ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 2, February 2016**

# Harnessing Big Data Analytics for Improving Information Security Practices

R. Kumar

Department of Computer Science, Jamia Millia Islamia (Central University), New Delhi, India

**ABSTRACT :** Information Security incidents are so normal today that ventures and governments are confronting the most ruinous risk environment presently.In nature these security threats are very real, and the stakes are higher than ever on today's informationcentric and interconnected networked world of individuals and organizations. And with the time and recent technologies these attacks are becoming more dynamic, more malicious and more complex.However, organizations must first need to understand the sources of these security threats, along with to know exactly what they're up against, including the origins, variations and methods of attack effectively. But the main problem is that most IT departments are unattended of the most highly complex threat data current available, and to place it into a meaningful context. With today's large variety of incoming attacks, it can be extremely difficult to detect and analyze everchanging threats, much less to turn collected data into insights that consistently identify the most dangerous threats and then take action on those insights intelligently.Big Data analytics is continuously perceiving as a "transformative" innovative technology for security incidents analysis prompting to find threats, attacks in the regularly expanding volume of incident information produced from the system of in-house, versatile, cloud-based and social administrations.

**KEYWORDS***:* information security analytics, information security practices, big data information security analytics, Strategic intelligence, information security, information technology, information assurance, threatmanagement

## I. INTRODUCTION

Security incidents are so common today that enterprises and governments are facing the most destructive threat environment never evident before in the history of information science.In 2012, 93% of large organizations surveyed had suffered at least one security breach in the recent year according to research undertaken by PricewaterhouseCoopers and Infosecurity[1]. Information's storage, use, or communication is the basis of the modern organizationconducting their businesses as almost everything the organisation is and does involves information processing. *Information security* is a broad term that essentially refers to the practice of protecting information and the ways in which it is used to serve the goals of an organisation [2]. The importance of information security is widely acknowledged as the information significance for the proper and effective functions of organisations[3] [4].

Disruptive computing trends greatly increase productivity and business agilityhowever conjointly herald a number of new risks and uncertainty. In addition the increasing willingness to share more and more data through mobility and social networks giving birth to new technologies which are capable to capture more data about data. Such a massive amount of structured or unstructured data and the growing business around it have at least one assured outcome, the need for effective information security practices.

Moreover, some statistics from the digital universe is as follows:

- The proportion of information within the digital universe that needs protection is growing quicker than the digital universe itself, from but a 3rd in 2010 to over four-hundredth in 2020
- Only regarding 50% the data that needs protection has protection. That might improve slightly by 2020, as a number of the better-secured data classes can grow quicker than the digital universe itself, however it still means the quantity of unprotected knowledge can grow by an element of 26.
- The size, nature and origin of the information, moreover because the organizations that manage it, are often represented and listed employing a big selection of criteria.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

Traditionallypoint security products used for IT security defences but as these products are effective for protecting against specific threats such as firewalls limit access to networks, anti-virus software detects malware on given devices and encryption protects stored data. However now advanced cyber security threats are becoming progressively more multifaceted and can often only be detected by looking at data from manifold sources. These various sources includes the logs from point security products, information about IT systems and the data that is used to store knowledge of users and their rights and other contextual information. A correlated view of all this data enables unforeseen attacks to be thwarted as they happenand effectively improve base security.

Therefore, some advanced approach are required that must be able to detect these threats by correlating information from a wide range of sources, including point security products etc. Most organisations already have much of the required data to achieve this but the tools still needed to intelligently process such vast amount of data. This objective has led to the emergence of next generation SIEM (security information and event management) tools. These enable the real time correlation of IT intelligence data and for many advanced threats to be foiled or pre-empted that would have been previously undetectable.

So much criminal activity and political activism has now been displaced from the physical world to cyber space, or at least extended to cover both, that IT security employees are now in the frontline when it comes to ensuring that the businesses they serve have the ability to function and that their continued good reputation is ensured. To this end they must be enabled with the tools that give them a broad insight into IT infrastructure, applications and user activity to protect their business from attacks tomorrow that no one can envisage today.

