A Comparative Study of Different Fuzzy Classifiers for Cloud Intrusion Detection Systems' Alerts

Saeed M. Alqahtani
School of Computer Science
PhD Student, ASAP and LUCID Groups
Nottingham University
Email: psxsa22@nottingham.ac.uk

Robert John
School of Computer Science
ASAP and LUCID Groups
Nottingham University
Email: robert.john@nottingham.ac.uk

***Abstract*—The use of Internet has been increasing day by day and the internet traffic is exponentially increasing. The services providers such as web services providers, email services providers, and cloud service providers have to deal with millions of users per second; and thus, the level of threats to their growing networks is also very high. To deal with this much number of users is a big challenge but detection and prevention of such kinds of threats is even more challenging and vital. This is due to the fact that those threats might cause a severe loss to the service providers in terms of privacy leakage or unavailability of the services to the users. To incorporate this issue, several Intrusion Detections Systems (IDS) have been developed that differ in their detection capabilities, performance and accuracy. In this study, we have used SNORT and SURICATA as well-known IDS systems that are used worldwide. The aim of this paper is to analytically compare the functionality, working and the capability of these two IDS systems in order to detect the intrusions and different kinds of cyber-attacks within $MyCloud$ network. Furthermore, this study also proposes a Fuzzy-Logic engine based on these two IDSs in order to enhances the performance and accuracy of these two systems in terms of increased accuracy, specificity, sensitivity and reduced false alarms. Several experiments in this compatrative study have been conducted by using and testing ISCX dataset, which results that fuzzy logic based IDS outperforms IDS alone whereas FL-SnortIDS system outperforms FL-SuricataIDS.**

***Index Terms*—Cloud Computing; IDS; Fuzzy Logic; Snort; Suricata; ISCX dataset.**

I. Introduction

With the fast growing digital technology, computer networks have been extensively developed and deployed world widely, allowing the users to communicate with each other [1]. In this modern age, digital communication is no more a big task and thus every single internet user can have an access to on-line information pool or can interact with anyone without worrying about the distance between them. According to the statistics reported in [2], the total estimated internet users in the year of 2016 are approximately over 3,424,971,237, which is 46.1% of the world population. Hence, we can conclude that internet has now become a part of modern age. Computer networks are being attacked every day and therefore they are unreliable and unsafe, which means that the users may experience malicious activities and may lose their privacy, personal data or any other important information that is available on-line, depending on the nature of attacks. For a normal user, this may not possess any real concern, but for people and firms which want their data to be private. Similarly, corporate offices, banks, hospitals, law enforcement organisations, emails services providers and millions of other organisations take extreme care of their privacy and availability of their services on-line [3],[4],[5].

Computer network attack, also known as Cyber-Attack was defined by Waxman that refers to any unwanted or unethical activity that is intended to disturb, alter or hit someones privacy or to steal others important data either secretly or publically [6]. These types of attacks are usually performed by anonymous hackers and it is very difficult to recognise the hackers or to catch them [7]. Cyber-attacks are performed using multiple methods such as, secretly installing spy software in the targeted systems [8], secretly attempting to log in the targeted system successfully [9] or secretly monitoring the internet traffic of the targeted system [10]. Cyber-attacks include, but are not limited to Malware, Phishing, Password Attack, Denial-of-Service (DoS) Attack, Man in the Middle (MITM) Attack, Drive by Downloads, Malvertising, Rogue Software and many more [11].

II. Related Work

Cyber-attacks are the modern age way of warfare accessing and exploiting private and secret data of a country, and therefore the cyber war has taken over the nuclear war in this modern age [12]. Thus, many international rules have been created by the law enforcement agencies including USAs [13], [14], [15]. It has also attracted the attentions of many researchers and a lot of work has been done in literature to protect the systems from cyber-attacks such as, inventory of authorised and unauthorised software and devices, to make configurations of hardware and software secure, to install intelligent firewalls, to install anti-malware software, to develop intrusion detection systems and to develop malware defensive systems. The most followed strategy to prevent such kind of cyber-attacks is the development of intrusion detection system (IDS) [16].

IDS systems are basically hardware or software systems that are deployed along with the main systems to monitor all the digital activities and the incoming as well as outgoing network traffic. These systems are made intelligent enough to detect the malware or suspicious activities by monitoring the whole

system; and therefore, they produce alarms or reports against such activities. IDS system acts as a firewall and keeps the main system safe from the malwares. Hence, it is deployed along with almost every critical system that is exposed to threats, making the organisation reliable and trustworthy.

The capability of IDS systems to detect the suspicious activities depends on how they have been developed. Stronger the IDS system would be, safer the main system would be, leading the organisation to win the trust of its clients. Moreover, IDS systems are consistently upgraded due to the fact that the cyber-attacks are becoming crucial and stronger day by day. A great deal of work has been done in literature in making intelligent and strong IDS systems. For example, reputation services have been added in the IDS systems. These services gather information about the suspicious protocols, IP addresses, domains and finally make a decision that either the traffic is malicious or not [17]. Transforming the wired IDS systems to wireless systems has also increased the safety level of the critical systems [18]. With the fast growing HTTPS traffic, the SSL traffic inspection feature has also been added in the IDS systems to stay up to date [19]. Klir stated that fuzzy logic has been widely used in the IDS systems because it helps increase the intrusion detection rates and thus significantly strengthens the IDS systems [20].

The paper of [21] presents a new Fuzzy-Genetics based hybrid approach that is considered to be superior then previously developed Genetic-Algorithm (GA) based approaches which do not have high capability of intrusion detection. The proposed approach adds in the GA based system, an ability to change according to the networking environment, to handle the noise and to detect intrusions in the system with significant accuracy. It is based on two major steps, including GA algorithm as an initial step to produce subset of the communication features by using traditional dimensional reduction technique and the next step as defining a set of fuzzy logic rules such as trapezoidal fuzzy sets that allow complete membership over all ranges. This approach has been tested by KDD Cup 1999 Dataset, and results show that the intrusion detection rate accuracy is above 90% whereas the false positive rate is below 1%.

