Insider Collusion Distinguish Using Intrusion Detection Techniques

MALATHI
M.Phil Research Scholar
Government Arts College (Autonomous)
Salem (DT) – 636 007
Tamilnadu

Ms. M. MALATHI, MCA, M.Phil.,
Assistant professor in Computer science
Government Arts college (Autonomous)
Salem (DT) – 636 007
Tamilnadu

Insider attacks are those originating from a trusted node that had initially passed all the authentication steps to access the network and then got compromised. Insider-related research involving the distribution of kernel-based data mining is limited, resulting in substantial vulnerabilities in designing protection against collaborative organizations. Specifically, if more of our assets are going to reside in the cloud, and as increasingly our lives, enterprises and prosperity may depend upon cloud, it is imperative that we understand the scope for insider attacks so that we might best prepare defenses. we are going to use String Matching algorithm. A substring is a sequence of consecutive contiguous elements of a string, we will denote the substring starting at i and ending at j of string. If the string has same value then only user can send and receive message .It check all types of string for authorized. A prefix of a string S is a substring that starts at position 0, and a suffix a substring that ends at |S|-1. A proper prefix of a S is a prefix that is different to S. Similarly, a proper suffix of S is a suffix that is different to S. The + operator will represent string concatenation.

Key Words:Mail submission agent(MSA), Mail user agent (MUA), Simple Mail Transfer Protocol(SMTP), Message-Digest(MD),Secure Hash Algorithm 1(SHA-1),Type-based Multiple Access(TBMA), Collective spatial keyword query (CoSKQ).

1. INTRODUCTION

Data mining the extraction of hidden predictive information from large databases, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future

trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems. Data mining tools can answer business questions that traditionally were too time consuming to resolve. They scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations.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. In this exisiting system used Insider-related research involving the distribution of kernel-based data mining is limited, resulting in substantial vulnerabilities in designing protection against collaborative organizations.

Homomorphism encryption algorithm, Prior works often fall short by addressing a multi factorial model that is more limited in scope and

implementation than addressing insiders within an organization colluding with outsiders. Such a pragmatic model considers the insider as the key player in sharing data with an attacker, who can then recover the original data from the intermediary kernel values of the SVM model. This attack is more realistic because the attacker needs only to obtain a few data entries rather than the entire database from an organization to successfully recover the rest of the private data. In our proposed system used string matching for detect the error.

2. RELATED WORK

Efficient data modeling and querying system for multi-dimensional spatial data suggested by **W. Li and C. X. Chen** using 3D dimensions to handle a set of attributes in multidimensional spatial data [1]. These techniques do not manipulate complex spatial objects. It provides only clustering, data modeling and their relationships.

Hybrid index structures for location-based web search are used to find the web content. **Y.zhou et.al** integrated both inverted files and R*-trees, to handle the textual and location aware queries.The main advantage is user can aware of the location [2]. The problem is that search engines only return the most relevant pages to users, so for those unpopular locations, the search results may contain very few correct results.

Keyword search on spatial databases describes about the keyword present on spatial database**. I.De Felipe et.al** answered a query for keyword search in most relavant and efficient search in top k keyword queries [3]. A drawback of the IR2-Tree described above is that the same signature length is used for all levels which leads to more false positives in the higher levels, which have more 1's.

Progressive computation of the mindist optimal location query by **D. Zhang et.al** suggests an idea to solve the mindist problem of optimal query location in spatial databases [4]. It is a challenging task for candidates to find an exact answer for particular theorem or concepts The main disadvantages is it doesn't suitable for practical problems.

Another related queries involve in Efficient Retrieval of the Topk Most Relevant Spatial Web Objects is a new kind of top-k query that used both location proximity and text relevancy. **G.Cong et.al** developed the framework for inverted file for both text retrieval and the R-tree for spatial proximityquerying [5]. As a result it achieves scalable and excellent performance.

Collective spatial keyword query (CoSKQ) which is to find a set of objects in the database such that it covers a set of given keywords collectively and has the smallest cost.

Cheng Long et.aldeveloped the maximum sum cost and diameter cost [6]. Result-faster and improves the factors rate. Unfortunately, existing exact algorithms have severe scalability problems and existing approximate algorithms, though scalable, cannot guarantee near-to-optimal problem and address the above issues.

