Automatic Detection Of Insider Threat In E-mail System Using N-gram Technique

Aishwarya Potu[1], Snehal Mane[2], Akshay Kondhalkar[3], Pooja Talathi[4], Prof. Aparna Hambarde[5]

[1,2,3,4]BE Student, Dept. of Computer Engineering, KJCOEMR, Pune, India
[5]Assistant Professor, Dept. of Computer Engineering, KJCOEMR, Pune, India

---

Abstract - *The organizations must implement the privileged account security to limit the damage that is done by insider and to reduce the risk of insider threats which is offered by the insider threat protection. In present years, the traditional cybersecurity safeguards against the insider threat had proved ineffective. Therefore for detecting insider threat an effective approach is necessary. Documents are classified in order to detect and prevent leakage of sensitive data n-gram frequency is used. The proposed system for detection of sensitive data using N-gram technique if the user sending sensitive information through email to outside network, it will get verified first on server-side using SHA, Threshold based system and N-GRAM. If it is found after verification a user is leaking sensitive data then that threat gets blocked.*

***Key Words*: Cyber security, *Insider* threat, Data leakage, N-gram, Threshold frequency, SHA.**

1. INTRODUCTION

A former or current employee, contractor, or business partner who has or had authorized access to the organization's network, system,or data is defined as Insider. Whenever an insider intentionally or unintentionally misuses access to negatively affect the integrity, confidentiality, or availablility of the organization's critical information or systems is defined as Insider threat.

Insiders can constitute a considerable threat to your organization. They may create many obstacles to security measures using their knowledge of and access to your proprietary systems. Devise strategies by CERT Researches help us to detect and prevent insider threats and respond should an insider unintentionally or intentionally cause harm to your critical assets.

Security professionals have to be constantly looking over their shoulder to see what threats and attacks are coming next. They have to not only lookout for external threats, but also inside their organization for malicious contractors who have authorized access, rogue employees, former employees who still have privileged access to business critical systems, and even employees at risk of causing unintentional abuse.

Key Benefits:

- Make sure that only authorized users are able to access powerful privileged accounts.
- Users are prevented from being able to gain unapproved elevated privileges.
- By establishing strict accountability over the use of privileged accounts by tracking who accessed what accounts and what actions were taken.
- By generating a detailed, tamper-proof audit trail of all privileged account and improve forensic analysis.
- By improving forensic analysis and by generating a detailed, tamper-proof audit trail of all privileged account activity
- Detect rapidly and be alerted on anomalous activity that could signal an inside attack in-progress

2. Related Works

Based on the concept of anomaly detection,the systematic approach for the insider threat detection and analysis is presented[1]. Within the organizations, the log data of each user and for each role that is collected helps in construction of tree-structured profile[2]. The Similarity between existing classified documents and regular documents is used to measure the effictiveness of N-Grams[3]. The dataset which contains actual formely sensitive information annotated at paragraph granularity is analysed. Complexity of Big Text Security Classification is cleared by Automated Classification Enabled by Security Similarity (ACESS)[4]. An overview of DBSAFE, a system for protecting data from insider threat is presented[8]. A use case driven mining framework is proposed with an adaptive detection algorithm to identify the potential malicious threats[6].

3 PROPOSED SYSTEM

The main purpose of the system is to detect the suspicious mail sent by the insider threat. Using SOAP (Simple Access Object Protocol) the mail is being sent from client side. When that mail is being received on the server side, JAVA API is used to send email to particular receipt. A login Id and Password is given to every authorized registered client to access to the system. Database is maintained on the server side which contains all the sensitive word list, files and documents. Admin maintains the email logs, manage block list, manage white list and black list. The admin has the authority to block the user if any illegal activities are caught.

3.1 TECHNIQUES

3.1.1 N-gram Technique

The N-Gram technique is used to break each word in the document into smaller character N-grams.