The paper of [22] discusses a very major and most common security challenge such as blackhole attack in Mobile Ad hoc Networks (MANETs) and also presents the corresponding solution by utilising the strengths of fuzzy logics. The proposed approach is comprised of four stages including fuzzy parameter extraction where the initial parameters are extracted based on the incoming network traffic, fuzzy computation which calculates the fidelity level on the bases of extracted parameters where the fidelity level defines the intrusion level of the packet, fuzzy verification module where a decision is made that either the blackhole attack exists or not; finally, an alarm module generates an alarm in case of blackhole identification. The approach has been applied on routing protocol and is simulated by varying the input parameters such as, mobility of nodes and traffic speed. It has been found that the blackhole attack detection is more accurate and the false detection ratio was also very low.

The paper of [23] introduces different kinds of attacks on internet including, Probe Assaults, DoS Attacks, R2L Attacks, U2R Attacks, Checking Attacks, Dissent of Service Attack, Infiltration Attack and describes the kinds intrusion detection systems that are, grouping, example mining, information mining procedures, computerised reasoning systems and delicate registering methods. The paper also presents a fuzzy logic enabled, oddity based intrusion detection system that is developed using information mining procedures to increase the intrusion detection rate as well as accuracy. The proposed approach has been divided into four steps including, classification of preparing information where the interested information is gathered, strategy for era of fuzzy guidelines where all the fuzzy sets are generated, fuzzy choice module, where a decision is taken about the nature of incoming traffic and the last step is to find the suitable order for a test information where the final decision is taken that either the incoming packet is assaulted or not. The approach is applied in a network and tested by introducing different kinds of attacks and results shows that the assaults detection rate is very high as well as accurate.

Denial of Service (DoS) attack has been addressed in the paper of [24] where such a study utilises fuzzy logic based intrusion detection approach in order to deal with this attack. The proposed approach leverages the fuzzy logic by applying it over an already developed IDS System with an aim to improve the detection rate of such attack whereas the IDS system is based on MCA-based DoS attack detection system. The MCA-based system works on triangle based MCA-based technique that involves the extraction of geometrical correlation of the mutually exclusive features. The proposed approach is tested by exposing it to KDD CUP99 data set and results indicate that the DoS attack detection rate has been considerably improved after applying fuzzy logic.

Several different tools are also available that perform Intrusion Detection. For instance, Security Onion system has the capability to monitor vLANs and virtualised networks but it cannot be used as an intrusion prevention system [25], [24]. OSSEC system can generate real-time alarms and has the capability of monitoring the files integrity [26]. OpenWIPS-NG system is dependent on network interfaces, devices, servers and other infrastructures [24]. BRO system is an alternative to Security Onion but has more defined rules to detect the malicious activities [27]. Among all IDS systems, Snort is considered to be the most efficient tool that performs real-time protection, real-time traffic analysis, protocol analysis, content matching, packets logging on IP networks and possesses many kinds of attacks detection capability [28].

This study discusses Snort and Suricata as IDS systems, which are namely SnortIDS/SuricataIDS, its strengths and capabilities, a demonstration of SnortIDS/SuricataIDS system using ISCX datasets[29] and finally proposes a new technique to increase the SnortIDS/SuricataIDS malicious activities detection rate by utilising the strengths of fuzzy logic. ISCX datasets are sets of malicious activities that are offered to the

IDS systems to analyse the capability of IDS systems to detect them [30]. If the detection rate is high and accurate, we can conclude that the IDS system is stronger enough to be used for live traffic. Several other datasets are also available for testing IDS systems such as, KDD CUP-1999, but they are not realistic [31].

Similarly, the statistical parameters that are used to describe the overall performance of $MyCloud$ as well as the capability of the underlying IDS system are: (1) specificity and (2) sensitivity. Specificity, commonly termed as true negative rate, is a parameter whose value represents the proportion of negatives that have been correctly identified as true. On the contrary, sensitivity, commonly termed as true positive rate, is a parameter whose value represents the proportion of positives that has been correctly identified as true.

III. An investigative Approach

A. *ISCX Dataset*

ISCX dataset is provided with a set of complete traffic of real-time network, carefully acquired for the applications which include web browsing (HTTP, HTTPS), mails (SMTP, POP, IMAP) and file sharing (FTP). It is simulated to provide real time network traffic for IDS from which IDS can detect different anomalies in the pattern of traffic, and generate different alerts ISCX data set is traffic of 7 days of activity of an agent that contains these five types of traffic, that needs to be analysed, 1) Normal Traffic 2) Infiltrating Network from Inside 3) HTTP Denial of Service 4) Distributed Denial of Service 5) Brute force SSH This traffic is divided into 7 days of real-time traffic, each day file ranging from 4 GBs to 23.4 GBs. In order to analyse such a traffic, we had to run SnortIDS/SuricataIDS in offline mode, which have their limits as they cannot read a trace file greater than 200MBs which varies depending on the system. Therefore, the only option was to split the per day files into different small files, which then read by them; and thus, they can provide alerts. An important feature of IDS is, in a single run, can read multiple files provided in the folder while maintaining states of previous connections. Until now, we have ISCX Dataset which is categorised with respect to dates in folder, and is split, ranging 50+files/day -450+files/day.

B. $MyCloud$

The hardware and software that were utilised to run $MyCloud$ in our experiment are 4 PCs that were used in this experiment as two of them running ESXi5.5 servers, one was run for vCenter Server and the last one was run Active Directory, vShield Server and vCloud namely $MyCloud$. We also used the following software: VMware Workstations for IDS Server, IPS Server, and Syslog Server, VMware Cloud Suite that includes vCenter Server, vShield Manager, Active Directory and vCloud, VMware Convertor Machines (VMs) in order to deploy above servers into $MyCloud$. For efficiency matter, router and switch were used. After performing several attempts, Table 1 shows the requirements to build a virtual cloud as we identified these servers to be the configuration of choice to run $MyCloud$. Subsequently, we performed installation and configuration for the above servers and the following machines: vIDS Server, vIPS Server, vSyslog Server, vAttcker, vCenter Server, vShield Manager, Active Directory and vCloud as well as two of ESXi5.5 servers.

