ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/290061926

The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set

Article *in* Information Security Journal A Global Perspective · January 2016

DOI: 10.1080/19393555.2015.1125974?tab=permissions

CITATIONS
23

READS
2,522

2 authors:

Nour Moustafa
UNSW Canberra
53 PUBLICATIONS **692** CITATIONS
SEE PROFILE

Jill Slay
La Trobe University
146 PUBLICATIONS **1,580** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

A flow aggregator module for analysing network traffic View project

I am developing a new Geometric AREA analysis technique for detecting zero-day attacks based on the methodology of anomaly detetion View project

All content following this page was uploaded by Jill Slay on 08 October 2017.

The user has requested enhancement of the downloaded file.

Information Security Journal: A Global Perspective

ISSN: 1939-3555 (Print) 1939-3547 (Online) Journal homepage: http://www.tandfonline.com/loi/uiss20

The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set

Nour Moustafa & Jill Slay

To cite this article: Nour Moustafa & Jill Slay (2016): The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set, Information Security Journal: A Global Perspective, DOI: 10.1080/19393555.2015.1125974

To link to this article: http://dx.doi.org/10.1080/19393555.2015.1125974

Published online: 11 Jan 2016.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
http://www.tandfonline.com/action/journalInformation?journalCode=uiss20

Download by: [UNSW Library] **Date:** 12 January 2016, At: 15:45

INFORMATION SECURITY JOURNAL: A GLOBAL PERSPECTIVE
http://dx.doi.org/10.1080/19393555.2015.1125974

Taylor & Francis
Taylor & Francis Group

The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set

Nour Moustafa and Jill Slay

School of Engineering and Information Technology, University of New South Wales at the Australian Defence Force Academy, Canberra, Australia

ABSTRACT

Over the last three decades, Network Intrusion Detection Systems (NIDSs), particularly, Anomaly Detection Systems (ADSs), have become more significant in detecting novel attacks than Signature Detection Systems (SDSs). Evaluating NIDSs using the existing benchmark data sets of KDD99 and NSLKDD does not reflect satisfactory results, due to three major issues: (1) their lack of modern low footprint attack styles, (2) their lack of modern normal traffic scenarios, and (3) a different distribution of training and testing sets. To address these issues, the UNSW-NB15 data set has recently been generated. This data set has nine types of the modern attacks fashions and new patterns of normal traffic, and it contains 49 attributes that comprise the flow based between hosts and the network packets inspection to discriminate between the observations, either normal or abnormal. In this paper, we demonstrate the complexity of the UNSW-NB15 data set in three aspects. First, the statistical analysis of the observations and the attributes are explained. Second, the examination of feature correlations is provided. Third, five existing classifiers are used to evaluate the complexity in terms of accuracy and false alarm rates (FARs) and then, the results are compared with the KDD99 data set. The experimental results show that UNSW-NB15 is more complex than KDD99 and is considered as a new benchmark data set for evaluating NIDSs.

KEYWORDS

Feature correlations; multivariate analysis; NIDSs; UNSW-NB15 data set

Downloaded by [UNSW Library] at 15:45 12 January 2016

1. Introduction

Because of the ubiquitous usage of computer networks and the plurality of applications running on them, cyber attackers attempt to exploit weak points of network architectures to steal, corrupt, or destroy valuable information (DeWeese, 2009; Eom et al., 2012; Vatis, 2001). Consequently, the function of a NIDS is to detect and identify anomalies in network systems (Denning, 1987). NIDSs are classified into Misuse based (MNIDS) and Anomaly based (ANIDS) (Lee, Stolfo, & Mok, 1999; Moustafa & Slay, 2015a; Valdes & Anderson, 1995). In MNIDS, the known attacks are detected by matching the stored signatures of those attacks (Lee et al., 1999; Vigna & Kemmerer, 1999). While ANIDS creates a profile of normal activities, any deviation from this profile is considered as an anomaly (Ghosh, Wanken, & Charron, 1998; Valdes & Anderson, 1995). Several studies stated that the MNIDS can often accomplish higher accuracy and lower FAR than the ANIDS (Lee et al., 1999), but the ANIDS has the ability of detecting novel attacks (Lazarevic et al., 2003). Therefore, the ANIDSs are becoming a necessity rather than the MNIDS (Aziz et al., 2014; Bhuyan, Bhattacharyya, & Kalita, 2014; García-Teodoro, Díaz-Verdejo, Maciá-Fernández, & Vázquez, 2009).

