Biologically Inspired Risk Assessment in Cyber Security using Neural Networks

Eng. Ionită Mihai-Gabriel[1,*], Prof. Victor-Valeriu Patriciu[2]
[1,2]*Military Technical Academy, Bucharest,Romania*
[*]Corresponding author (E-mail : mihai_ionita01@hotmail.com)

***Abstract*— The only suitable option for risk assessment systems designed with real-time constraints in mind, in the present seems to be the one based on attack graphs. Even though it is not computationally feasible for every circumstance, it is elastic enough for the usual use case. Paper [8] proposes an interesting approach based on attack graphs. An attack graph is a graph that represents all possible sequences of the attacker's actions that lead him/her to the established goals. These action sequences are also called attack traces. The main disadvantage of this approach is its high computational complexity. Thus, attack modeling needs to represent not only the sequences of actions, but also the attack impact, as well as how countermeasures can mitigate this impact and at which cost. However, the human body, based on the immune system, calculates risk every second for offering the correct manner of immune response to foreign threats, without hampering normal cell operation. Why don't we use this behavior in cyber defense systems? Applying Matzinger's danger theory, with the key concept of a distress signal, involved in risk assessment seems to be the logical decision, due to its lite resource consumption and categorical nature, which rapidly defines an attack surface, when correlating information gathered from local agents dispersed on protected hosts. As a proof of concept in favor of supporting this idea, a feed-forward backward-propagating neural network was setup to correlate threat data from agents installed on remote protected hosts. This intelligent system assesses the risk of a cyber-attack taking place and bringing the defense systems to an alarmed state in a timely manner, which can help offer a quick response against an attacker.**

***Keywords*-danger theory; neural networks; cyber security; attack vector; risk assessment; early warning systems.**

I. INTRODUCTION

No country, industry, community or individual is immune to cyber risks. Cyber-attacks are a certain event in this ever-changing world we live in.

Dynamic risk assessment can be used by an intelligent autonomous cyber defense system to learn the optimal action sequence for recovering from given cyber defense risk situations. These actions can be setup as a policy. Different policies can be created for different event patterns. Commercial network management could then use such policies and security products to implement selected mitigating actions automatically, as risk states are sensed by distributed agents.

The Tallinn Manual defines Network Defense as: "Actions taken through the use of computer networks to protect, monitor, analyze, detect and respond to unauthorized activity within information systems and computer networks." This definition is a broad one, leaving room for speculation. This is a positive aspect from the point of view of a cyber-security defense system, which is able, from a legal perspective to do more for protecting its assets.

When it comes to assessing risk in the cyber defense domain, it all comes down to probable events, that are going to happen with some probability and certain events, which have already happened [1]. This clear separation of events facilitates „drawing" the decision tree, and choosing between proactive measures or reactive ones. The proactive measures are those methodical actions taken for minimizing the probability of a negative event to take place, such as a Denial of Service (DoS) attack. For example, setting up a network Unified Threat Management (UTM) equipment for blocking a certain IP address if the number of connections reaches a certain threshold, can be regarded as a proactive measure against DoS attacks. The reactive measures are those taken against a negative event that already took place. The role of these steps is to minimize the aftermath of that unwanted event, for example a malware injection in one of the protected systems. A reactive measure in such a case could be the isolation of the network segment on which the infected host resides.

When designing a cyber-defense system for an architecture the size of Facebook's, which reaches 650.000 read/write checks per second, at peak [2] building decision trees is very computationally intensive, thus computationally infeasible.

The following section describes the architecture of a EWS and the main approaches for creating such a system with an accent on the ones based on AI. Also there are a few comparisons between the main types described. Section III of the paper describes the proposed approach, and the underlying architecture, for a EWS based on Intelligent Threat Assessment. Section IV describes the work in progress for extending this model to a production one.

II. EARLY WARNING SYSTEMS

Every large national cyber defense entity has an Early Warning System (EWS) setup. Moreover, if they do not, they are certainly working on one. At the fast pace at which cyber-attacks take place, doing cyber security, and more specifically, cyber defense without an EWS could lead to failure, when coming into highly targeted attacks which use specially crafted

978-1-4799-2385-4/14/$31.00 ©2014 IEEE

0-day exploits that abuse of previously unknown vulnerabilities.

Unfortunately, the author's opinion is that all these systems have similar, if not identical, functions to that of a well calibrated Security Information and Event Management (SIEM). In the best case, this system is set up as a top layer SIEM that collects data form the inferior levels and correlates the based on precisely written directives or signatures.