Querying Spatial Patterns-Spatial data are commonly used in many scientific and commercial domains such as geographical information systems and gene/protein expression profiles.

V. Singh, A. Bhattacharya, and A. K. Singhgiveasolution to this problem of querying significant sub regions on spatial data provided as raster images [7]. This paper design a scoring scheme to measure the similarity of subregions.

Chao-Chin Chou et al (2007) says to propose a peer to peer communication protocol in this process proposed without user query to analyze multiple path with multiple query based on single based query search [8]. It is used to avoid the passive attacks. In this process used probabilistic algorithm to avoid the flooding attacks in the data transmission process. The optimal path between the two nodes decides the hop count information. A **middleware protocol** is design by the MAPCP protocol. This middleware protocol lies between the application and network protocol. In this approach avoid the flooding attack between the peer to peer networks.

Security vulnerability issues in wireless sensor networks: a short survey uses algorithm such as shared keys, encryption, secure data aggregation, spins: security protocols for sensor networks, tiny sec, defending against dos attacks, defending against attacks on routing protocols[10]. The disadvantages in this paper are wireless sensor networks often operate in a resource constrained environment. Optimal resource utilization is main objective of wsn. but wireless sensor networks are equally vulnerable to security attacks. Ensuring security in a hostile

operational environment of wsn is a hurricane task. However, advantages in this paper are to provide comprehensive information on types of attacks wireless sensor network is exposed to and possible methods of countering such attacks effectively. The motto here is to help novice researchers with objective to work on security challenges in wireless sensor network environment.

In detecting and counteracting statistical attacks in cooperative spectrum sensing uses algorithm of novel bayesian method, joint estimation approach[12]. The disadvantages are cooperative spectrum sensing is vulnerable to misbehaving nodes, which might report wrong sensing data, either because of malfunctioning radio, or intentionally as malicious users. The problem of trustworthiness in cooperative spectrum sensing has been addressed in some recent works. For example, the method proposed in assigns a "reputation value" to secondary users. The reputation of each secondary user is updated at every time slot, based on the level of agreement of the considered secondary users and the majority of secondary users[13]. This method relies on the existence of a number of trusted secondary users. If the reputation of a certain user is below a threshold, its sensing report is ignored. However, advantage is joint estimation approach outperforms traditional cooperation schemes based on exclusion of the unreliable nodes from the spectrum sensing process, and that it nearly achieves the performance of an ideal maximum likelihood estimation if attack probabilities remain constant over a sufficient number of sensing time slots. The proposed method improves the robustness of cooperative spectrum sensing against misbehaving secondary users, which may send wrong sensing reports in order to artificially increase or reduce the throughput of a cognitive network.

In decentralized hypothesis testing in wireless sensor networks in the presence of misbehaving nodes, expectation maximization algorithm used. The disadvantages in cooperative spectrum sensing may also have more than one class of unreliable nodes. While some malicious users may send false data in order to gain unfair access to the channel, others may be sending incorrect data due to the malfunctioning of their sensing terminal. It also point out that while a collaborative crn may consist of at most tens of radios, a sensor network may comprise of hundreds or thousands of nodes. Advantages of proposed algorithm significantly outperform the reputation-based methods in classification of the nodes as well as the detection of the hypotheses[14]. The estimated operating points are compared to the cramer- rao lower bound which shows the efficacy of the proposed method.

Optimal distributed detection strategies for wireless sensor networks used type-based distributed detection algorithm. The disadvantage of hard-decision fusion incurs a loss in error exponent. Detection performance degrades under the more stringent tpc. Histogram fusion and soft-decision fusion no longer achieve the optimal centralized exponent. The advantages of existing system, the computation of likelihood ratios are shifted from the nodes to the decision center and the nodes can be completely oblivious of the observation statistics.

The optimality of histogram fusion, coupled with the use of dumb sensors, makes type based decentralized detection a very promising strategy for large-scale, energy- and cost constrained wireless sensor networks. furthermore, in view of the zero-rate property of types, under relaxed latency requirements, type-based detection can be performed as a background process, consuming a vanishing amount of network bandwidth and secure type-based multiple access uses novel **(type-based multiple-access)TBMA** protocol called secure TBMA.

3. SYSTEM MODEL

3.1. String Matching

String Matching have a high detection rate by which you can easily detect a hacker.

