A Clustering based Intrusion Detection System for Storage Area Network

Garima Singh
Department of CSE
Jaypee University of Information Technology
Waknaghat, Solan, H.P. (India)

Anubhav Patrick
Department of CSE
Jaypee University of Information Technology
Waknaghat, Solan, H.P. (India)

Lucky Rajpoot
Department of CSE
Jaypee University of Information Technology
Waknaghat, Solan, H.P. (India)

ABSTRACT
A storage area network (SAN) is a high-speed and widely used special-purpose network that interconnects different kinds of storage devices with associated data servers on behalf of a larger network of users. SAN security is a specialized field dealing with issues related to the storage industry, it follows the same established principles found in all modern IT security. Therefore, it requires a continuous process of evaluating SAN environment's current state of security against the constant changes brought about by innovations in technology and an increase in awareness concerning security issues. This paper is all about intrusion detection in storage area network, and more important, how to detect and prevent suspicious activity of an unauthorized user by maintaining an audit record. This paper proposes an approach to detect an intrusion attack by clustering (k-mean) (to identify groups of similar behaved object, i.e. malicious and non-malicious activity), classification technique (to classify all data into particular class categories).

General Terms
Storage area network (SAN), Fibre channel, Intrusion detection system, Audit record, Clustering, Classification technique, K-means, One rule algorithm

Keywords
SAN observer, IDS manager, host agent module, SAN observer module, manager module

1. INTRODUCTION
A SAN is a network of storage devices which is designed specifically to support efficient data storage, management and transmission between computational and storage elements of a system [2]. SAN often uses Fibre Channel or SCSI infrastructure for data transmission. SAN concept is quite similar to network attached storage (NAS) with the major difference that SAN works on block level while NAS works on file level. The major advantage of SAN is it provides consolidated storage by combining and virtualizing several distinct storage devices and a single, large data storage pool [3]. SAN removes dedicated links between servers and storage devices and replaces it with a single high speed network of storage devices [4]. Another major advantage of SAN is it reduces the inter-dependency between computing and storage resources and provides reliability [3]. IDS is a security mechanism used to prevent and deter security breaches or intrusions. IDS works in conjunction with other security measures like firewall, antivirus, antimalware etc. IDS are usually characterized in three types [1]- (i) host based IDS which reside on the host locally that needs to be protected against attacks; (ii) network based IDS which is centralized and monitors the whole network as a whole; and (iii) hybrid IDS which shares the characteristics of both (i) and (ii). To detect intrusions IDSs use two different approaches [1]- (i) Rule based approach in which there are pre-determined rules of what can be considered as a legitimate act and what cannot.` The rules are unambiguously and explicitly stated and generated beforehand the actual intrusion detection process; (ii) Anomaly based approach in which the user's or process' behavior is tested against a standard profile to detect any deviation from the expected behavior.

SAN based IDS is a relatively unexploited field in SAN research. SAN involves a network of storage devices that is available to an entire enterprise or organization which can be accessed through LAN or WAN. Since SANs are often connected to servers, clients and other networked entities through unreliable and often unsecure data networks so there are serious security concerns. A SAN can be subjected to intrusions, DOS attacks, hacking, virus and worm attacks etc. These security threats are even more compounded since SANs usually store organizational secrets and critical data at a centralized location and its failure can jeopardize the whole business. So threats from insiders and outsiders attackers cannot be underestimated and we need to fortify the SAN with mechanisms to tackle them. IDS provide a security cover to SAN against masquerading entities which hide their true identity and legitimate entities which accessing data are beyond their privileges.

Clustering is a process of creating groups or clusters based on similarities between objects. Clustering is a form of unsupervised machine learning and it does not require that class label of each object must be known in advance. Clustering methodology is categorized under following major types [5]: partitioning methods, hierarchical methods, and model based methods, methods for high dimensional data and constraint based clustering. Clustering is used for data mining, pattern recognition, image processing, security techniques, data analysis etc.

We have proposed a clustering based IDS for SAN which uses K-means clustering technique for intrusion detection. The rest of the paper is divided as follows: Section 2 gives an overview of related work in security domain of SAN which particularly focuses on intrusion detection. Section 3 provides our proposed approach for intrusion detection in SAN. Section 4 gives results and comparative analysis of our

approach with existing approaches. Section 5 provides future direction of our work and conclusion of the research.

