A Bayesian Network Model for Risk Management in Cyber Situation

Bode Moyinoluwa A., Alese Boniface K., Thompson Aderonke F. and Iyare Otasowie

Abstract:- **The transformation in e-commerce with the advent of other internet based activities has increased the impact of Cyber-attacks on organizations and nations at large. However, tracking these attacks and determining risks posed ensure the overall security in cyber situation and invariably enhances communication and economic growth. Consequently, this paper presents a robust predictive model based on Bayes Theorem, capable of distinguishing between bad connections called threats or attacks, and good which is known as normal connections. The model analyses the state of network traffic and determine its risk using the KDD Cup 1999 dataset. In conclusion, comparison of the Bayes model with Association Rule Mining model shows efficient performance, and an improved performance with a Genetic Algorithm technique**

Index-Terms: Cyber-attacks/threats, Cybercrime, Bayesian Network, Cyber Situation, **Risk**

I. INTRODUCTION

Cyber security, which is also referred to as information security, is the protection of information against unauthorized disclosure, transfer, or modifications, whether accidental or intentional. Information security is the major challenge to gains of Information Technology (IT) world. Information security is required at all levels – personal, corporate, state and country. Handling cyber threats deal with both uncertain and imprecise information. However, Lipschutz [1] emphasized that security demands certainty. This certainty means, acceptance of a fact without doubt; that is, acceptance of fact with absolute confidence.

In IT security, a lot has to do with certainty about the present and future, the efficiency of the political, economic, strategic and tactical tools that the liberal society produces to be successful rather than certainty about the figure(s) of the enemy and possible threats.

Manuscript received March 20, 2014; revised April 20, 2014.

Bode Moyinoluwa Abidemi. is with Engineering Materials Development Institute Akure, Nigeria. Phone: +2348034819325 Email: bodemoyinka@gmail.com

Alese Boniface Kayode. is with the Department Computer Science, Federal University of Technology, Akure, Nigeria. Phone: +2348034540465. Email: bkalese@futa.edu.ng

Thompson Aderonke Favour-Bethy is with the Department Computer Science, Federal University of Technology, Akure, Nigeria. Phone: +2348034540465. Email: afthompson@futa.edu.ng

Iyare Otasowie. is with the Department Computer Science, Federal University of Technology, Akure, Nigeria. Phone: +2347033513174. Email: oiyare@futa.edu.ng

Realizing this task, liberal societies need opportunities and risks. Risk analysts are of the opinion that the cost of eliminating a risk is infinite [1]. In this perspective, as opined, we can never be totally secure. Thus, security is dynamically seeking to establish its markers of certainty and fixity, which are themselves always moving. However, Alese et al., [2] states that new risk factors and challenges to data and communications networks are evolving as rapidly as the spread of high-speed internet infrastructure. Among these compelling problems are: computer worms and viruses, organized criminal activity, weak links in the global information infrastructure: and hacker-activists and protestors have proven themselves capable of temporarily disrupting ICT-based services of governments and international organizations.

The International Telecommunication Union (ITU) defined cyber security as the prevention of damage, unauthorized use, exploitation, and if needed the restoration of electronic information and communications systems with the information content. This is in order to strengthen the confidentiality, integrity and availability of these systems. In general, cyber security threats are increasing rapidly, the incidents range from defaced websites to theft of large volumes of intellectual property and money, to Internet crimes. Quantifying this problem led to the 2011 Norton Cybercrime Report, surveying nearly 20,000 people with children of 8-17 years inclusive from 24 countries around the world. It was observed that every day there are twice as many cybercrime victims as new born babies [3].

There has been tremendous growth in the past two decades in computing power and explosive applications of computing devices. In the same manner, there has equally been increase in security breaches in/with IT infrastructures by organized crime group and nation/state sponsored adversaries.

Unfortunately, current cyber defence capability is still at an infancy state. It is quite common for an enterprise to rely its information security on a few knowledgeable, but overwhelmed analysts and a collection of tools that may provide some useful defence against known or past attacks. The old knowledge or tool may not guide against new exploits from attackers [4].

Advanced cyber threats have established a stealthy, persistent presence on many computer networks and they adapt to evade cyber defences. There is now ample evidence that despite significant strides toward building secure, trustworthy systems, advanced cyber adversaries are successful using other means of attack to compromise our information systems. In this threat environment, cyber

ISBN: 978-988-19252-0-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

situation awareness is a very challenging problem. A promising direction is to take advantage of the recent advances in trusted computing. For example, we may gain high confidence in the trustworthiness of data gathered for cyber situational awareness by protecting them using a Trusted Platform Module (TPM). Nevertheless, substantial research is necessary to guarantee the successful use of trusted computing technologies to support cyber situational awareness [5].