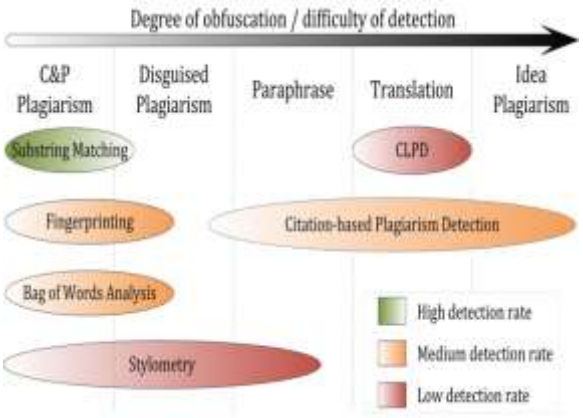

Fig: 3.1.1 String Matching

A string is a sequence of characters. In our model we are going to represent a string as a 0-indexed array. So a string S =”Galois” is indeed an array [‘G’, ’a’, ’l’, ’o’, ’i’, ’s’]. The number of characters of a string is called its length and is denoted by |S|. If we want to reference the character of the string at position i, we will use S[i].

A substring is a sequence of consecutive contiguous elements of a string, we will denote the substring starting at i and ending at j of string S by S [i...j].

A prefix of a string S is a substring that starts at position 0, and a suffix a substring that ends at |S|-1. A proper prefix of a S is a prefix that is different to S. Similarly, a proper suffix of S is a suffix that is different to S. The + operator will represent string concatenation.

ADVANTAGES

- It is Efficient compared to previous Method.
- Propose privacy preserving schemes to counter the attacks.

4. METHODOLOGY

4.1 METHODOLOGY DESCRIPTIONS:

- Authority Key Identification
- Message Transfer
- Reducing the number of the insiders
- Terminate Intruder

4.1.1 Authority Key Identification

Using String Matching, we match the particular id with password then only message transfer to other end for particular user. Other end user, login and then only they can study the private message from user identification. Every user has a separate random key.If that intruder not have that separate key, then that user unable to view message and send that message. Using MD5 we can terminate intruder without having key Value that intruder can’t view or send message Data.

4.1.2 Message Transfer

A message transfer agent receives mail from either another MTA, a mail submission agent (MSA), or a mail user agent (MUA). The transmission details are specified by the Simple Mail Transfer Protocol (SMTP). When a recipient mailbox of a message is not hosted locally, the message is relayed, that is, forwarded to another MTA. Every time an MTA receives an email message, it adds a received trace header field to the top of the header of the message thereby building a sequential record of MTAs handling the message. The process of choosing a target MTA for the next hop is also described in SMTP, but can usually be overridden by configuring the MTA software with specific routes.

4.1.3 Reducing the number of the insiders

In information security, intruder detection is the art of detecting intruders behind attacks as unique persons. This technique tries to identify the person behind an attack by analyzing their

computational behavior. This concept is sometimes confused with Intrusion Detection techniques which are the art of detecting intruder actions.

4.1.4 Terminate Intruder

By Using **MD-5(Message-Digest)**algorithm Receiver gets the message and extracts the encrypted message digest . Then he computes his own message

digest of the received message . He also decodes received message digest with sender's public key and gets decoded message digest . Then he compares both message digests When both message digests are equal, the message was not modified during the data transmission.

SHA-1 (Secure Hash Algorithm 1) SHA-1 forms part of several widely used security applications and protocols, including use within other cryptographic algorithms and protocols, for the protection of sensitive unclassified information. Nobody has been able to break SHA-1, but the point is the SHA-1, as far as Git is concerned, isn't even a security feature. It's purely a consistency check. The security parts are elsewhere, so a lot of people assume that since Git uses SHA-1 and SHA-1 is used for cryptographically secure stuff, they think that, Okay, it's a huge security feature. It has nothing at all to do with security

4.2 ALGORITHM DESCRIPTIONS

4.2.1 Cryptography techniques:

The proposed approach used a point of multiplication and it is difficult to compare the logarithmic problem. For security purpose signature generation and signature verification is used in this process. The different types of components are used in this process. The private key, public key, set of operations and domain parameters. Private Key indicates the random numbers, Public keys indicate point on curve, set of operations denotes mathematical curves and the domain parameters denote constant variables.

Proposed system used a secure hash algorithm for mathematical calculation and it will give more authentications compare to RSA and blowfish algorithm because the private key and pubic key is more secure. The sender and receiver initialize the private and public key the curved path is also analyzed in this approach.

