Analysis of AI based Intrusion Detection Systems

Kapil Varma, Vivek Rajaram
Operating Systems (COEN 283)
Computer Science and Engineering
Santa Clara University
California, US
kvarma@scu.edu, vrajaram@scu.edu

Abstract—**This paper studies the evolution of intrusion detection systems (IDS) and usage of AI and possible improvements using different sub fields of AI. We evaluate the various AI algorithms used for network-based IDS and compare between the two widely used IDS - Snort and Suricata. Suricata showed better performance with low packet loss and more alert. We also analyse and suggest possible branches of AI that will improve the performance of IDS.**

Keywords—Intrusion Detection System; Suricata; Snort; Artificial Intelligence (AI); Machine Learning (ML); Recommender System;

I. INTRODUCTION

Intrusion detection is defined as the process of monitoring the events occurring in a computer system or network and analysing them for signs of intrusions. Denning and Neumann were the first to recommend an intrusion detection mechanism to find the reasons of intrusions within a secure computing framework [1]. As certain types of intrusions can be identified through a detailed analysis of information contained in the system's audit trial, initial intrusion detection were done manually by anticipating network anomalies in the consoles.

Intrusions are classified into two categories: External Intrusions caused by unauthorised users of the systems and Internal Intrusions caused by authorised users using system in unauthorized way and beyond the access. Accordingly, approaches used for intrusion detection are broadly divided into two techniques: Signature based detection and Anomaly based detection. In signature-based detection, the "signatures" of the malicious packets are stored in databases and compared with incoming packets. If the signature of an incoming packet matches with the signature in the database, it is flagged as malicious. There is no need of learning about the network traffic in signature based system and works best in the case of known attacks. In anomaly-based detection, behaviour of the network traffic is saved and if incoming packet deviates from the saved model it is marked as malicious. As they use a statistical model of normal behaviour all unknown, attacks are also detected in anomaly-based detections unlike signature-based detection.

Intrusion Detection Systems are classified into Host based IDS and Network based IDS. Host based IDS are systems that monitor the device on which they are installed, whereas, Network based IDS are placed at certain points within a network in order to monitor traffic from and to devices within the network. We have studied and analysed the methodologies and available Network based IDS in this paper.

With the exponential growth of network bandwidth, task of intrusion detection demanded a substantial improvement in both speed and accuracy. The best way to achieve this is to apply artificial intelligence (AI), with its sub-fields – machine learning (ML) – where main idea is learning from data. In this paper analysis of AI algorithms, and performance comparisons of Network IDS like Snort and Suricata is done which uses AI technologies in implementation.

II. ANALYSIS OF AI TECHNOLOGIES

As traditional IDS's are mainly signature-based, detecting only known attacks, their biggest problem is the inability to detect new or variant attacks [2]. Potential solution for solving this problem is AI, or more specifically Machine Learning (ML); a branch of AI that provides computer the ability to learn without being explicitly programmed for particular

purpose. ML deals with the algorithms that learns from previous inputs and make predictions and decisions depending upon them. ML helps to solve three basic problems of detecting and classifying the data, which are classification, regression and clustering. Classification means identifying group membership, regression involves estimating or predicting a response, while in clustering a set of inputs is divided into groups, where members have similar characteristics [2].

The major factors of AI used for detecting intrusions are explained and compared below:

a) Datasets and evaluation

A bunch of databases that contains possible signature of malicious packets is termed as datasets. KDD'99 and DARPA are the two widely accepted and public datasets with 5 million training connection records classified into intrusion and non-intrusion records. These datasets also consists different testing records with known and unknown intrusion records.