II CYBER SITUATION AWARENESS (CSA)

Situation Awareness (SA) is the field of study concerned with perception of the surroundings and derivative implications critical to decision makers in complex, dynamic areas such as military command and security. Situation awareness focuses on what is known about the past and present situations, [6,7, 8]. Given the positive outcomes of SA, it is now being applied to cyber space, so the concept Cyber Situation Awareness. According to Endsley [9], situation awareness begins with perception. Perception provides information about the status, attributes, and dynamics of relevant elements within the environment. It also includes classifying information into understood representations and provides the basic building blocks for comprehension and projection. Without a basic perception of important environmental elements, the odds of forming an incorrect picture of the situation increase dramatically. It has long been recognized that logical relations in computer attack conditions are important to consider in security analysis [10, 11]. Statistical models have been used for this security analysis in these analysis in the past. [12, 13] extended Endsley's Model by adding a fourth level, which is called Resolution. Resolution results from drawing a single course of action from a subset of available actions. .

Vulnerability analysis [14] and IDS alert correlation [15] have been used to model such relations. While these types of logical relations are important, they cannot account for the uncertainty in cyber security analysis. An example is zero-day vulnerabilities, which have enabled a large number of intrusions into enterprise networks. One cannot make a deterministic judgment on whether a piece of software contains zero-day vulnerability, but has to consider this possibility in security defence. However, there is a fundamental limitation in solving the uncertainty problems in cyber security using statistical models alone. Attackers do not play by rules. They adapt and do not typically follow a statistical pattern, as demonstrated by various forms of evading techniques [16]. Information security continuous monitoring is deals with as maintaining on-going awareness of information security, vulnerabilities, and threats to support organizational risk management decision. The terms continuous and on-going in this context mean that security controls and organizational risks are assessed and analyzed at a frequency sufficient to support risk-based security decisions to adequately protect organization information [17].

A high network traffic deviating from the statistical norm gives a valuable hint on potential problems, and the confidence level on the true causes of such alerts can be statistically described as false positives and false negatives. It is important to account for the statistical differences in various assertions confidence level. For example, compared with the anomalous high network traffic, a net flow filter that shows communication with known Botnet controllers is a more confident assertion on attacker activity. A simple and effective model for such statistical differences on assertion confidence will help in tackling the uncertainty problem [18]. Li [19] sought to ease the difficulty encountered in cyber situation awareness by the bottom-up transformations of low-level data to meaningful information, from information to actionable knowledge and from knowledge to trustworthy intelligence. These can assist human analysts understand the current situation and project future situations while they can exhibit unique analysis that surpass the most advanced security analysis software tools. He stressed the need to transfer such expertise into automated cyber situation awareness software tools. This top-down transformation can be achieved via knowledge engineering techniques. The author also suggested that the use of mainstream approaches such as Bayesian network to solve uncertainty management and Trusted Computing approach.

Ning [20] noted several years of research on intrusion detection and prevention has shown that overcoming attackers is not an easy task. The author noted the use of trusted computing to reduce uncertainty which is the high confidence in the trustworthiness of data gathered for cyber situational awareness by protecting them using a TPM.

In summary, both deterministic logics and statistical models are valuable tools in cyber defence, but neither alone is sufficient to tackle the uncertainty challenge. Combining the two, however, will likely yield a reasoning method much more powerful than their sum. A reasoning framework that accounts for both logical relations and confidence differences among the various assertions will be the key in handling uncertainty in cyber security.

III COMMON CYBER-ATTACKS/THREATS

Amos [3] classified cyber threats into three major areas. They are cyber espionage, Internet crimes, and cyber warfare.

A Cyber Espionage

Cyber espionage or cyber spying is the act or practice of obtaining secrets without the permission of the holder of the information. It may be personal, sensitive, proprietary or classified information from individuals, competitors, rivals, groups, governments and enemies for personal, economic, political or military advantage. It involves using various methods on the Internet, networks or individual computers through the use of cracking techniques and malicious software including Trojan horses and spyware. Also, it may be the criminal handiwork of amateur malicious hackers and software programmers. Cyber spying typically involves the use of such access to secrets, classified information, and control of individual computers or whole networks for a strategic advantage, for psychological, political, physical subversion activities and sabotage. More recently, cyber

ISBN: 978-988-19252-0-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

spying involves analysis of public activity on social networking sites such as Facebook and Twitter [21].

B Internet Crimes

According to Moore [22], computer crime refers to any crime that involves a computer and a network. The computer might have been used in the commission of a crime, or it may be the target. Net crime refers to criminal exploitation of the Internet. Halder *et al.* [23] defines cybercrimes as offences that are committed against individuals or groups of individuals with a criminal motive to intentionally harm the reputation of the victim, cause physical or mental harm to the victim directly or indirectly. This is done through the use of modern telecommunication networks such as Internet (Chat rooms, emails, notice boards and groups) and mobile phones (SMS/MMS). Such crimes may threaten a nation's security and financial health. Issues surrounding these types of crime have become high-profile, particularly those surrounding cracking, copyright infringement, child pornography and child grooming. There are also problems of privacy when confidential information is lost or intercepted, lawfully or otherwise.