Evaluating the efficiency of any NIDS requires a modern comprehensive data set that contains contemporary normal and attack activities. McHugh (2000), Tavallaee et al. (2009), and Moustafa and Slay (2015a) stated that the existing benchmark data sets, especially, KDD99 and NSLKDD, negatively affect the NIDS results because of three major problems. Firstly, lack of modern low footprint attack fashions, for instance, stealthy or spy attacks that change their styles over time to become similar to normal behaviors (Cunningham & Lippmann, 2000; Tavallaee et al., 2009). Second, the existing data sets were created two decades ago, indicating that the

CONTACT Nour Moustafa nour.abdelhameed@student.adfa.edu.au School of Engineering and Information Technology, University of New South Wales at the Australian Defence Force Academy, Northcott Drive, Campbell, ACT 2600, Canberra, Australia.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/uiss.

© 2016 Taylor & Francis

Downloaded by [UNSW Library] at 15:45 12 January 2016

normal traffic of the existing benchmark data sets is different from the current normal traffic because of the revolution in networks speed and applications (McHugh, 2000). Third, the testing set of the existing benchmark data sets has some attack types which are not in the training set; this means that the training and testing set have different distribution (Tavallaee et al., 2009). The difference in the distribution persuades classifier systems to skew toward some observations causing the FAR (Cieslak & Chawla, 2009; Tavallaee et al., 2009).

In the light of the above discussion, to address these challenges the UNSW-NB15 data set has recently been released (Moustafa & Slay, 2014, 2015b). This data set includes nine categories of the modern attack types and involves realistic activities of normal traffic that were captured with the change over time. In addition, it contains 49 features that comprised the flow based between hosts (i.e., client-to-server or server-to-client) and the packet header which covers in-depth characteristics of the network traffic.

A part of UNSW-NB15 data set was decomposed into two partitions of the training and testing sets to determine the analysis aspects. The goal of the three aspects is to evaluate the complexity of the training and testing sets. First, the Kolmogorov-Smirnov Test (Justel, Peña, & Zamar, 1997; Massey, 1951) defines and compares the distribution of the training and testing sets; skewness (Mardia, 1970) measures the asymmetry of the features; and kurtosis (Mardia, 1970) estimates the flatness of the features. The reliability of results can be achieved when these statistics are approximately similar to the features of the training and testing sets. Second, the feature correlations are measured in two perspectives: (1) the feature correlations without the class label, (2) and the feature correlations with the class label. To achieve the first perspective, Pearson's Correlation Coefficient (PCC) (Bland & Altman, 1995) is used. Gain Ratio (GR) method (Hall & Smith, 1998) is utilised to achieve the second perspective. Third, five existing techniques, namely, Naïve Bayes (NB)(Panda & Patra, 2007), Decision Tree (DT) (Bouzida & Cuppens, 2006), Artificial Neural Network (ANN) (Bouzida & Cuppens, 2006; Mukkamala, Sung, & Abraham, 2005), Logistic Regression (LR) (Mukkamala et al., 2005), and Expectation-Maximisation (EM) Clustering (Sharif, Prugel-Benett, & Wills, 2012) are executed on the training and testing sets to assess the complexity in terms of accuracy and FARs. Further, the results of this data set are compared with the KDD99 data set (KDDCUP1999, 2007) to identify the capability of the UNSW-NB15 data set in appraising existing and novel classifiers.

The objective of the paper is to analyse the UNSW-NB15 data set statistically and practically. First, in the statistical aspect, the distribution of data points specifies the suitable algorithms of classification. To be clear, if a data set follows Gaussian distribution, many statistical algorithms, for instance, HMM and Kalman filter are used. However, if a data set does not fit Gaussian distribution, other algorithms, for example, particle filter and mixture models are applied. Second, in the practical aspect, the adoption of the best attributes decrease false alarm rates and reduce the execution costs. For that purpose, feature correlations with label and without label are demonstrated.

The rest of this paper is organised as follows: Section 2 describes the UNSW-NB15 data set. Section 3 discusses the training and the testing sets extracted from this data set. Section 4 discusses the statistical mechanisms used on the two sets. Section 5 presents the feature correlation methods. Section 6 identifies the classification techniques which are involved to evaluate the complexity of the KDD99 and UNSW-NB15 data sets. Section 7 presents the experimental results of the statistical techniques, the feature correlations, and the complexity evaluation. Finally, section 8 provides a conclusion to the paper and examines the future research area.

