Asian Journal of Information Technology 13 (12): 733-738, 2014
ISSN: 1682-3915

Heuristic Based Hybrid Network Intrusion Detection System: A Novel Approach

N. Priya and S. Vasantha
Coimbatore Institute of Technology, Coimbatore, India

Abstract: Intrusion detection is a vital component in any service hosted environment. Intrusion in a network or system involves malicious activities such as information theft, denial of service attacks, etc. Many Intrusion Detection Systems (IDS) are developed and implemented successfully in the protection layer of the service hosted environment; even then the organizations suffer from data loss, injection attacks, back door attacks, etc. Due to intrusions, hundreds of intranet networks and thousands of websites report attacked each year and the number has not reduced in past corresponding days. Some intrusions/attacks use stealthy evasion techniques to evade the IDS and access the hosted services. An efficient IDS identifies all possible intrusions that are intended to complement the existing security measures and deter any such intrusions from accessing the hosted environment. In this study, several intrusion detection techniques implemented in network environments are discussed and different evasion approaches used by attackers to bypass the Intrusion Detection System are listed. Also, a novel hybrid intrusion detection approach to identify known and unknown attacks is proposed in which the unknown intrusions are classified from known intrusions by a series of emulation technique in a dedicated virtual machine.

Key words: Intrusion Detection System, anomaly, signature, evasion, attacks

INTRODUCTION

Intrusion literally means the illegal activity of entering, seizing or taking possession of network or a system. The intrusion can also be an interruption that disturbs the privacy of any host of a network or the network itself. Such intrusions can be external or internal. The external threats are monitored by firewall which does not have control over the internal intrusions happening within the network. Some of the possible internal intrusions are password cracking, sniffing unencrypted or clear text traffic, etc. to exploit the system vulnerabilities. These prolonged intrusions have led to the emergence of a new system called Intrusion Detection System (IDS) and its techniques are used to detect the possible intrusions in the network.

Intrusion detection allows any organizations to protect their systems from the attacks that flow from the increasing network connectivity and relies on information systems. Given the severity and nature of modern network security threats, the question for security professionals should not be whether to use intrusion detection but which intrusion detection features and capabilities to use.

Hybrid network intrusion detection architecture is proposed in this study by combining two different IDS techniques which pave way for efficient and high performing intrusion detection. Also the fundamental limitations on the accuracy of one type of IDS technique is shown and combined it likely with the on demand heuristic based approach for optimized intrusion detection. A framework is formulated that allows unified analysis and emulation of incoming data stream on a legitimate virtual machine to access the service hosted by the provider.

Dororthy Denning proposed an Intrusion Detection Model (Denning, 1987) for real time expert systems that list out various forms of intrusions ranging from outsider attacks to inside malicious intruders. This model also ensures that the intrusion detection rate is increased and the false alarm rate is reduced. Hu *et al.* (2009) has focused on anomaly intrusion detection using HMM training based on system calls. Though, the false alarm rate is higher the training cost is reduced by 50% without affecting the performance of the anomaly detection. Patcha and Park (2007) has summarized the survey on statistical based, machine learning based and data mining based anomaly detection system. The comprehensive survey on anomaly detection and the recent technological trends in hybrid intrusion detection is not the only fold but open problems and challenges in this area are identified and mentioned in this study. An Anomaly Intrusion Detection System was proposed by Nakkeeran *et al.* (2010) which comprises of detection modules in a layered manner to detect the anomalies. This method proves to reduce the false positive alarms compared to other methods. Ahmed and Traore (2005)

Corresponding Author: N. Priya, Coimbatore Institute of Technology, Coimbatore, India

proposed an Intrusion Detection Model based on biometrics in which the results were more accurate in profile computing than the statistical methods since they were based on biological behaviours of users.