This approach can only warn security experts of an attack only during the actual process of exploiting a resource. A sensitive system could warn the security experts even when unusual activity, involved in scanning or information gathering, is detected. On the other hand, if the system is very sensitive, this could lead to many false alarms, which could be overwhelming for a human operator when it comes to the huge number of events reported at state level, or even a geographical province. Worse, taking into account highly targeted attacks, which do not trigger any signature based analyses or any heuristics based components, the security experts may well be alerted of something that was a successful attack, which has just finished, and the attacker already using daisy-chaining methods has covered his tracks and left with critical information. In the worst case scenario an attacker using a stealth reconnaissance approach and Intrusion Detection System (IDS) evading techniques could exfiltrate highly confidential information from the organization without setting off any alarms. The hacked organization will find about the attack from the news headlines.

A. Early Warning Systems based on Artificial Intelligence

As stated above, the systems currently used today do not have much of use over a correctly set up SIEM with evenly distributed sensors in the right networks and on the protected hosts that generate important alerts. The above-depicted approach also has a few weak points when implementations are deployed in areas with high real-time constraints.

As in article [3] the shortcomings of these systems are as follows:

- "Internet telescopes and monitoring systems strongly rely on the use of the dark address space. Although this is efficient for the detection of worms, network scans, etc., target-oriented attacks are hard to be recognized. [3]" Furthermore, even if attacks were recognized, it would be very hard to attribute an illicit action to an attacker, if he would use a segment of a "darknet". These network segments are connected only between highly trusted and verified peers, usually called "friends". They use the Tor anonymity protocol for communication or even custom protocols for file sharing and other actions.
- "Misuse detection is realized in particular by means of Deep Packet Inspection (DPI) and the evaluation of header information. DPI does not scale well with massive bandwidth levels, such as those at the Internet backbone. [3]" In this instance, the problem can be solved by distributing the inspection to units placed lower in hierarchy, which would introduce significant latency in packet routing. Another approach would consist in modifying an artificial intelligence (AI) algorithm for processing this data in an optimized manner for increased performance and superior detection rates.
- "One of the most important sources for information is the evaluation of flow data. All of the systems in use strongly rely on the evaluation of sFlow which is a sampling technology and therefore not able to provide 100% accurate results. [3]" Apart from this loss of resolution caused by sampling data, there are different flow sampling techniques that sometimes are not interoperable, such as Juniper's jFlow and Cisco's netFlow. The sampling approach is used again, because of the high load such a monitoring task puts on the system. In addition, there are high storage costs that involve capacity and performance, which have to be taken into account when everything is taken to a national or continental scale.
- "The inherent division between network and host-based indicators is a weakness of the current approaches. Currently, there is no known robust system that effectively correlates these disparate data streams. [3]" Very important pieces of data can surface when connecting and correlating these types of data. Putting these together can be done only if intelligent agents are networked, or they all report in near real time to a central entity that correlates the information for further reference.
- "Anomaly detection is only realized in subnets and it is extremely difficult to profile <<normal>> behavior with any level of identity. [3]" This phenomenon is due to the difficulty of processing high volumes of traffic and information when near-real-time constraints are in effect. Furthermore, this problem can be alleviated by using smart algorithms or by correlating distributed sensor information in a pyramidal approach, correlating it at every level, but this would not deliver a useful result when timing constraints are imposed.
- "The operation on a non-interoperable security infrastructure, which furthermore is not homogenous, containing stovepipe systems, and application and task specific <<security silos>> is a shortcoming of state of the art approaches. [3]" Unfortunately, this remains one of the primary concerns of modern cyber defense systems. The incapacity of collecting data from legacy implementations has been addressed in several ways but none is ideal. Some approaches involve connecters especially created for data gathering from these systems, but this means losing money and time for developing software that will be replaced by smarter systems, with more complex requests.

B. Existing system implementations

The Early Warning and Intrusion Detection System Based on Combined AI Methods (FIDeS) project aims at developing an advanced, intelligent assistance system for detecting attacks

from the Internet in both local area networks (LAN) and wide area networks (WAN) as early as possible.

Conventional Intrusion Detection Systems (IDS) and in particular IDS systems used for anomaly detection trigger a high false positive rate or do not detect attacks at all, this latter category is named false negative.

Complementary to anomaly-based IDS, the system presented in paper [4] develops an early warning system based upon heterogeneous methods of Artificial Intelligence (AI). This system supports a security officer in analyzing attacks and carrying out appropriate counter measures. The project focuses on concrete instructions during the attack rather than mere intrusion detection.

On the lowest layer of the FIDeS system, the data traffic of a network is tapped and analyzed by a SIEM and other external tools, such as the Internet Analysis System. This external tool, the Internet Analysis System (IAS) analyzes local communications in specific subnetworks and creates a global perspective of the Internet by bringing together the large number of local events. The functions of the Internet Analysis System can be divided up into the four segments of pattern formation and creation of a knowledge base, description of the actual status, alarm signaling and forecasting [5].