2. Description of the UNSW-NB15 data set

The UNSW-NB 15 data set (Moustafa & Slay, 2014, 2015b) was created using an IXIA PerfectStorm tool (IXIA PerfectStormOne Tool, 2014) in the Cyber Range Lab of the Australian Centre for Cyber Security (ACCS) (Australian Center for Cyber Security (ACCS), 2014) to generate a hybrid of the realistic modern normal activities and the synthetic contemporary attack behaviors from network traffic. A tcpdump tool (tcpdump tool, 2014) was used to

Downloaded by [UNSW Library] at 15:45 12 January 2016

capture 100 GB of a raw network traffic. Argus (Argus tool, 2014), Bro-IDS (Bro-IDS Tool, 2014) tools were used and 12 models were developed for extracting the features of Tables 1, 2, 3, 4 and 5, respectively. These techniques were configured in a parallel processing to extract 49 features with the class label. After finishing the implementation of the configured techniques, the total number of records, 2,540,044, were stored in four CSV files. The records and the features of the UNSW-NB15 data set are described in-depth as follows.

Table 1. Flow features.

No.	Name	Description
1	*Srcip*	Source IP address
2	*Sport*	Source port number
3	*Dstip*	Destination IP address
4	*Dsport*	Destination port number
5	*Proto*	Protocol type (such as TC, UDP)

Table 2. Basic features.

6	*state*	Indicates to the state and its dependent protocol (such as ACC, CLO and CON).
7	*dur*	Record total duration
8	*sbytes*	Source to destination bytes
9	*dbytes*	Destination to source bytes
10	*sttl*	Source to destination time to live
11	*dttl*	Destination to source time to live
12	*sloss*	Source packets retransmitted or dropped
13	*dloss*	Destination packets retransmitted or dropped
14	*service*	Such as http, ftp, smtp, ssh, dns and ftp-data.
15	*sload*	Source bits per second
16	*dload*	Destination bits per second
17	*spkts*	Source to destination packet count
18	*dpkts*	Destination to source packet count

Table 3. Content features.

19	*swin*	Source TCP window advertisement value
20	*dwin*	Destination TCP window advertisement value
21	*stcpb*	Source TCP base sequence number
22	*dtcpb*	Destination TCP base sequence number
23	*smeansz*	Mean of the flow packet size transmitted by the src
24	*dmeansz*	Mean of the flow packet size transmitted by the dst
25	*trans_depth*	Represents the pipelined depth into the connection of http request/response transaction
26	*res_bdy_len*	Actual uncompressed content size of the data transferred from the server's http service

Table 4. Time features.

27	*sjit*	Source jitter (mSec)
28	*djit*	Destination jitter (mSec)
29	*stime*	record start time
30	*ltime*	record last time
31	*sintpkt*	Source interpacket arrival time (mSec)
32	*dintpkt*	Destination interpacket arrival time (mSec)
33	*tcprtt*	TCP connection setup round-trip time, the sum of 'synack' and 'ackdat'
34	*synack*	TCP connection setup time, the time between the SYN and the SYN_ACK packets
35	*ackdat*	TCP connection setup time, the time between the SYN_ACK and the ACK packets

Table 5. Additional generated features.

36	*is_sm_ips_ports*	If *srcip* (1) equals to *dstip* (3) and *sport* (2) equals to *dsport* (4), this variable assigns to 1 otherwise 0
37	*ct_state_ttl*	No. for each *state* (6) according to specific range of values of *sttl* (10) and *dttl* (11)
38	*ct_flw_http_mthd*	No. of flows that has methods such as Get and Post in http service
39	*is_ftp_login*	If the ftp session is accessed by user and password then 1 else 0
40	*ct_ftp_cmd*	No of flows that has a command in ftp session
41	*ct_srv_src*	No. of records that contain the same *service* (14) and *srcip* (1) in 100 records according to the *ltime* (26)
42	*ct_srv_dst*	No. of records that contain the same *service* (14) and dstip (3) in 100 records according to the *ltime* (26)
43	*ct_dst_ltm*	No. of records of the same *dstip* (3) in 100 records according to the *ltime* (26)
44	*ct_src_ ltm*	No. of records of the *srcip* (1) in 100 records according to the *ltime* (26)
45	*ct_src_dport_ltm*	No of records of the same *srcip* (1) and the *dsport* (4) in 100 records according to the *ltime* (26)
46	*ct_dst_sport_ltm*	No of records of the same *dstip* (3) and the *sport* (2) in 100 records according to the *ltime* (26)
47	*ct_dst_src_ltm*	No of records of the same *srcip* (1) and the *dstip* (3) in in 100 records according to the *ltime* (26)

