Securing Trustworthy Three-tier Grid Architecture with DDoS Attack Defense Mechanism

P. Varalakshmi, S.Thamarai Selvi, S. Monica and G. Akilesh

Department of Information Technology, MIT, Anna University
E-mail : varanip@gmail.com

ABSTRACT

Grid is an emerging technology that aims at utilizing resources efficiently and effectively. A three-tier architecture consisting of Service Providers, Brokers and Regional Resource Administrators is proposed. Consumers submit service requests and policy constraints to the RRA. Each entity has a trust value associated with them which is computed based on their behavior. The three-tier architecture ensures trustworthy resource selection in a grid environment. This architecture is secured by building a mechanism to detect and counter Distributed Denial of Service (DDoS) Attacks. The DDoS defense consists of sensory registers that capture and analyze the traffic characteristics, short-term memory for local detection of DDoS attacks and a long-term memory for collaboration. The results show that the solution is capable of detecting all the genuine packets and discarding attack packets once trained. Also, the wastage of resources is minimized which can be put to effective use in the grid.

1. INTRODUCTION

A Distributed Denial of Service (DDoS) attack can be characterized as a large-scale, coordinated attack that is launched indirectly through multiple compromised hosts (called zombies) on victim network resources, with the purpose of preventing legitimate users from using those resources [4]. DDoS attacks consume resources rendering them unavailable to legitimate users thus mitigating the usefulness of the grid. A DDoS attack on commercial SUN grid in March 2006 brought down the entire grid to its knees within hours of its launch necessitating a login procedure change. A DDoS defense mechanism must be able to distinguish attack packets from legitimate ones with high accuracy and minimal resource consumption. In a grid, the availability of services is of paramount importance and the effect of a DDoS attack could be annihilating. The complexity of DDoS problem suggests that the solution will require a collaborative defense mechanism.

1.1 Evolution of the architecture

In a grid environment, *Service Providers* (SP) provide services which are consumed by *consumers*. In order to make service discovery easier as well as to monitor transactions, a new entity called *broker* is introduced in the architecture. The presence of multiple brokers provides redundancy in the architecture and ensures the availability of services even if one broker crashes. SPs and consumers are registered with more than one broker.

Consumers present their request to a broker which calls for a suitable SP based on the service request as well as policy requirements. The broker also forwards the request to other brokers in the same domain as well as other domains who in turn pick the most suitable SP registered under them. The broker may then choose the best SP from the nominated set of most suitable SPs (chosen by it as well as other brokers) based on various parameters such as trust,

satisfaction of policy requirement, service queue length, QoS etc. However, as a broker gains monetary benefit from each individual transaction, it tends to favor an SP directly registered under it. This leads to a problem called as *biased brokers*. In order to overcome this problem, a three-tier-architecture is adopted here for the grid system.

In the three-tier architecture, neutral entities called Regional Resource Administrators (RRA) are responsible for publishing and discovery of services. Brokers register themselves under the RRA. Consumers present their request and policy-constraints to an RRA which picks the most suitable SP for the transaction. Since RRA are not benefited monetarily from individual transactions, they can maintain the neutrality.

2. RELATED WORK

The architecture proposed in [1] is extended in this paper. A number of mechanisms have been proposed by research community to counter DDoS attacks. Paper [11] proposes using traps for malicious attackers called honey-pots, for mitigating service level DDoS attacks. During IP spoofing, packets with new randomly generated addresses appear in the network. A novel scheme to detect an attack by monitoring the increase of new IP addresses has been proposed in [7]. A sequential change-point detection algorithm has also been proposed in the same. Path Identification has been presented in [9] which is basically a packet marking approach in which a path fingerprint is embedded in each packet that enables a victim to identify packets coming from the same path, regardless of source IP spoofing. A user authentication at IP level by periodically stamping each IP packet with a unique, identity based signature of the user is enforced in [12]. This allows spoofed packets to be detected as soon as they arrive at the source authenticating router by detecting discrepancies between signatures in the packet headers and their declared IP address or by checking special markings. However a specialized router is necessary to deploy this mechanism which may not be cost efficient of scalable. Paper [13] proposes a signal-processing approach to identify and detect shrew attacks by examining the frequency domain characteristics of incoming traffic flows to a server. DDoS can also be treated as a congestion-control problem and solved by discarding packets selectively based on a score [14]. A method to automatically fingerprint and identify repeated attack scenarios has been proposed in [15]. The fingerprint is based on spectral characteristics of attack stream which are hard to forge. A three level hierarchy consisting of traffic monitors, local and global analyzers has been proposed in [5] to collaboratively detect DDoS attacks using machine learning techniques.