The most important function of this external tool, in accordance with the present's paper subject is that of forecasting events. The researchers have developed some extensions for the IAS, one of which allows making certain long-term and short-term forecasts. The long-term forecast can be conducted with or without seasonal impacts. One of the discoveries is that the trend of the use of a technology is clearly noticeable without considering seasonal impact. The use of "linear regression" offers the greatest accuracy for long-term forecasts. In the case that there is some heavy noise in the data methods of smoothing are more precise. In the area of short-term forecasts, the consideration of seasonal aspects is very important, for instance day- and night changes, lunch breaks, working day and holidays. Their conclusions show that when breaking down to an interval of hours the method of "linear regression" is preferred. The shorter the interval the more exact the "Holts-Winters" method turns out to be, which should be used for intervals of minutes [5].

III. Early Warning Systems based on Intelligent Threat Assessment

The proposed approach in this paper consists of using an intelligent method of calculating the probability of certain cyber-attacks to happen, based on risk assessment. This will be calculated periodically and when certain limits are exceeded, the system will trigger an early warning so that human operators are prepared. This can also be extended to autonomic agents who could stop forensic actions in favor of increasing detection sensitivity, when the risk of a cyber-attack is high. This action could help preserve resources for detecting, stopping and healing the affected systems.

As a proof of concept, the following threat estimates will be used [6]:

- Probabilities of an event to happen will be setup based on seven levels.
- Harm will be quantified using a six level scale
- The risk of a probable event having negative consequences will be defined using 6 levels of thresholds from zero to thirty
- Risk is defined as follows:

$$\text{Risk} = (\text{Probability} \times \text{Harm})^{(\text{Distress\_ signal} + 1)} \quad (1)$$

The local agent alongside its risk analysis result transmits the distress signal referred in equation (1). This distress signal is inspired from Matzinger's Danger Theory [9]. This theory implies that there is no need to attack all the foreign cells detected after a self-non-self-analysis which appears to be the normal functioning of the human immune systems that does not attack transplanted organs for example, even if they would be considered as foreign. These distress signals, on the other hand are sent when a cell dies an unnatural death. Coming to EWS this distress signal appears when more than two distributed agents located in close proximity signal the same threat. This enables the defense system to concentrate investigation and healing resources in specific areas, where risk is greater.

Table 1 – Probability Estimation

Probability	***Definition***	***Scale***
Negligible	Unlikely to occur	0
Very Low	2-3 times every 5 years	1
Low	Later than once per year	2
Medium	Later than once every 6 months	3
High	Later than once per month	4
Very High	Sooner than once per month	5
Extreme	Sooner than once per day	6

Table 2 - Impact Determination

Harm	***Definition***	***Scale***
Insignificant	No impact	0
Minor	No extra effort required to repair	1
Significant	Tangible harm / extra effort to repair	2
Damaging	Significant expenditure of resources required Damage to reputation and confidence	3
Serious	Extended outage / loss of connectivity Compromise of large amounts of data or service	4
Grave	Permanent shutdown – Complete compromise	5

Table 3 - Risk Assessment

Scale	***Definition***
0	Null
1-3	Low
4-7	Medium
8-14	High
15-19	Critical
20-30	Extreme

Table 4 - Risk determination

Asset	**Threat**	**Vulnerability**	**Probability**	**Harm**	**Risk**
Network info	Scanner	Open ports	5	0	Null
User accounts	Password cracking	Weak passwords	3	4	High
System integrity	Exploiting a vulnerability	Unpatched system	3	4	HIGH
Data exfiltration	Malware install	0-day vulnerability	2	6	HIGH
System availability	Botnet - DDoS	Link contention	3	5	HIGH

For automated intelligent risk determination that leads to a fully functional EWS, a neural network will be used, as in Fig. 1.

The proposed architecture implies a Feed-Forward Backward-Propagating neural network based on two input layers, ten hidden layers and one output layer. The training was done using 1000 input values and 1000 output values captured from a network of sensors formed by local agents, based on the processed security events. The training was done using the Levenberg-Marquardt method. Performance was calculated using the Mean Square Error approach.

