Global Journal of Computer Science and Technology: H Information & Technology
Volume 17 Issue 4 Version 1.0 Year 2017
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Intrusion Detection System based on Ant Colony System

By H. Fathima & Dr. A.Syed Musthafa
Ksrcas-Periyar University

Abstract- Challenge of designing and building of the current Network Intrusion Detection System not only improves the ability for discriminating the improper internet behaviors, but also considers the plenty of computer resources which will be cost during the analysis of the network packets and behaviors. During the establishment of a fast intrusion detection system, the major purpose of the research is to intensify the data handling capacity when the network management system faces the mass network behavior. In the research, the modules stored in the network packets of the intrusion detection system are analyzed, and then the network flow data by applying the clustering algorithm based on the ant colony system is classified. Finally, a kind of algorithm which can remove repetitive computation is designed so that the flow can accelerated and the velocity can be distinguished. Experimental results show that the proposed fast clustering algorithm can significantly reduce the original computation time while sacrificing or promoting a very small accuracy, and then the computation speed of the intrusion detection system can be accelerated.

Keywords: intrusion detection system, clustering algorithm, ant colony system.

GJCST-H Classification: I.4.6

Strictly as per the compliance and regulations of:

Intrusion Detection System based on Ant Colony System

H. Fathima[α] & Dr. A.Syed Musthafa[σ]

Abstract- Challenge of designing and building of the current Network Intrusion Detection System not only improves the ability for discriminating the improper internet behaviors, but also considers the plenty of computer resources which will be cost during the analysis of the network packets and behaviors. During the establishment of a fast intrusion detection system, the major purpose of the research is to intensify the data handling capacity when the network management system faces the mass network behavior. In the research, the modules stored in the network packets of the intrusion detection system are analyzed, and then the network flow data by applying the clustering algorithm based on the ant colony system is classified. Finally, a kind of algorithm which can remove repetitive computation is designed so that the flow can accelerated and the velocity can be distinguished. Experimental results show that the proposed fast clustering algorithm can significantly reduce the original computation time while sacrificing or promoting a very small accuracy, and then the computation speed of the intrusion detection system can be accelerated.

Keywords: intrusion detection system, clustering algorithm, ant colony system.

I. Introduction

Owing to the network attack and the improper behavior, the loss of the internet work environment is up to 140 million dollars in 2004. Although the loss has descended into 130 million dollars in 2005, these problems still caused great damage to the internet work performance. When the network bandwidth and flow is substantially developing, analyzing the network behavior and finding out the improper/attacking behaviors by the Intrusion Detection System [2-4] have been becoming very difficult and important. From the network safety perspective, the research is discussed, and then finds out the improper network behavior and illegal network attack by analyzing the network flow and user's network behavior. Finally, the improper net is managed and controlled. From the data mining perspective, Intrusion Detection System is discussed. The major challenge of the research is to rapidly and real-time identify the normal and abnormal network behavioral characteristics. The major two goals are the necessary computation resources and the accuracy distinguished by the network behavior. According to the detection method, the system design is classified [4-5] as follows:

a) Misuse Detection System

The research is to collect the known attacking signatures from the network flow, and then establish a database to store these signatures. These signatures in the monitoring network are compared. If those signatures are similar to the attacking signatures or the same network behavior, it can judge that it belongs to an improper network behavior. The major advantage of the system is to improve the high accurate resolution. The system cannot rapidly and effectively find out the new-type intrusion behavior for it cannot real-time update the attack signature style.

b) Anomaly Detection System

The research is usually to find out the normal network behavior style signature and then establish a normal network behavior model. If the flow is different to the normal behavior style collected from the system in the monitoring network, it will be judged as the improper/attack behavior. The research [3-5] shows that the major advantage of the system can fast find the new type intrusion behavior. The disadvantage of the system is that its precision is not so insufficient that much inaccurate warning information can be produced. C. Hybrid Detection System The system is to achieve more accurate intrusion behavior judgment by combining the signatures of the improper detection system with the signatures of the anomaly detection system, or integrating multifarious detection strategies. Many machine learning and data mining technology have been widely applied to the system, and then the accuracy of the system can be improved. Although the system occupies the merits of the above all systems, but its demerits are that it needs to cost many computation resources when it conducts different kinds of the detection procedures and the judging procedures with the use of the machine learning. The major purpose of the research is to reduce the necessary computation time in the intrusion detection system when the network behavior does not affect the judgment of the module's accuracy. It has its real time in the face of the more scaled flow information. Firstly, the research will design a fast algorithm for it can shorten the computation time which is spent in analyzing and judging the modules, and then improve the system's performance. In addition,

Author α: Teacher, Department of Computer Science, K.S.Rangasamy Matriculation & HSS, K.S.R.Kalvi Nagar, Thokkavadi (P.O), Namakkal-637215. e-mail: fathi.fathimahussain@gmail.com