3. SYSTEM ARCHITECTURE

In a grid environment where the availability of services is important, detection and defense against DDoS attacks is a key issue. The three-tier grid architecture consists of Service Providers (SP), brokers and Regional Resource Administrators (RRA) in three different tiers. Consumers are treated as external entities. This paper proposes an intuitive mechanism based on Atkinson-Shiffrin's cognitive memory model to detect DDoS attacks. Defense against these attacks is provided by designing a firewall that filers the attack packets based on rules defined dynamically based on the current traffic pattern.

3.1 Atkinson-Shiffrin's Memory Model

Atkinson-Shiffrin's cognitive memory model is one of the models used to represent the human brain and extensively studied in the field of Artificial Intelligence. Sensory information stored in the human brain is linked to neurons. Information in some cells can be preserved only for a short term and these constitute the short-term memory (STM). There are other cells in the human brain which can store information for a long duration, in the order of years. Such cells make up the long term memory (LTM).

The Atkinson-Shiffrin's model (Fig. 1) consists of a three-layered structure of memory consisting of sensory registers, STM (also called active memory) and LTM. The sensory registers are large capacity storage that can save information with high accuracy. However, they also decay at a fast rate to keep provisions for entry of new information. STMs are

fragile but can hold information with significant strength for quite sometime. As they have a fast access-time they are used during inference generation as well. Part of the information stored in the STM is copied into the LTM. LTMs have large capacity and can hold information for a longer duration, sometimes spanning years. This model can be compared with the model of memory systems in computers.

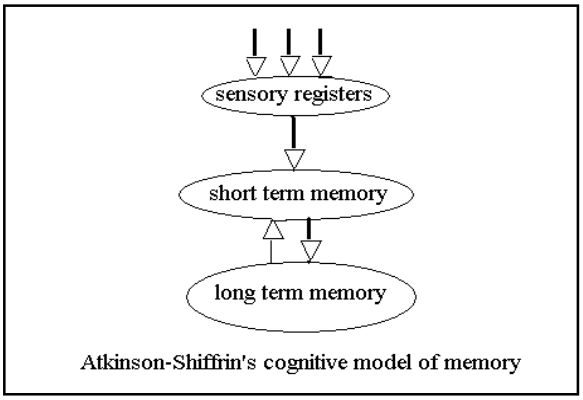

Fig. 1: Atkinson-Shiffrin's cognitive memory model

This paper defines registers and STM at the various entities in the grid environment (RRA, broker and SP). However the long term memory is present only at RRA and high-level correlation as well as collaboration takes place at the RRA. The rules for defense are also initiated at the RRA level and sent to the individual entities.

3.2 Extended System Architecture

In the existing three-tier architecture, a network of Anomaly Detection Agents (ADA) is incorporated which is responsible for DDoS defense. An overlay network is defined over the existing grid which designates some RRA as super-peers and all others as peers as shown in Fig. 2. An ADA that is placed at the perimeter of each RRA does the processing for attack detection at the global level. The super peers store information required for collaboration as well as control over the entire architecture. The information is stored locally in the local databases and globally at the central databases. In order to provide redundancy, more than one copy of central databases is present-one at each ADA at the super-peer RRA.

