INTRUSION DETECTION SYSTEM IN A HYBRID NETWORK USING DATA MINING-RESISTANCE SYSTEM

HASINA A. RAZZAK. A. KARIM1, SUPRIYA P. PANDA2 & M. V. RAMANA MURTHY3

1Research Scholar, Department of Computer Science & Engineering-Manav Rachna International Institute for Research and Studies, Faridabad, India
2Professor, Department of Computer Science and Engineering, Manav Rachna International Institute for Research and Studies, Faridabad, India
3Professor, Department of Math & Computer Science, Osmania University, Hyderabad, India

ABSTRACT

Data mining methods can be very useful to add to the Intrusion Detection System (IDS) to identify common and suspicious activity patterns and help managers protect sensitive data and all information contained in businesses or organizations. This paper attempts to analyze the types of network attacks and the existing IDS drifts based on Data Mining techniques. In recent days, the growth of science, information technology, and communications, in particular, has resulted in a rise in the quantity of digital data. These enormous quantities of information are no longer able to handle standard (statistical) techniques. Data Mining (DM) has emerged and has proved to be one of the most successful alternatives for analyzing large quantities of data. Data Mining has been developed with the assistance of many fields, including Statistics, Databases, and Artificial Intelligence. This paper focuses on tracking IDS in a hybrid network using the Genetic Algorithm (GA), by applying essential functions of IDS technologies, standard detection methodologies as signature-based and anomaly-based methods of detection.

KEYWORDS: Data Mining, Intrusion Detection System, Hybrid Intrusion Detection System, Peer to Peer Network (P2P), Adaptive Consistency Maintenance Algorithm (ACMA), Top Caching (TC) algorithm & Genetic Algorithm (GA)

INTRODUCTION

As network system security is becoming increasingly important as more and more sensitive information is being stored and exploited online. Intrusion Detection Systems (IDS) has thus become a critical technology to improve the safety of these networks. Most IDS are signature-based.

Such systems match the activity being tracked on the device to known attack signatures.

The problem with this approach is that these IDSs fail to recognize to detect new attacks or attacks with unknown signatures. There is a lot of effort and new researches to develop new algorithms in data mining based approaches to discover detection models for IDS.

These models generalize to detect unknown attacks from both known attacks and normal behaviour. In a smarter and more automated system the application of good IDS can also be generated. IDS enhancing detection performance and generalization capability of DM-based IDS are some of the system implementation challenges.
The system should work in real-time, as intrusions take place, for an effective IDS implementation to minimize the false alarms, focus on the detection of intruders, useless inputs, and faster and more accurate detection results with high accuracy. Proposed approach of Relating-IDS is comparing with the Accuracy, Efficiency, and Usability (AEU) using the Genetic Algorithm (GA) with the detecting system (A.Kumar, H. Chandra Maurya, and R. Misra 2013).

![Figure 1: (GA) General Processing (Parry Gower Majeed and Santosh Kumar, 2014)](image)

GA as shown in Fig. 1, is a group of computational models depends on some principles of evolution and natural selection (Parry Gower Majeed and Santosh Kumar, 2014). These algorithms change the problem in a specific domain into a model by using a “chromosome-like data structure and evolve the chromosomes using selection, recombination, and mutation operators”. In the field of computer network security, it is mainly used for finding optimal solutions to a specific problem. So IDS are designed to detect system attacks and classifies system activities into normal and abnormal subjects. DM-based IDS model generalizes and detects both known and unknown attacks. The main issues that come in front of IDS are Accuracy, Efficiency and Usability (AEU). By applying robust system, the IDS can be used to detect all sorts of new attacks, which can be achieved using machine learning techniques with high efficiency and reducing the false alarm rate. Here, the proposed system compares the accuracy between two resistance systems to examine their False-Accuracy and Predictive Accuracy. Ultimately, the evaluation indicates that the Relating-IDS method in a distributed system is a good intrusion detection and forecast technique.