Author σ: Assistant Professor, Department of Information Technology, K.S.R.College of Technology, K.S.R.Kalvi Nagar, Thokkavadi (P.O), Namakkal-637215.

it completes a high-performance intrusion detection system. The system not only tests all developed algorithms, but also is applied in the real network management system. In Chapter two (the related researches), the related researches and design frameworks about the intrusion detection system are introduced, and then the intrusion detection system based on the ant colony optimization is discussed. In Chapter three (fast ant colony clustering algorithm), the conception of the designed fast clustering algorithm and the design details of the method are introduced. In Chapter four, the system design illustrates the network environment applied in the established intrusion detection system and its practical design. In Chapter five, the experimental results introduce the experimental results and the related analysis. In Chapter six, it includes the conclusion and the future research fields.

c) Intrusion Detection System

The present intrusion detection system research not only considers the accuracy judged by the network behavior, but also its extensibility in the face of the mass network flow data. Figure1 designs the framework [2] for an intrusion detection system. The framework is divided into the above several basic modules and an Audit Database, Monitored Entity module is in charge of determining the necessary monitoring network behavior; Audit Collection module is in charge of collecting the data from the host or network, its data is used for Analysis and Detection module; Audit DB is in charge of storing the mass collected data. Analysis and Detection is the core of the system. The intrusion detection system judges whether the network flow is normal or improper behavior in the module with users' designed algorithm. The part of the research, as previously stated, can be divided into improper detection system, anomaly detection system and hybrid system. Configuration Data module is in charge of setting up the corresponding strategies for the normal or improper flow and other related setup program. Reference Data module stores the relevant data related to the intrusion detection, and it can be used as the foundation for Analysis and Detection module makes the judgment. Active/ Processing Data module is in charge of storing the routine data. Alarm module is in charge of the intrusion detection warning information in the output system, and then transmits the related information to Entity Security Authority and Monitored Entity modules. Entity Security Authority module is in charge of authenticating whether the network behavior is normal, and then transmits the information to Monitored Entity modules. A more effective intrusion detection system can make the webmaster manage the anomaly behavior of the network. These applications include: packet retrieval and filter; packet discrimination; signature comparison and other researches [7-15]. Many researches analyzes and discriminates the network flow information with the high-efficient hardware framework for analyzing the signatures of the packet content is very time consuming and expensive. Improving the algorithm performance is another method to improve this kind of systems. For example, According to the style comparison algorithm, it develops into a fast comparison packet signature method. As to the information retrieval technology, data mining technology and heuristic algorithm, the mass data analysis has remarkable progress and achievements in recent years. Partial researches begin to import the above technologies into the intrusion detection system research, and then it can design a more effective intrusion detection system. It improves the judging precision to the packet, as well as intensifies the computing performance.

The intrusion detection system should analyze mass information in the large-scale network environment. The first bottleneck can be happened in the data compilation, supervisory and related modules. Partial network flow can increase the computing resources as well as the storage space without being real-time analyzed. The research [6] can solve this kind of problems by the sampling method. Analysis and Detection module is another link which would affect the resolution and performance of the intrusion detection system for many current methods could not effectively influence mass data. Considering the system's performance and the real-time problems, it uses the machine learning method and trains the classifiers with the existing data. The topic of many researches in recent years is to speed up the network behavior discrimination with the previously established classifiers while it conducts the system. The intrusion detection system based on the machine learning cannot effectively and real-time train the classifiers again for the traditional machine learning method would consume much computation time and need oversize internal memory. The intrusion detection system's achieved effects are subject to a large limitation for it cannot dynamically update the classifiers. B. Ant Colony Optimization applied in the Intrusion Detection System The performance of the traditional regulation judgment or the discrimination method based on the statistics is insufficient in the face of the mass network flow data. It solves these problems with the use of heuristic algorithm [4]. Many achievements have proved its feasibility and effectiveness. Many researches attempt to use it as the judging algorithm of the intrusion detection system for the effect of the heuristic algorithm based on the ant behavior is remarkable. These researches can be divided into two kinds: ant-based clustering algorithm (ACA) [16-17] introduced by Deneubourg and ant colony optimization (ACO) [18-19] introduced by Dorigo.