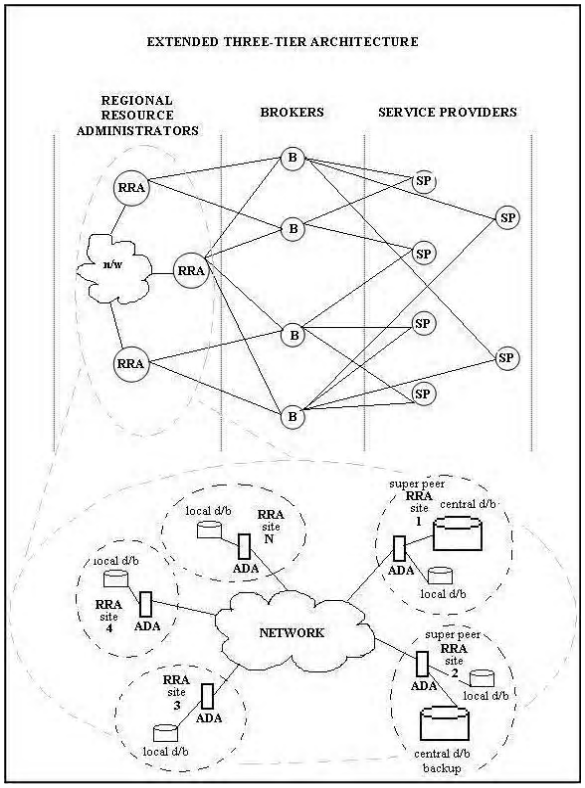

Fig. 2: Extended three-tier architecture

Here, DDoS detection takes place at two levels-at the STM as well as the LTM. In order to monitor the traffic passing through the architecture as well as the detection performed at the various ADA, a centralized control is required. The super-peer RRA provides centralized control over the network. Hence, the centralized databases are maintained at the ADA associated with the super-peer RRA. An ADA associated with a super-peer RRA thus performs superfluous tasks in addition to those performed by the other ADA.

Grid is essentially distributed in nature. Providing centralized control on the other hand, introduces a single point of failure. To overcome this disadvantage, redundancy is brought in by allowing more than one RRA to act as super-peer. Each ADA at the super-peers contains identical, synchronized copies of the central databases.

An ADA is capable of running multiple threads. It can also be extended to provide the full functionality

of an Intrusion Detection System. For example, a separate thread may be introduced to allow the ADA to collaboratively detect and filter spam messages. Such mechanisms can be used to further secure the three-tier grid architecture.

3.3 DDoS attack models

In this paper, flooding attacks are considered. During flooding attacks, the number of packets with similar packet attributes increases in number. This is the underlying principle of the architecture.

A mechanism to detect TCP SYN flood attacks was proposed in [10] which used the ratio of number of TCP SYN packets to the number of TCP FIN and RST packets. Paper [5] builds on this model to detect SYN flood attacks using machine learning techniques. A three-level hierarchy of traffic monitor, local and global analyzer has been proposed that uses spatial and temporal correlation to detect an attack.

In this paper, an intuitive mechanism based on a cognitive model of the human brain has been proposed to detect the attacks. Unsupervised learning techniques in the form of Self-Organizing maps are used to detect an attack by capturing traffic characteristics and processing packet attributes. A set of attributes are monitored and per value frequency is computed to define appropriate rules for packet discarding based on the attribute values. For this purpose, an ADA is installed at each entity (RRA, broker, SP) in the three-tier architecture.

4. DDOS DEFENSE MECHANISM

An intuitive hierarchical defense mechanism has been proposed. Registers and STM are implemented at all RRA site while LTM is present only at super-peer RRA.