C Cyber Warfare

Cyber warfare refers to politically motivated hacking to conduct sabotage and espionage. It is a form of information warfare sometimes seen as analogous to conventional warfare [24]; although, this analogy is controversial for both its accuracy and its political motivation. U.S. government security expert Richard A. Clarke, in his book Cyber War [25] defines cyber warfare as actions by a nation-state to penetrate another nation's computers or networks for the purposes of causing damage or disruption. The Economist describes cyberspace as the fifth domain of warfare, Jeffrey [26] and Lynn [27] states that as a doctrinal matter, the Pentagon has formally recognized cyberspace as a new domain in warfare, this has become just as critical to military operations as land, sea, air, and space.

IV SYSTEM ANALYSIS AND DESIGN

Cyber defenders do not know who the attackers are or their location. The Intrusion Detection Systems (IDS) large number of false positives brings significant uncertainty to the true interpretation of IDS alerts. Also, there are still false negatives where some attacks may not be reported by any IDS sensor. There are plenty of zero-day vulnerabilities in application software, however, there is no way to discern which software can be exploited by an attacker.

The uncertainty challenge exists in all the phases of cyber situation awareness: prior security risk management, real-time intrusion detection, and posterior forensics analysis. The nature of uncertainty in these three aspects is slightly different. In risk management, there is uncertainty about the likelihood that vulnerability exists in a piece of software, the chances that vulnerability can be exploited successfully, the possibility that a user may succumb to social engineering, and so on. This type of uncertainty is static and reflects various inherent risks in a system while uncertainty in intrusion analysis is dynamic uncertainty, since they are mostly related to dynamic events [28].

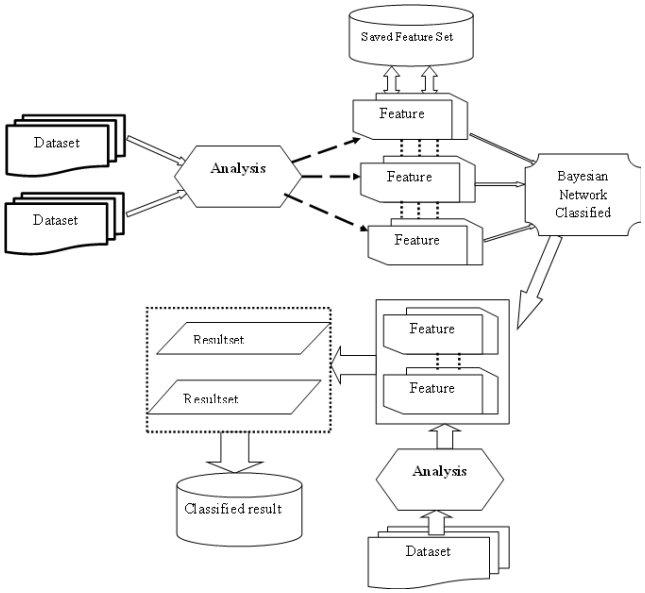

Fig 1: Architecture for Bayesian Network Risk Management System

A System Architecture

Fig 1 shows the architecture of the risk management system using Bayesian Network approach. The architecture shows the different constituent that make up the system. The files that contain records (dataset) obtained from cyber network environment which forms the training set for the system. The training sets are analyzed to ensure consistency and availability of the required dataset. The content of the dataset are then categorized into features for each field that constitute a valid information entry of network activity, creating a pattern that will be used for training the Bayesian Network classifier.

The classifier creates a relationship between the features that serves as a model for the system that will be used for establishing the result for an input cyber information dataset. The testing of the Bayesian classifier is carried out with a test dataset which is analyzed into feature set and the corresponding output forms the result set where the test dataset have been classified with the correct label (whether it is a threat or non-threat). The performance and outcome of the result set determines whether the Bayesian classifier can identify uncertainty and risk in a cyber-environment.

B A Cyber Attack Scenario

The cyber-infrastructure in a corporate network typically consists of a webserver and a fileserver that are protected by two firewalls in the Demilitarized Zone (or DMZ) (where the DMZ separates the external network (Internet) from the company's internal LAN network) [29, 30].

ISBN: 978-988-19252-0-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

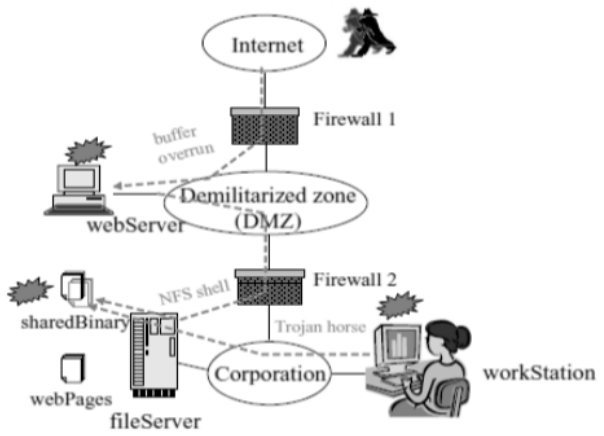

Fig 2: A Simple Scenario of a Cyber-Attack [30]