There are four possible outcomes of any IDS. TP (True Positive), TN (True Negative), FP (False Positive) and FN (False Negative). Depending on these outcomes, we can find the values of detection rate (DR) and overall accuracy (OA).

$$DR = [TP / (TP + FN)] * 100\%$$

$$OA = [(TP + TN) / (TP+TN+FP+FN)] * 100\%$$

IDS with higher values of DR and OA are more efficient.

b) Neural Network

Neural network methodology helps detection of the fault packages using the technique called as feed-forward. It incorporates error back propagation method to represent both linear and non-linear relations and it has capacity to learn these relationships directly from the data being modelled. The performance of this model tested on the DARPA 1998 dataset was a DR of 77% with 2.2% FP.

c) K- nearest Neighbours

K-nearest neighbours (kNN) is one of the most simple and traditional non-parametric supervised learning algorithms [2]. A cluster is made by including the nodes of Kth nearest distance and considered as a group. Different metrics are used to decide the K factor so that clusters with similar elements can be created. As kNN works well in dynamic environments where frequent updates in dataset is performed, therefore it can be used to detect newer attacks on a network. kNN based IDS achieved DR of 90.28% (k=1) on KDD'99 dataset.

d) Naïve Bays

Naive Bayes (NB) is a supervised learning algorithm based on Bayesian theorem with the "naive" assumption of dependence between every pair of features. Classification is made by combining prior probabilities and likelihood, to form a posterior probability using the so-called Bayes' rule [2]. It is suitable for particularly dimensionally large inputs. NB based IDS and achieved OA of 94.90% on KDD'99 dataset.

e) Decision Tree

Decision tree (DT) is a supervised learning algorithm based on flowchart-like structure in which internal node represents a "test" on an attribute, each branch represents the outcome of the test and each leaf node represents a class label [2]. Depending on previous traversed path from root to leaves, classification rules are defined. On dataset, KDD'99 OA of 92.60% was achieved.

f) Support Vector Machine

Support vector machine (SVM) is a supervised learning algorithm based on concept of decision planes that define decision boundaries. An SVM constructs the hyperplane, or a set of hyperplanes in a high-dimensional space, that separates all data points of one class from those of the other class. The best hyperplane for an SVM is the one with the largest margin between the two classes [2]. It achieved the OA of 86.79% on KDD'99 dataset. Usually combination of SVM and DT gives performance than the individual ones.

While comparing the performance factors of previous experiments we can infer that SVM has the best OA and kNN has the worst OA on KDD'99 dataset. Whereas NN was found better at generalisation but not that efficient in detection new attacks. In comparison to all factors, DT proved to be the most efficient in both generalizations and detecting new attacks.

III. ANALYSIS OF SNORT AND SURICATA

There are many open source IDS tools, we will be discussing about two popular network intrusion detection system.

a) Snort
b) Suricata

a) Snort

Two of most commonly used open source IDS systems are Snort and Suricata. They share many common rule sets. Snort makes use of threat intelligence and writes snort rules that would detect emerging threats.

Structure of snort is divided into multiple components. The packet capture module, detection rules, pre-processor and alert output can be configured separately. [7] Packet Capture library: It will capture packets from different network interfaces. Packet Decoder: It inspects packet headers and examines for any peculiarities. Packet is then decrypted for further processing.

- Pre-processor: Its puts together the TCP stream and decrypts HTTP URI.
- Detection Engine: It applies rules to packets to check the packages against various options in the snort rule files.
- Logging and Alerting system: Used for logging the alerts to an output file.
- Output Plugins: It gives the final output by analysing logs and alerts.
- Features: Flexibility, live and real time, modular detection engine.