CLASSIFICATION OF INTRUSION DETECTION TECHNIQUES

Intrusion Detection System monitors computer systems and analyses the network traffic for possible hostile attacks originating from outside the organization and for system misuse or attacks originating from inside the organization. An Intrusion Detection System can detect the security breaches that are not found by the boundary devices such as firewalls due to lack of intelligence of internal attacks and help in maintaining proper security and defending against those internal attacks. Depending on the monitoring location, the IDS can be classified in to two types (Ye *et al*., 2002):

- Host based Intrusion Detection System (HIDS)
- Network based Intrusion Detection System (NIDS)

Host based Intrusion Detection System (HIDS): The host based Intrusion Detection System runs on individual host to monitor the activity of the user and the system resources. The Intrusion Detection System running on the host machine monitors the state of the system, various processes running in the host, system critical files, configuration logs and system logs. The accuracy of detecting an attack is greater with fewer false positives is the key attribute of the HIDS since the system uses system logs containing the events that have actually occurred. Also, the speed of execution in detecting the intrusion is more because of the HIDS System's direct interaction with the host system. Based on the system call HMM training (Hu *et al*., 2009) has analysed the host based anomaly intrusion detection. Moreover, the feature that HIDS sensors reside on the host system helps reduction in any additional hardware for the implementation of the system in any host system.

Network based Intrusion Detection System (NIDS): The Network based Intrusion Detection System tries to detect the malicious activity by monitoring inbound and outbound network traffic. The suspicious data packets that have slipped through the firewall can be monitored by the NIDS. Network based IDS are more adaptable to cross platform environments. It uses packet snuffing technique to pull data packets from TCP/IP or similar protocols instead of analysing the information that originates and resides on an integrated system. Some of the main intrusion problems like bandwidth theft, denial of service and unauthorized outside access can be easily captured by the NIDS. With NIDS, the cost of implementation of separate sensors at each host system node is no more likely. Since, NIDS are operating system independent they can be installed on any network where multiple platform middleware systems are connected. Depending on alarm triggering mechanism/techniques, the IDS can be classified in to two types (Escamilla, 1998; Debar *et al*., 1999):

- Signature based Intrusion Detection System
- Anomaly based Intrusion Detection System

Signature based Intrusion Detection System: The signature based IDS monitors the network traffic and checks the incoming patterns to identify the intrusion. This is also known as misuse based Intrusion Detection System. This method is used mainly to detect the known attacks. The signature based IDS maintains a knowledge base which has predefined patterns of attacks known as signatures. As and when the network traffic is analysed, the incoming packets are compared with the signatures in the knowledge base and if any match found an alarm or alert is generated, suggesting that the system is intruded. Hence, it is prerequisite for the knowledge base to contain all possible patterns of signatures in order to signal for the possible intrusions in the network. The signature based IDS has high precision detection rate of known attacks and that a small variation in the pattern of the known attack may go unnoticed.

Anomaly based Intrusion Detection System: The concept behind anomaly based detection is not to learn what is anomalous but to know what a normal behaviour is. The anomaly based IDS detects the intrusions based on the assumption that attack behaviour is different from that of normal behaviour, i.e., it detects the unusual behaviour patterns rather than signatures. The abnormal or unusual behaviour is termed as anomaly.

Normal usage pattern or behaviour is stored as independent profiles in the knowledge base. Upon the arrival of the incoming data, the anomaly based IDS compares the incoming pattern with the pattern present in the knowledge base. When there is a considerable deviation from the normal behaviours of the incoming data an alarm is triggered and the data is flagged anonymous.

Normal behaviour of the user, host, network and applications has to be defined clearly in anomaly knowledge base. The initial time to define the normal behaviour is called the training period and the defined

normal behaviour is called as profile. The justification of using this approach for anomalous detection in intrusion system is described by Qayyum *et al.* (2005). Based on the different approaches it uses the anomaly based IDS can be further classified as follows:

Statistical based approach: Here, the IDS uses statistical information to find the intrusions. Whenever there is a deviation of the incoming pattern from the normal usage pattern (Qayyum *et al.*, 2005) the analyser counter value crosses the threshold values for specific profiles. The advantage of this approach is unknown attacks are easily determined even without the prior knowledge about the attacks. The disadvantage of this approach is it requires complete statistical information, proper threshold value must be finalized and high rate of false positives are generated. Considering the payload length a statistical model (Wang and Stolfo, 2004) is built to find any deviation from the normal activity. Different models used in this approach are as follows (Denning, 1987):