C Bayesian Model

Thomas *et al.*, [31] express Bayesian inference as a rational engine for solving such problems within a probabilistic framework which is the heart of most probabilistic models. Bayesian model emerges from Bayes' theorem when stated in terms of abstract random variables; Bayes' theorem is an upshot of probability theory. Assume two random variables, A and B. One of the principles of probability theory also called chain rule, allows joint probability to be written in terms of these two variables particular values a and b; P (a, b), as the product of the conditional probability that A takes on value a given B takes on value b, P (a|b), and the marginal probability that B takes on value b, P (b). Thus, it can be written as:

$$P(a,b) = P(a|b)P(b) \quad 1$$

Using factorization with the choice of B rather than A, the joint probability is written as:

$$P(a,b) = P(b|a)P(a) \quad 2$$

From Equations 3.0 and 3.1 that P (a|b)P(b) = P(b|a)P(a), thus

$$P(a|b) = \frac{\mathrm{P(a|b)P(b)}}{\mathrm{P(a)}} \quad 3$$

This expression is Bayes' theorem, it indicates the computation of the conditional probability of b given a, from the conditional probability of a given b.

The following illustrates some of the characteristics of Bayesian model for cyber security. Consider an attack on a system in a network, the system may become at risk of any form of attack as a result of the use of network resources, an event represented by the variable Denial of Service (DoS) attack (denoted by D). Such an attack can cause damage to systems or lead to denial of service, an event represented by the variable teardrop (denoted by TD). The DoS attack might result from status flag of connection, represented by the variable SF (denoted by S) or connection protocol, represented by the variable http (denoted by H). It is reasonable to assume that a network user is at risk of a Probe attack, an event represented by the variable IMAP (denoted by I). All variables representation are binary; thus, they are either true (denoted by T) or false (denoted by F). The condition probability table (CPT) of each node is listed besides the node. In this example the parents of the variable DoS are the nodes SF and http. The child of DoS is teardrop, and the parent of IMAP is Probe. Following the Bayesian Network (BN) independence assumption, several independence statements can be observed in this case. For example, the variables http and SF are marginally independent, but when DoS is given they are conditionally dependent. This relation is often called explaining away.

When http is given, SF and DoS are conditionally independent. When DoS is given, teardrop is conditionally independent of its ancestors http and SF. The conditional independence statement of the BN provides a compact factorization of the Joint Probability Distributions (JPD). Rather than factorizing the joint distribution of all the variables by the chain rule, i.e.

$$P(S,H,I,D,TD) = P(S)P(H|S)P(I|H,S)P(D|I,H,S)P(TD|D,I,H,S) \quad 4$$

The BN defines a unique JPD in a factored form, i.e.

$$P(S,H,I,D,TD) = P(S)P(H)P(I|S)P(D|H,S)P(TD|D) \quad 5$$

Note that the BN form reduces the number of the model parameters, which belong to a multinomial distribution in this case, from $25 - 1 = 31$ to 10 parameters. Such a reduction provides great benefits from inference, learning (parameter estimation), and computational perspective. The resulting model is more robust with respect to bias-variance effects. A practical graphical criterion that helps to investigate the structure of the JPD modelled by a BN is called d-separation. It captures both the conditional independence and dependence relations that are implied by the Markov condition on the random variables.

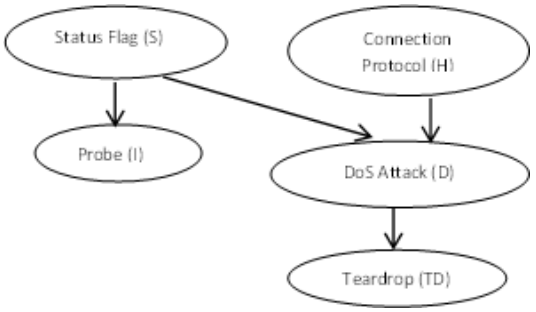

Fig 3: A Conceptual Bayesian Network

D Inference via Bayesian Network

Given a scenario in Fig 3, one might consider the diagnostic support for the belief on the occurrence of a Flag, given the observation that the network suffers from the risk of teardrop. Such a support is formulated as follows:

$$\mathrm{p(S = T|TD = T)} = \frac{\mathrm{p(S = T, TD = T)}}{\mathrm{p(TD = T)}} \quad 6$$

Where,

$$\mathrm{p(S = T, TD = T)} = = \sum_{\mathrm{H,I,D,S} \in \{\mathrm{T,F}\}} \mathrm{p(S = T)p(H)p(I|S = T)p(D|H, S = T)p(TD = T|D)} \quad 7$$

And

ISBN: 978-988-19252-0-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

$$p(TD = T) = \sum_{H,I,D,S \in \{T,F\}} p(S)p(H) \times p(I|S)p(D|H,S) \times p(TD = T|D) \quad 8$$

