An Approach of Data Mining in Security Information and Event Management: A Survey

Drashti Bhavsar[1] Hiral Chhaniyara[2] Krunal Joshi[3] Jagrati Shekhawat[4]
[1,2]PG Scholar [3,4]Assistant Professor
[1,2,3,4]Department of Computer Engineering
[1,2,3,4]Sal Institute of Technology and Engineering Research, Ahmedabad, India

Abstract— Security Information and Event Management (SIEM) systems are today a main ingredient of complex enterprise networks. SIEM associates Security Information Management (SIM) and Security Event Management (SEM). It highlights the effect of the technology on the whole system, even though the focus is on security. The first limelight is on analysis and reporting of log data and long-term storage while the second limelight on real-time monitoring and notifications. The basic role of SIEM in data infrastructure, its classification in specific cloud environment, and technical requirements for SIEM solution implementation into a cloud environment correlate to individual cloud distribution models. Some researchers would rather speak of 'SIEOM', adding the O for "opportunity". We will see how various data mining techniques can be used in security information and event management system to upgrade the efficiency of the system.

Key words: Data Mining, Security Information Event Management System

I. INTRODUCTION

In many organizations, security policies or business regulations require that security events are monitored and that security logs are reviewed to identify security issues. Information captured in security logs is often critical for reconstructing the sequence of events during investigation of a security incident, and monitoring security logs may identify issues that would be missed otherwise. The problem is that the amount of information generated by security devices and systems can be vast and manual review is typically not practical. Security event management (SEM or SIM-security information management) aims to solve this problem by automatically analyzing all that information to provide actionable alerts. In a nutshell, security event management deals with the collection, transmission, storage, monitoring and analysis of security events.

When implemented correctly, a security event management solution can benefit a security operations team responsible for monitoring infrastructure security.

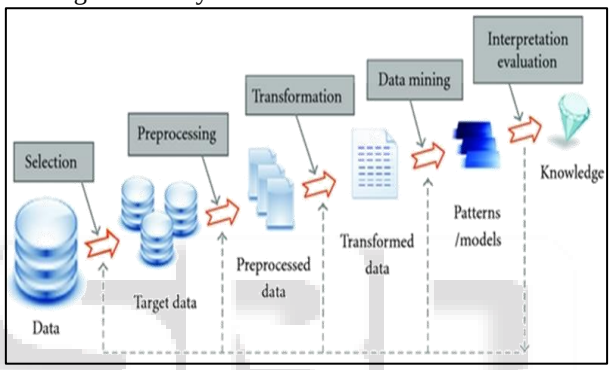
The object of information security management is the protection of these systems, whereas security information and event management (SIEM) addresses those information management tasks which focus on the short term handling of events, as well as on the long term improvement of the entire information security architectures. This is carried out based on those data which can be logged and collected within the enterprise information security infrastructure.

II. DATA MINING BASICS

Data mining is the process of digging through large volumes of data and extracting previously unidentified and potentially useful information. By finding out useful patterns and trends about different aspects of the company, businesses can come up with new strategies that are helpful in gaining competitive advantage.

Most companies already collect and refine massive quantities of data. Data mining techniques can be implemented rapidly on existing software and hardware platforms to enhance the value of existing information resources, and can be integrated with new products and systems as they are brought on-line.

He following diagram shows the process of knowledge discovery



Some people don't differentiate data mining from knowledge discovery while others view data mining as an essential step in the process of knowledge discovery. Here is the list of steps involved in the knowledge discovery process

A. Data Cleaning

In this step, the noise and inconsistent data is removed.

B. Data Integration

In this step, multiple data sources are combined.

C. Data Selection

In this step, data relevant to the analysis task are retrieved from the database.

D. Data Transformation

In this step, data is transformed or consolidated into forms appropriate for mining by performing summary or aggregation operations.

E. Data Mining

In this step, intelligent methods are applied in order to extract data patterns.

F. Pattern Evaluation

In this step, data patterns are evaluated.

G. Knowledge Presentation

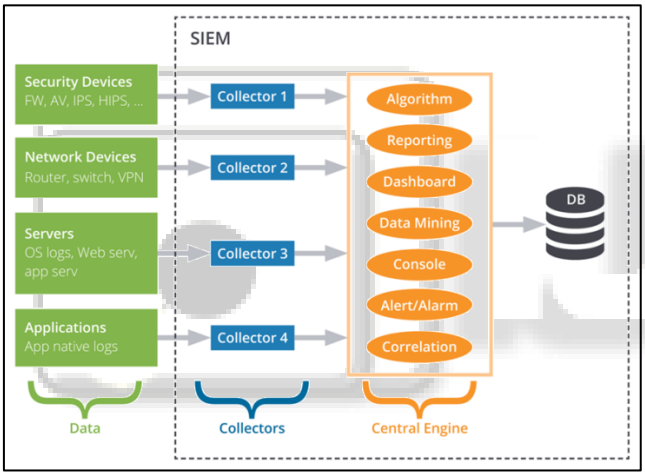
In this step, knowledge is represented.

