Wild-Inspired Intrusion Detection System Framework for High Speed Networks

\((\varphi|\pi)\) IDS Framework

Hassen Sallay, Adel Ammar, Ouisssem Ben Fredj, Mohsen Rouached, Hafidh Faires, Khalid Al-Shalfan, Majdi Ben Saad
Research and Consultancy Unit, Aman Security Team
Department of Computer Sciences, Imam University
Riyadh, Saudi Arabia

{Hassen_Sallay, Ammar, Ben Fredj, Rouached, Kshalfan, Bensaad}@amansystem.com

Abstract—While the raise of the Internet and the high speed networks made information more easy to acquire, more fast to exchange and more flexible to share, it also made the cybernetic attacks and crimes more easy to perform, more accurate to hit the target victim and more flexible to conceal the crime evidences. Although we are in an unsafe digital environment, we often feel safe. Being aware of this fact and this fiction, we draw in this paper a security framework aiming to build real time security solutions in the very narrow context of high speed networks. Such framework is called \((\varphi|\pi)\) since it is inspired by the elephant self-defense behavior which yields \(\pi\) (7 security targets over 22 security tasks).

Keywords-component; Intrusion Detection; High Speed Network; Monitoring; Framework

I. INTRODUCTION

Our digital environment has a fact and a fiction. The fact is that mice, snakes, bats, camels, foxes and wolfs are there. Mice that have no aim except the corruption, snakes spouting venom everywhere, bats loving work in the dark, hatful camels looking for revenge, foxes and wolfs using cunning and Rogan to hit victims and conceal the crime. The fiction is that we often feel safe. Some behaves as peacock, proud of its security arsenal and infrastructure even if the attacks rain cats and frogs. Some other behaves as ostrich only burying the head in the sand. Some other behaves as spider, protect themselves by a security infrastructure as weak as a spider web is weak. Being aware of this fact and this fiction, several efforts have been conducted in the literature. We survey briefly in the following some of these efforts from both the industrial and academia sides.

Form the industrial side, several real security platforms provide an integral security solutions. This performed by merging between different techniques known as hybrid IDS. We cite here CheckPoint IPS based mainly on Confidence indexing, Cisco IPS and BreachGate WebDefend based on behavior and statistical analysis, DeepNines BBX IPS, AirDefense Guard and BarbedWire IDS based on protocol analysis and data correlation[1]. From academia, we cite the misuse based IDS Snort Inline and Snort with SPADE anomaly plug-in. Snort is largely considered as the de facto IDS [2]. BRO, from Lawrence Berkeley National Laboratory, which is compatible with snort and it includes semantic analysis at the application layer [3], while EMERALD, from SRI laboratory, considers rule-based discovery and Bayesian networks [4]. Some other are based only on anomaly based techniques such that SPADE, Prelude, nPatrol and Mazu profiler [5]. Other research projects form the academia side yielding several interesting IDS framework and systems such as MINDS IDS from University of Minnesota [6], Orchids from ENS de Cachan [7], Intelligent IDS from Mississippi State University, GIDRE from University of Granada, Genetic Art- IDS from Northwestern University [1] and anagram form Columbia University [8]. We note that the commercial systems basically tend to use well mature known techniques by enhancing their implementation issues while the research systems tendency to use much more innovative techniques. Both sides use a large spectrum of techniques such as statistical methods, clustering techniques, diversification, Bayesian inference, genetic algorithms, payload modeling through n-grams, stochastic modeling, fuzzy logic, data mining and neural network.