To protect a 128 bit AES key, the RSA Key Size takes 3072 bits but this takes 256 bits key size. Each and every user has a public and private key the public key is used for encryption/signature verification and the private key is used for decryption/signature generation.

These proposed scheme consists of the following 3 phases

1. Key pair generation
2. Signature generation
3. Signature verification

5. CONCLUSION

In this paper proposed the solutions to the problem of keyword search in multi-dimensional data and a methodology called search based on random projections and hashing.
Based on this index developed search that finds
a set of points that searches results with better efficiency. Our results show that search is faster than state-of-the-art tree-based techniques, with performance improvement.

6. FUTUREWORK

There are several avenues for future research. First will investigate other selection criteria for cluster merging in our proposed hierarchical co-clustering framework. Second will study methods for improving the layer-wise optimization scheme used in our hierarchical technicals are used. Third will extend our proposed framework to cluster more than two types of data.

REFERENCES

[1] W. Li and C. X. Chen, “Efficient data modeling and querying system for multi-dimensional spatial data,” in Proc. 16th ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst., 2008, pp. 58:1–58:4.

[2] Y. Zhou, X. Xie, C. Wang, Y. Gong, and W.-Y.Ma, “Hybrid index structures for location-based web search,” in Proc. 14th ACM Int. Conf. Inf. Knowl.Manage., 2005, pp. 155–162.

[3]. I. De Felipe, V. Hristidis, and N. Rishe, “Keyword search on spatial databases,” in Proc. IEEE 24th Int. Conf. Data Eng., 2008, pp. 656–665

[4] D. Zhang, Y. Du, T. Xia, and Y. Tao, “Progressive computation of the min-dist optimal-location query,” in Proc. 32nd Int. Conf. Very Large Databases, 2006, pp. 643–654.

[5] G. Cong, C. S. Jensen, and D. Wu, "Efficient retrieval of the top-k most relevant spatial web objects," Proc. VLDB Endowment, vol. 2,pp. 337–348, 2009

[6] C. Long, R. C.-W. Wong, K. Wang, and A. W.-C. Fu, "Collective spatial keyword queries: A distance owner-driven approach," in Proc. ACM SIGMOD Int. Conf. Manage. Data, 2013, pp. 689–700

[7] S. Jiang, Y. Liu, Y. Jiang, and Q. Yin, "Provisioning of Adaptability to Variable Topologies for Routing Schemes in MANETs," IEEE J. Selected Areas in Comm., vol. 22, no. 7,pp. 1347-1356, Sept. 2004.

[8] G. Chakrabarti and S. Kulkarni, "Load Balancing and in Mobile Ad Hoc Networks," Ad Hoc Networks, vol. 4,pp. 186-203, 2006.

[9] Z. Shen and J.P. Thomas, "Security and QoS Self-Optimization inMobile Ad Hoc Networks," IEEE Trans. Mobile Computing, vol. 7,pp. 1138-1151, Sept. 2008.

[10] S. Ibrahim, K. Sadek, W. Su, and R. Liu, "Cooperative Communications with Relay-Selection: When to Cooperate and Whom to Cooperate With?" IEEE Trans. Wireless Comm., vol. 7, no. 7,pp. 2814-2827, July 2008.

[11] A. Bletsas, A. Khisti, D.P. Reed, and A. Lippman, "A Simple Cooperative Diversity Method Based on Network Path Selection," IEEE J. Selected Areas in Comm., vol. 24, no. 3, pp. 659-672, Mar.2006.

[12] J. Cai, X. Shen, J.W. Mark, and A.S. Alfa, "Semi-Distributed User Relaying Algorithm for Amplify-and-Forward Wireless Relay Networks," IEEE Trans. Wireless Comm., vol. 7, no. 4, pp. 1348-1357, Apr. 2008.

[13] L. Feeney, B. Cetin, D. Hollos, M. Kubisch, S. Mengesha, and H.Karl, "Multi-Rate Relaying for Performance Improvement inIEEE 802.11 WLANS," Proc. Fifth Int'l Conf. Wired/Wireless Internet Comm., 2007.

[14] D. Lin and R. Morris, "Dynamics of Random Early Detection, "Proc. ACM Special Interest Group Data Comm. (SIGCOMM),1997