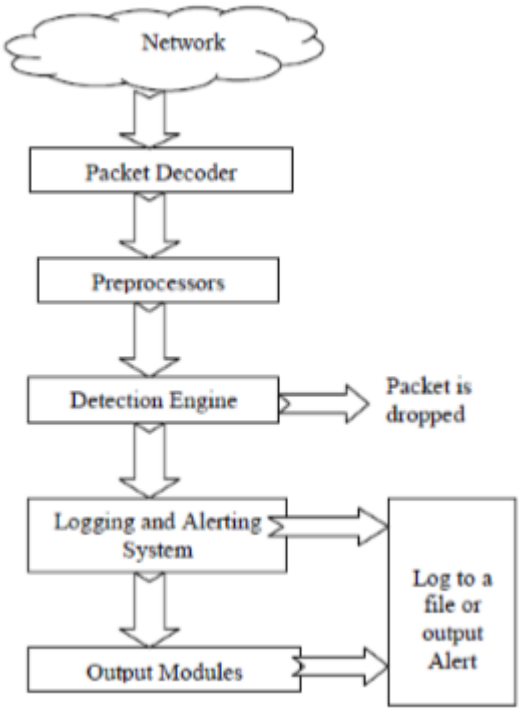

Fig 1

Some of the advantages of Snort are it can be easily installed on any type of network. The technology used by snort adopts a broad adoption rate, which helps to adapt it to emerging threats faster. Example: A snort rule was available to monitor for the vulnerability at the centre of the Equifax breach. Snort can also act as an intrusion prevention system [7]. As it is open source it enables security professionals around the world to develop customized rules and contribute them to community's knowledge.

c) Suricata

Suricata is a somewhat newly discovered open-source IDS. It is evolved by Open Information Security Foundation (OISF). Suricata is suitable for many operating systems like Mac, UNIX, Linux, Free BSD, Windows [7].

Like snort, Suricata is a rule-based system. It offers compatibility with Snort rules and has introduced multithreading [7]. Suricata also incorporates the Lua scripting language, which makes Suricata adapt to complex threats. The structure of Suricata is shown in Fig 2.

1) Capture Module: After the device has been initialised, it collects packets and sends them to Suricata. Suricata then acts as a thin wrapper around the data provided, making it compatible with the link type decoders [7].
2) Decode Module: Decodes the packets captured by the capture module into Suricata supported data structure [7].
3) Detect Module: Following tasks such as loading all signatures, initializing detection plug ins and running all the packets through all the rules, are taken care by detection module [7].

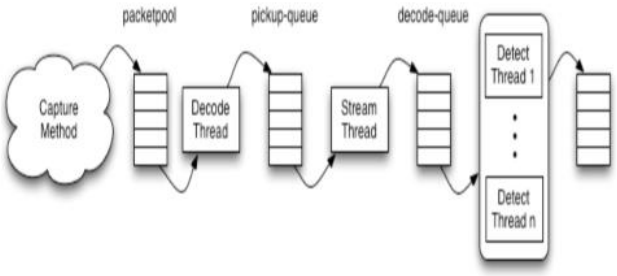

Fig 2

Features of Suricata include high performance, inclusion of intrusion prevention system and network security monitoring.

Multi-threading provides the ability to process more rules across faster networks with larger traffic volumes. It incorporates Lua scripting language, which provides greater flexibility to create rules that identify complex conditions. Suricata also supports all operating systems [9]

General Comparison of Snort and Suricata can be seen in Table 1. Snort and Suricata support most of the popular operating systems. Both IDS systems are licensed under GNU (General Public License). Under this license, the software can be download for free and it is open source. All user has the freedom to make any changes.

Parameters	Snort	Suricata	Bro
Supported Platform	Win, MacOS, Unix	Win, MacOS, Unix	Unix like system, MacOS
License	GNU GPL V2	GNU GPL V2	BSD
IPS feature	Yes	Yes	No
PGP signed	Yes	Not Applicable	No
Support to high speed network	Medium	High	High
Configuration GUI	Yes	Yes	No
Offline Analysis	Yes for multiple files	Yes for single file	Yes for single file
Threads	Single Thread	Multithreaded	Single Thread
IPV6	Yes	Yes	No
Installation and Deployment	Easy	Easy	Difficult

Table 1 [1]

Example of a Suricata and Snort rule is shown in Fig 3.

```
alert tcp $EXTERNAL_NET any -> $HOME_NET 21 (msg:"FTP adm
scan"; flow:to_server,established; content:"PASS ddd@\0A\";
reference:arachnids,332; classtype:suspicious-login; sid:353; rev:6;
```

Fig 3

IV. RESULTS

From the results we obtained by looking at the previously executed experiments which are shown in the below tables we have made the following analysis.