2. RELATED WORK

Researchers of [6] have given a new security framework for SAN which focuses on intrusion detection and tolerance against attacks. Their approach involves cooperation among various components of SAN including SAN volume controller (SVC) and SD (storage disks). The intrusion detection process is based on an important property called compromise independence which ensures that the IDS is able to discover intrusions even after the system is compromised. They have used distributed rules for intrusion detection. The effectiveness of IDS depends on the rules selected to detect intrusions. In [7], two IDS approaches have been proposed. The first approach is a real time IDS in which each block that is being transmitted from or to SAN is evaluated in real time for possible signs of intrusions. This approach provides continuous protection against attackers. The second approach proposed works on file level. In this approach there is little dependency between storage components of SAN and IDS. The IDS works as a service and there is no requirement to modify existing components of SAN. Adam G. Pennington et al., [8] have researched on IDS for storage systems. They presented various kind of types of attacks on storage systems that come under purview of storage IDS. Their proposed IDS is rule based and is embedded in an NFS server. They have used Tripwire-style rules and specified with two additional rules for intrusion detection. Each NFS operation is tested against IDS rules and an alert is generated in case a suspicious activity is detected. In [9], researchers have focused on wireless SANs (WSAN) which are an evolution of traditional SANs. They have proposed an intrusion detection and tolerance scheme for managing the transactions in WSAN. Their devised scheme can detect intrusions and specifies how to process transactions even if the system has been compromised. They have used flex transaction model given by [10] to efficiently secure transactions and Predicate Transition Net to model flex transactions. In [11], the researchers have raised some common issues regarding development of IDS for storage systems including SAN. They have demonstrated how current centralized and distributed IDS like CIDTS (Cooperative Intrusion Detection and Tolerance System) are unfit for SAN environment due to various reasons including processing and memory constraints, lack of cooperation among SAN entities, inaccuracy of system to detect intrusions etc. The researchers have also given a detailed account of various attacks and vulnerabilities that exists while designing security mechanisms for storage systems. P.Mahalingam et al. in [12] have provided a security framework for SAN. They have discussed the basic architecture of SAN from security perspective and enumerated various level of threats and risks surrounding a modern day SAN. They have also specified various security practices and mechanisms that are currently applicable to SAN to overcome those security threats. Lastly they devised a security framework incorporating several security mechanisms including IDS for SAN. Their proposed framework integrates security mechanisms in all the stages of development of SAN. Thus their security framework provides an all-round security for confidential enterprise data that is stored in a SAN.

3. PROPOSED SCHEME

The operation performed by an unauthorized user will definitely vary from an authorized user .Thus by observing the behavior of the object or transaction, it can be verified that whether the user is authorized or the transaction is valid or not. An audit record is a fundamental tool for detecting the behavior of an object. Now, the objects which are related to security concern are chosen and clustered with the help of K-mean clustering technique and then classified into particular classes and passed it to the Administrative manager.

3.1 Audit Record

Audit record must contain the following fields:

3.1.1 Subject

A subject is an end user but might also be process acting on behalf of users or group of users.

3.1.2 Action

Action involves operations performed by the subject on or with an object (SAN); for example, login, read, write, perform I/O, execute.

3.1.3 Object

Receptors of actions. The example involves a storage area network

3.1.4 Exception Condition

Defines which, if any, exception condition is raised on returned.

3.1.5 Resource Usage

A list of quantitative elements in which each element gives the amount used of some resource. Example, number of lines printed, number of records read or written, processor time, I/O used session elapsed time.

3.1.6 Timestamp

It's a unique time and date stamp used for identifying when the action took place.

3.2 Measures for Unauthorized Activity

3.2.1 Login and session activities

Login frequency by day and time, Frequency of login at different locations, Time since last login, Elapsed time per session, Quantity of output to location, session resource utilization, password failures at login, Failure to login from specified terminals

3.2.2 Command and program execution activities

Execution frequency, Program resource utilization, Execution denials.

3.2.3 File access activities

It involves Read, write, create, delete frequency, records read and written, failure count for read, write, create and delete

3.3 Architecture

The proposed approach is based on Intrusion Detection System (IDS) to detect the intrusion attack, which involves a machine learning technique (clustering technique). Figure 1 shows the proposed architecture. It consists of one or more SANs consisting of a variety of storage devices connected to an IDS server via WAN. Each SAN also consists of a local SAN observer which observes the activities taking on SAN devices.