TABLE I: $MyCloud$ Lab Specifications

Requirements		Specifications	
1st PC	1st ESXi Server	Process	Intel(R) Core(TM) i7 CPU
		RAM	16 GB
		System Type	64-bit Operation System
		Hard Disk	1TB
		IP address	192.168.1.48
2nd PC	2nd ESXi Server	Process	Intel(R) Core(TM) i7 CPU
		RAM	16 GB
		System Type	64-bit Operation System
		Hard Disk	1TB
		IP address	192.168.1.49
3rd PC	vCentre Server	Process	Intel(R) Core(TM) i7 CPU
		RAM	16 GB
		System Type	64-bit Operation System
		Hard Disk	1TB
		IP address	192.168.1.50
4th PC	Active Directory	Process	Intel(R) Core(TM) i7 CPU
		RAM	16 GB
		System Type	64-bit Operation System
		Hard Disk	1TB
		IP address	192.168.1.51
	vShield	IP address	192.168.1.54
	vCloud	IP address	192.168.1.66

C. *IDS Systems*

There is a great deal of open source Intrusion Detection tools available. The use of these tools depends on the user or administrator. Some of them for monitoring hosts and others are for the networks connecting them to identify the latest threats. The IDS systems: Snort[32] and Suricata[33] were utilised for comparison purposes as they are considered one of the most effective and accurate open source tools. In this study, we implemented these tools in order to pre-process the fuzzy classifiers: FL-SnortIDS/FL-SuricataIDS.

SnortIDS is an open source, rule based Intrusion Detection System provided by Cisco. It is now also being used as Intrusion Detection and Prevention System. SuricataIDS is also another open source IDS system that has been developed by a foundation i.e., Information Security Foundation (OISF).

Both the above mentioned IDSs are widely used around the globe making any network infrastructure safe and reliable by detecting and resisting the well-known cyber-attacks or malwares by evaluating the incoming network traffic. These both IDSs use rule-based language and their working can be classified into four major stages which are packet decoding, packet preprocessing, intrusion detection and alerts generation. Alternatively they also possess other important features including packet logging and packet sniffing. These IDSs are usually deployed right next to firewall or gateway router.

These IDSs makes decisions about the activities either to be regular or malicious, on the bases of some predefined rules. These rules have been set by the respective community and are applied for the evaluation of incoming network traffic. With the ever growing on-line communication technologies, the network traffic is becoming more and more complex day by day; hence the results obtained by applying such predefined rules and keeping track of the changes is a very tiresome effort and might become outdated up to some extent.

D. IDS Fuzzy Classifier

Once SnortIDS/SuricataIDS demonstrated the experimental results against ISCX dataset, it concludes that the false detection rate is high enough that it cannot be ignored; and thus, it requires a serious attention. In order to deal with this issue, IDS fuzzy classifiers were built for these IDS called FL-SnortIDS/FL-SuricataIDS. The fuzzy logic based IDS approaches have been presented in this section which refurnishes the alerts generated by the SnortIDS/SuricataIDS systems; and then it takes extra-cautious decisions that either the incoming traffic is actually a regular traffic or malicious. These approaches enhance the performance and accuracy of these two systems considerably. In terms of increased accuracy, specificity and sensitivity and reduced false alarms. The alerts generated by SnortIDS are not categorised in any manner, which may help us identify the real threats vs. alerts generated by bad network or sometimes a simple mistake in credentials that can cause an alert. Thus, these alerts need to be categorised by the types of attack they represent. The alerts generated by SuricataIDS is much like SnortIDS that is a list of long unsorted lines, which is very difficult for any network administrator to understand. Therefore, it is very important to learn to read the log provided by SnortIDS or SuricataIDS so the attack classifications can be arranged as desired.

After extensive analysis of the alert files which were generated by SnortIDS/SuricataIDS, these alerts were programmatically categorised on the basis of alert classification. Unknown Traffic alert of SnortIDS contains 46% of alerts. This alert was being generated against HTTP_INSPECT rules, where size of transferred data was not the same as already communicated.

For SuricataIDS system, it shows that GENERIC Protocol Command Decode alert contains 97% of alerts. These alerts were being generated against HTTP_INSPECT and TCP_INSPECT rules, where size of transferred data was not the same as already communicated or the window size was different. There are many reasons for these alerts to be generated. It may be due to a bad network, or wrong configuration of HTTP server, but as the communication between server and client is established legitimately, so these are the alerts we can remove from the alert files of SnortIDS or SuricataIDS, as these are not the work of any intruder. It is just some server error. The table II shows the alerts classified in both systems and removing any unwanted and false alerts which are green coloured .

Potentially Bad Traffic alert generated by SnortIDS is 18% of the alerts. This alert was being generated by an FTP server that used to generate an extra reset flag to make sure the connection was terminated, a services hosted on servers like AKAMI and such servers generate extra resets making sure that connection is terminated, where SnortIDS deals it as an unknown connection packet as SnortIDS has already removed that connection from its memory. Hence, SnortIDS classifies this alert as Potentially Bad Traffic. One more reason for the alert to be generated for SuricataIDS is an ill configure FTP server, which was generating an extra reset flag to make sure the connection was terminated, a services hosted on servers like AKAMI will cause these issues, where SuricataIDS deals it as application error packet as SuricataIDS has already removed that connection from its memory. Hence, SuricataIDS will generate per packet threat. The table below shows the alerts before removing any unwanted and false alerts.

Similarly working on the alert files for both systems: SnortIDS/SuricataIDS, the following types of alerts were discarded by carefully analysis of the traffic of ISCX Dataset. This exercise is always done by network administrators when installing new IDS. We configured the rules of IDSs with respect to the traffic but SnortIDS/SuricataIDS were not a network aware IDS, hence the administrators cannot remove some rules randomly. For this reason, we used a fuzzy logic controller to carefully remove the rules. The unwanted alerts types for SnortIDS/SuricataIDS are shown in the table below

These alerts for both systems: SnortIDS/SuricataIDS were generated mostly due to ill-configured services. Some alerts were being generated due to network congestion and drop packets. Besides these alerts, all other alerts posed a real threat to network and devices by injecting some kind of malware, or trying to access password protected files.