- Markov Model or Marker Model
- Multivariate Model
- Time Series Model
- Statistical Moments or Mean and Standard Deviation Model
- Operational model or threshold metric

Machine learning based approach: Here, the IDS are trained on the system under investigation over a period of time. Based on the machine learning techniques, the newly gathered information helps in changing the execution strategy (Patcha and Park, 2007). This approach is still at evolving phase and can be used in combination with the statistical based approach to yield better accuracy in results. The models which can be used in this approach are as follows (Cannady and Harrell, 1996; Khan *et al.*, 2007):

- Bayesian networks
- Genetic algorithm
- Fuzzy logic
- Neural network
- Support vector machines

Data mining based approach: This approach is used to inspect a stream of data. The analysed data stream can be used extract the features of attacks, identify the category to which the attack belongs, classify them and group the attacks. The rules are applied to detect the intrusions in the system. The execution time in this approach increases exponentially as the number of entries in the knowledge base increases. The methods used in this approach are as follows (Pei *et al.*, 2004; Khan *et al.*, 2007; Chandola *et al.*, 2009; Lee and Stolfo, 1998):

- Clustering
- Association rule discovery
- Classification

Knowledge based approach: This approach is used for both signature based IDS and anomaly based IDS. Here, the threat is identified with the information from the stored knowledge data base. The accuracy of detecting the intrusions using this approach is greater compared to other approaches. The knowledge base of the system will be updated frequently in order to effectively use this approach (Debar *et al.*, 1999). The main advantage is that this approach possesses a very low false positive alarm rate and that this approach is robust and scalable. In the techniques mentioned below either the attack with set of goals and transitions are described (Porras and Valdes, 1998) or the system containing a set of rules and fact that describes the attack (Lunt and Jagannathan, 1988). These techniques have a list of actions that have to occur for the completion of an intrusion:

- State transition analysis
- Expert systems
- Signature analysis
- Petri nets

LIMITATION OF THE EXISTING APPROACHES

IDS evasion techniques are developed by hackers on day to day basis and hence many IDS techniques are prone to evade of specific intrusions under various circumstances. For example, a new breed of stealthy intrusion data may go unnoticed under signature/anomaly based unique IDS techniques. Some of the IDS evasion techniques are discussed below.

Path obfuscation: This technique involves replacing part of the URL or HTTP sequence with generic sequences so that the data has different appearance but same meaning. Some of the generic sequences used are backslashes (\), single dot sequences (/./), double-dot sequences such as /../ to further obfuscate paths.

Polymorphic buffer overflow attack: Buffer overflow attacks overflow a buffer with excessive data. This type of attack allows an attacker to gain the privileges of the system that are granted to the application that is being attacked by running on remote shell of the computer.

Hex encoding: Web servers may identify the hex encoding but not the IDS. Many of the web service calls use hex encoding for invoking the hosted services. The hex encoding for space is 20% and for s is 73%. Unfortunately most IDS can not identify hex encoding and they need to integrate with analysis-based signatures to decode the data.

Double hex encoding: In-order to identify some obfuscation attempts, thorough http protocol will perform two rounds of hex decoding on entire URL paths. Attempts like this are very difficult for the IDS to figure out since encoded url's are encoded again for intrusion.

Directory transversal attacks: Some intrusion data contains "../" strings to access secure directories apart from web server directories. Even though many IDS has signatures to detect the transversals, attackers replace the slash with the Unicode equivalent, "%co%af".

Hence, there exists a strong need for the organizations to implement an IDS technique which is of hybrid in architecture and efficient in capturing and logging the new breed of stealthy intrusions.

If the existing IDS architecture is hybrid, i.e., combination of new technology/existing IDS techniques, the intrusions would absolutely have least or no chance to get through the security system in-order to access the hosted environment.

HEURISTIC BASED HYBRID NETWORK INTRUSION DETECTION SYSTEM

In this study, we propose a new IDS detection system called Heuristic based Hybrid Network Intrusion Detection System (HH-NIDS). The proposed system comprises of two clusters namely anomaly and heuristic detection clusters of which the heuristic cluster is invoked on demand, i.e., whenever the anomaly cluster is inefficient in identifying the new incoming threats/intrusions, heuristic node will be invoked.