## II. BIG DATA ANALYTICS

Big Data analytics is being recognizing as a ‘transformative’ technology for security event analysisleading to detect threats in the ever-increasing volume of event data generated from in-house, mobile, and cloud-based services. Investigation and collection activities also terming to a group of techniques to collect and analyse vast amount of information using advanced technology, rather than human agents. For the existence of multidimensional data set, to extract the features and also to remove the redundant and inconsistent features that affect classification, information gain and genetic algorithm has been combined to select the significant features [5]. This method shows better accuracy when features are selected than individually applied. Author uses two class classification methods in terms of normal or attackwithout any misclassification [6].

The Big Data analytics opportunity is touting as group of methodologies/techniques which is capable of driving actionable insight and timely enterprise decisions from the amalgamation of different types of data – unstructured, semi-structured, and structured data. While each type of data alone delivers value, together they are useful to provide true insight leading to intelligence. Unstructured data consists of the ideas and concepts that people communicate every day, in emails, PowerPoints, phone conversations, texts, tweets and videos. Machines, such as the small wireless electronic sensors appearing on virtually every new product now a day generated Semi-structured data. The data what customers traditionally store in databases, such as credit card transactions, customer relationship data, or account information recognized as structured data. This is further complicated by the very volume of this data (now routinely measured in Petabytes) and the velocity of its growth.

## III. INFORMATION SECURITY ANALYTICS

An important perimeter defenceto reducing the overall threat landscape is never enough as the advanced human or machine driven attacks are getting through undetected. As most of the time these undetected attacks also then go unnoticed, often for a very long period of time, embedding themselves within a network, collecting what they need, or simply waiting for an opportune time to cause havoc.

Many organizations have the practice to conduct information security risk assessments on the basis of infrequent events occurring somewherebetween quarterly and yearly [7]. As the relevant information collected within a limited time frame a detailed security assessment can become overwhelming. Furthermore, only within a limited time gathered information can only be assumed as a status update of theinformation security environment related to the organization[8]. Hulme notes that “A riskassessment conducted on the first day of the month can be quite different than

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 2, February 2016**

the same assessment conductedseveral weeks later;" and that risk can be most effectively minimized by keeping "eternally vigilant" [9].It is also important and essential for organisations to maintain record of past security incidents and identified problem indicationsfrom that information [10]. If the organizations, are not residualconsiderate and retentive then vital risk-pertinent data about current developments and long-term trends can be missed out. While such type of data could afford them advance warning by means of predictive analysis related to impending incidents.

However, the pursuit for meaningful, reliable information security data is very complex, and often daunting. But there is need to develop tailored security analytics methodologies or techniques that deliver insightful and actionable information. Key points need to be covered in this direction include how to define success, identify needs, and develop and integrate data sources. Further actions are needed to analyze key security information to make smarter security decisions. Therefore, security analytics can be considered a very effective tool needed to get reliable, actionable security data and intelligence.

## IV. BIG DATA SECURITY ANALYTICS

Information security is becoming a big data analytics problem, where massive amounts of data needs to be correlated, analyzed and mined for meaningful and logical patterns.This is the fact that cyber advanced attacks are harder to defend against as Data Breach Investigations Report from Verizon Business found that breaches are taking longer to discover, with 85% taking weeks or more in 2011, up 6% over the previous year. In the same report it is also stated that 92% of security breaches were discovered by the third parties vendors not by the victim organizations. In addition only 4% were discovered by active internal methods that include log monitoring,antivirus controls, intrusion detection and prevention systems. Just 1% of organizations found security breaches by reviewing and analyzing log records despite the fact that 84% had log evidence available for forensic investigation. Further verizon business states that among all nearly methods available for security breach detection log analysis is more effective [11].

To overcome the issue of less accuracy in unsupervised method which uses a huge set of data as pre-labeled training data, a semi-supervised algorithm is used [12]. For Euclidean distance and statistical both properties of clusters, Fuzzy Connectedness based Clustering approach is evaluated [13]. It facilitates the discovery of any shape and detects not only known but also its variants. The semi supervised learning mechanism is used to build an alter filter to reduce the false alarm ratio and provides high detection rate [14]. Where the features of both supervised and semi supervised learning are same in nature.