III. OVERVIEW OF SIEM

The underlying principle of a SIEM system is that relevant data about an enterprise's security is produced in multiple locations and being able to look at all the data from a single point of view makes it easier to spot trends and see patterns that are out of the ordinary. SIEM combines SIM (security information management) and SEM (security event management) functions into one security management system.

A SIEM system collects logs and other security-related documentation for analysis. Most SIEM systems work by deploying multiple collection agents in a hierarchical manner to gather security-related events from end-user devices, servers, network equipment -- and even specialized security equipment like firewalls, antivirus or intrusion prevention systems. The collectors forward events to a centralized management console, which performs inspections and flags anomalies. To allow the system to identify anomalous events, it's important that the SIEM administrator first creates a profile of the system under normal event conditions.

A. Centralized Logging Infrastructure



At its core, a SIEM provides:

1) Event and Log collection

This may come in many forms, especially with in-house applications.

2) Layered Centric Views or Heterogeneous

This is usually in the form of dashboards or "views," referred to as a bird's-eye view.

3) Normalization

A two-part function. This includes translating computerized jargon to readable data to be displayed, and mapping data to user- or vendor-defined classifications/characterizations. This is sometimes referred to as "field mapping."

4) Correlation

This essentially gives the data context and forms relationships based on rules, architecture and alerts. This should be either historical or real-time.

5) Adaptability (Scalable)

This dumbs down to being able to speak the language regardless of source vendor, format, type, change or compliance requirement.

6) Reporting and Alerting

This may be used to not only show value to executives but also provide automated verification of continuous monitoring, trends and auditing. Some would argue that the auditing aspect is an essential function but the SIEM alone does nothing – like a retired general with no troops or a SQL instance with no tables or data.

7) Log Management

Allowing the capability for storing event and logs into a central location, while also allowing the application of compliance storage or retention requirements. (Again, many would argue this is a separate function, and I would disagree.)

IV. NECESSARY SOLUTION QUALITIES

While big data platforms can be used for improved information security data mining, there are several properties unique to the security discipline. Any solution—whether SIEM, BI, big data, or something else down the road—needs to be able to take these special qualifiers into account:

A. Speed, Speed and Speed

Unlike a traditional business case—say, identifying long-term buying patterns among specific demographics—most information security use cases that need data now if they're going to be useful.

B. Data from Everywhere

Really advanced security functions, such as discovering new attacks, require visibility into not only multiple types of assets—perimeter network control devices (e.g., firewalls), perimeter security data (e.g., IDS/IPS), servers, databases, and applications—but multiple levels of data as well. Events, system state changes such as altered configurations, network traffic, and more can all be required to get value out of data mining for information security.

C. Discovering and Alerting on Abnormalities, not just Signatures.

This has been the holy grail of security monitoring for some time, and it's a problem that still hasn't been fully solved. Signature-based detection is a useful part of security in depth, but it's not a solution for discovery of real-time threats, zero-day activities, and carefully planned and orchestrated attacks that utilize layers of technology infrastructure. Without this capability, all the speed in the world won't help us get past the square we've been stuck at for quite a few years now.

V. BENEFITS OF DATA MINING

There are many benefits of data mining. For example:

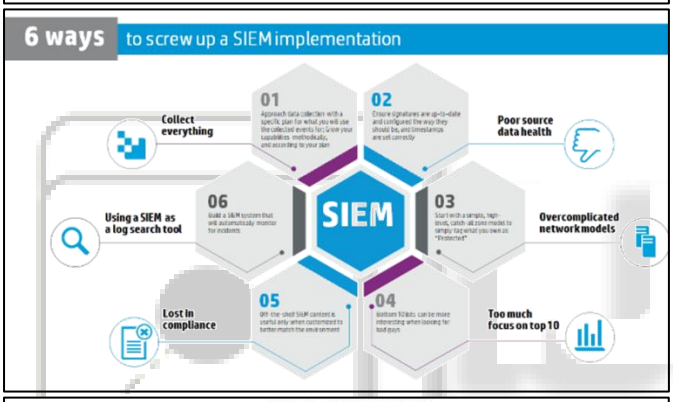
In finance and banking, data mining is used to create accurate risk models for loans and mortgages. They are also very helpful when detecting fraudulent transactions.