While the aforementioned systems provide interesting functionalities, they only partially satisfy the narrow constraints exhibited by the high speed network context mainly the real-time criterion. Recently, new approaches and solutions addressing the huge amounts of transferred network data and increasing speeds of today's networks were proposed. [9,10,11,12,13,14] aim to accelerate the speed of detection process by using specialized hardware mainly based on FPGA (Field-Programmable Gate Array) technology. Liberator project [15] introduce the COMBO platform with the FlowMon probe and the IDS probe implemented by FPGA. [16,17] proposed solution based on traffic load balancing between different IDS sensors. In [16] the clustering technique was used. The load balancer distributes the traffic among cluster nodes on a predefined policy. The authors proposed a logic ensured maximum utilization of cluster resources by exchanging state information, load sharing, reducing data loss and performing recovery evaluation procedure to maximize overall efficiency. [17] present the design and implementation of a new approach for anomaly detection and classification
over high speed networks. Mainly the proposed approach is based on two phases. The first one is a data reduction phase through flow sampling by focusing mainly on short lived flows and the second step is then a random aggregation of some descriptors such as a number of SYN packets per flow in two different data structures. [18,19, 20,21] consider the problem of real time and scalable stateful intrusion detection for high-speed networks. Basically the authors state that this problem cannot easily be solved by optimizing the packet matching algorithm utilized by a centralized process or by using custom-developed hardware. Instead, there is a need for a parallel approach that is able to decompose the problem into sub-problems of manageable size to master the complexity and improve the scalability. In this direction, [18] presents a novel parallel matching algorithm for the signature-based detection of network attacks. The algorithm is able to perform stateful signature matching and has been implemented only using off-the-shelf components. [19] proposes a novel agent-based distributed IDS which integrates the desirable features provided by the distributed agent-based design methodology with the high accuracy and speed response of the Principal Component Classifier. The author state that the simulation results demonstrate a satisfactory linear relationship between the degradation of response performance and the scalability of the system. [22,23] deal with the real time detection and alert management efficiency. [22] design an intrusion detection system architecture and improved pattern matching arithmetic based on protocol analysis and theorization machine on Frete. The proposed architecture can use network processor to collect and analyze data in network, which enhance the speed and efficiency in intrusion detection system. [23] proposes a distributed intrusion alert fusion scheme based on multiple keywords and routing infrastructure. The experimental results show that their scheme has well scalable, and can achieve significant improvement in load balancing.

We observe that even if some interesting generic standard management frameworks and technologies have been proposed, they need an adaptation and an instantiation for the IDS process in high speed networks. We believe that improving efficiency, real-time and scalability of DIDS in HSN passes through specifying a dedicated integrated security management framework. Such framework makes it easier to work with complex technologies and it ensures its good and flexible implementation. The intended framework should couple both technical and management plans. The goal of the work undertaken in our group is to build such a framework to manage DIDS entities. Mainly our target is to reach a security infrastructure that exhibits an elephant behavior i.e. a security solution that is robust, stable and flexible as an elephant is. Moreover, the elephant is distinguished by its high level of intelligence, interesting security behavior (i.e. it does not attack others and at the same time it does not permit to be attacked), methods of communication and complex social structure [24,25]. The paper is organized as follow. Section 2 specifies the intended framework. Section 3 discusses the framework implementation issues. Section 4 presents our work in progress. Finally some conclusions together with an outlook for future work are given in section 4.

II. (Φ[Π]) FRAMEWORK SPECIFICATION

Following we depicts some similarities of our environment with the elephant one:

A. Environment characteristics

- Elephant inhabits a diverse array of habitats including tropical forests, savannas, grasslands, and woodlands. Vs. Our environment contains a variety of network topologies, technologies and services.
- Nearly an elephant’s day is spent in feeding. It consumes grasses, small plants, bushes, fruit, twigs, tree bark, and roots. Vs. Nearly a network day is spent in feeding (incoming and outgoing traffic). It provides several services such that FTP, HTTP, etc.
- Elephant can drink up to 212 L (55 gal.) of water in less than five minutes. Vs. With High Speed Network we can reach a throughput of 40 gigabits/s.
- In contrast to most mammals, elephant lungs are directly attached to the chest wall and therefore rely on direct muscular action to expand the lungs. Vs. In contrast to most networks, HSN has a DMA and programmable network cards and we write directly into the memory skipping out the NIC and operating systems latencies.

B. Security Threats

- An elephant calf’s first year of life is its most vulnerable time with mortality rates exceeding 30%. Vs. A network service’s first days of deployment is its most vulnerable time with the most high rate of hacking.
- African elephants are threatened by poaching for ivory largely more than Asian ones due to their overall smaller tusk size. Vs. More crucial services and sites are more susceptible to attacks than less crucial ones.
- Anthrax is one of the most fatal diseases impacting elephants. It is a bacterium that causes high fever, shivering, ulcers, and swellings. Vs. Malicious attacks are the most fatal diseases impacting network and computers. They cause high troubling of the network, make unstable the service and provoke congestion in the network links.
- This disease may be spread through contaminated water or soil. Vs. This disease is spread through contaminated packet in the network traffic.