Defining big data security analytics has to start at a high level. Big data security analytics is simply a collection of very large and complex sets of security data that it becomes difficult or impossible to process using traditional security data processing techniques/tools and on-hand database management tools.A co-training framework to leverage unlabeled data to improve intrusion detection was proposed [15] [16]. This framework provides lower error rate than single view method and thereby incorporating an active learning method to enhance the performance. Further from a security perspective view there are two distinct issues first, securing the organisation and its customers' information in a Big Data context; and second using Big Data techniques to analyse, and even predict, security incidents.There are three basic characteristics on which Big Data Security Analytics Solutions (BDSAS) distinguish themselves based upon:

• *Scale:*BDSAS must have the power to gather, process, associate degree store terabytes to petabytes of data for an assortment of security analytics activities.

• *Analytical flexibility*:BDSAS should offer users with the flexibility to act, query, and associate visualize the volume of data in an assortment of the way.

• *Performance:*BDSAS should be engineered with acceptable calculate design to method data analytic algorithms and complicated queries then deliver ends up in a suitable timeframe.

Taking the idea a step further, Big Data style analysis can be a answer to the challenge of detecting and preventing advanced persistent threats massively occurring. These techniques could also play a key role in helping to 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.

Practice of ignored the logs if no incident occurs can be harmful. Big Data provides the opportunity to consolidate and analyse logs automatically from multiple sources rather than in isolation. This could provide insight that individual

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 2, February 2016**

logs cannot, and potentially enhance intrusion detection systems (IDS) and intrusion prevention systems (IPS) through continual adjustment and effectively learning "good" and "bad" behaviours appearing in the correlated data of logs.

Integrating information from physical security systems, such as building access controls and even CCTV, could also significantly enhance IDS and IPS to a point where insider attacks and social engineering are factored in to the detection process. This procedure presents significantly the possibility of more advanced detection of security fraud and criminal activities.

## V. POTENTIAL SOURCES FOR INFORMATION SECURITY DATA

Before collecting and analysis of security data it is very necessary to identify the important sources of data. Although these resources can be hundreds in number but potential one few some. However resources are identified which can be termed potential for information security data collection [17]. ***Entity:*** some important and specific entities are identified from which security data can be collected for the analysis and outcome for the preventive techniques.***Issues/observations:*** some specific issue/observations are also discussed which needs to be address significantly while analysing data from potential sources.***Potential sources:*** a number of significant and potential data sources are identified to get security data for suspicious activities.

***Network/host traffic***

- Traffic anomalies to/from these servers
- Distribution of Protocols
- Usage of Encryption techniques
- Suspicious activities to and from source and destinations

***Potential sources***

- SIEM,
- Application monitoring
- Networking monitoring

***Web transactions***

- Surveillance of suspicious or deceptive activities in sensitive and high value applications and assets involved in web transactions.

***Potential sources***

- Web server logs
- Web application server logs
- Authentication data
- Transaction monitoring
- Network session data

***Infrastructure***

- Change or manipulation of the server
- State of vulnerableinfrastructure
- Updation/modificationof configuration settings
- Policy compliance

***Potential sources***

- II assets,
- Configuration management
- Vulnerability management
- Grc systems,

***Information***

- Type of data stored in the information system
- Transition state of data of the information system
- Kind of data processed by the information system
- Status of the information related to the various regulations
- Value Status of IP address such as high, medium or low

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 2, February 2016**

***Potential sources***

- Data classification
- DLP
- GRC systems

***Identity***

- Recognition of logged in users
- Status of the privileges assigned to the different users
- Place and resource in which users logged in
- Usage of other assets by the users

***Potential sources***

- Asset management
- Authentication data
- Server logs
- SIEM
- Microsoft active directory
- Networking monitoring

Moreover, in order to obtain the actionable intelligence either for early threat detection or for historical analysis to track advanced threats over time,organizations need to better collect and analyse the reams of security-related information that is generated within their networks[18]. In addition correlation of this collected security information with external intelligence feeds that include real-time threat intelligence information required for further actionable analytics. The more information that is collected, the greater the chance of finding malicious signals that point to security incidents that are occurring or that have occurred, improving an organisation's ability to reduce its overall security risk throughout its infrastructure, network, applications, data and stakeholders.