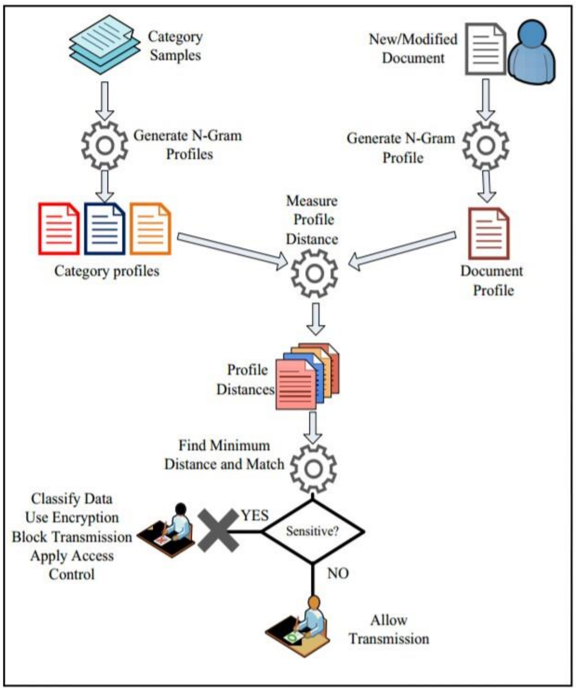


Fig-1: Word N-gram classification process.

The smaller character n-grams are rearranged based on the N-gram frequency to create an N-gram profile. The created N-gram profile are compared with existing category N-gram profiles. The document are classified under the category with the smallest distance measure. Word N-gram classification process This include testing one-word and two-word N-gram sizes and different document sizes.

3.1.2 SHA

United States National Security Agency has designed SHA-1 (Secure Hash Algorithm 1) is a cryptographic hash function . A 20-byte hash value known as a message digest is produced in SHA-1. A SHA-1 hash value is a hexadecimal number i.e 40 digits long. It works by transforming the data using a hash function using an algorithm that consists of bitwise operations, modular additions and compression functions. A fixed size string produced by hash function looks nothing like the original. To be one-way functions the algorithms are designed i.e. once they are transformed into their respective hash values then it is virtually impossible to transform them back into the original data. A common application of SHA is to encrypting passwords, as the server side only needs to keep track of specific users hash value, rather than the actual password. SHA algorithms are designed with increasingly stronger encryption in response to hacker attacks.

3.1.3 Threshold Frequency

Threshold is a value. We associate the Threshold to a statistic. Data is collected for that statistic, then it is compared with the associated Threshold value. If the collected data value does not warrant the Threshold value, then it indicates that this type of data might lead to poor performance of the network or device. Set up a Threshold value along with a level such as the maximum value, the minimum value, and equal value. When the collected value exceeds the threshold value, notification, which we receive is in the form of a Threshold Event. An event is an occurrence of any action. Whenever a threshold value has exceeded, a threshold event is generated. Also, every Threshold event is associated with a Severity to denote how critical the situation is. The total count number of sensitive word frequency is compared with threshold value, the log is maintained and the appropriate action is taken, i.e. in email system ,to block the mail or send it.

3.1.4 Jaccard Index

The Jaccard similarity index (sometimes called the Jaccard similarity coefficient) compares members for two sets to see which members are shared and which are distinct. It's a measure of similarity for the two sets of data, with a range from 0% to 100%. The higher the percentage, the more similar the two populations. Although it's easy to interpret, it is extremely sensitive to small samples sizes and may give erroneous results, especially with very small samples or data sets with missing observations.

The formula to find the Index is: Jaccard Index = (the number in both sets) / (the number in either set) * 100

3.2 Software Technologies

The system uses Java Language for development of the project. To compile Java program ,JDK development kit is used. NetBeans provides an integrated development environment for Java.

3.3 Serialized Database

Java is collecting APIs use data structure like list and vectors, then we declare our own classes using this data structure. Example a class employee to hold all employee information then this class precompiled and called within a Java application as a library. The object of the class converted into bytes so that it can be stored on the hard drive. For this we are using serialization where the object can be read or write to the file.