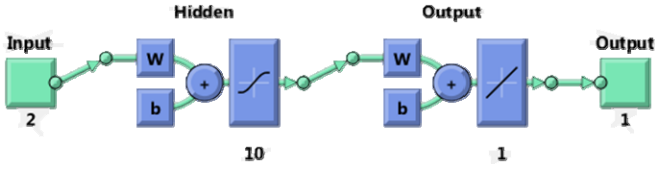

Figure 1. The proposed architecture based on Feed Forward Backward Propagating Neural

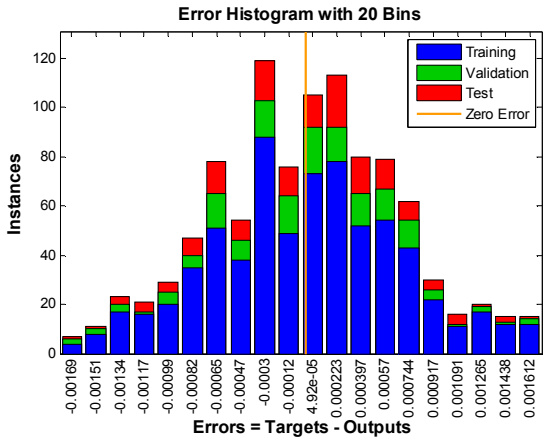

Figure 2. Training state of the Neural Network

Table 5 – Comparing calculated results to those obtained from the Neural Network

Asset	Determined risk using Neural Net	Probability	Harm	Calculated Risk
Network info	0.0026	5	0	Null – 0
User accounts	12.0014	3	4	High – 12
System integrity	12.0013	4	3	HIGH – 12
Data exfiltration	12.0009	2	6	HIGH – 12
System availability	15.0007	3	5	HIGH - 15

As it can be seen from table 5, results obtained from the neural network in the second column are comparable to those obtained by manually calculating the proof-of-concept example in the fifth column of the same table.

IV. CONCLUSIONS AND FUTURE WORK

Some systems that generate attack graphs can be made easier to calculate by using reinforcement learning or by centralizing all the computations on neural networks that exhibit unparalleled speed and synchronous decision taking abilities, alongside parallel learning abilities. These neural networks can be used in software as described in the above section, or in hardware implementations for increased speed and accuracy.

In addition, Danger Theory, even if not confirmed in the immunologic community, seems to be a promising candidate for inclusion of some primitives based upon the human immune system into the Cyber Defense world, as demonstrated in the proposed architecture presented in this paper.

REFERENCES

[1] Luc BEAUDOIN, Nathalie JAPKOWICZ and Stan MATWIN. Autonomic Computer Network Defence Using Risk State and Reinforcement Learning. CCDCOE 2009 – 1415. http://www.ccdcoe.org/publications/virtualbattlefield/17_BEAUDOIN%20Autonomic%20Computer%20Network%20Defence.pdf

[2] Tao Stein, Erdong Chen and Karan Mangla. Facebook Immune System. Microsoft Research. http://research.microsoft.com/en-us/projects/ldg/a10-stein.pdf

[3] Golling, Mario and Björn, Stelte. Requirements For A Future EWS Cyber Defence In The Internet Of The Future. 2011 3rd International Conference on Cyber Conflict. 2011. https://www.ccdcoe.org/publications/2011proceedings/RequirementsForAFutureEWSCyberDefenceInTheInternetOfTheFuture-Golling-Stelte.pdf.

[4] Stefan Edelkamp, et al. Early Warning and Intrusion Detection based on Combined AI Methods. TZI, Universitat Bremen. http://www.tzi.de/~edelkamp/secart2/papers/Fides.pdf.

[5] Malte Hesse and Prof. Norbert Pohlmann. Internet Analysis System (IAS). Germany: Institute for Internet Security. https://www.internet-sicherheit.de/fileadmin/docs/publikationen/Internet_Analysis_System_13_10_08.pdf.

[6] I. Bica, I. Livadariu, "Analysis of the recent cyberspace attacks" The 2nd International Conference on Security for Information Technology and Communications SECITC'09, ISBN 978-606-505-283-3, Bucharest, Romania, 2009. http://www.secitc.eu/about/previous-conferences/secitc-2009/

[7] Uwe Aickelin and Steve Cayzer. The Danger Theory and Its Application to Artificial Immune Systems. ICARIS 2002, pp 141-148, Canterbury, UK, 2002. http://arxiv.org/abs/0801.3549v3

[8] Igor Kotenko and Andrey Chechulin. A Cyber Attack Modeling and Impact Assessment Framework. 5th International Conference on Cyber Conflict. 2013. www.ccdcoe.org/publications/2013proceedings/d1r2s3_kotenko.pdf

[9] P. MATZINGER. Essay 1: The Danger Model in Its Historical Context. Scandinavian Journal of Immunology, Volume 54, Issue 1-2, pages 4-9, July/August 2001. http://onlinelibrary.wiley.com/doi/10.1046/j.1365-3083.2001.00974.x/pdf.

[10] I. Bica, A. Furtuna, „DC++ and DDOS Attacks", Proceedings of the 13th World Multi-Conference on Systemics, Cybernetics and Informatics (WMSCI 2009), ISBN 978-1-934272-62-6, Orlando, Florida, USA, 2009. http://www.iiis.org/CDs2009/CD2009SCI/SCI2009/PapersPdf/S167TT.pdf