Table I Attack Type and Category Description

ID	Types of Attack	Category
1	Back	DoS
2	buffer_overflow	u2r
3	ftp_write	r2l
4	guess_passwd	r2l
5	Imap	r2l
6	Ipsweep	Probe
7	Land	DoS
8	Loadmodule	u2r
9	Multihop	r2l
10	Neptune	DoS
11	Nmap	Probe
12	Perl	u2r
13	Phf	r2l
14	Pod	DoS
15	Portsweep	Probe
16	Rootkit	u2r
17	Satan	Probe
18	Smurf	DoS
19	Spy	r2l
20	Teardrop	DoS
21	Warezclient	r2l
22	Warezmaster	r2l

E Database Design

The design of the system dataset employs a relational database model where the first name on the stack of box represents the name of the table and the other fields represents the field names in the table. The Training_set and Testing_set represent the table for keeping information about the training dataset used to train the system and the testing dataset for evaluating the system respectively. The field name refers to the name of the file containing the dataset, the field_size represents the number of fields (column) in the dataset. The record_size field hold the information about the number of rows (records) in the dataset and the summary field records the details about the data attribute that forms each fields in the dataset, all successfully registered dataset (training set or testing set) is assigned a unique number ID to identify the dataset for use by the system. The result_set table holds information about the output by the system for a given test dataset, trained by a given training set. The testing_id and training_id refer to the corresponding training set and testing set respectively that was used. Name field refers to the name of the file containing the result of the system after classification of the testing set, confusion_matrix represent the confusion matrix of the classification class label and the result_summary holds the information of the error rate and percentage of correctly classified information and other result from classification.

F Bayesian Classifier

The Bayesian classifier interface is used for training the system and testing the performance of the classifier. The testing dataset section provides a section for selecting and evaluating the performance of the classifier. At end of the classification of the test dataset, the result is display in the classifier result area in Fig 4. The summary displays the confusion matrix of the Dataset.

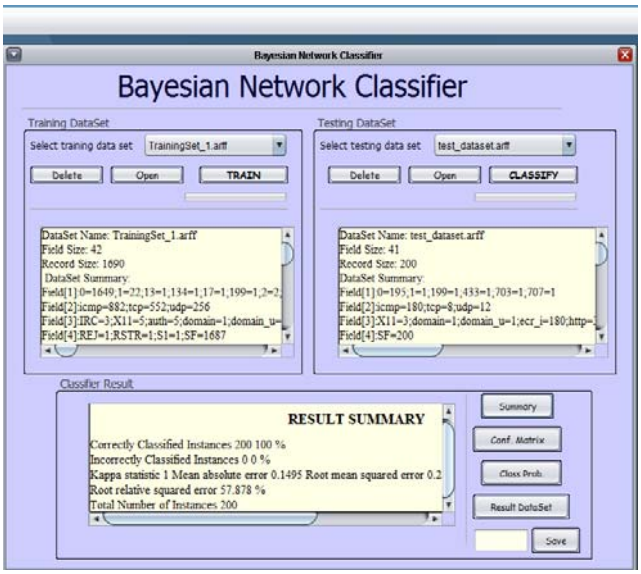

Fig 4: Classifier Result Interface

V EXPERIMENTAL SETUP AND RESULTS

The application software was executed and tested on the Windows 7, 4GB RAM, 350GB HDD, Intel dual Core T4200 2.0GHz Processor. The distribution of training dataset which is made up of 10,949 records consisting the following distribution and their corresponding degree of occurrences: back:783(7.15%), buffer_overflow:9(0.08%), ftp_write:4(0.04%), guess_passwd:0(0.00%), imap:11(0.10%), ipsweep:16(0.15%), land:9(0.08%), loadmodule:8(0.07%), multihop:10(0.09%), neptune:2099(19.17%), nmap:25(0.23%), perl:4(0.05%), phf:2(0.02%), pod:142(1.3%), portsweep:287(2.62%), rootkit:9(0.08%), satan:736(6.83%), smurf:3837(35.04%), spy:6(0.05%), teardrop:199(1.82%), warezclient:184(2.34%), warezmaster:20(0.25%), normal: 2549(17.18%).

A Classification Probability

Classification probability for attack represents the probability of occurrence of each attack type for a given data set.

$$P_A = \frac{\sum T_A}{\sum n} \quad 9$$

Where n is the total number of dataset, T_A is the total number of occurrence of a type of attack in the same dataset and P_A is the probability of classification of the attack type. Therefore, Table II below shows the classification probability of attack and normal network traffic dataset obtained from the training dataset classification

ISBN: 978-988-19252-0-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Table II Classification Probability

ID	Type of Attack	Probability
1	Back	0.0715134
2	buffer_overflow	0.000822
3	ftp_write	0.0003653
4	guess_passwd	0.0000000
5	Imap	0.0010047
6	Ipsweep	0.0014613
7	Land	0.000822
8	loadmodule	0.0007307
9	multihop	0.0009133
10	Neptune	0.191707
11	Nmap	0.0022833
12	Perl	0.0003653
13	Phf	0.0001827
14	Pod	0.0129692
15	portsweep	0.0262124
16	Rootkit	0.000822
17	Satan	0.0672208
18	Smurf	0.350443
19	Spy	0.000548
20	Teardrop	0.0181752
21	warezclient	0.0168052
22	warezmaster	0.0018267
23	Normal	0.2328066