C. Identification of security Design requirements

- Although elephant is the largest of all land animals. It is very tactile in nature. It uses all parts of their body to interact with one another in all forms of behavior. Behaviors includes defensive, exploratory, and anti-predator. yields. Scalability, Flexibility & Versatility: the ability to remain efficient when the traffic and security operations increase. And the ability to add, remove, update and cooperate the different security components inside the network and perform a variety of functions according to different security goals.
Elephants survived the Asian tsunami in 2004. The complete lack of vision did not hinder the blind elephant's ability to fulfill its leadership role. **Yields.** *Survivability & Fault Tolerance:* the ability to resist and survive in the crisis situations and critical constraints. And the ability to continue functioning even if some security components were defected.

Moreover the elephant’s enables secure movement over uneven terrain and swampy ground. **Yields.** *Adaptability, accuracy & stability:* the ability to adjust and tuning the security system in order to cope with a variety of situation and context in real time.

Elephants are very autonomous and they may live in small group with hierarchal ranking social structure. Their trunks are very powerful capable of uprooting an entire tree trunk, tearing down heavy branches, and lifting weights in excess of 250 kg. **Yields.** *Autonomy & Robustness:* the ability to self protection without complete relying on third party.

Leaders protect the front and rear of the herd. More docile bulls serve as stabilizing members within the group. Hierarchical roles are reestablished and readjusted whenever a male leaves or enters the group. **Yields.** *Specialization & Self organization:* the ability of specification of security roles and their distribution and their readjustment through automatic and self-configuration process.

Although primarily solitary in nature, bulls will associate with non natal family units (family units to which they are not related). Bulls do not have preferences for specific family units and will randomly move to different groupings daily and even hourly looking for reproductively receptive females. **Yields.** *Integration with the legacy & Productivity:* the ability to be open and interact with heterogeneous environments to maximize the production through searching for more fertile techniques and technologies.

Elephants have benific impact by creating trails that other animals and humans use to travel. The tusks are used to dig wells, generating multiple water sources throughout the habitat. As elephants have large size, this helps bring down vegetation, which then become accessible to smaller species. **Yields.** *Accessibility and sharing:* the ability to have benefic impact on computer security society by opening new research tracks for the security community. By opening the code and sharing the product for free, security solutions become accessible for individual people and small institutions.

**D. Identification of security Tasks and components**

Elephants have the largest brain of any land mammal, highly developed cerebrums involved in movement and muscle coordination, large temporal lobes which facilitate memory. Elephants have excellent long-term memory (for decades) and are capable of remembering experiences for long periods of time. The elephant’s trunk is an extension of the upper lip and nose. It functions for grasping, breathing, feeding, dusting, smelling, drinking, lifting, sound production/communication, defense/protection, and sensing. Elephants’ trunks and keen sense of smell are used to survey the environment. Through this process, elephants are capable of locating water sources up to 19.2 km (12 mi.) away and can even determine the reproductive status of distant elephants. The eyes of an elephant are located on the sides of the head and therefore provide better peripheral vision (angle of vision extending from the sides to the rear), rather than binocular vision (eyes located on the front of the face, in which fields of vision overlap, creating depth perception). The tsunami’s approaching vibrations were detected by the elephant’s feet and alerted them to the approaching storm. The soles of elephant feet, assisting in the detection of seismic vibrations. Elephant skin is sensitive to touch, detecting insects and changes in its environment. To supplement the diet, elephants will dig up earth to obtain salt and minerals. The tusks are used to churn the ground. The elephant then places dislodged pieces of soil into its mouth, to obtain nutrients. The tusk are used to chisel hard objects, cutting and shaving branches, a task which is very important for both the elephant and the environment. More than 13 species of elephants are now listed as endangered or critically endangered. Finally, the elephant uses all parts of their body to interact with one another in all forms of behavior. Mainly, Elephant body is very powerful capable of delivering a forceful blow in self-defense [24, 25].

a) **Brain- Trunk- Feet (B-T-F) component**

The B-T-F component should improve the accuracy of detection, permits a prioritization of packet per severity and correlation of alerts, makes application context detection and assists to the Control decision. The main idea is to integrate the probabilistic models, neural network, recurrent neural network to provide an efficient adaptive real time learning. Basically, the main tasks of this component are

1. **Memorizing** all IDS related data and prepare them to the learning process. We complete incomplete data (dataset or logs etc) by using simulated data techniques.

2. **Correlating** between the different IDS data types (alerts, attacks, etc), basically we should correlate alerts generated by several IDS types (anomaly based, misuse based and specification based IDS).

3. **Learning** inclemently about different security aspects. Mainly we should identify the main important features by using the Neural Network, reduce the number of features without losing their meanings and impacts by using Principal Component Analysis and integrate the probabilistic models, neural network, recurrent neural network to provide an efficient adaptive real time learning.