**Importance of Data Mining in Network Security**

Today, the world is witnessing continuous development in all areas of life, and the Internet has matured an essential element of daily lives where it is indispensable. To keep pace with these developments, computers, and networks are must, and data available in these networks throw an insight into modernity in science and the world. These data and information should be provided with complete protection against intruders and hackers. Data protection can be defined as a process involving the collection of data and trading it correctly and free from vandalism. Usually, all the personal and business data that is shared on the computer networks every day are vulnerable to attack and infiltration, so security and protection of this data and information are some of the most critical aspects of communication. There is no single specific way to protect networks against intruders fully. Network Technologies are improving in network security (NS) and evolving as attack and defence techniques progress (Monali Shetty and N.M. Shekokar, 2012).

Therefore, data mining is one of the fastest-growing fields in computer science majors, and its popularity has come from the growing need for tools to help in analyzing and understanding the vast amounts of data which are produced daily by various on a network. Tools and software have emerged with data mining to assist in exploration huge and increasing amount of data to access knowledge in databases. It can be said that Data Mining is the Discovery of Knowledge is the process of analyzing data from different perspectives and extracting relationships between them and summarizing them to useful information by using a set of sophisticated tools. Some of these tools include regular statistical ...
There are two types of knowledge in light of this huge amount of data: the first is Online analytical processing (OLAP), and the second is data mining (DM), i.e., mining, both types of data warehouse analytical tools. However, OLAP data appeared before DM, which depends on a multi-dimensional presentation emphasizing the efficiency and speed of satisfying the desires of the receivers (Eltabakh, M., 2010). Fig. 2 shows the difference between OLAP and prospective data in data warehouses.

The Data Mining basic elements are Data, Information, Knowledge, and Data warehouses, where

- Data are facts, figures, and text that can be processed by the computer.
- Information, that are models and relationships between those data that constitute useful information.
- Knowledge, inferring Previous information can be converted into knowledge about patterns Historical or future outlook.
- Data warehouses that are used in time analysis, knowledge discovery, and decision-making.

It is specifically designed for data extraction, processing and representation, which are provided appropriately for these purposes and store a huge amount of data that may be from various sources, e.g., several databases from numerous models (Eltabakh, M., 2010).

Network Security and Intrusion Detection System

Network security (NS) is a collection of different activities, such as the policies, constraints, and user practices adopted to stop the intruders and data bandits. NS prevents from carrying out exploits and threats by illegal access into the network and hence the misuse of any malicious modification in data or Denial of Service (DoS). The NS should be for both data as well as for hardware and must be controlled by officially authorized administration. It was essential to put an end to these risks and to have tools and appropriate, effective techniques to prevent these hackers and malicious programs from entering the computers of users and from catching up with the networks. Several technologies and concepts have emerged to combat these risks. Because threats are becoming increasingly advanced, and it also implies that safety technologies need to be smarter, users became more focused on safeguarding and protecting information and data from the physical protection of operations to prevent damage or theft to the hardware, to prevent damage or theft to the data processed and to prevent service interruptions (A.Kumar, H. Chandra Maurya, and R. Misra 2013).
Secured Network Properties

The main purpose of IDS is to give alerts to administrators whenever there is an attempt to penetrate or infiltrate the network system by intruders and hackers. The type of signature on comparison with the malicious signature table defines the misuse of the Network Intrusion Detection System (NIDS). If a continuous activity matches a signature in that table, an alert will be generated to detect the occurrence of the intrusions and then respond to prevent and stop them as appropriate.

Secrecy: The message must be sent in the form of encryption so that no one can read the message until and unless it is encoded in such a way that the only person desired can understand and read the content of the message received.

Authentication

It can be described as “The method in which the certificate provided is estimated to those on file in a database of certified users’ information on a local operating system or within an authentication server”. If the credentials are equivalent, the process is completed, and the user is granted permission for access.

Message-Integrity

In encryption, it is called Sent or transmitted message validity, the one-way hash function is the most common method used, where the whole message contents are converted with secret symbols and creates a new message that it is difficult to change or reverse, with the guarantee that the message contents are not manipulated or altered (B. A. Forouzan, 2013).

Network Threat: Denial-of-Service (DoS)

This is a well-known attack on internet security. In B. A. Forouzan’s view, aim is to slow down the server by sending a lot of fake texts to get the system into a busy mode and not react to the client’s server-crash or delete the response message from the server. This leads to delay service and traffic crowding in the response from the sites causing demanding user’s access due to high congestion and forces the client to think that the server out of reach, or to interrupt the whole system service, it creates a heavy load on the central server (Simon Edwards 2012).