4.1 Level 1 - Registers

The registers are used to capture and analyze the traffic characteristics. A set of attributes 'A' is chosen and their values are mapped in a hashmap data structure called the *memory table*. In case another packet with the same value for a particular attribute arrives, the count of that value is incremented. For a packet, with different value for an attribute, collision is resolved by chaining. Thus, the memory table contains values and frequencies of occurrence of those values for the specific set of attributes in 'A'. A score is computed per packet based on the frequency of attribute values. The idea is that packets causing a flooding attack tend to increase the frequency of a certain set of attribute-value pairs. If the attribute-value pairs present in a packet correspond to that in the set, then the frequency of those values in the memory table will increase. An upper limit is fixed on the frequency of a value. This limit varies dynamically based on traffic characteristics.

Initially, the score of a packet is initialized to zero. Every time an attribute-value pair of a packet causes the frequency to exceed the limit, its score is incremented. This score is compared with the threshold and discarded if it is above the threshold. Otherwise, the packet is considered genuine.

The memory table is refreshed periodically by making the registers 'forget' the old attribute values. This is achieved by uniformly decreasing the frequency of all attribute-value pairs present in the memory table. The memory table is refreshed every T_s second where T_s is a short interval that is chosen based on the bandwidth of the traffic.

4.2 Level 2 – Short Term Memory

The STM is used to detect an attack locally. The STM at an entity copies the data from the local register to its hashmap and computes a per attribute score. The per-attribute score is computed as a function of standard deviation of values and the fraction of packets arriving with that attribute.

$$\omega = \frac{\sigma * \rho a}{\rho t} \qquad ...(1)$$

Here ρ_a is the fraction of packets containing an attribute, ρ_t is total number of packets that have arrived in the interval T_s, ω is the Per-attribute score and σ is standard deviation of the values obtained in that interval. This attribute score reflects the pattern in the packet header. Standard deviation indicates the deviation of the packet characteristics from the existing experience. If most packets show similar

deviation, then the experience has to be changed to reflect that. This ensures that the defense is adaptive.

It also stores the current traffic characteristics which are used to modify the frequency limit, the refresh rate and the threshold dynamically. Other statistics such as total number of packets received, number of attack packets, and number of genuine packets are also stored at the STM. The scores from each STM are sent to the LTM every T_p seconds i.e., after each epoch. The epoch period is computed as $T_p = k * T_s$, where k is any integer, and this is common to all the STM present in the architecture.

4.3 Level 3 – Long Term Memory

The LTM is present at each super-peer ADA site and stores information for the long run. The LTM is also responsible for the collaboration between the RRA. Every T_p seconds, data from various STMs arrive at an LTM. The statistics present have to be stored at the LTM and a new threshold has to be computed that acts as the global discarding threshold. This global threshold is sent to all RRA which choose the minimum of locally computed threshold and the global threshold as the new threshold. The score of all packets arriving after this update is compared with the new threshold.

To compute global threshold, unsupervised learning in the form of self-organizing maps (based on Kohonan maps) may be used. Self-Organizing networks can learn to detect regularities and correlations in their input and adapt their future responses to that input accordingly. Self-organizing maps learn to recognize groups of similar input vectors in such a way that neurons physically near each other in the neuron layer respond to similar input vectors. Thus, self-organizing maps learn both the distribution (as do competitive layers) and topology of the input vectors they are trained on. It is based on the winner-take-it-all notion where the 'winner' is the neuron that yields the maximum value when computing a dot product of input and weight vectors.

The thresholds from various RRA form the input matrix to the neural network. The weight matrix reflects the current traffic characteristics. The output is a single value that is the global threshold.

4.4 Informal Algorithm

The mechanism of the proposed DDoS attack defense is as follows:

Choose a specific set of attributes, 'A' for analyzing the packets. Create a memory table indexed by attribute names. This abstracts the sensory registers. The following light-weight processes are executed in parallel at each RRA:

LWP 1: For each packet, capture the packet header. The registers remember the values of attributes in 'A'. Any collision due to different values in the memory table is resolved by chaining. The packet header is compared with the already existing patterns in the memory table.

If the frequency of pattern occurrence exceeds a threshold, then the probability of it being an attack packet is high.