4. **Predicting** different security related behaviors. Mainly we should predict (1) Traffic behavior (2) Attack behavior (3) Flow behavior (connections) (3) Packets/flow containing attacks per service (4) Packets/flow containing specific attack or class of attack (5) Specific Service behavior and (6) Packet/Flow containing tunneled attacks. The prediction model should improve real time and accuracy of
IDS detection process by scanning deeply the predicted suspected malicious packets

5. **Assisting** the administrator decision by providing attack classes, alert priorities (security severity level) and service behavior status. Integrate in real time the manager decision, his new policies, service and traffic behavior to update the learning and decision process in real time.

The BTF data input is a set of datasets describing (1) the Traffic (2) the Attacks (3) the Alerts (4) the Application payload (5) the Services behavior or a combination of the 5 previous inputs. The input can be get from KDD data set, IDSheild data set or through building a new data sets based on (RootFu and CTF) traces.

b) **Trunk-Eyes- Mouth (T-E-M) component**

The T-E-M component should implement an efficient sniffing and adaptive load balancing strategies to provide a real time and accurate detection process through a deep analyzing of packets at network and host sides. It should ensure a near real time scanning of the entire environment (network and hosts), perform some penetration testing for the detected vulnerabilities, updating the detection rules based on an accurate vulnerabilities scanning report. Finally, it should provide a real time alerting in case of an attack detection. Basically, the main tasks of this component are:

1. **Feeding** from the network through an efficient data sniffing by optimizing sniffing and preprocessing by several techniques such as optimization for a specific network service (i.e HTTP), using a filtered input network data-flow and opt to a real time configuration in order to get an optimization for a specific HW/SW configuration

2. **Splitting** the sniffed data to make an efficient adaptive real time load balancing between the different IDS sensors. Architectural techniques implementing load balancing based on new techniques such as choquet integration and dirichlet process to improve real time and efficiency of detection process.

3. **Scanning** hosts and network to assess the security vulnerabilities. Specializing the detection process to our specific platform through a network and host vulnerability scanning process using scanning tools such as Nessus and Openvas etc.

4. **Perceiving** the network security status by performing a penetration testing scenarios to improve the accuracy of the vulnerability report. Penetration testing techniques by using some well known tools such as metasploit etc.

5. **Detecting** the malicious attacks in the packet and connection on both incoming and outgoing traffic. The detection algorithm and strategies should be assisted by BTF component mainly we can improve real time and accuracy of IDS sensor detection by giving more care to the predicted suspected malicious packets

6. **Inspecting** the tunneled and/or encrypted packet at the host end to improve the accuracy of the detection process. We should also make a context aware (i.e application layer detection in the network side for the tunneled and encrypted traffic.

7. **Alerting** the security administrator, in real time, of all detected attacks. Specifying alert format and alerting strategies to notify the administrator of a detected attack.

The TEM data input is a set of (1) network traffic sniffed packets (2) network traffic sniffed connections (3) the vulnerabilities report (4) the misuse IDS detection rules (5) the network anomaly behavior specification or a combination of the 5 previous inputs. The input can be get from a specific network card interface sniffing software traces such as wireshark traces, vulnerability scanning reports and snort and bro rules and anomalies files.

c) **Tusks-Skin (T-S) component**

The T-S component should improve the efficiency of the detection process by including the specification based detection ingredient in the framework. Through formal methods techniques, we can detect the deviation against a correct behavior of network services specification. Once detected, a security rule should be generated and published. Basically, the main tasks of this component are:

1. **Specifying** different service policies which describe the normal behavior that should be admitted in the network. For such purpose we should cope with the standard service protocols specification and the compliant with the ISO security guidelines and recommendations. The specification will be translated to event calculus by a specific generator. The main idea is to use the formal methods to discover the new attacks.

2. **Preprocessing** the data of different logs (traffic, alert etc.) to translate them to event calculus data logs. We use some techniques to solve the problem of data log incompleteness. We use also some rewriting techniques to reduce the size of the logs to improve the efficiency of time analyzing in the next step.

3. **Sensing** new discovered attacks by deeply analyzing investigating of the mismatching with the preprocessed logs against the policy specification. Mainly we use the spike theorem prover which has the ability to trace the mismatches accurately. A Data mining process could be also investigated.

4. **Generating** new security policies remedying these discovered attacks. This task should be performed through an automatic or in worst case semi automatic (i.e. assisted by administrator) process.