3.4 Working

The flow of the system is when a client sends a mail, at the server side the mail is examined if any sensitive information is leaked with reference to the database maintained on the server side. Using SHA, n-gram technique and threshold technique the email content is compared with sensitive file stored at server side. If the complete sensitive file is sent as it is then by SHA algorithm it is detected and the mail is blocked. In case if some changes are made in the sensitive file and then sent then by n-gram algo the sensitive information is checked. In mail subject if any sesitive words are detected which are stored at server side and if the threshold is greater then mail gets blocked. If sensitive information caught, the mail will be disapproved and blocked. The insider threat login access is also denied later i.e. the account of the user is deactivated. Only admin has the authority to activate the user's account. If mail does not contain any sensitive information, the mail is approved and sent with the common mail Id of an organization at the server side.

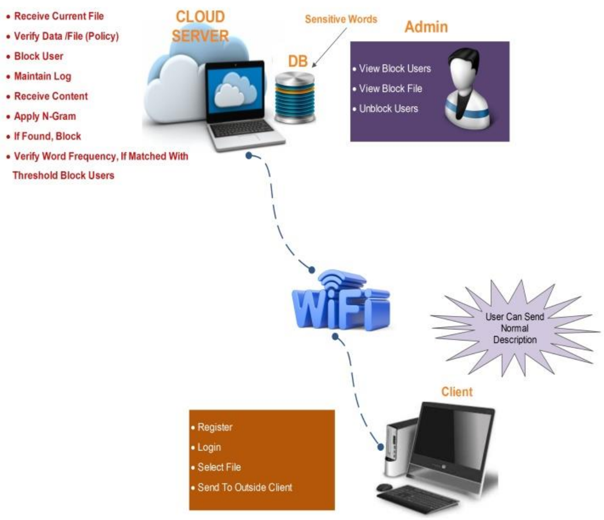
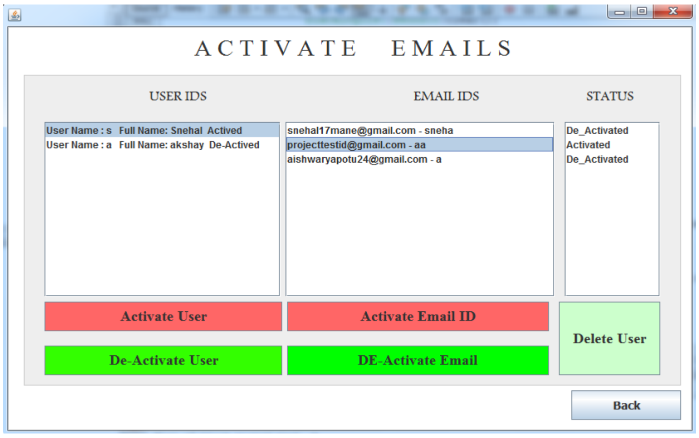


Fig-2: Proposed system scenario

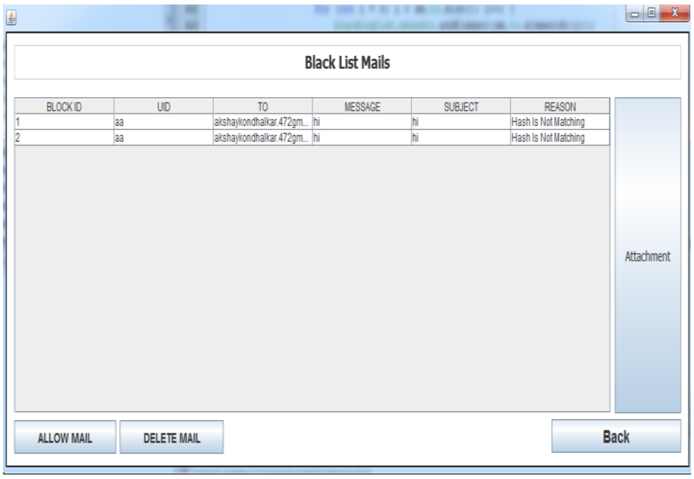
3.5 Implementation

3.5.1 Server Side


ACTIVATE EMAILS

USER IDS	EMAIL IDS	STATUS
User Name : s Full Name: Snehal Actived	snehal17mane@gmail.com - sneha	De_Activated
User Name : a Full Name: akshay De-Actived	projecttestid@gmail.com - aa	Activated
	aishwaryapotu24@gmail.com - a	De_Activated