IV. How Does Fuzzy Classifier Work

First of all, we have a fuzzifier that makes the alerts generated by SnortIDS or SuricataIDS into understandable alerts. Fuzzifier classifies the alerts into different categories. These categorised alerts are the inputs of the FL controller where on the basis of alert types; these alerts are further categorised as threat or false alerts. We have a basic minimum amount of 3 alerts per generated alert, to call it an illegal activity e.g. Network policy dictates, a user gets 3 passwords attempts per day over domain. Thus, if in case, a user mistakenly put the password wrong, an alarm is generated but it is not a threat because he/she is a legitimate user. If the retries count increased to 3, then the user gets blocked for that day. This means that if the total numbers of attempts to log in by a user are greater than allocated retries, it will be considered as a potential threat and will be presented on the threat screen; due to the fact that an authorised user can never miss hit the password thrice and still be unblocked. The whole process of accurate threat detection has been divided into three major stages: 1. Alert classification 2. Threat detection 3. Threat severity

The initial stages intends to refurnish the already generate alerts, as generated by the typical SnortIDS/SuricataIDS systems. This stage helps increase the accuracy of true threat detection and mitigates the inaccuracy of false threat detection. Afterwards, these classified alerts are passed through the threat detection engine which detects the potential threats. Finally, we checked the total number of potential threats generated against single activity such as, login. It helps us differentiate from alert and threat. For instance, if this number exceeds three, which is the predefined threshold, the potential threat is marked as a genuine threat; otherwise it is considered as a mistake and thus ignored.

TABLE II: Alerts Classified for IDS Systems Including Unwanted Alerts Types

No	Alert Types	Number of Alerts		Ratio	
		SnortIDS	SuricataIDS	SnortIDS	SuricataIDS
1	Network Trojan	2075	2119	0.82645	0.618417
2	Access to a Potentially Vulnerable Web Application	10	2	0.003983	0.000584
3	Attempted Administrator Privilege	44887	336	17.878	0.09806
4	Attempted Denial of Service	28	27	0.011152	0.00788
5	Attempted Information Leak	16208	220	6.455467	0.064206
6	Attempted User Privilege Gain	5	3	0.001991	0.000876
7	Detection of a Network Scan	1	None	0.000398	None
8	Detection of a Network Scan	5	None	0.001991	None
9	Executable Code was Detected	115	None	0.045803	None
10	Generic Protocol Command Decode	11661	335570	4.644447	97.93404
11	Information Leak	3	None	0.001195	None
12	Misc Attack	2011	445	0.800959	0.129871
13	Misc activity	3	4	0.001195	0.001167
14	Not Suspicious Traffic	215	215	0.085632	0.062746
15	Potential Corporate Privacy Violation	9838	2755	3.918367	0.80403
16	Potentially Bad Traffic	45610	596	18.16596	0.173939
17	Unknown Traffic	116665	None	46.46638	None
18	Unsuccessful User Privilege Gain	4	2	0.001593	0.000584
19	Web Application Attack	1730	34	0.68904	0.009923
20	Detection of a Non-Standard Protocol or Event	1	None	0.000398	None
21	(Null)	None	321	None	0.093682

V. Experimental Results

A. *Methodology*

Final results for all these systems were compiled and the comparison of these systems was done on the basis of these matrices:

1) Numbers of threats detected (Accuracy)
2) False positives and false negatives ratio per system (False Alarms Ratio)
3) Sensitivity Ratio
4) Specificity Ratio
5) Threat Detection Rate (DR)

Accuracy of any system is determined by the ratio of true positives and true negatives detected vs. all connections; this provides us with a matrix that how accurate threats and non-threats are differentiated. It can be calculated by:

$$AccuracyRatio = \frac{(Numberofcorrectassessment)}{(Numberofallassessments)}$$

False Alarm ratio tells us how many connections are falsely categorised as threats or legitimate connections. False Alarm Ratio can be measured by the following equation:

$$FalseAlarmRatio = \frac{(Numberoffalsepositiveassessment)}{(Numberofallnegativeassessment)}$$

Sensitivity ratio tells us that our IDS detected how many threats vs. actual threats, while specificity ratio tells us our IDS treated legitimate connections as threats vs. all clean traffic. Sensitivity and specificity of a system can be measured using the following equation:

$$SensitivityRatio = \frac{(Numberoftruepositiveassessment)}{(Numberofallpositiveassessment)}$$

$$SpecificityRatio = \frac{(Numberoftruenegativeassessment)}{(Numberofallnegativeassessment)}$$

Threat detection rate is the rate of detection of threats per system, classified as low, medium, and high. In this study, our aim was to identify the performance of which of these systems: SnortIDS, SuricataIDS, FL-SnortIDS/FL-SuricataIDS is better than others. In order to do this, we set two hypotheses based on the comparison matrices above. The first hypothesis was designed for sensitivity, specificity, and accuracy while the other one was for the false alarm ratio. The first hypothesis was as a follows;

- Null Hypothesis : Performance of two methods are identical (i.e. $\mu 1 = \mu 2$).
- Alternative Hypothesis : Performance for one method significantly improves over other methods (i.e. $\mu 1 > \mu 2$).

For false alarm ratio, we set the following hypothesis;

- Null Hypothesis : False Alarm ratio of two methods are identical (i.e. $\mu 1 = \mu 2$).
- Alternative Hypothesis : False Alarm ratio for one method significantly lesser than the other methods (i.e. $\mu 1 < \mu 2$).