## VI. INFORMATION ANALYTICS TECHNIQUES AND INFORMATION SECURITY PRACTICES

Intelligence analysis of information can be defined as "A formal process which attempts to find and measure relations among variables. Although at times it may draw heavily on mathematics and numeric procedures, it is a logical and not a mathematical process". Intelligence analysis provides a unique way of thinking that integrates possible scenarios or complements incomplete information with data that might have been isolated but is linked because of specific characteristics. This analysis needs to be very extensive and its outcome can extend entire program of identification of security incidents and their preventions schemes. For keeping in mind the significance of analysis of security information, a number of analysis techniques are explained along with possible uses:

- **Accumulation:**Storing data inone place; impliessome notion ofretrievability. Example:Construction of knowledge bases (KBs) that store cases and procedures regarding InfoSec incidents.Keeping InfoSec configuration diaries that illustrate specific ways of setting upequipment.

- **Aggregation:** Many data points brought together into a smaller set which is usually more easily accessed.Example: Use of reporting tools like Critical Watch that can trim the number of events and produce a simplified vulnerability report. Integration of specific data from different platform into custom reports designed by purpose-specific InfoSec analysts.

- **Analysis:**Aanalyze hundreds of different structured and unstructured data types. Support for diverse types of data is critical to enabling the types of wide-ranging investigations typically conducted by security analysts. Example: Use of data collection selection tools that identify possible threats based on readings and indicators.Use of methodologies like COBIT to map out specific vulnerable areas of an InfoSecInfrastructure.

- **Mix**: Passing of data around to a variety of managers looking for possible links. The data is often not well ordered.Example: Creation of task groups for the solution of issues that require IT and InfoSec joint

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 2, February 2016**

cooperation. Determination of Operating System hardening procedure by IT and InfoSec analysts aimed to prevent external attacks

Organizations will be able to detect patterns by collecting, analyzing and correlating information of diverse formats from multiple sources. However it would not be possible to discover such important and usable patterns, if they were looking at each source in isolation. For example, a security incident might have been happened if an event log indicating that a password has been reset but does not by itself. But when that reset is correlated with help desk tickets and it can be seen that no ticket has been raised for that event, it could indicate that someone has made that change without authorization. Similarly, suppose a user logged in to the network from an office location when there is no entry record for the same in physical means. Correlation of such information of employee network login events with physical access records could flag suspicious behavior resulting that another user is impersonating them. Correlating and analyzing log and event data from multiple sources provides the overall picture of the entire landscape so that gaps in security can be weeded out.

More extensive protection can be resulted with the broader aspiration of real-time mitigation of security threats. Supplementing the analytics across a wide range of data sources during an attack provides more insight needed for extensive protection. Examples are:

- Identifying uncommon traffic between servers, which may be a characteristic of undetected malware looking out data stores
- Matching data egress from a device with access records from a suspicious IP address, user or location
- Preventing non-compliant movement of information, serving to get an employee who may be not following the rules
- Associating IT security events with physical security systems, up authentication and authorization of known employees accountable for maintenance of plant infrastructure
- Identifying uncommon access routes, as regularly some databases accessed via bound applications and circuitously by users.