Illegal Access

The popular term for this is hacking. It is defined as when “The system proprietor unintentionally gains entry by a customer who does not have the permission to connect to or use a system. It is gaining or securing access into any computer, storage medium, network, program, system, file, client area, or other private/private repository, without the prior permission of the owner”. Here are some categories of network attacks in Fig.3:

Figure 3: categories of network attacks(Manu Bijone, 2016)

Hybrid Peer-to-Peer (P2P)

This pattern aims to present an intrusion detection module based on Hybrid Peer-to-Peer network system this is a pattern of Network, where all the computers are connected in a system so that they communicate via ordinary intermediate nodes in a
network for sharing the files, data and other information. Moreover, each node acts and operates as a client and a server rather than one centralization server system, where connected together in such a way that all the nodes are at equal responsibility and can share files and provide data at any request to other nodes, and also can ask for any information from the others, and not contain any device that is treated as the Main Server. It means each device performs both the client and server simultaneously. Applying Adaptive Consistency Maintenance Algorithm (ACMA) maintains time-to-time polls of the owner of the file, which keeps updating the file because of the lowest number of replica consistency. Moreover, Top Caching (TC) algorithm is used to help improve the implementation, cache configuration, and fully distribute the most common information (Kalaivaniet al. 2012).

**Proposed Relating Model**

**Definition and Formulation**

Networks are vulnerable to intrusion, and hackers have become a major problem on the Internet. Many intrusion detection systems (IDS) have been proposed to detect these breaches and to prevent steal, and manipulate information. The aim is to provide a better mechanism to improve data transmission security in a hybrid network. Hybrid Intrusion Detection and Prediction (HIDP) provide the ability to detect and predict all security threats. The detection system monitors abnormal traffic patterns, incidents, and then reports suspicious events to administrators, and it is impossible to predict any unknown incidents to prevent them. The process of prediction is a smart procedure that learns to distinguish unidentified threats and adapts the scheme. Fig. 4 portrays the overview of Proposed System for transmission data.

![Figure 4: The Overview of Proposed System for transmission data](image)

**The Architecture of the System**

On the basis of Darwin’s concept of the fittest survival, Genetic Algorithms (GA) attempt to arrive at optimal and appropriate solutions to complex issues, GA attempts to preserve the qualities of the parents’ generation and transmitting them to the generation of children to obtain a strong generation.

“The Proposed system is based on Resistant System (RS), which is a biological pattern or model that applies to many difficulties in security and protection problems in networks. In GA, the population is modified frequently in the search area along with genetics operators and looking for the best fitness response. In the GA algorithm, a group of random individuals is created of virtual values, and over subsequent productions, the population evolves towards the most favorable response. On the other hand, an extended variant of GA with various population sets is the Connecting Presage (CP) algorithm”. An Intrusion Detection Prevention architecture has been proposed for relating-IDP as shown in Fig. 5 below:
This structure designs a detector that connects incidents for a group of common features. The purpose is to get a group/set of "detectors/sensors" that can track the entire traffic and pickup/detect the unknown incidents and predict all the strange occurrences. So, practically every group concentrates only on a specific category of events such as DoS, R2L. The detection probability and the rate of false-accuracy in the suggested work are compared with many known processes where those comparisons emphasize the merits of Relating IDS systems to detect intrusion.

System Parameters and Evaluation

This system would create a detector, that joins a group of events with common features, of “64- bits string includes real traffic or existing incidents as follows”:

\[
T_a = (h_1, h_2, h_3, \ldots, h_n)
\]

\[
T_b = (p_1, p_2, p_3, \ldots, p_n)
\]

Traffic = \( T_a + T_b \)

Where, \( T_a \): describes a set of incidents (\( h_i \)) along with members (\( a \)). It is level of threshold,

\( T_b \): describes a set of normal traffic (\( p_i \)) along with members (\( b \)).

\[
D = [D_i]_{\alpha \in n} = \sum_{i=1}^{n} d_{i,j} \quad \alpha \in n
\]

\( (D_i) \) shows a group of detectors: (\( d_{i,j} \)) is a detector/sensors. And (\( n \)): the greatest number of detectors.