2.1. Attack types

Attack types can be classified into nine groups:

(1) **Fuzzers**: an attack in which the attacker attempts to discover security loopholes in a program, operating system, or network by feeding it with the massive inputting of random data to make it crash.

(2) **Analysis**: a type of variety intrusions that penetrate the web applications via ports (e.g., port scans), emails (e.g., spam), and web scripts (e.g., HTML files).

(3) **Backdoor**: a technique of bypassing a stealthy normal authentication, securing unauthorized remote access to a device, and locating the entrance to plain text as it is struggling to continue unobserved.

(4) **DoS**: an intrusion which disrupts the computer resources via memory, to be extremely

Downloaded by [UNSW Library] at 15:45 12 January 2016

busy in order to prevent the authorized requests from accessing a device.

(5) **Exploit**: a sequence of instructions that takes advantage of a glitch, bug, or vulnerability to be caused by an unintentional or unsuspected behavior on a host or network.

(6) **Generic**: a technique that establishes against every block-cipher using a hash function to collision without respect to the configuration of the block-cipher.

(7) **Reconnaissance**: can be defined as a probe; an attack that gathers information about a computer network to evade its security controls.

(8) **Shellcode**: an attack in which the attacker penetrates a slight piece of code starting from a shell to control the compromised machine.

(9) **Worm**: an attack whereby the attacker replicates itself in order to spread on other computers. Often, it uses a computer network to spread itself, depending on the security failures on the target computer to access it.

2.2. Features

Features are categorized into five groups:

(1) **Flow features**: includes the identifier attributes between hosts (e.g., client-to-serve or server-to-client), as reflected in Table 1.

(2) **Basic features**: involves the attributes that represent protocols connections, as shown in Table 2.

(3) **Content features**: encapsulates the attributes of TCP/IP; also they contain some attributes of http services, as reflected in Table 3.

(4) **Time features**: contains the attributes time, for example, arrival time between packets, start/end packet time, and round trip time of TCP protocol, as shown in Table 4.

(5) **Additional generated features**: in Table 5, this category can be further divided into two groups: general purpose features (i.e., 36–40), whereby each feature has its own purpose, according to protect the service of protocols, and (2) connection features (i.e., 41–47) that are built from the flow of 100 record connections based on the sequential order of the last time feature.

To label this data set, two attributes were provided: *attack_cat* represents the nine categories of the attack and the normal, and *label* is 0 for normal and otherwise 1.

3. Training and testing set distribution

A NIDS data set can be conceptualized as a relational table (T) (Witten & Mining, 2005). The input to any NIDS is a set of instances (I) (e.g., normal and attack records). Each instance consists of features (F) that have different data types (i.e.,$\forall f \subset \{R \cup S\}$, where $\forall f$ means each feature in T, R is real numbers and S denotes characters). It is observed that NIDS techniques face challenges for using these features because no standard format for feature values (e.g., number or nominal) is offered(Shyu et al., 2005). In statistical perspective, T is a multivariate data representation which is codified in Definition 1.

Definition 1: Let $I_{1:N} \in T, I_{1:N} = \{f_{ij} \in F\}$, $Y_{1:N} = \{c_i \in C\}$, where $i, j = 1, 2 \ldots, N$. Suppose F is *iid* (independently and identically distributed). Defining $I_{1:N}$ and $Y_{1:N}$ as a column-vector, as given in Eq. (1).

$$I_{1:N} = \begin{bmatrix} f_{11} & f_{12} & \ddots \\ f_{21} & f_{22} & f_{ij} \end{bmatrix}, \; Y_{1:N} = \begin{bmatrix} c_1 \\ c_i \end{bmatrix} \tag{1}$$

such that I represents the observations of T,Y is the class label (C) for each I, N is the number of instances, F denotes the features of I.