Our approach for testing the ISCX dataset against 4 systems is to compare the two independent results of each sensitivity, specificity, false alarm ratio and accuracy for SnortIDS vs FL-SnortIDS, SuricataIDS vs FL-SuricataIDS, SnortIDS vs SuricataIDS, SnortIDS vs FL-SuricataIDS respectively. As an essential criteria, we checked for the normality assumption with Shapiro Test for each of the category above and figure out that none of our sample data does satisfy the normality assumption, so we applied then the non-parametric test for two sample comparison for each category above viz. Mann-Whitney Test. Mann-Whitney Test was used to compare two population means that come from same population by using this equation.

$$U = n_1 n_2 + \frac{n_2(n_2+1)}{(2)} \sum_{i=n_1+1}^{n_2} R_i$$

where,

n_1 : sample size of sample 1

n_2 : sample size of sample 2

Ri : Rank of sample (whose rank is greater)

TABLE III: Generated Alerts Classification on $MyCloud$

	DoS	Probe	R2L	U2R
SnortIDS	2103	142247	61818	46915
SuricataIDS	2146	339344	816	343
FL-SnortIDS	2103	1846	16208	44909
FL-SuricataIDS	2146	9	225	338

For detection rate, our approach was to calculate the detection rate number of threats detected vs total stream and then get them categorised in high, medium, low priority classes. This will be calculated overall of each system. We then normalised the 3 steps of detection rate from 0-1. After getting these values for each system, we obtained a final result for each system. We defined the threshold for law, medium, and high as a follows;

$$high > 0.2$$
$$0.2 > medium > 0.01$$
$$low < 0.01$$

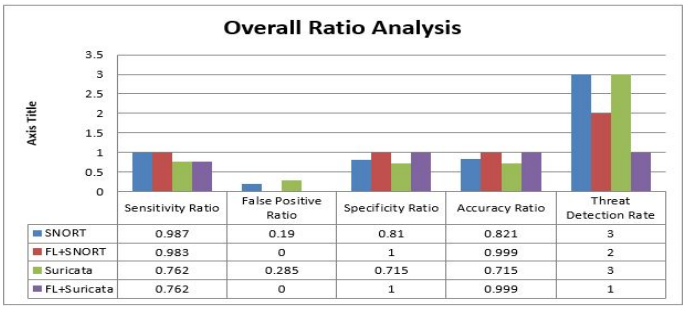
B. *Descriptive Statistics*

Figure 1 shows the overall results of both IDS and FL Systems. SnortIDS system analysed the total of 1268735 connection streams of the ISCX Dataset, out of which SnortIDS generated 251,074 alerts connections for SnortIDS and 342,649 alerts connections for SuricataIDS. These alerts for both systems contain malicious or anomaly alerts. The numbers show that SnortIDS classifies 19.78% of traffic as malicious while SuricataIDS classifies 27.01% of traffic as malicious. With regards to FL based IDS systems,

As it can be seen in the figures below the overall results of all systems: SnortIDS, SuricataIDS, FL-SnortIDS/FL-SuricataIDS. The first figure shows that IDS systems analysed the total of 1268735 connection streams of the ISCX Dataset, out of which SnortIDS generated 251,074 alerts connections for SnortIDS and 342,649 alerts connections for SuricataIDSout while FL-SnortIDS generated 65,066 alerts connections and 2743 alerts connections for FL-SuricataIDS. These alerts for all systems contain malicious or anomaly alerts. The numbers show that SnortIDS classifies 19.78% of traffic as malicious while SuricataIDS classifies 27.01% of traffic as malicious, while in case of FL-SnortIDS these numbers reduces to 5% and when FL is applied on SuricataIDS this number is less than 1%.

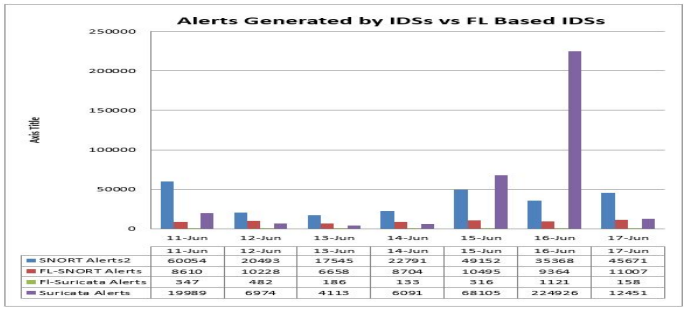
Figure I shows the ratio analysis for these four systems in terms of sensitivity, specificity, accuracy and false alarm. The detection rate tells the network administrator that at what rate the alerts are generated the greater the detection rate means the higher numbers of alerts are generated. In both the cases on average more than 20% of traffic is marked malicious, generating a high detection rate.

The total generated alert types for these IDS systems were 34 alert types: 19 for SnortIDS and 15 for SuricataIDS. The alerts then were classified into attack classifications. 203 of which for SnortIDS and 152 for SuricataIDS. Based on these



	Sensitivity Ratio	False Positive Ratio	Specificity Ratio	Accuracy Ratio	Threat Detection Rate
SNORT	0.987	0.19	0.81	0.821	3
FL+SNORT	0.983	0	1	0.999	2
Suricata	0.762	0.285	0.715	0.715	3
FL+Suricata	0.762	0	1	0.999	1

Fig. 1: Overall Ratio Analysis on $MyCloud$



	11-Jun	12-Jun	13-Jun	14-Jun	15-Jun	16-Jun	17-Jun
SNORT Alerts2	60054	20493	17545	22791	49152	35368	45671
FL-SNORT Alerts	8610	10228	6658	8704	10495	9354	11007
Fl-Suricata Alerts	347	482	186	133	316	1121	158
Suricata Alerts	19989	6974	4113	6091	68105	224926	12451

Fig. 2: $MyCloud$ Alerts Generated by IDS vs FL Based IDS

classifications, attacks were prioritised based on its priority. This priority shows how dangerous this attack can be for $MyCloud$, 1 being highest and 4 being the lowest. We went a step further to categories these alerts for each system into four attack classes that are DoS, Probe, U2R and R2L. In terms of fuzzy classifiers, the total generated alert types for them were 9 types: 5 for FL-SnortIDS and 4 for FL-SuricataIDS. The alerts then were classified into attack classifications. 77 of which for FL-SnortIDS and 46 for FL-SuricataIDS. Based on these classifications, attacks were prioritised based on its severity 1 as a high, 2 as a medium and 3 as a low. We then categorised these alerts classifications for each system into four attack classes that are DoS, Probe, U2R and R2L. Figure 2 shows the number of these alerts in each day for each dataset.