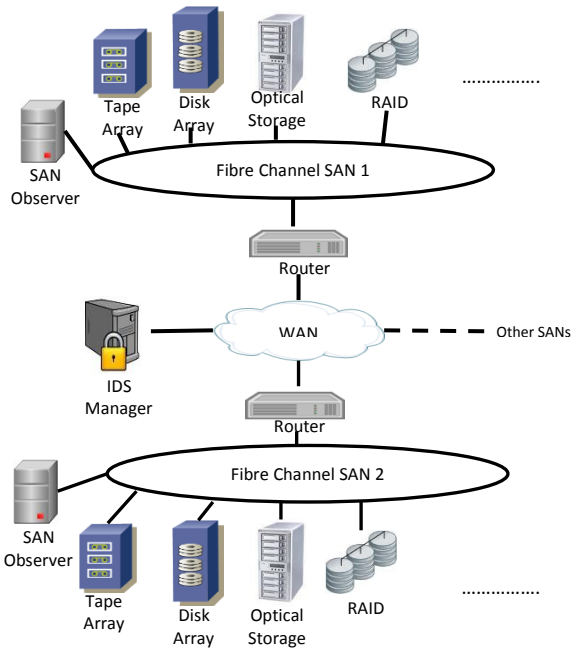

Fig 1: Architecture of clustering based SAN IDS.

Following are the major modules of IDS entities as shown in figure 2:

SAN Host agent module: Its purpose is to collect data and process on the basis of above measures and transmit these to the Administrative manager.

SAN Observer: Its work is to check Execution frequency, program resource utilization and reports the results to the Administrative manager.

Administrative module: It receives reports from the above two modules and processes and correlates these reports to detect unauthorized access

The proposed scheme is not dependent of any operating system and system auditing implementation. The agent collects each audit record produced by the native audit collection module. A filter is implemented that retains only those commands that are of security interest. Records are then formatted in a standardized format as the SAN host audit record (HAR). After this, clustering technique is used for analyzing the commands for suspicious activity. At the lowest stage, the agent scans for suspected events that are of interest independent of any past events. Examples include failed record accesses, accessing system files, and changing a file's access control .At the next higher level, the agent looks for anomalous behavior of an individual user based on a historical profile of that user, such as the number of programs accessed or executed, number of records accessed, authorized for accessing the records etc. When suspicious activity is detected, an alert is sent to the Administrative manager. The Administrative manager module includes an expert system that can draw inferences from received data. The manager can also query individual systems for copies of SAN HARs to correlate with those of other agents.

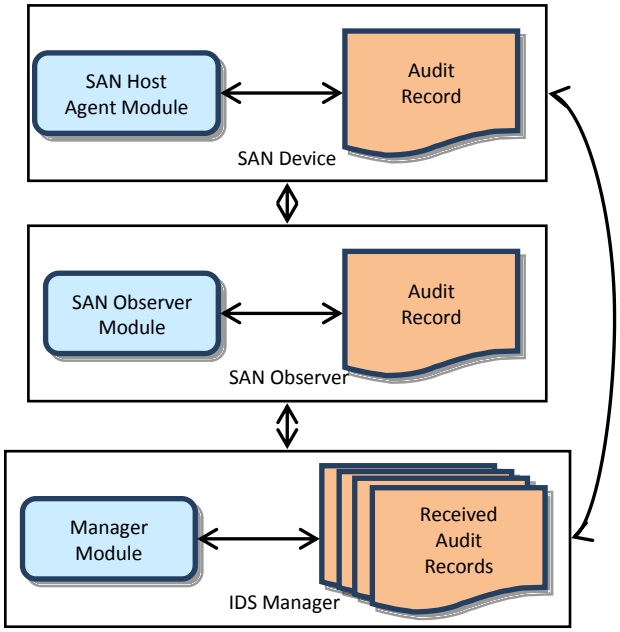

Fig 2: Modules of IDS entities.

3.4 Intrusion Detection Using Clustering

We group similar data objects based on their behaviors by utilizing a K-Means clustering as a pre-classification component.

Network intrusion classes are divided into five main classes, which are Denial of service attack (DoS), distributed denial of service attack (DDoS), Probe, User-to-remote, Remote-to-local, legitimate activity(normal activity).

The main goal to use K-Means clustering technique is to divide and group data into the normal and attack classes. K-Means clustering methods partition the input dataset into k-clusters according to an initial value known as the seed-points in each cluster's centroids. The mean value of each cluster is called cluster centroids. In this case, we choose k= 6 in order to cluster the data into six clusters (C1, C2, C3, C4, C5, C6). Each input data will be assigned to the closest centroid by Euclidean distances method between the input data points and the centroids. New centroids will be generated for each cluster by calculating the mean values of the input set assigned to each cluster and the process is repeated until the result has reached convergence.

The K-Means algorithm works as given below:

1. Consider initial centers of the given K clusters. Then, repeat step 2 through 3 until the cluster membership converges.