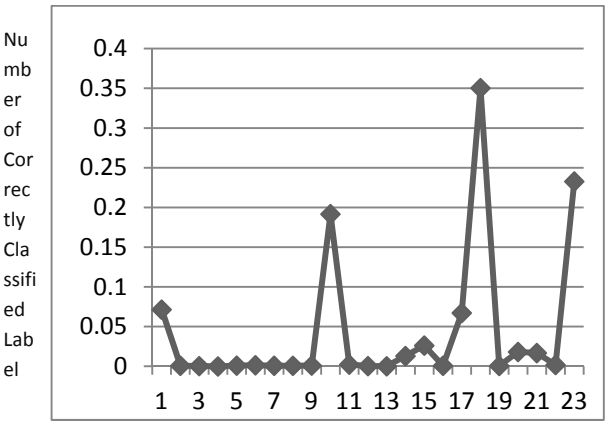

Fig 5: Graph of the Probability Distribution for Types of Attack

Fig 5 above shows the probability for the distribution of the occurrence of attacks that form the training dataset. The graph shows the probability occurrence (on the y-axis) and the corresponding attack identified by the ID in Table 4.2 (on the x-axis). The graph shows that Neptune, Satan, Smurf and Normal have the highest number of occurrence in the dataset.

VI SYSTEM VALIDATION

A Association Rule Mining

A supervised data learning technique known as Association Rule Mining was used to build a classifier for detecting some denial of service attacks. Olasehinde [32] defines Association rule mining as follows:

Let $I = \{i_1, i_2, \ldots\ldots\ldots, i_n\}$ 10

be a set of *n* binary attributes called *items.*

Let $D = \{t_1, t_2, \ldots\ldots\ldots\ldots, t_n\}$ 11

be a set of transactions called the *database*.

Each transaction in *D* has a unique ID and contains a subset of the items in *I*. A *rule* is defined as an implication of the form

$X \rightarrow Y$ where $X, Y \subseteq I$ and $X \cap Y = \emptyset$ 12

Table III: Traffic Data Sample [32]

Transaction ID	Protocol	Service	Flag	Attacks/Label
Traffic 1	Udp	Smtp	Sf	Teardrop
Traffic 2	Tcp	Http	Sf	Smurf
Traffic 3	Udp	Http	So	Neptune
Traffic 4	Icmp	Private	Sf	Teardrop
Traffic 5	Tcp	Smtp	So	Land

The author considered network traffic data in Table III. The best result of classification was obtained after pruning process, the pruning process include: removal of all rules with confidence less than 50%; duplicate rules; all identical rules pointing to difference attacks and all one attribute rules were not considered for classification. Table 4a and 4b below shows the confusion matrix for the comparison of Bayes network classification with association rule mining from five attributes combination pruned rules for DoS attack.

Table IVa: Confusion Matrix Obtained the Prune Rules

Predicted as Actual	(Olasehinde [32]				
	Neptune	Smurf	Pod	Teardrop	Land
Neptune (16)	16 (100%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
Smurf (264)	0 (0.00%)	264 (100%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
Pod (20)	0 (0.00%)	0 (0.00%)	20 (100%)	0 (0.00)%	0 (0.00)%
Teardrop (99)	0 (0.00%)	0 (0.00%)	0 (0.00%)	99 (100%)	0 (0.00%)
Land (1)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	1 (100%)

Table IVb: Confusion Matrix Obtained the Prune Rules

Predicted as Actual	Bayes Network Proposed				
	Neptune	Smurf	Pod	Teardrop	Land
Neptune (16)	16 (100%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
Smurf (264)	0 (0.00)%	264 (100%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
Pod (20)	0 (0.00%)	0 (0.00%)	20 (100%)	0 (0.00%)	0 (0.00%)
Teardrop (99)	0 (0.00%)	0 (0.00%)	0 (0.00%)	99 (100%)	0 (0.00)%
Land (1)	0 (0.00%)	0 (0.00)%	0 (0.00%)	0 (0.00%)	**1 (100%)**

ISBN: 978-988-19252-0-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

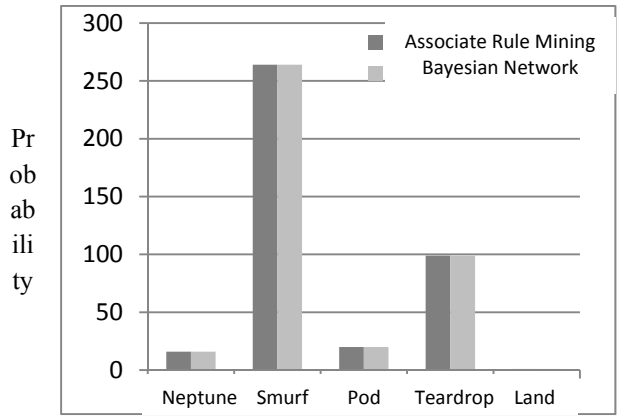

Fig 6: Correctly Classified Label from Tables 4a and b

Fig 6 above shows equal performance of bar chart for validation of the system using Bayesian Network and Association Rule Mining.