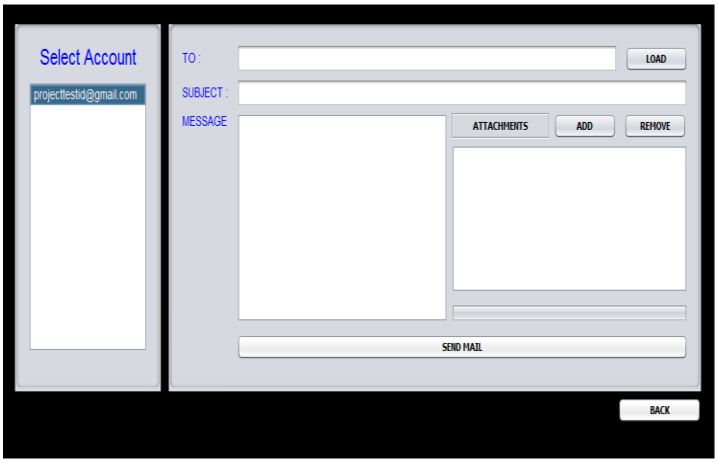
Fig-3: Admin activate email


Black List Mails

BLOCK ID	UID	TO	MESSAGE	SUBJECT	REASON
1	aa	akshaykondhalkar.472gm...	hi	hi	Hash Is Not Matching
2	aa	akshaykondhalkar.472gm...	hi	hi	Hash Is Not Matching

Fig-4: Black mail list

3.5.2 Client Side


Select Account

projecttestid@gmail.com

Fig-5: Mail page

4 CONCLUSION

An application is developed to detect insider threat in Email system. In this system when the user is sending sensitive information to outside networks, it will get verified first on server-side using three techniques SHA, N- GRAM and Threshold based system. The log is maintained by Admin and only admin has the authority to activate the access to the system to all the users. After verification if found that sensitive information is leaking , then insider threat gets blocked automatically.

REFERENCES

[1] "Automated Insider Threat Detection System Using User and Role- Based Profiling Assessment", Philip A. Leg, Oliver Buckley, Michael Goldsmith, and Sadie Creese, 2015

[2] "Caught in the Act of an Insider Attack: Detection and Assessment of Insider Threat ", Philip A. Legg, Oliver Buckley, Michael, Goldsmith and Sadie Creese ,Cyber Security Centre, University of Oxford, UK. , 2015

[3] "Word N-gram Based Classification for Data Leakage Prevention", Sultan Alneyadi, Elankayer Sithirasenan, Vallipuram Muthukkumarasamy, Faculty of Science, Environment, Engineering and Technology, Griffith University Gold Coast Campus, Australia, 2013

[4] "Automated Big Text Security Classification", Khudran Alzhrani, Ethan M. Rudd, Terrance E. Boultand C. Edward Chow, University of Colorado at Colorado Springs Department of Computer Science Vision and Security Technology (VAST) Lab, 2015

[5] "Visualizing the Insider Threat: Challenges and tools for Identifying malicious user activity", Philip A. Legg, 2015

[6] "Mining Software Component Interactions to Detect Security Threats at the Architectural Level" , Eric Yuan And Sam Malek, 2016.

[7] "AD2: Anomaly Detection on Active Directory Log Data for Insider Threat Monitoring", 2016

[8] "A Model-Based Approach to Predicting the Performance of Insider Threat Detection Systems", Shannon C. Roberts, John T.Holodnak, Trang Nguyen, Sophia Yuditskaya, Maja Milosavljevic, William W. Australian MIT Lincoln Laboratory, 2016

[9] "Cyber security for Product Lifecycle Management A Research Roadmap", 2015.

[10] "Dynamic Defense Strategy against Advanced Persistent Threat with Insiders", Pengfei Hu, Hongxing Li Hao Fu, Derya Cansever and Prasant Mohapatra, Department of Computer Science, University of California, Davis, USA

[11] "Modeling Insider Threat Types in Cyber Organizations", Eunice E. Santosa,c, Eugene Santos Jr. b,d, John Korah a,e, Jeremy E. Thompson b, Vairavan Murugappana, Suresh Subramaniana, Yan Zhao B, 2017