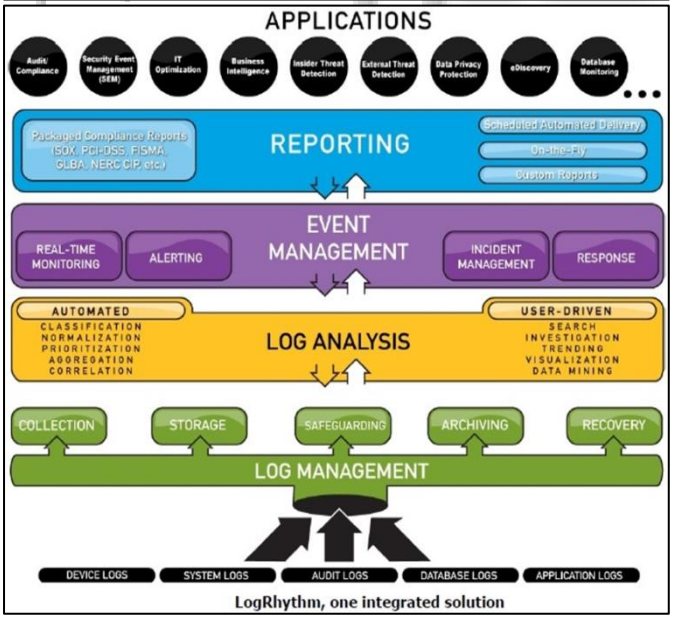
- In marketing, data mining techniques are used to improve conversions, increase customer satisfaction and created targeted advertising campaigns. They can even be utilized when analyzing the needs in the market and coming up with ideas for completely new product lines.

This is done by looking at historical sales and customer data and creating powerful prediction models.

- Retail stores use customer shopping habits/details to optimize the layout of their stores in order to improve customer experience and increase profits.
- Tax governing bodies use data mining techniques to detect fraudulent transactions and single out suspicious tax returns or other business documents.
- In manufacturing, data discovery is used to improve product safety, usability and comfort.

Strengths	Weakness
Extensive Log collection support, mature event correlation, categorization and reporting.	Complex deployment & configuration leading to increased time to go-live
Comprehensive User Behavior Analytics is possible using ArcSight. This is due to their collaboration with Securonix.	Mostly suited for Medium to Large Scale deployment
Application view using HP Fortify technology has added contextual data to Application monitoring	Requires skilled resources to manage the solution
Highly customizable platform enabling advanced correlation and analytics	Steep learning curve for Analysts & Operators
Highly Available & Scalable Architecture supporting Multi-tier & Multi-tenancy	Longer ROI
More and more Web based administration, monitoring etc is becoming a focus.	





VI. CONCLUSION

This study shows that how data mining can be used in SIEM system. This paper firstly introduces the related knowledge, logging infrastructure, overview ,necessary solutions of SIEM system One of the areas we are exploring for future research is how we can use other data mining technique like classification, clustering to enhance the system capacity.

REFERENCES

[1] J. W. Seifert, "Data Mining and Homeland Security: An Overview,"CRS Report, pp. 1-1, Jan. 2007.
[2] M. S. Chen and J. H. Philip, "Data Mining: An Overview from a Database Perspective," IEEE Trans on knowledge and dataengineering, vol. 8, no. 6, pp. 1-1, Dec 1996.
[3] S. Yuan and C. Zou, "The Security Operations Center Based on Correlation Analysis."
[4] E. E. Eljadi and Z. A. Othman, "Anomaly Detection for PTM's Network Traffic Using Association Rule," in Proc. of 2011 3rdConference on Data Mining and Optimization DMO, June 2011
[5] J. Han and M. Kamber, "Data Mining Concepts and Techniques. Second Edition," The Morgan Kaufmann Series in Data ManagementSystems
[6] D.F. Carr, "Security Information and Event Management". Baseline, No. 47, 2005, p. 83.
[7] B. Gilmer, "Firewalls and security", Broadcast Engineering, Vol. 43, No. 8, 2001, pp. 36-37.
[8] A. Williams, "Security Information and Event Management Technologies", Siliconindia, Vol. 10, No. 1, 2006, pp. 34-35.
[9] R. Gabriel, T. Hoppe, A. Pastwa, and S. Sowa, "Analyzing Malware Log Data to Support Security Information and Event Management: Some Research Results", Proc. First International Conference on Advances in Databases, Knowledge, and Data Applications (DBKDA 2009), IEEE Press, Mar. 2009, pp. 108-113, doi: 10.1109/DBKDA.2009.26.
[10] Damian Hermanowski, "Open Source Security Information Management System Supporting IT Security Audit" published in IEEE, year 2015.
[11] Igor Anastasov, DancoDavcev, "SIEM Implementation for Global and Distributed Environments" published in IEEE year 2014.
[12] Domingos, C., Gavalda, R. & Watanabe, O. (2002). Adaptive Sampling Methods for Scaling Up Knowledge Discovery Algorithms. Data Mining and Knowledge Discovery 6: 131-152.