VI. Comparative Analysis

Based on the experimental $MyCloud$ datasets, we have conducted 5 comparisons: SnortIDS vs FL-SnortIDS, SuricataIDS vs FL-SuricataIDS, SnortIDS vs SuricataIDS, Snor-

tIDS vs FL-SuricataIDS, and SnortIDS vs SuricataIDS vs FL-SnortIDs vs FL-SuricataIDS respectively.

A. SnortIDS vs FL-SnortIDS

Figure 3 states the alternative hypothesis as the true location shift is greater than 0. In sensitivity, the level of significance was greater than 0.5 ($p - value > 0.05$). Hence, we do not have sufficient evidence to reject our null hypothesis i.e. sensitivity performances on both methods are the same. This is can be clearly seen from the box-plot visualisation as well as Mann-Whitney test that there is no difference in performance of sensitivity between two methods. For specificity and accuracy, the level of confidence was $p - value < 0.05$, and therefore, we have sufficient evidence to reject our null hypothesis i.e. specificity and accuracy performances for SnortIDS is better than FL-SnortIDS. For false alarm performance, the true location shift is less than 0 and the level of confidence was $p - value < 0.05$. Therefore, we have sufficient evidence to reject our null hypothesis i.e. false alarm ratio for FL-SnortIDS is lesser than the SnortIDS.

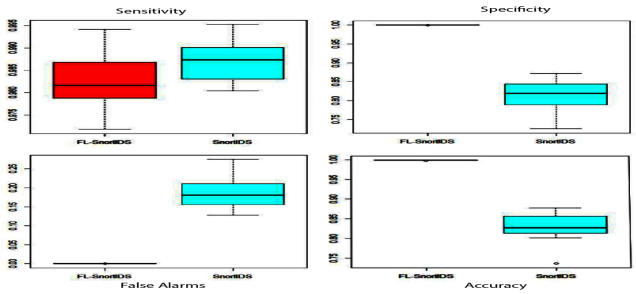


Fig. 3: SnortIDS vs FL-SnortIDS

B. SuricataIDS vs FL-SuircataIDS

Figure 4 states the alternative hypothesis as the true location shift is greater than 0. In sensitivity, $p - value > 0.05$, hence we do not have sufficient evidence to reject our null hypothesis i.e. sensitivity performances on both methods are identical. The box-plot visualisation and Mann-Whitney test show that there is no difference in performance of sensitivity between two methods. For specificity and accuracy, the level of confidence was $p - value < 0.05$, and therefore, we have sufficient evidence to reject our null hypothesis i.e. specificity and accuracy performances for FL-SnortIDS is better than SnortIDS. For false alarm performance, the true location shift is less than 0 and the level of confidence was $p - value < 0.05$. Therefore, we have sufficient evidence to reject our null hypothesis i.e. false alarm ratio for FL-SuricataIDS is lesser than the SuricataIDS.

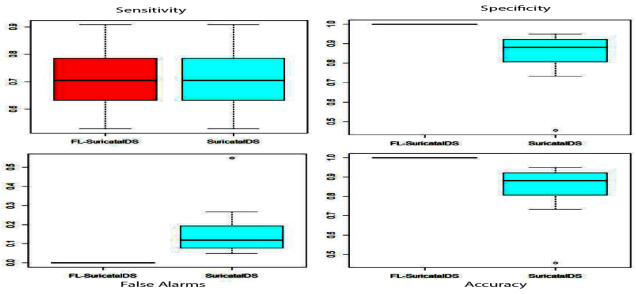


Fig. 4: SuricataIDS vs FL-SuricataIDS

C. SnortIDS vs SuircataIDS

Figure 5 states the alternative hypothesis as the true location shift is greater than 0. In sensitivity, $p - value < 0.05$, hence we have sufficient evidence to reject our null hypothesis i.e. sensitivity performance for SnortIDS is better than the SuricataIDS. The box-plot visualisation and Mann-Whitney test show that sensitivity performance is better for SnortIDS is better than the SuricataIDS. For specificity and accuracy, the level of confidence was $p - value > 0.05$, and therefore, we do not have sufficient evidence to reject our null hypothesis i.e. specificity and accuracy performances for SnortIDS are similar to SuricataIDS. For false alarm performance, the true location shift is less than 0 and the level of confidence was $p - value > 0.05$. Therefore, we do not have sufficient evidence to reject our null hypothesis i.e. false alarm ratio for SnortIDS is same as of SuricataIDS.

D. FL-SnortIDS vs FL-SuircataIDS

Figure 6 states the alternative hypothesis as the true location shift is greater than 0. In sensitivity, specificity and accuracy, $p - value < 0.05$, hence we have sufficient evidence to reject our null hypothesis i.e. sensitivity, specificity and accuracy performances for FL-SnortIDS are better than FL-SuricataIDS. The box-plot visualisation and Mann-Whitney test show that sensitivity, specificity and accuracy performances are better for FL-SnortIDS than FL-SuricataIDS. For false alarm performance, the true location shift is greater than 0 and the level of confidence was $p - value < 0.05$. Therefore, we have sufficient evidence to reject our null hypothesis i.e. false alarm ratio for FL-SuricataIDS is lesser than FL-SnortIDS.