5. **Sharing** the learned expertise with other parties by publishing the discovered attacks with their security patches through the different publishing techniques and channels.

The TS data input is a set of (1) Security requirements (2) Security assumptions (3) Network specification/ model abstract (4) IDS rules files (5) Event logs.

d) **Body (B) component**
The B component deals with the overall management aspects in order to satisfy the real time and scalability issues. Mainly it should collect management data, analyze it and perform some control measures to implement the self defense process. Basically, the main tasks of this component are:

1. **Collecting** different data management and integrate them in unified information model. An optimized management signaling protocol dedicated for HSN should be defined. Mainly we should opt to web based management platform.

2. **Storing** efficiently the gathered data by using the HSN capabilities and putting it in an easy to use format in order to improve the its real time processing.

3. **Self-Controlling** through an adaptive autonomous self organization and predefined strategies. The control signaling protocol should be lightweight and real time oriented. Defending measures should be triggered automatically in near real time. These measures and strategies will be improved incrementally by learning from dirichlet process which describes the overall management plan status.

4. **Assisted-Controlling** the different system component to perform defense by coordinating the different security system through an administrator assistance. Some Management architectural techniques should be investigated to satisfy the near real time defending constraint. This collaboration and coordination between all components should be valid and correct. For that purpose some formal methods checking for the correctness and validity will be used. The administrator should be equipped by a friendly GUI compliant with the HCI rules and techniques to perform his assisted control tasks.

3. **Self-Evaluating** through a simulation techniques, testing and integration techniques and continuous performance analysis process of the overall integrated platform implementing the \((\varphi|\pi)\) framework.

The B data input is a set of (1) Management requirements (2) Security policies (3) Management algorithms and strategies (4) Benchmarks.

III. Work in Progress and Futur Work

In [27], we exposed the different architectures used in High Performance Computing (HPC), the common high-speed networks, the programming models, the communications models, and the communication libraries. In [28], we stated the need of HPC for Distributed IDS and we discussed the design requirements of the system. We studied the mapping of the different requirements over the software and hardware features of HSN. We proposed several recommendations for the design of IDS over HSN, starting from the communication protocol and the programming model that should be adopted, to the way the system should handle the communication flow, the memory management and the data transfer between IDS sensors. In [29], we presented a formal based approach to intrusion detection in the context of high speed networks. We proposed the global architecture of our approach and detailed its components. We also showed how the proposed formalism can be used. In [30], a security model as well as an architecture able to perform automated and procedural security safeguards are proposed. In [31], we presented an optimized scalable distributed architecture which is about 10 times quicker than the centralized architecture. The solution is based on switch-based splitting approach that supports intrusion detection on high-speed links by balancing the traffic load among different sensors running Snort placed in each point of access to the internet. In [32], we designed and implemented a p2p platform dedicated to share different types of files such as vulnerability report files, security policy rules, intrusion detection rules, attack alerts files and security reports about the network security status, in a decentralized manner making easy the exchange of these files and the sharing of the security information involved by the DIDS process. [33] presents some benchmarking of the well known intrusion detection system Snort as well as an integrated monitoring platform dedicated to manage the different components of our architecture developed in [31]. [34] presents the application of an original feature selection method based on neural networks to the problem of intrusion detection. We apply this method to a case study. We show its advantages compared to some existing feature selection approaches, and we measure its dependence to the network architecture and the learning database. In [35] we have presented the ingredients to propose a real time IDS management framework for high speed networks; REST technology and JSON data exchange format. After enumerating the requirements to be taken into account, we have reviewed the existing exchange protocols and their data formats. Then, we have focused on the REST and JSON. Motivations of our choice and an illustrative are also exposed. Ongoing work consists of refining the proposal, developing an automatic parser from IDEMF to JSON, defining the REST based exchange protocol, and considering a concrete application to show the feasibility of the whole framework. In [36] we assumed that our internet flood is influenced by several criteria and the goal was to find the continuous time function for aggregation to calculate a total average score in real time to help administrator to scan the suspected packets.

We proposed a new hierarchical model with a Dirichlet process adapted to the analysis of temporal trajectories analysis which are governed by a stochastic differential equation in random environment and the fuzzy integrals based on the Choquet integral which is an extension of the Lebesgue integral and of the balanced sum.

As future work, we plan to more investigate the framework themes which are modeling the brain theme (BTF component), architectural them (TEM), modeling the mind theme (TS component) and the management theme (B component) to develop an efficient real-time distributed intrusion detection system for high speed network.

REFERENCES
