' information in a Big Data context; and using Big Data techniques to analyse, and even predict, security incidents.

One of the first data sets we were given to analyse was a 3TB data set from a customer. It was every packet in and out of their 100Mbps internet connection for 6 weeks. It contained approximately 500,000 attacks. Making sense of this volume of information is incredibly difficult with current tooling. Even Network Security Monitoring (NSM) tools have difficult with this size of data. However it's not just size and scale. No existing toolset allowed you to provide the same level of context. Packetpig allows you to join together information related to threats, sessions, protocols (deep packet inspection) and files as well as Geolocation and Operating system detection information

Introduction

Interruption identification is the examination of system activity to recognize interlopers on your system. Most interruption recognition frameworks (IDS) search for marks of known assaults and recognize them progressively. Packetpig is distinctive. Packetpig dissects full bundle catches – that is, logs of each and every parcel sent over your system – sometime later. As opposed to existing IDS frameworks, this implies utilizing Hadoop on full bundle catches, Packetpig can identify 'zero day' or obscure adventures on authentic information as new endeavors are found. Which is to say that Packetpig can figure out if interlopers are as of now in your system, for to what extent, and what they've stolen or manhandled.

Packetpig is a Network Security Monitoring (NSM) Toolset where the 'Huge Data' is full bundle catches. Like a Tivo for your system,

through its mix with Snort, p0f and custom java loaders, Packetpig does profound bundle assessment, document extraction, highlight extraction, working framework recognition, and other profound system examination. Packetpig's examination of full parcel catches concentrates on giving however much setting as could be expected to the expert. Setting they have never had. This is a 'Major Data' opportunity.

Security

In the world of computer essentiality, many businesses already use Big Data for marketing and research, yet may not have the fundamentals right – particularly from a security perspective. As with all new technologies, security seems to be an afterthought at best.

The stealing and using of another's data is not new for us. Big Data breaches will be big too, with the potential for even more serious reputational damage and legal repercussions than at present. A growing number of companies are using the technology to store and analyze petabytes of data including web logs, click stream data and social media content to gain better insights about their customers and their business.

In the modern world of computer era, most of the organizations already struggle with implementing these concepts, making this a significant challenge. We will need to identify owners for the outputs of Big Data processes, as well as the raw data. Thus data ownership will be distinct from information ownership – perhaps with IT owning the raw data and business units taking responsibility for the outputs. With the outlook of securing data a very few organizations are likely to build a Big Data environment in-house, so cloud and Big Data will be inextricably linked. As many businesses are aware, storing data in the cloud does not remove their responsibility for protecting it - from both a regulatory and a commercial perspective. Techniques such as attribute based encryption may be necessary to protect sensitive data and apply access controls (being attributes of the data itself, rather than the environment in which it is stored). Many of these concepts are foreign to businesses today.

Set up Big Data for Security

The deployment of Big Data for fraud detection, and in place of security incident and event management (SIEM) systems, is attractive to many organizations. The overheads of managing the output of traditional SIEM and logging systems are proving too much for most IT departments and Big Data is seen as a potential savior. There are commercial replacements available for existing log management systems, or the technology can be deployed to provide a single data store for security event management and enrichment.

Taking the idea a step further, the challenge of detecting and preventing advanced persistent threats may be answered by using Big Data style analysis. These techniques could play a key role in helping detect threats at an early stage, using more sophisticated pattern analysis, and combining and analysing multiple data sources. There is also the potential for anomaly identification using feature extraction.Today logs are often ignored unless an incident occurs. Big Data provides the opportunity to consolidate and analyse logs automatically from multiple sources rather than in isolation. We know that organisational silos often reduce the effectiveness of security systems, so businesses must be aware that the potential effectiveness of Big Data style analysis can also be diluted unless these issues are addressed. At the very least, Big Data could result in far more practical and successful SIEM, IDS and IPS implementations. As a result, information classification becomes even more critical; and information ownership must be addressed to facilitate any reasonable classification.

Professional Outlook

With the help of technology we may deploy many authorities and arrange set up to save and secure our data. In originality, Big Data is more about the processing techniques and outputs than the size of the data set itself, so specific skills are required to use Big Data effectively. There is a general shortage of specialist skills for Big Data analysis, in particular when it comes to using some of the less mature technologies.

Suppliers

suchas :

Cloudera, MapR, Hortonworks and IBM offer training courses in Hadoop, offering organisations the opportunity to build their in-house skills to address Big Data challenges. Before leaping into this brave new world, companies must be clear about what they are actually trying to achieve, otherwise their investment will be wasted. Otherwise to steal any data through any common system or in person will be so much easy and the shuffling and using illegally may be extended in computer field.

Conclusion:

However the security term is not so peculiar but its need in the world of computer is imperial. Big Data expands the boundaries of existing information security responsibilities and introduces significant new risks and challenges. In the current scenario the security of data has become the most essential object for any organization and in this continuation many software security systems are being invented.

References

[1] A. Rowstron et al. Nobody ever got fired for using hadoop on a cluster. In HotCDP, 2012.

[2]. E. Chickowski, "A Case Study in Security Big Data Analysis," Dark Reading, 9 Mar. 2012.

[3] E. Ryvkina et al. Revision processing in a stream processing engine: A high-level design. In ICDE, 2006.

[4]. J. François et al., "BotCloud: Detecting Botnets Using MapReduce," Proc. Workshop Information Forensics and Security, IEEE, 2011, pp. 1–**6**.

[5]. N. Pansare, V. R. Borkar, C. Jermaine, and T. Condie. Online aggregation for large mapreduce jobs. PVLDB, 2011.

[6] P. Upadhyaya, Y. Kwon, and M. Balazinska. A latency andfault-tolerance optimizer for online parallel query plans. InSIGMOD, 2011.

[7]. P. Giura and W. Wang, "Using Large Scale Distributed Computing to Unveil Advanced Persistent Threats," Science J., vol. 1, no. 3, 2012, pp. 93–105.

[8]. T. Dumitras and D. Shou, "Toward a Standard Benchmark for Computer Security Research: The Worldwide Intelligence Network Environment (WINE)," Proc. EuroSys BADGERS Workshop, ACM, 2011, pp. 89–96.

[9]. T.-F. Yen et al., "Beehive: Large-Scale Log Analysis for Detecting Suspicious Activity in Enterprise Networks," to be published in Proc. Ann. Computer Security Applications Conference (ACSAC 13), ACM, Dec. 2013.

[10] T. White. Hadoop: The Definitive Guide. 2009.

*****