Traffic Rate	TCP			CPU		
	Suricata	Snort	Bro	Suricata	Snort	Bro
50	5.0	11.0	10	9.0	12.0	17
100	9.0	12.0	9.0	11.0	16.0	19.0
200	16.0	15.0	10.0	27.0	18.0	19.0
300	23.0	16.0	30.0	34.0	25.0	20.0
400	22.0	26.0	31.0	44.0	29.0	20.0
500	25.0	32.0	39.0	43.0	29.0	27.0
600	38.0	44.0	40.0	54.0	45.0	33.0
700	45.0	48.0	40.0	67.0	52.0	37.0
800	47.0	50.0	42.0	68.0	56.0	36.0
900	48.0	52.0	43.0	69.0	59.0	37.0
1000	54.0	60.0	43.0	73.0	68.0	43.0
2000	58.0	62.0	46.0	81.0	70.0	55.0

Table 2 [6] CPU Usage

From Table 2: We see that, for TCP traffic, the CPU usages of Snort and Suricata is increased rapidly with increasing traffic rates. Similar trend is shown for UDP traffic, but here Suricata has the highest CPU usage for higher traffic rates.

Traffic Rate	TCP			UDP		
	Suricata	Snort	Bro	Suricata	Snort	Bro
50	0.0	0.0	0	1.0	0.0	0
100	0.0	0.0	0.0	15.8	0.0	0.1
200	0.0	0.0	0.0	17.7	0.0	0.1
300	0.0	0.7	9.0	17.8	0.0	5.1
400	0.0	0.9	9.6	18.0	5.4	8.0
500	0.4	10.1	12.1	18.4	8.3	10.5
600	0.4	10.1	12.5	18.5	15.2	13.6
700	0.7	17.1	18.3	18.7	19.7	15.9
800	8.5	15.6	22.7	22.7	25.4	15.4
900	9.4	30.5	22.9	21.8	27.8	17.6
1000	33.6	35.4	24.2	20.1	28.4	21.5
2000	43.3	45.2	31.4	24.1	30.9	33.8

Table 3 [6] Packet Loss

From Table 3: Here we see that Suricata has a better trend of less packet loss in both types of traffic. Packet loss of Snort is comparatively higher.

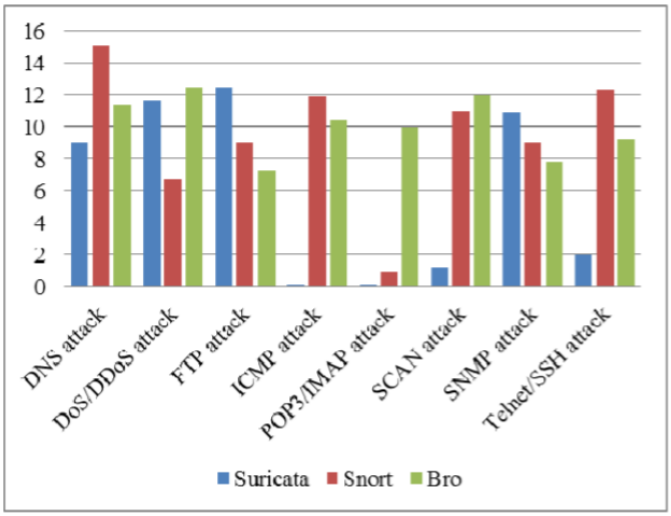

Fig 4 Packet loss of all Attacks

- From Fig 4: This graph compares the number of packet loss for different attack using Suricata Snort and Bro IDS systems. We see that packet loss is high for DNS, DoS, FTP and SNMP attacks for all three IDS. Suricata has the smallest packet loss for ICMP, POP3/IMAP, SCAN, Telnet/SSH [6].