***Improved Information Security Practises***As a wide range of sources are available for collection of security data presented in structured, semi-structured and unstructured form. With the availability of advance data analytics techniques an intelligence analysis can be conducted on security data which further resulted in more extensive information security practices related to the following activities:

- To identify control weaknesses
- To identify lower cost control
- To quantify threat levels
- To quantify incident indicators
- To prevent incidents happening
- To gather intelligence
- To prevent accidental or unlawful destruction or loss

## VII. CONCLUSION AND FUTURE WORK

A security intelligence platform should enable organisations to harness and make sense of massive volumes of security data with big data analytics capabilities. The system should be capable of not solely to defend against advanced threats and attacks and illicit activity in real time however additionally perform continuous real time monitoring and automated historical correlation so that threats are discovered in an exceedingly timely manner. In addition an intelligent security analytics system should also store all data for forensic and compliance purposes for the better forensic analysis to ascertain how an attacker gained a foothold on the network. The developed information security analytics platform must integrate and interface with multiple information sources as well as other security and IT controls in use in the organisation.Such intelligence will boost stake holders' ability to determine most relevant and severe threats, attacks and vulnerabilities in present and future, allowing them to better priorities their defenses and fine-tune their information security policies and practices.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 2, February 2016**

## REFERENCES

1. http://www.pwc.co.uk/audit-assurance/publications/uk-information-security-breaches-survey-results-2012.jhtml
2. Whitman, Michael E., and Herbert J. Mattord. 2004. *Management of information security*. Boston, Mass.Thomson Course Technology
3. Baskerville, R. 1991. "Risk Analysis: an interpretive feasibility tool in justifying information systems security." *European Journal of Information Systems* 1, no.2: 121-130.
4. Shedden, Piya, A. B. Ruighaver, and Atif Ahmad. 2010. "Risk Management Standards - The Perception of ease of use." *Journal Of Information System Security* 6, no. 3: 23-41.
5. Sethuramalingam S. Hybrid feature selection for network intrusion. Int J ComputSciEng 2011; 3(5):1773–9.
6. Mrutyunaya Panda, Ajith Abraham, ManasRanjan Patra. A hybrid intelligent approach for network intrusion detection. In: International conference on communication technology and system design, procedia engineering, vol. 30; 2011. p. 1–9.
7. Rees, J, and J Allen. 2008. "The State of Risk Assessment Practices in Information Security: An Exploratory Investigation." *Journal of Organisational Computing and Electronic Commerce* 18, no. 4: 255-277.
8. Schmittling, Ron. 2010. "Performing a Security Risk Assessment." *ISACA Journal*, Vol. 1: 1-7.
9. Hulme, George V. "Getting at Risk." In *Management of Information Security*, by Michael E. Whitman and Herbert J. Mattord, 307-308. Boston: Thomson Course Technology, 2004
10. Ahmad, Atif, Justin Hadgkiss, and A.B. Ruighaver. 2012. "Incident response teams – Challenges in supporting the organisational security function." *Computers & Security* 31, 643-652.
11. http://www.verizonbusiness.com/resources/reports/rp_data-breach-investigations-report-012_en_xg.pdf
12. Gao Xiang, Wang Min. Applying semi-supervised cluster algorithm for anomaly detection. In: IEEE international symposium on information processing, no. 3; 2010. p. 43–5.
13. Qiang Wang, VasileiosMegalooikonomou. A clustering algorithm for intrusion detection. In: International conference on data mining, intrusion detection, information assurance, and data networks, security, 5(12), 2005, p. 31–8.
14. Chien-Yi Chiu, Yuh-Jye Lee, Chien-Chung Chang. Semisuper-vised learning for false alarm reduction. In: Industrial conference on data mining, no. 10; 2010. p. 595–605
15. Ching-Hao, Hahn-Ming L, Devi P, Tsuhan C, Si-Yu H. Semisupervised co-training and active learning based approach for multi-view intrusion detection. In: ACM symposium on applied computing, no. 9; 2009. p. 2042–7.
16. T. S. Bharati; R. Kumar; "Secure Intrusion Detection System for Mobile Adhoc Networks", Proceedings of the 9thINDIACom; IEEE Conference ID: 35071, Delhi, INDIA, March 11-13, 2015, pp 855-859
17. R. Kumar; "Revisiting Security Vulnerabilities: Web Applications Perspective", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, No. 6, June 2013, pp 1653-1659
18. R. Kumar; "Mitigating the Authentication Vulnerabilities in Web Applications through Security Requirements", World Congress On Information and Communication Technologies (WICT 2011), Mumbai, December 11-14, 2011 (IEEE Xplore).