1. Randomly create initial solutions
2. While the termination criterion is not met

3. For each ant i
4. For each pattern x ∈ X
5. Calculate the distance of x to all the centroids, denoted ij c for j 1, 2....k.
6. Assign x to the cluster the centroid of which is nearest to x.
7. Local update each path and move ant i
8. End
9. Global update each path
10. End
11. Detect the set of patterns R that are static and that are within a predefined radius r to its centroid.
12. Compress the set of patterns R into a single pattern r and remove R: that is X = ∪ X r { } and X X R.
13. End
14. Output result

The basic conception of ACA mainly uses ants randomly move in the specific range. It classifies the objects in terms of the object's picking-up and dropping methods. Unlike ACA, ACO is regarded as a kind of the artificial ant algorithm, but its major conception is derived from the real ant's foraging behavior model. There are three major procedures: establishing the separation procedure, the pheromone updating procedure and the alternative procedure zone to search. In the research [20-21], Tsang and Kwong uses the ant colony clustering algorithm based on ACA to establish the intrusion detection system, it is called ant colony clustering model (ACAM). In their researches, it measures the clustering results by adding the regional entropy and average similarity methods. Tsang and Kwong use the tournament selection method to add ACA's searching range, except for improving the probability computation method in which the ant picks up and drops the data in each node. In the KDD99 data test, the achieved results in ACAM are superior to k-means, SOM, ACA and other algorithms. In the research [22], Ramos and Abraham use the technologies in the intrusion detection system are as follows: linear genetic programming, decision trees and support vector machine. In the research, Ramos and Abraham introduce my modules and links in the intrusion detection system. It illustrates that it can use principal component analysis (PCA) or genetic algorithm to help ACA reduce the number of the data signatures and then speed up its computing time. In the research [23-24], Banerjee develops an intrusion detection mechanism based on ACO algorithm, and then applies it in the sensor network environment; it is called the intrusion detection based on emotional ants (IDEAS). Using the IDEAS to alter ACO mechanism and join into emotional ant and emotion template for exchanging information, and then detect anomaly place to find out the improper behavior in the sensor network. In the research [25], Gao etl uses ACO to integrate the support vector technology and then improve the intrusion detection system's performance.

II. Fast Ant Colony Clustering Algorithm

Conception From the previous research [26-27], we find that most searching procedure in the heuristic algorithm has repetitive computations in the procedure of solving many complex problems for the partial segment solution have achieved to the optimal or final states before the searching procedures finished. The time for the partial solving segments reach to the final result is successively different. The early segment solution of reaching the final result can become the repetitive computation in the subsequent searching procedure. If it can find out and remove these repetitive computations in the searching procedure, it is unnecessary for us to calculate the repetitive contents again. In the situation, the whole searching computation time can be largely reduced. The research deigns a fast ant colony clustering algorithm with the use of the conception, it Is called Fast Ant Colony System; FACS. The algorithm can make the intrusion detection system have the performance in the face of the mass flow data by reducing the necessary calculated quantity and maintaining the solution quality. B. Method Design Figure 2 illustrates that the research judges the clustering algorithm of the modules with the use of the packet. The algorithm is to reduce the computation time on the basis of Ant Colony System and pattern reduction algorithm. Firstly, FACS randomly produces the initial solution, and then initializes pheromone form. Similar to the traditional ACS clustering algorithm, it passes each input data x (network flow data) in the searching procedure of each ant. Data x must be compared with centroid (ij c) in all clusters. It conducts x the clustering affiliation, k represents the species quantity of the flow. We can put these data into their affiliated classifications by calculating the distance from each data to centroid. When each flow data is judged and put in its affiliating clusters by the algorithm, FACS will conduct the regional updating to the passes routes, and then move the ant to the next data node. When the ant goes through all nodes, FACS will conduct all regional updating. Detect and compress proceeds before the finish of each round. Detect procedure in the FACS is in charge of judging the input data, and later it can repeat the computational data. The current used method is used to make x data node in all ants affiliate to the same clustering, and then it judges the x data node as the repetitively computational data node. It removes and compresses the current analyzed datasets. In the successive FACS iterative procedure, the compressed data node will be avoided by the FACS's computation procedure. Many methods can reduce the repetitively computational procedures, such as the distance from x to centroid, the

clustering group in the affiliating points, the regional and the whole regional updating. Therefore, the procedure can reduce much computation time. We firstly establish an intrusion detection system as the developed algorithm which tests the research in the practical part of the system. The following chapter introduces the system's design and practice, and further illustrates the design conception and method used in the packet judging modules. C. Network Environment The network topology framework in the research is as the figure 3 shown, the regional network and the internetwork connect the router or network switch. All input or output network flow will pass through the network equipment (router or network switch). Take the figure as the example, the network switch remains a copy of the input and output network flow to IDS and it offers analysis flow packet information and maintains the network quality. It can collect the necessary testing data flow, analyze the later designed packet and conduct the test and analysis of the packet judging the modules through the method. It can further make the judgment module have more high accuracy with the effective training.