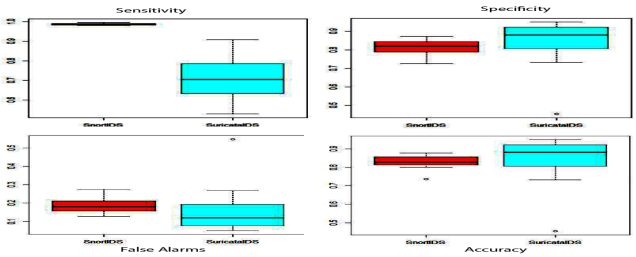


Fig. 5: SnortIDS vs SuricataIDS

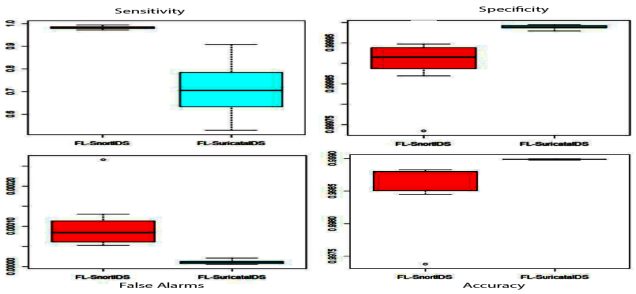


Fig. 6: FL-SnortIDS vs FL-SuricataIDS

TABLE IV: Final Comparison

Systems	Sensitivity	Specificity	False Alarm Ratio	Accuracy
SnortIDS vs FL-SnortIDS	same	FL-SnortIDS	FL-SnortIDS	FL-SnortIDS
SuricataIDS vs FL-SuricataIDS	Same	FL-SuricataIDS	FL-SuricataIDS	FL-SuricataIDS
SnortIDS vs SuricataIDS	SnortIDS	Same	Same	Same
FL-SnortIDS vs FL-SuricataIDS	FL-SnortIDS	FL-SuricataIDS	FL-SuricataIDS	FL-SuricataIDS

E. Final Results

We did pair-wise comparison as it can be seen in the graphical representation of all four methods in figure 7. On analysing the below result, we can see for the first three comparisons we have clear results:

- FL-SnortIDS is better than SnortIDS
- FL-SuricataIDS is better than SuricataIDS
- SnortIDS is better than SuricataIDS

For the fourth comparison between FL-SnortIDS vs FL-SuricataIDS, we found FL-SnortIDS is better in terms of sensitivity while the other criteria are other way round. So to come up with the conclusion, the graph of specificity, false alarm ratio and accuracy and also the descriptive statistics show that there was a difference in these criteria; yet it is not too much comparing to the criterion of sensitivity. Therefore, the FL-SnortIDS is better than FL-SuricataIDS to get false alarm rather than not getting the alarm when actually it should.

- FL-SnortIDS is better than FL-SuricataIDS (Based on sensitivity performance).
- SnortIDS is better than FL-SuricataIDS (Based on sensitivity performance).

Combining results of all five category viz. sensitivity, specificity, False Alarm ratio, accuracy and detection rate, we have the following result ranked according to their performance:

1) FL-SnortIDS which detects the threat with Medium detection rate.
2) SnortIDS which detects the threat with High detection rate.
3) FL-SuricataIDS which detects the threat with Low detection rate.
4) SuricatIDS which detects the threat with High detection rate.

VII. Conclusion

The focus of this research was to understand and find the best approach towards cloud security and its availability. We initially found the best rated open source intrusion detection systems on which we could run a simulated dataset and find where these systems lack or supersede others. The Information Security Centre of Excellence (ISCX) provided the dataset what was required. The data consisted of 7 days activity carefully simulated to run on network intrusion detection systems and check the performance of system against the data. The proposed approach was simulated to demonstrate the higher level of accuracy, sensitivity and specificity achieved. The substantial decrease in false alarms was also achieved. By using fuzzy technique, unwanted alerts were removed while the others were categorised into 4 types of cyber-attacks; DoS, R2L, U2R and Probe. This improvement on SnortIDS/SuricataIDS were named to be FL-SnortIDS/FL-SuricataIDS respectively.

Results showed that the capabilities of IDSs were considerably increased after applying fuzzy logic over the alerts generated by any of the IDS systems. In particular, the main focus of this study was on the comparison between alerts generated by typical SnortIDS/SuricataIDS and similarly the alerts generate by Fuzzy Logic based SnortIDS/SuricataIDS system. Experimental results showed the attainment of satisfactory detection rates based on the recent and most evaluated benchmark ISCX dataset on intrusions. The statistical values of accuracy, sensitivity, specificity and false alarm ratios justified that fuzzy logic based SnortIDS works the best than any other IDS system. These results were further analysed using tools such as Mann-Whitney Test. These analyses showed these results:

- FL-SnortIDS is better than FL-SuricataIDS
- FL-SnortIDS is better than SnortIDS
- FL-SuricataIDS is better than SuricataIDS
- SnortIDS is better than SuricataIDS

This goes a long way in understanding different emerging attacks and techniques used by network or forensic analyst to try to determine and restrict the intrusion in their networks. It can be seen that fuzzy logic along with the legacy intrusion detection systems yields better results and facilitates the network administrators to mitigate the issues.

New Genetic algorithms are being developed and extensive researches are being carried out to analyse the huge amount of data being transported over the networks. In the future, FL Based IDS incorporated with genetics algorithm can be designed and implemented. A network aware IDS is the only solution for the ever changing network traffic. Both SnortIDS and SuricataIDS systems with the current design will not be able to understand the changing networks and complex attacks.