The value of the Match Strength Factor (\( S \)) should be found to select the members with the best fitness, (\( S \)) can be obtained:

\[
S(x, y) = \sum_{j=1}^{1} \begin{cases} 0 & \text{if } z_j \neq y_j, \alpha \neq x \\ \frac{\theta g(z_j, y_j)}{\theta e} & \text{if } \theta e 
\end{cases}
\]

“The Probability of Detection Ratio (PDR) and False Accuracy Ratio (FAR) have been introduced for evaluation of the proposed system, as the two evaluation parameters. The probability of detection shows the successful detection of a real incident where the probability of false accuracy ratio refers to the probability of selecting regular traffic rather than an incident or vice-a-versa”.

Figure 5: Intrusion Detection Prevention architecture
The system evaluated the ability for detection/prediction of RS and proposed Relating systems in operational mode. A set of dedicated detectors collect the final representative detectors, which are created for training incidents including the “DOS, R2L, U2R, and Probing”. The dedicated sensors identify the training disturbances or incidents in each type, where the representative sensors are ready for identifying/detecting and predicting all the training and testing incidents. As a result, in both assessments show that detectors/sensors related to proposed model are more effective and powerful in both identification and forecast relative to the RS scheme, and is a good intrusion detection and forecast technique. Whatever the Intrusion Detection System is adopted, the attackers can create or develop the new kind of intrusion technique or hacking systems to attack the data during transmission and exchange of information or data. Therefore, the database must update with all the new types of attacks, intrusion, threats, and advanced malware.

Implementation of Hybrid Intrusion Detection System (HIDS):

The implementation of the proposed model “Relating-IDS system” has been done using the MATLAB software platform using JAVA and NS Tool and supporting MATLAB Distributed Computing Engine (MDCE). Also, the Knowledge Discovery in Databases (KDD) records comprises five classes: “Normal, DOS, R2L, U2R, and Probing”. KDD process main objective is to extract information from data in the context of great databases, and this can be done by using different Data Mining algorithms such as GA to identify the suspicious incidents in our case. In the Database (DB), the testing events in the database have a distinct likelihood of allocation from events of practice. To identify the intrusion “Monitoring Tool-Wireshark” (Shilpi Gupta & Roopal Mamtora, 2012), has been used to capturing the real-time network traffic by the help of Snort and show the analyzed details of the captured packet. A network packet analyzer presents captured packet data in as much detail as possible, and it has ability to capture traffic from many different network media types, including Ethernet, Wireless LAN, Bluetooth, USB, and more with full information related to the captured packet. It is the predefined rule that allows observing the network status as shown in Fig.6. (HaroonIqbal, SameenaNaaz, 2019).

![Figure 6: The network status(Shilpi Gupta & Roopal Mamtora, 2012)](image)

CONCLUSIONS

After the comparison between the proposed model and some of the well-known models like “Maximum likelihood Gaussian classifiers, Incremental Radial Basis Function (IRBF), and Hypersphere algorithm (HYP)”, are prepared using a predictive accuracy metric (PAM) to demonstrate the effectiveness of the model, the result shows that the merits of the proposed model are more successful in Detection of suspicious incidents with low False Accuracy Rate (FAR) and high Probability of Detection (PD). By calculating the value of the PD and the FAR in the proposed work system, It is found that:

(PD) of Relating-IDP = 96.35 (in DoS class)

(FAR) of Relating-IDP = 0.04123 (in DoS class).

From above it is clear that by evaluation, the results determine that the proposed Relating System is a prosperous
system for Intrusion Detection and Prediction System (IDPS) in a distributed system. The idea of the applied technical model in the research work confirms that the robustness of the proposed system for the detection of suspicious events with almost less fitness comparing with the other available techniques.

Future Scope

To enhance the security of the information in the network i.e., to continue the improvement in the IDS in the future it may say, that creating a standard test data set with “high fitness and reduce false alarm” for the GA or any other algorithm for IDS and applying it to the environment is very important. To improve the detection, the specification of parameters for the GA should be defined during the tests. Also, full Network Intrusion Coverage must be available; several deployments are leading to incomprehensible intrusion and network vulnerabilities. Another way to enhancing the system ability would be by developing some dataset which represents modern unknown attack and to evaluate the anomaly-based Network Intrusion System.