IDS	DNS Attack	DoS/DDoS Attack	FTP Attack	ICMP Attack	POP3/IMAP Attack	Scan Attack	SNMP Attack	Telnet/SSH Attack
Suricata	1221	1189	982	1898	505	1275	921	879
Snort	882	884	913	1700	396	1007	831	760
Bro	948	975	958	1872	442	1391	917	960

Table 4 [6] Number of Alerts for Attacks

- From Table 4: This table gives data for the number of alerts when the IDS are under all eight attacks. The results show higher number of alerts for 'all rule sets' as one packet payload may fit more than one rule and some attacks can have similar payload.

V. CONCLUSION

Intrusion detection systems are considered one of the best technologies to detect threats and attacks. Two of the major open source IDSs, Snort and Suricata have been analysed and compared. The results show that, Suricata utilizes higher number of computational resources (CPU usage) but it also processes a higher number of packets per seconds using multithreading. Because of multithreading, Suricata also requires more memory (RAM) than Snort and Snort is not capable of handling the networks with bandwidth beyond 300 Mbps. Thus, we can conclude that Suricata outperforms Snort, even when it uses single core.

While the analysis of these IDS systems, we came across an upcoming IDS system known as 'Bro'. This intrusion detection system provides the facility of experimentation, which in turn helps in developing new ways to look at the data. As it has not been adapted on a larger scale, many improvements to it are still possible. One of the possible ways of improving the performance of IDS systems is by constructing a hybrid system. We can achieve this by implementing the features of Suricata such as multithreading into Bro.

Another way of improvement is by the use of artificial intelligence based techniques like Machine Learning for making the systems capable of detecting and predicting the status of unknown packet-ids. Analysing different algorithms of ML for their Detection Rate and Overall Accuracy, we conclude that Decision Tree algorithms works better than all other algorithms. However, a research in hybrid algorithms, by cascading it one after the other will yield a better output.

One of the ways to improve existing IDS system using artificial intelligence is by combining signature based rules with anomaly based detection profiles and giving this as an input to a recommender system which in turn would recommend the suitable algorithm for any specific type of network and even the most matching dataset to reduce the learning time.

VI. REFERENCES

[1] Building Recommender Systems for Network Intrusion Detection Using Intelligent Decision Technologies, Mrutyunjaya Panda.

[2] Artificial Intelligence in Network Intrusion Detection, Poojitha.G Naveen kumar .K JayaramiReddy.P

[3] Intrusion Detection System- Types and Prevention, B.Santos Kumar, T.Chandra Sekhara Phani Raju, M.Ratnakar, Sk.Dawood Baba, N.Sudhakar.

[4] Intelligent intrusion detection systems using artificial neural networks, Alex Shenfielda, David Dayb, Aladdin Ayeshb.

[5] An Implementation of Intrusion Detection System Using Genetic Algorithm Mohammad Sazzadul Hoque, Md. Abdul Mukit and Md. Abu Naser Bikas

[6] Evaluation Studies of Three Intrusion Detection Systems under Various Attacks and Rule Sets, Kittikhun Thongkanchorn, Sudsanguan Ngamsuriyaroj, Vasaka Visoottiviseth

[7] Comparative Study and Analysis of Network Intrusion Detection Tools, Dhanashri Ashok Bhosale, Vanita Manikrao Mane.

[8] Statistical Analysis of Different Artificial Intelligent Techniques applied to Intrusion Detection System, Hind Tribak, Blanca L. Delgado-Marquez, P.Roj, O.Valenzuela, H. Pomares, I. Roj.

[9] A Realistic Experimental Comparison of the Suricata and Snort Intrusion-Detection Systems, Eugene Albin and Neil C. Rowe.

[10] Intrusion Detection using Artificial Neural Network, Poojitha.G Naveen kumar. K JayaramiReddy.P.