B Genetic Algorithm (GA)

Mohammad *et al.*, [33], presented an Intrusion Detection System (IDS) using GA. GA was applied, to efficiently detect various types of network intrusions. This approach uses evolution theory in order to filter the traffic data thus, reducing its complexity. KDD Cup '99 benchmark was used to implement and measure the system performance.

GA Algorithm for IDS system

Input: Network audit data (for testing), Precalculated set of chromosomes
Output: Type of data.
1. Initialize the population
2. Crossover Rate = 0.15, Mutation Rate = 0.35
3. While number of generation is not reached
4. For each chromosome in the population
5. For each precalculated chromosome
6. Find fitness
7. End for
8. Assign optimal fitness as the fitness of that chromosome
9. End for
10. Remove some chromosomes with worse fitness
11. Apply crossover to the selected pair of chromosomes of the population
12. Apply mutation to each chromosome of the population
13. End while

GA uses evolution and natural selection of chromosome-like data structure and evolve the chromosomes using selection, recombination and mutation operators.

Table Va: Confusion Matrix Obtained GA

		Predicted Label		
		Mohammed et al. [34]		
Actual Class		Probe	DoS	U2R
	Probe	2963(81.9%)	654(18.1%)	2(0.005%)
	DoS	432(0.19%)	228489(99.8%)	1(0.0004%)
	U2R	21(29.2%)	8(11.1%)	43(59.7%)

Table Vb: Confusion Matrix Obtained Bayesian Network

		Predicted Label		
		Bayesian Network (BN) Proposed		
Actual class		Probe	DoS	U2R
	Probe	3106(85.8%)	511(14.2%)	2(0.005%)
	DoS	432(0.19%)	228490(99.8%)	0(0.0)
	U2R	11(15.2%)	3(4.2%)	58(80.6%)

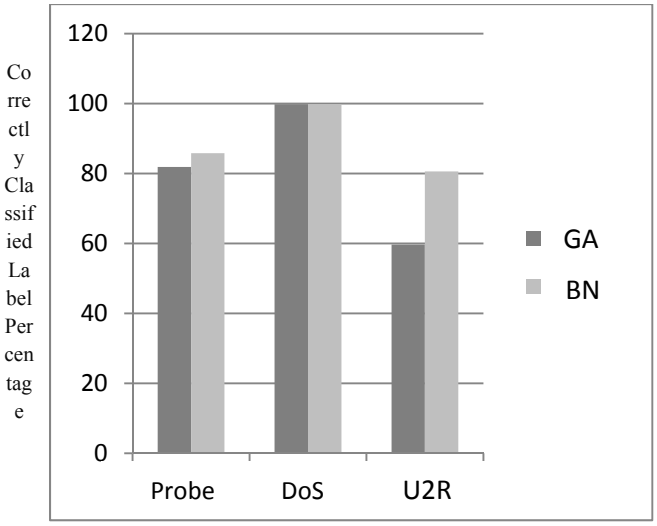

Fig 7: Correctly Classified Label from Tables Va and b

Fig 7 shows the correctly classified attribute (labels) between GA technique and BN. The graph shows that BN has a better performance because there is a slight improvement in the number of probe and U2R attacks correctly classified.

VII CONCLUSION

This research work has implemented Bayesian network to classify normal and abnormal attacks from network traffic using uses KDD Cup '99 dataset. It was observed that the Bayesian approach to cyber situation awareness is suitable large data sets application and interpretation. Therefore, the Bayesian network is an enhanced classification model that is suitable to implement in areas of text classification and spam filtering.

REFERENCES

[1]. Lipschutz, R. (2000) After Authority: War, Peace and Global Politics in the 21st Century,
[2]. Alese, B. K., Gabriel A. J. & Adetunmbi A. O. (2011); Design and Implementation of Internet Protocol Security Filtering Rules in a Network Environment, International Journal of Computer Science and Information Security (IJCSIS), Vol. 9, No. 7, July 2011
[3]. Amos Nungu (2012) Cyber Security in Tanzania: Roles and Responsibilities, For Cyber Security Mini-Conference, Dar Es Salaam Institute of Technology (DIT), June 26, 2012
[4]. Jajodia S., Liu P., Swarup V., & Wang C. (2010). Cyber Situational Awareness: Issues and Research, Springer, New York