REFERENCES

1. A. Forouzan, and F. Musharaf (2012), “Computer Networks A Top-down Approach,” Tata McGraw Hill.
2. A. Forouzan (2013), “Data Communication and networking (Global Edition)”, Tata McGraw Hill.
3. A. Kumar, H. C. Maurya, and R. Misra (2013), “A Research Paper on Hybrid Intrusion Detection System”, International Journal of Engineering and Advanced Technology (IJEAT), ISSN: 2249 – 8958, Volume-2, Issue-4, pp. 294-297.
4. BEATRICE, B. ANUJA, E. KIRUBAKARAN, and KLEODEV WINS. “ARTIFICIAL NEURAL NETWORK AND PARTICLE SWARM OPTIMIZATION HYBRID INTELLIGENCE FOR PREDICTING CUTTING FORCE DURING HARD TURNING OF H13 TOOL STEEL WITH MINIMAL CUTTING FLUID APPLICATION.” International Journal of Mechanical and Production Engineering Research and Development (IJMPERD) 8. 4, Aug 2018, 923-932.
5. E.Kalaivani,J.Selva Kumar (2012), “An Efficient Caching Scheme and Consistency Maintenance in Hybrid P2P System”, IRACST – International Journal of Computer Networks and Wireless Communications (IJCNWC), ISSN: 2250-3501, Vol. 2, No. 1, pp. 74-78.
6. Eltabakh, M. (2010). OLAP & Data Mining. Available at: http://web.cs.wpi.edu/~cs561/s12/Lectures/IntegrationOLAP/OLAP and Mining.pdf.
7. KAUR, RAVINDER, PANKAJ SHARMA, and MEENAKSHI SHARMA. “SECURE IMAGE SHARING WITH ENCRYPTION OF BLOWFISH, AES AND RC4 HYBRID ALGORITHM WITH HASH FUNCTION.” International Journal of Computer Science Engineering and Information Technology Research (IJCSEITR) 4. 2, Apr 2014, 143-148.
8. HaroonIqbal,SameenaNaaz, (2019), “Wireshark as a Tool for Detection of Various LAN Attacks “,International Journal of Computer Sciences and Engineering (IJCSE),Vol.-7, Issue-5, May 2019 E-ISSN: 2347-2693,pp. 838- 837.
9. Karen Scarfone and Peter Mell, (2012), “Guide to Intrusion Detection and Prevention Systems (IDPS)Draft”, Special Publication 800-94 Revision 1 (Draft), Recommendations of the National Institute of Standards and Technology.
10. Khan, ASIF ULLAH, and B. H. U. P. E. S. H. Gour. "Stock market trends prediction using neural network based hybrid model." International Journal of Computer Science Engineering and Information Technology Research 3.1 (2013): 11-18.
11. M.Bijone (2016), “A Survey on Secure Network: Intrusion Detection & Prevention Approaches.” American Journal of Information Systems. Vol. 4, No. 3, pp. 69-88.
12. MonaliShetty, Prof. N.M.Shekokar, (2012), “Data Mining Techniques for Real Time Intrusion Detection Systems”, International Journal of Scientific & Engineering Research Volume 3, Issue 4.
14. ISSN 2229-5518.

15. Parry Gowher Majeed and Santosh Kumar, 2014, “Genetic Algorithms in Intrusion Detection Systems: A Survey” International Journal of Innovation and Applied Studies, ISSN 2028-9324 Vol. 5 No. 3, pp. 233-240.

16. Charan, P. I. Y. U. S. H., T. A. H. S. I. N. Usmani, and SYED HASAN Saeed. “A Comprehensive Study of Various on Demand Routing Protocols for MANETs.” International Journal of Electronics and Communication Engineering 4 (2015): 1-12.

17. Shilpi Gupta and Roopal Mamtora, 2012, “Intrusion Detection System Using Wireshark”, International Journal of Advanced Research in Computer Science and Software Engineering Research Paper Available online at: www.ijarcsse.com, Volume 2, Issue 11, ISSN: 2277 128X, pp. 358-363.

18. Sumitha Thankachan and Suchithra (2017), “Data Mining & Warehousing Algorithms and its Application in Medical Science - A Survey”, International Journal of Computer Science and Mobile Computing A Monthly Journal of Computer Science and Information Technology ISSN 2320–088X IMPACT FACTOR: 6.017 IJCSMC, Vol. 6, Issue. 3, pp.160 – 168.