a) System Design and Experiment

Figure 4 is the sketch map of the system module framework. The system developed in the research is based on the Linux exercise system of the openly original code. The retrieval method of the packet is to start up the Promiscuous mode in the network card, and then require the network card to receive all packets from the network exchangers. It makes the packet capturing modules on the basis of the glib and data link layer socket. Therefore, Packet Capturing is used to receive the complete picket contents, dismantle the received, retrieve the external information of the packet and record the flow information. The achievement of the packet external information uses the external Protocol segments in the IP, the total length segments of the packet, the external source port and destination port in the TCP and UDP and other information as the flow classified references. The Traffic Clustering is designed on the basis of FACS algorithm and uses the flow to cluster. When it is abnormal, the system will initiatively send the anomaly notification to the network managers. At present, the anomaly notification module sends the anomaly notification by the E-mail format. The intrusion detection system includes packet capturing, packet encapsulation, packet collection, flow classification and anomaly notification module. The performance of the Traffic Clustering module has a great influence on the intrusion detection system.

In the system performance test experiment, we use the famous testing data KDD99 [28] to test, and use its results to improve the system's performance. The research conducts the numerical attributes regulation in the preprocessing phase and make its values situate in [0:1] range. If the non-numerical attributes are the same in the algorithm operating procedure, it sets its difference as 0; If the non-numerical attributes are different, it sets its difference as 1. KDD99 has 4,898,431 data in total; the normal data is 972,781 which is about 20% of the total data. In the research, we adopts the probability sampling method as shown in the table 1, its probability sampling is from 1% to 5%. It produces five datasets used in the experiments from DS1 to DS5. The related statistics of the datasets is as shown in the table 1. It samples 8,298 data to conduct the experiment in DS1; the normal flow information has 9,618 data. The data number from DS2 to DS5 is gradually increasing. KDD 99 is the most often used as the dataset in - $\beta = 100\% \beta \beta \beta \Phi \Psi \Psi \Delta \times$ (1) Formula 1 improves AS's and ACS's computation time respectively for calculating the FAS and FACS algorithm developed in the research in the part of the performance measurement. $\beta \Phi$ Represents FACS's computation time; $\beta \Psi$ represents ACS's computation time and $\Delta\beta$ represents the improving computational time percentage. Table 2 analyzes experimental data. The retrieval KDD99 data in divided into 10 clusters (k=10). Precision (P) and recall(R) [29] is used for measuring the data judging accuracy conducted by all algorithms. Time represents computation time; $\Delta\beta$ represents the saving computation time percentage. All testing repeats 30 times. AS represents Ant System [19]; ACS represents Ant Colony System [18]; FAS and FACS represent AS, ACS and the proposed speeding strategy. From the experimental results, we can find that the designed speeding strategy can effectively reduce about 37%to 60% computation time on the basis of the ant optimization in the sacrifice of the slight accuracy. Partial experimental results show that FAS and FACS can improve the discriminated accuracy, except for reducing large computation time. For example, ACS's precision is 95.29; recall is 97.44, FACS can improve precision and recall up to 96.94 and 98.41. From the size of data quantity perspective, the accuracy of ACS and FACS will not be descended with the growing of the data quantity. When the data quantity is becoming big, the shortened computation time has increasing trend. It will have the essential assistance to the intrusion detection system in the face of the mass data.

III. Conclusion

The paper speeds up the analysis and the judgment of the modules in the intrusion detection system by the use of the designed fast clustering algorithm. The intrusion detection system can analyze and manage the large network flow behavior by reducing the computation time of the modules. In the practical application, it firstly establishes a prototype of the intrusion detection system in order to test and analyze the related data. In the theoretical design, it makes several algorithms based on the heuristic

computation, and then applies them in the analysis and the judgment of the modules in the intrusion detection system. As to the research on improving the intrusion detection system's performance, we design a pattern reduction algorithm on the basis of the ant optimization which can reduce the repetitive and unnecessary computations. The experimentally simulation results show that the proposed method in the research can effectively reduce the original computation time for analyzing and judging modules on the basis of the ant optimization without losing overdue correct rate. In the future research, we will regard the algorithm developed in the research and the ant optimization applied in the intrusion detection system as the foundation. It further designs and modifies the current algorithm in order to effectively detect the repetitive computation and reduce large computation time.

References References Referencias

1. Z. Jiangtao, H. Hejiao, and W. Xuan, "Resource provision algorithms in cloud computing: a survey," J. Network Comput. Appl. *64*, 23–42 (2016).CrossRefGoogle Scholar.
2. P. M. Vdovin and V. A. Kostenko, "Algorithm for resource allocation in data centers with independent schedulers for different types of resources," J. Comput. Syst. Sci. Int. *53,* 854 (2014).MathSciNetCrossRefMATHGoogle Scholar.
3. I. A. Zotov and V. A. Kostenko, "Resource allocation algorithm in data centers with a unified scheduler for different types of resources," J. Comput. Syst. Sci. Int. *54,* 59 (2015).

Global Journals Inc. (US) Guidelines Handbook 2017

www.GlobalJournals.org