2. Now, select a new partition by assigning each data to its closest cluster centers.

3. Compute new clusters as the centroids of the clusters.

With the help of this technique, classification rules on particular tested objects (attributes) are generated by One-Rule based on the value of only a single attribute. One-Rule approach choose attribute with lowest error rate as its "one rule". A proportion of instances that do not belong to the majority class of the corresponding object (attribute) value will contribute to the error rate.

The One-Rule algorithm works as follows:

1. From the clustered classes, create a rule set for each value of each attribute predictor as in step i, ii, iii and iv.

i. Count how often each value of the target class appears.
ii. Next, find the most frequent class.
iii. Make a rule set assign that class to this value of attribute predictor.
iv. Compute the total error occurs in the rules set for each attribute predictor.

2. Find the best attribute predictors which have a smallest total error and this as a classification rules.

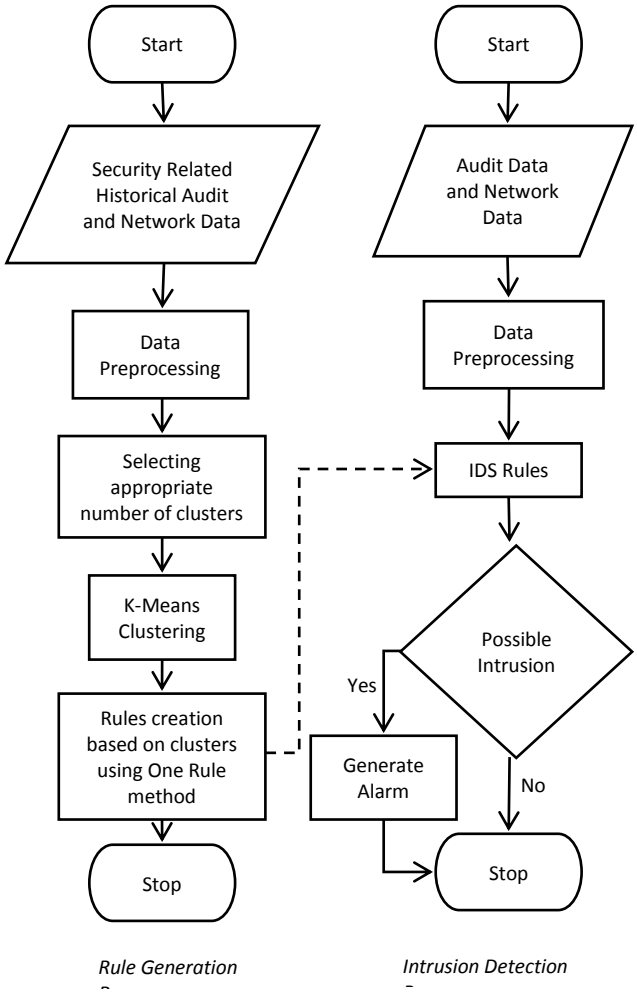

Fig 3: Flow chart of rule creation through clustering and intrusion detection.

Figure 3 above demonstrates the steps of our proposed IDS for SAN using clustering technique.

4. RESULTS AND COMPARATIVE ANALYSIS

Multiple experiments were performed in order to determine the efficiency of our proposed scheme. The KDD cup*99 dataset [13] is chosen for evaluation of intrusions in SAN environment. The initial experimental and analytical results show that our technique works satisfactorily for various types of attacks. Our proposed technique works better or on-par with several earlier techniques since K- Means clustering is being used for pre-classification and then One Rule classification is being used to reduce the time complexity. The clustering techniques is being used as a pre-classification component for the purpose of grouping similar data items into their respective classes which helps to produce better results as compared to only one rule classifier.

We have also done a comparative analysis of various popular security related researches in the field of SAN especially focusing on intrusion detection and summarized the results in table 1.

Table 1.Comparative analysis of existing techniques

Methods	Features	Advantages	Disadvantages
TH-MSNS (Tsing Hua Mass Storage Network System)[14]	Network architecture and design, mass storage, information storage, FCP.	High adaptability, high efficiency, high scalability, and high compatibility and is easy to maintain.	No Proper parallel scheduler, no optimized I/O routes
An automatically tuning IDS (ATIDS)[15]	Attack detection model, classification, data mining, intrusion detection, learning algorithm, model-tuning algorithm, self-organizing map (SOM)	Automatically tune the detection model, tuning procedure is simple, the detection model is represented in the form of rule set which can be easily understandable and controlled.	System behaviour changes drastically if tuning is delayed, take much time in identifying false prediction.
Network-based Intrusion Detection and Prevention System (IDPS)[16]	DPS (Intrusion Detection and Prevention System), machine learning, web application, online detection, internet worm detection	Provides user friendly GU I with web application, all sniffer can be managed by network administrator	Centralized managed server
Intrusion Tolerant Transaction Management Model[9]	Transaction, Flex, Compensation, Attack, Intrusion tolerance, PTN, WSAN.	Provides wireless storage, flexibility and mobility, Malicious transaction can be undone	Does not support interconnected WSAN environment
Real time IDS [7]	Real time intrusion	At block storage level	Tightly coupled with