Functionality and components of anomaly cluster: At first phase, the incoming data packets which enter the network from the end users are examined in the anomaly block. Here, the system uses anomaly based IDS technique to filter out the possible intrusions using knowledge based approach. The various components involved in the anomaly based IDS technique is shown in Fig. 1.

Data source: Here, the stream of data entering the network is segregated in to individual packets and sent to anomaly engine for analysis.

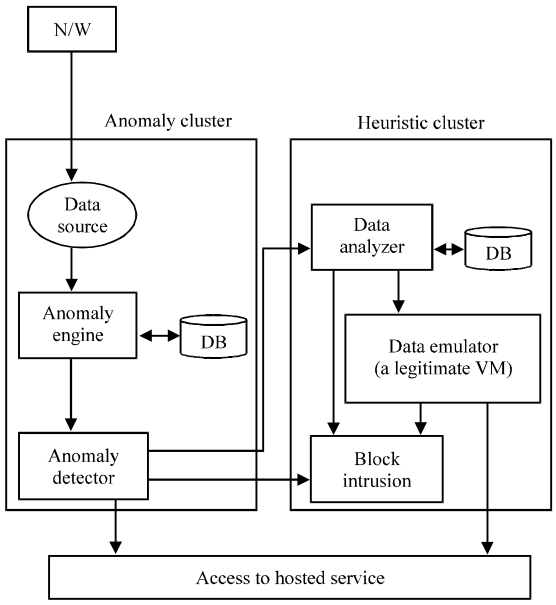

Fig. 1: Components of heuristics based hybrid network Intrusion Detection System

Anomaly engine: Here, the segregated data pattern from data source is compared with the normal data pattern present in anomaly Data Base (DB) and the comparison results are passed on to the next level.

Anomaly detector: Here, the pattern comparison results are analyzed and decision is made to route the incoming data either to the heuristic cluster for further analysis or to the service hosted environment. Whenever, the incoming data packets are examined to be abnormal and flagged as anonymous by the anomaly detector, the data is denied of access to the hosted environment.

Functionality and components of heuristic cluster: Since, the anomaly cluster filters out the intrusion data from normal data using knowledge bases approach there exists a possibility for new breed of stealthy intrusions to evade the anomaly cluster. In the heuristic cluster, the segregated data packets from anomaly cluster are examined and scanned for suspicious attributes or characters of possible intrusion data by a series of emulation techniques. The various components involved in heuristic based approach as shown in Fig. 1.

Data analyzer: Here, the system compares the patterns of the incoming segregated data with intrusion data present in database. The database contains list of previously emulated data sets which are considered threat. If there's any match between the patterns of the incoming data

packet and the intrusion data present in database, the incoming data packet is flagged as intrusion and the packet counter (Boolean System flag specific for each data packet) is set to 1. After examining the entire packets in the stream if the sum of all the packet counters is more than that of threshold value of the predefined intrusion instance, the data is considered intrusion and blocked for further access to the hosted environment.

Data emulator: Suppose if the incoming data stream does not matches with any intrusion signature data in the analyzer DB, the entire data stream is copied and emulated in a legitimate virtual machine which is an exact clone of the hosted service. Since, the false positive rate is high in the above mentioned data analyzer, it can be combined with dynamic emulation based method for optimized detection. The emulated results are traced and if found any abnormal actions on the legitimate VM, the data stream is dropped and blocked for further invocation.

The data emulator requires additional resources than data analyser or anomaly cluster because the emulator engine involves using a legitimate virtual environment which is the subset of exact copy of the service hosted by the provider.

CONCLUSION

Although, there exist various IDS techniques, each possess its own advantages and limitations which affect the performance of the entire network integrated system. The proposed heuristic based hybrid network intrusion detection system is a powerful extension to the network intrusion detection architecture. Since, the malicious intrusions are emulated in the integrated VM there exist much lower false positive rate and very least possibility for intrusion to access hosted environment.