LWP 2: Every T_s seconds, decay the registers by 'forgetting' the old patterns present in the memory table.

LWP 3: Every T_p seconds, after the elapse of an epoch, refresh the registers by computing new threshold.

For this, a miniature table is created and sent as an experience to the STM for pattern comparison with existing experiences. This is used for local detection of attacks.

The detection results, packet feature data and traffic statistics are sent to the LTM where temporal correlation of data is carried out to compute global thresholds. The global threshold is sent to all RRA who then choose the minimum of local and global thresholds as the threshold for the next epoch.

The new thresholds reflect the current load at each site as well as the nature of incoming traffic (attack or genuine). The statistics and thresholds of previous epochs are entered in the STM as 'experience'.

5. SIMULATION AND RESULTS OF SIMULATION

The three tier architecture is simulated with 100 consumers, 100 SPs, 50 brokers and 10 RRA.

Currently ten different service types with fixed criticality rates are considered for the purpose of simplicity; however this can be easily extended without loosing generality. The consumer-requests are generated randomly and submitted to the RRA. A DDoS attack is simulated by sending requests from external entities.

For simulation only the source address, destination address, protocol flag and QoS of the packet are extracted. A packet capturing tool JpCap is used in this simulation to capture the packets and access all its headers. A greedy approach is followed in the defense mechanism which allows an entity to receive and respond to requests till it reaches its resource utilization limits. A resource consumption table of each entity is maintained at the RRA for this purpose.

The effectiveness of DDoS mechanism is depicted by plotting the fraction of genuine packets that are present in the test set versus that detected by the DDoS defense mechanism for various time intervals (Fig. 3). The length of the interval equals the time elapsed between recordings of two successive experiences, also called the epoch. It can be inferred that the false classifications are kept to a minimum and that nearly all genuine packets are detected and allowed to pass.

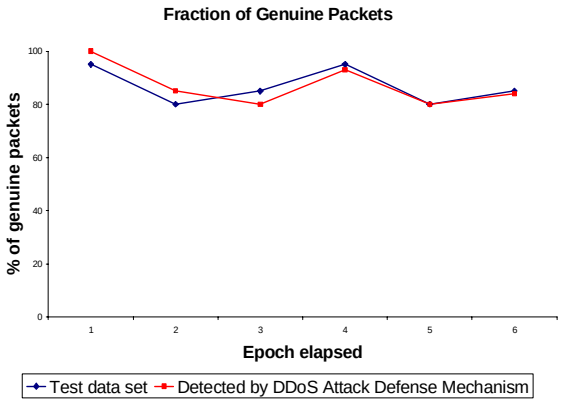

Fig. 3: Fraction of Genuine packets present vs. detected

Fig. 4 is the resource depletion graph that illustrates the percentage of resources that are wasted, i.e., consumed but not for problem solving. For simulation, threads in different machines were used to send requests non-stop effectively creating a DDoS attack. It can be seen that resource depletion without the defense mechanisms has a linear relationship with the number of messages.

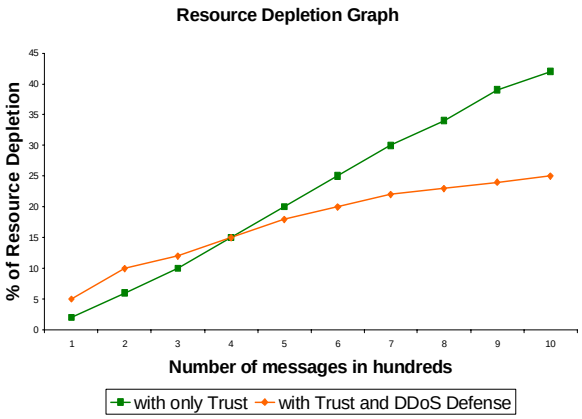

Fig. 4: Resource Depletion Graph

As the number of request and feedback messages increase, more resources are wasted. However with the defense mechanisms in place, although more resources are depleted initially over time, as more requests are processed the resource wastage stabilizes. This overhead is due to the threads that are alive throughout. Future work aims at minimizing resource depletion further.