	detection, block storage level	intrusions can be detected, compromised data can be recovered	SAN- a small change in SAN or SAN data requires significant modification in IDS
IDS loosely coupled with SAN [7]	Non real time IDS, file level storage	Any modification and enhancement to the storage system software not required, negligible impact on storage performance	No real time protection
Proposed Method	Intrusion detection system, Audit record, clustering, Classification technique	Provides secure and efficient intrusion detection system, less expensive, less time complexity because of classification rule.	The number of clusters or groups of attacks must be known in advance, centralized IDS server

5. CONCLUSION

Storage area network (SAN) provides an efficient, coherent and consolidated mechanism to store our personal and organizational data. Since SAN is actually a network of storage devices that is connected to clients and servers using LAN/WAN, so most of the security threats related to traditional networks also apply to SANs. In this research we have focused on intrusion detection in SAN which is becoming a very serious concern these days. We have developed a security framework which detects any malicious activity in a SAN and alerts the administrator in case of any suspicious activity. Our proposed IDS is a rule based IDS which uses K-means clustering and One-Rule method to generate intrusion detection rules. We have compared our proposed work with other state of the art SAN based IDS and came to a conclusion that our approach provides an efficient, cost and time effective solution against intrusion attacks. In the future we will like to improve our approach by incorporating new ideas and better algorithms for rule generation and intrusion detection. We will also like use a distributed architecture for IDS server rather than the current centralized one.

6. REFERENCES

[1] William Stallings, "Intruders" in Cryptography and Network Security, Fourth Edition, Prentice Hall, 2005, pp 565-594

[2] Jon Tate et al., "Introduction" in Introduction to Storage Area Networks and System Networking, IBM, 2012, pp 11-12

[3] Storage Area Network, en.wikipedia.org/wiki/ Storage_area_network

[4] Barry Phillips, "Have Storage Area Networks Come of Age?" in Computer Volume: 31, Issue: 7, 1998

[5] Jiawei Han and Micheline Kambler, "Cluster Analysis" in Data Mining Concepts and Techniques, Second Edition, Morgan Kaufmann Publishers, 2006, pp 383

[6] Yacine Djemaiel et al., "Dynamic detection and tolerance of attacks in Storage Area Networks" in 22nd International Conference on Advanced Information Networking and Applications – Workshops 2008

[7] Mohammad Banikazemi et al., "Storage-Based Intrusion Detectionfor Storage Area Networks (SANs)" in Proceedings of the 22nd IEEE / 13th NASA Goddard Conference on Mass Storage Systems and Technologies 2005

[8] Adam G. Pennington et al., "Storage-based intrusion detection: watching storage activity for suspicious behavior" in SSYM'03 Proceedings of the 12th conference on USENIX Security Symposium - Volume 12, Pages 10 – 10, August 4–8, 2003

[9] Yacine Djemaiel et al., "An intrusion tolerant transaction management model for Wireless Storage Area Networks" in Computer and Information Technology (WCCIT), 2013 World Congress, 22-24 June 2013

[10] A. Elmagarmid Y. Leu and N. Boudriga, "Specification and execution of transactions for advanced database applications", Technical report, Purdue University, 1990

[11] Sandeep Abhang et al., "Design issues of 'Vulnerabilities and Suspicious behavior detection system' in Storage Area Network (SAN)" in International Journal of Recent Trends in Engineering, Vol 2, No. 4, November 2009

[12] P.Mahalingam et al., "Enhanced Data Security Framework for Storage Area Networks" in 2009 Second International Conference on Environmental and Computer Science

[13] "KDD Cup Data," http://kdd.ics.uci.edu/databases/ kddcup99/kddcup99.htm

[14] Jiwu Shu et al., "Design and Implementation of an SAN System Based on the Fiber Channel Protocol" in IEEE Transactions On Computers, Vol. 54, No. 4, April 2005

[15] Zhenwei Yu, "An Automatically Tuning Intrusion Detection System" in IEEE Transactions On Systems, Man, And Cybernetics—Part B: Cybernetics, Vol. 37, No. 2, April 2007

[16] Ekgapark Wonghirunsombat et al., "A Centralized Management Framework of Network based Intrusion Detection and Prevention System" in 10th International Joint Conference on Computer Science and Software Engineering (JCSSE), 2013