RECOMMENDATIONS

Future scope of this research work is to implement the proposed heuristic engine in a distributed network framework. At the same time, the current approach can also be applied in the cloud environment where the resources are shared across multiple platforms. Overwhelming intrusions/threats in the cloud environment can be drastically reduced if the users accessing cloud data services are monitored and analyzed properly. Improvisation of the proposed research work for applicability to cloud environment would effectively monitors and analyzes the newly developed attacks/intrusions and nullifies its effects.

REFERENCES

Ahmed, A.A.E. and L. Traore, 2005. Anomaly intrusion detection based on biometrics. Proceedings of the 6th Annual IEEE SMC Information Assurance Workshop, June 15-17, 2005, West Point, New York, USA., pp: 452-453.

Cannady, J. and J. Harrell, 1996. A comparative analysis of current intrusion detection technologies. Proceeding of the 4th Technology for Information Security Conference, May 13-16, 1996, Houston, TX., USA., pp: 1-17.

Chandola, V., A. Banerjee and V. Kumar, 2009. Anomaly detection: A survey. ACM Comput. Surv., Vol. 41, No. 3. 10.1145/1541880.1541882.

Debar, H., M. Dacier and A. Wespi, 1999. Towards a taxonomy of intrusion-detection systems. Comput. Networks, 31: 805-822.

Denning, D.E., 1987. An intrusion-detection model. IEEE Trans. Software Eng., SE-13: 222-232.

Escamilla, T., 1998. Intrusion Detection: Network Security beyond the Firewall. John Wiley and Sons, New York, USA., ISBN-13: 9780471290001, Pages: 368.

Hu, J., X. Yu, D. Qiu and H.H. Chen, 2009. A simple and efficient hidden markov model scheme for host-based anomaly intrusion detection. IEEE Network, 23: 42-47.

Khan, L., M. Awad and B. Thuraisingham, 2007. A new intrusion detection system using support vector machines and hierarchical clustering. VLDB J., 16: 507-521.

Lee, W. and S. Stolfo, 1998. Data mining approaches for intrusion detection. Proceedings of the 7th USENIX Security Symposium, January 26-29, 1998, USENIX Association, Berkeley, CA., USA., pp: 79-94.

Lunt, T. and R. Jagannathan, 1988. A prototype real-time intrusion detection expert system. Proceeding of the IEEE Symposium on Security and Privacy, April 18-21, 1988, Oakland, CA., USA., pp: 59-66.

Nakkeeran, R., T.A. Albert and R. Ezumalai, 2010. Agent based efficient anomaly intrusion detection system in ad hoc networks. IACSIT Int. J. Eng. Technol., 2: 52-56.

Patcha, A. and J.M. Park, 2007. An overview of anomaly detection techniques: Existing solutions and latest technological trends. Comput. Networks, 51: 3448-3470.

Pei, J., S.J. Upadhyaya, F. Farooq and V. Govindaraju, 2004. Data mining for intrusion detection: Techniques, applications and systems. Proceedings of the 20th International Conference on Data Engineering, March 30-April 2, 2004, Boston, MA., USA.

Porras, P.A. and A. Valdes, 1998. Live traffic analysis of TCP/IP gateways. Proceedings of the Network and Distributed System Security Symposium, March 11-13, 1998, San Diego, CA., USA., pp: 1-13.

Qayyum, A., M.H. Islam and M. Jamil, 2005. Taxonomy of statistical based anomaly detection techniques for intrusion detection. Proceedings of the IEEE Symposium on Emerging Technologies, September 17-18, 2005, Islamabad, Pakistan, pp: 270-276.

Wang, K. and S.J. Stolfo, 2004. Anomalous payload-based network intrusion detection. Proceedings of the 7th International Symposium on Recent Advances in Intrusion Detection, September 15-17, 2004, Sophia Antipolis, France, pp: 203-222.

Ye, N., S.M. Erman, Q. Chen and S. Vilbert, 2002. Multivariate statistical analysis of audit trails for host-based intrusion detection. IEEE Trans. Comput., 51: 810-820.