6. CONCLUSION

This paper extends the existing trustworthy three-tier grid architecture by creating an overlay network of RRA for collaborative DDoS defense meachanism. An Anomaly Detection Agent (ADA) that runs the DDoS defense is placed at the perimeter of each RRA site. An intuitive three-level hierarchical model based in the Atkinson-Shiffrin's cognitive model of human brain has been proposed that consists of registers to capture traffic characteristics, a short term memory to perform local traffic analysis as well as to locally detect attacks and a long term memory for global analysis and collaboration. The graphs indicate the effectiveness of the defense mechanism in terms of detection as well as resource depletion. Future work consists of including inference engine to define dynamic firewall rules automatically based on the

goals defined at the LTM at ADA to counter Distributed Denial of Service attacks more effectively.

REFERENCES

1. P. Varalakshmi, S. Thamarai Selvi, A., Javed Ashraf and K. Karthick, "B-Tree based Trust Model for Resource Selection in Grid", *IEEE International Conference on Signal Processing, Communications and Networking 2007*
2. I. Foster, C. Kesselman and S. Tuecke, "The anatomy of the grid: Enabling scalable virtual organizations", *International Journal of Supercomputer Applications*, 2001.
3. Jimmy McGibney and Dmitri Botvich, "A Trust Overlay Architecture and Protocol for Enhanced Protection against Spam", *IEEE Second International Conference on Availability, Reliability and Security 2007*
4. Jelena Mirkovic, Peter Reiher, "A Taxonomy of DDoS Attack and DDoS Defense Mechanisms"
5. Kejie Lu, Dapeng Wu, Jieyan Fan, Sinisa Todorovic and Antonio Nucci, "Robust and efficient detection of DDoS attacks for large-scale internet", *Computer Networks: The International Journal of Computer and Telecommunications Networking, Elsevier*, December 2007
6. Mark E. Snyder, Ravi Sundaram, and Mayur Thakur, "A Game-Theoretic Framework for Bandwidth Attacks and Statistical Defenses"
7. Jelena Mirkovic and Peter Reiher, "D-WARD: A Source-End Defense against Flooding Denial-of-Service Attacks"
8. Yu Chen, Kai Hwang and Wei-Shinn Ku, "Collaborative Detection of DDoS Attacks over Multiple Network Domains" *IEEE Transactions on Parallel and Distributed Systems*
9. Abraham Yaar, Adrian Perrig and Dawn Song, "Pi: A Path Identification Mechanism to Defend against DDoS Attacks"
10. H. Wang, D. Zhang, K.G. Shin, "Detecting distributed Denial of Service attacks", *IEEE INFOCOM* 2002
11. Sherif M. Khattab, Chatree Sangpachatanaruk, Daniel Moss´e, Rami Melhem, and Taieb Znati, "Roaming Honeypots for Mitigating Service-level Denial-of-Service Attacks"
12. Michael Sirivianos, Ersin Uzun, Ines Viskic. "SANALDA: A Source Authenticating Network Architecture Limiting DoS Attacks"
13. Yu Chen, Yu-Kwong Kwok, and Kai Hwang, "Filtering of shrew DDoS attacks in frequency domain", Local Computer Networks, 2005.
14. Yoohwan Kim, Wing Cheong Lau, Mooi Choo Chuah and H. Jonathan Chao. "PacketScore: A Statistics-Based Packet Filtering Scheme against Distributed Denial-of-Service Attacks", IEEE *Transactions on Dependable and Secure Computing*, April-June 2006
15. Alefiya Hussain, John Heidemann and Christos Papadopoulos, "Identification of Repeated Denial of Service Attacks"
16. Jiangtao Li, Ninghui Li, XiaoFeng Wang, Ting Yu, "Denial of Service Attacks and Defenses in Decentralized trust Management", SecureComm 2006, August 2006.