ISBN: 978-988-19252-0-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

[5]. Barford P., Dacier M., Dietterich T. G., Fredrikson M., Giffin J., Jajodia S., Jha S., Li J., Liu
[6]. Alberts, D.S., Garstka, J.J., Hayes, R.E., Signori, D.A., 2001, Understanding Information Age Warfare. CCRP Publication Services, Washington, 319pp.
[7]. U. S. Army Field Manual 1-02, Washington, D.C.: Headquarters, Department of the Army, September 2004.
[8]. Eric S. Toner **(2014) Creating Situational Awareness: A Systems Approach** http://www.lbjfire.org/safety/situational-awareness Accessed 09 March, 2014
[9]. Endsley M. R, (1995); Toward a Theory of Situation Awareness in Dynamic Systems. *In Human Factors Journal*, Volume 37(1), pages 32-64.
[10]. Baldwin. R. (1989) Rule Based Analysis of Computer Security, Technical Report TR-401, MIT LCS Lab.
[11]. Farmer D. and Eugene H. (1999) The COPS Security Checker System. Technical Report, CSD-TR-993, Purdue University.
[12]. Axelsson S. (2000) A Preliminary Attempt to Apply Detection and Estimation Theory to Intrusion Detection, Technical Report, Chalmers University of Technology,Department of Computer Engineering, Goteborg, Sweden
[13]. McGuinness B. and Foy J. L. (2000) A Subjective Measure of SA: The Crew Awareness Rating Scale (CARS). *In Proceedings of the First Human Performance, Situation*
[14]. *Awareness, and Automation Conference*, Savannah, Georgia, USA
[15]. Ammann P, Duminda W, and Saket K. (2002); Scalable, Graph-based Network Vulnerability, Analysis, *In Proceedings of 9th ACM Conference on Computer and Communications Security*, Washington, DC.
[16]. Cheung S. (2003) Modeling Multistep Cyber-attacks for Scenario Recognition. In DARPA, Information Survivability Conference and Exposition (DISCEXIII).
[17]. Prahlad F. and Wenke L. (2006) Evading Network Anomaly Detection Systems: Formal Reasoning and practical Techniques. *In Proceedings of the 13th ACM Conference on Computer and Communications Security* (CCS 2006), Alexandria, VA.
[18]. Mell P., Karen S., and Sasha R. (2007) A Complete Guide to the Common Vulnerability Scoring System Version 2.0. Forum of Incident Response and Security Teams (FIRST).
[19]. Jason L., Xinming O., Raj R. (2009) Uncertainty and Risk Management in Cyber Situational Awareness Published in ARO Workshop on Cyber Situational Awareness. Albany: State University of New York Press
[20]. Li J.(2009), Uncertainty and risk management in cyber situational awareness. , In Sushil Jajodia *et al.*, editor, *Cyber Situational Awareness: Issues and Research* , chapter 4. Springer, Nov. 2009.
[21]. Ning P., Ou P., Song X., Strater D., Swarup L., Tadda V., Wang G., C., & Yen J. (2010) Cyber SA: Situational Awareness for Cyber Defense
[22]. Schiller Bill (2009), "Chinese ridicule U of T spy report - But government officials choose words carefully, never denying country engages in cyber-espionage", The Star(Toronto, Ontario, Canada),
[23]. Moore, R. (2005) "Cyber crime: Investigating High-Technology Computer Crime," Cleveland, Mississippi: Anderson Publishing
[24]. Halder, D., & Jaishankar, K. (2011) Cyber crime and the Victimization of Women: Laws, Rights, and Regulations. Hershey, PA, USA: IGI Global.
[25]. Jason Andress,.and Winterfeld, Steve. (2011). Cyber Warfare: Techniques, Tactics and Tools for Security Practitioners. Syngress.
[26]. Clarke, Richard A. (2010) Cyber War, Harper Collins Publishers
[27]. Jeffrey Carr. (2010). Inside Cyber Warfare: Mapping the Cyber Underworld. O'Reilly
[28]. Lynn, William J. III. (2010) "Defending a New Domain: The Pentagon's Cyber Strategy", Foreign Affairs
[29]. Jason Li, Peng Liu, and Xinming Ou. (2008), Using Bayesian Networks for cyber security analysis. Manusrcipt, 2008.
[30]. Ou, X., Boyer, W. F., & McQueen, M. A. (2006). A Scalable Approach to Attack Graph Generation. In Proceedings of the 13th ACM Conference on Computer and Communications Security (pp.336–345). Alexandria, VA: ACM. Proceedings of Twenty-Second AAAI Conference on Artificial Intelligence (pp. 1535 – 1541) Vancouver, British Columbia, Canada
[31]. Xie P., Li J. H., Ou X., Liu P., & Levy R. (2010). Using Bayesian Networks for Cyber Security Analysis
[32]. Thomas L. Griffiths, Charles Kemp & Joshua B. Tenenbaum (2006). Bayesian Models of Cognition
[33]. Olasehinde, O.O. (2012). Design and Implementation of Denial of Service (DoS) Detection System Using Association Rule, M.Tech Thesis submitted to the Department of Computer Science, The Federal University of Technology, Akure, Ondo State, Nigeria
[34]. Mohammad Sazzadul Hoque, Md. Abdul Mukit & Md. Abu Naser Bikas (2012). An Implementation of Intrusion Detection System Using Genetic Algorithm, International Journal of Network Security & Its Applications (IJNSA), Vol.4, No.2, March 2012, DOI: 10.5121/ijnsa.2012.4208 109

ISBN: 978-988-19252-0-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)