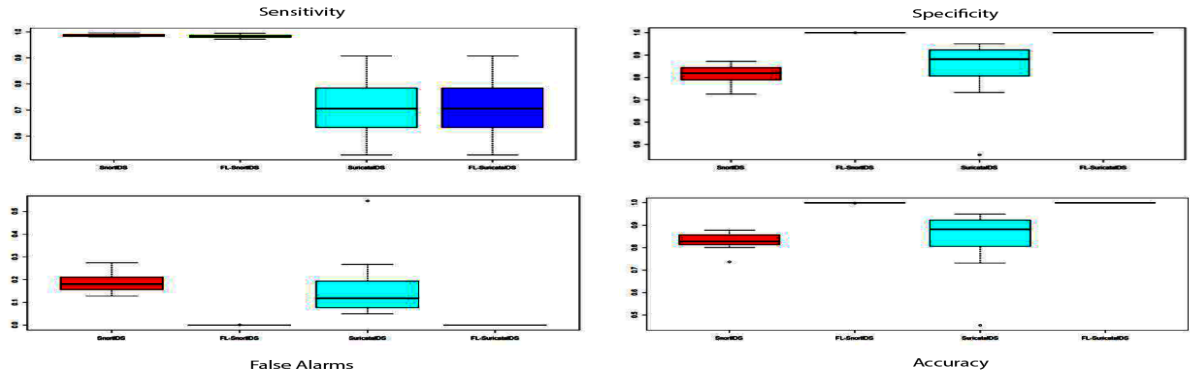


Fig. 7: Final Results for All Systems

REFERENCES

[1] F. Halsall and D. Links, "Computer networks and open systems," *Addison-Wesley Publishers*, pp. 112–125, 1995.
[2] I. L. Stats. (2016) Internet Users in the World. [Online]. Available: http://http://goo.gl/9zCfjv
[3] R. L. Rivest, L. Adleman, and M. L. Dertouzos, "On data banks and privacy homomorphisms," *Foundations of secure computation*, vol. 4, no. 11, pp. 169–180, 1978.
[4] P. M. Schwartz, "Internet privacy and the state," *Conn. L. Rev.*, vol. 32, p. 815, 1999.
[5] M. Talwar, "Security issues in internet of things," *International Journal on Emerging Technologies*, 2015.
[6] M. C. Waxman, "Cyber-attacks and the use of force: Back to the future of article 2 (4)," *Yale Journal of International Law*, vol. 36, 2011.
[7] S. Levy, *Hackers: Heroes of the computer revolution*. Penguin Books New York, 2001, vol. 4.
[8] A. Runthala, "Hacking: A threat to industrial work forces." *CURIE Journal*, vol. 3, no. 1, 2010.
[9] A. V. K. V. G. Puzmanova, Rita and A. A. Mikhailovsky, "Review of wi-foo: The secrets of wireless hacking," *Queue*, vol. 2, no. 8, pp. 70–70, 2004.
[10] L. Garber, "Denial-of-service attacks rip the internet," *IEEE Computer*, vol. 33, no. 4, pp. 12–17, 2000.
[11] K. Pipyros, L. Mitrou, D. Gritzalis, and T. Apostolopoulos, "A cyber attack evaluation methodology," in *Proc. of the 13th European Conference on Cyber Warfare and Security*, 2014, pp. 264–270.
[12] S. Shackelford, "From nuclear war to net war: analogizing cyber attacks in international law," *Berkley Journal of International Law (BJIL)*, vol. 25, no. 3, 2009.
[13] K. Burkadze, "Cyber security and international law," *Journal of Technical Science and Technologies*, vol. 4, no. 2, pp. 5–10, 2016.
[14] D. E. Graham, "Cyber threats and the law of war," *J. Nat'l Sec. L. & Pol'y*, vol. 4, p. 87, 2010.
[15] O. A. Hathaway, R. Crootof, P. Levitz, H. Nix, A. Nowlan, W. Perdue, and J. Spiegel, "The law of cyber-attack," *California Law Review*, 2012. [Online]. Available: www.californialawreview.org
[16] W. . Richland. (2016) Protecting Organizations from Cyber Attack. [Online]. Available: https://http://goo.gl/H8W6cu
[17] K. Hwang, S. Kulkareni, and Y. Hu, "Cloud security with virtualized defense and reputation-based trust mangement," in *Dependable, Autonomic and Secure Computing, 2009. DASC'09. Eighth IEEE International Conference on*. IEEE, 2009, pp. 717–722.
[18] P. Brutch and C. Ko, "Challenges in intrusion detection for wireless ad-hoc networks," in *Applications and the Internet Workshops, 2003. Proceedings. 2003 Symposium on*. IEEE, 2003, pp. 368–373.
[19] M. Augustin and A. Baláž, "Intrusion detection with early recognition of encrypted application," in *2011 15th IEEE International Conference on Intelligent Engineering Systems*. IEEE, 2011, pp. 245–247.
[20] G. Klir and B. Yuan, *Fuzzy sets and fuzzy logic*. Prentice hall New Jersey, 1995, vol. 4.
[21] T. P. Fries, "A fuzzy-genetic approach to network intrusion detection," in *Proceedings of the 10th annual conference companion on Genetic and evolutionary computation*. ACM, 2008, pp. 2141–2146.
[22] A. V. Katherine and K. Alagarsamy, "A fuzzy mathematical model for peformance testing in cloud computing using user defined parameters."
[23] D. N. P. S. Kumar and D. G. P. Ramesh, "Intrusion detection analysis by implementing fuzzy logic," 2016.
[24] S. Bezborodov *et al.*, "Intrusion detection systems and intrusion prevention system with snort provided by security onion." 2016.
[25] D. Burks. (2012) 229 Doug Burks Security Onion Network Security monitoring in minutes. [Online]. Available: https://www.youtube.com/watch?v=mazSRVFYmLQ
[26] R. Bray, D. Cid, and A. Hay, *OSSEC host-based intrusion detection guide*. Syngress, 2008.
[27] V. Paxson, S. Campbell, J. Lee *et al.*, "Bro intrusion detection system," Lawrence Berkeley National Laboratory, Tech. Rep., 2006.
[28] M. Roesch *et al.*, "Snort: Lightweight intrusion detection for networks." in *LISA*, vol. 99, no. 1, 1999, pp. 229–238.
[29] I. S. C. o. E. I. ISCX. (2012) UNB ISCX Intrusion Detection Evaluation DataSet. [Online]. Available: http://www.unb.ca/research/iscx/dataset/iscx-IDS-dataset.html
[30] A. Shiravi, H. Shiravi, M. Tavallaee, and A. A. Ghorbani, "Toward developing a systematic approach to generate benchmark datasets for intrusion detection," *Computers & Security*, vol. 31, no. 3, pp. 357–374, 2012.
[31] H. Chauhan, V. Kumar, S. Pundir, and E. S. Pilli, "A comparative study of classification techniques for intrusion detection," in *Computational and Business Intelligence (ISCBI), 2013 International Symposium on*. IEEE, 2013, pp. 40–43.
[32] Snort. [Online]. Available: https://snort.org/
[33] Suricata. [Online]. Available: https://suricata-ids.org