Proposition 1: A standard format for features (F) is prepared to have a same type (i.e., (number only)$\forall F \subset \{R\}$) to make the analysis of the data points easier. it assigns each nominal feature (S) to a sequence of numbers (i.e.,$\forall S \rightarrow R_{0:R}$, where$\{0 : R\}$ denotes as equence of numbers (Salem & Buehler, 2012). For instance, the UNSW-NB15 data set has three major nominal features (e.g., protocol types (e.g., TCP, UPD), States (e.g., CON, ACC) and services (e.g., HTTP, FTP)).This issue can be tackled by converting each value in these features into ordered numbers such as TCP = 1, UDP = 2 and so on.

Downloaded by [UNSW Library] at 15:45 12 January 2016

Table 6. A part of UNSW-NB15 data set distribution.

Category	Training set	Testing set
Normal	56,000	37,000
Analysis	2,000	677
Backdoor	1,746	583
DoS	12,264	4089
Exploits	33,393	11,132
Fuzzers	18,184	6,062
Generic	40,000	18,871
Reconnaissance	10,491	3,496
Shellcode	1,133	378
Worms	130	44
Total Records	175,341	82,332

Table 6 demonstrates the creation of the training and the testing sets from the UNSW-NB15 data set; a part of the data set records has been divided with an approximate 60%:40% ratio of the training and testing sets, respectively. To achieve the authenticity of NIDS evaluations, no redundant records among the training and testing set.

4. Statistical descriptive observations

In this section, the statistical analysis of the training and the testing sets attributes are elaborated to measure the relationship of the two set. The training and testing sets are prepared to in the form of Eq. (1), namely, TR_{I_N} and TS_{I_N}. Kolmogorov-Smirnov (K-S) test (Justel et al., 1997; Massey, 1951), Multivariate skewness and kurtosis functions (Mardia, 1970) are customized to examine the relationship and the distribution nature of the TR_{I_N} and TS_{I_N}. The values of the features are not in a confidence interval, the z-score transformation (Jain, Nandakumar, & Ross, 2005) is used to make a standard format for these values of the attributes.

4.1. Z-score transformation of the feature values

The attributes of the TR_{I_N} and TS_{I_N} have a large scale between the minimum and the maximum value. Therefore, it is extremely difficult to estimate their variance from the central tendency of the distribution. This issue is formulated as a multivariate problem of the varied distributions in Dilemma 1.

Dilemma 1: $\forall f$ involves in I_N has a wide range of values, such that

$$max(f_{ij}) - min(f_{ij}) < \infty \cap max(f_{ij}) \gg min(f_{ij}) \tag{2}$$

In Eq. (2), each feature values are not specified into a confidence interval, for instance [−1, 1], and the maximum value (i.e., $Max(f_{ij})$)of the feature (F) is much larger than the minimum value ($i.e., Min(f_{ij})$). This causes a noise distribution, because the smallest and the largest values highly deviate from their mean (M) (Cherkassky & Mulier, 2007).

Proposition 2: To tackle Dilemma 1, the z-score function is utilised as formulated in Eq. (3). It is a linear transformation to standardise the format of the f_{ij} values, this makes it easier to compare values in diverse distributions without changing the shape of the original distribution.

$$z_{ij} = \frac{f_{ij} - M}{\delta} \tag{3}$$

In Eq. (3), the z-score of each value in $I_{1:\ N}$ is calculated by subtracting $\forall f$ from its M, and then the value is divided by the standard deviation(δ) of $\forall f$ to measure how far away $\forall f$ from its M. The sign of the z-score demonstrates that the value of $\forall f$ is either below or above M. After the transformation into z-score, all the distributions of the f_{ij} values are standardised with $M = 0$ and $\delta = 1$.

4.2. Kolmogorov-Smirnov (K-S) test

The K-S test is used to decide the proper distribution of the features (Massey, 1951; Shyu et al., 2005). Let $\forall f$ has $x_1, x_2, \ldots, x_n$ be an ascending order sample, an empirical distribution function $F_n(x)$is the fraction of sample observations less than or equal to the value x, then the empirical distribution function is defined as:

$$F_n(x) = \begin{cases} 0, x < x_1 \\ {}^{k}/_{n}, x_k \le x < x_{k+1}, k = 1, 2, \ldots, n-1 \\ 1, x \ge x_n \end{cases} \tag{4}$$

The Kolmogorov distribution function is denoted as:

$$f(x) = \frac{\sqrt{2\pi}}{x}\sum_{n=1}^{\infty} e^{-(2n-1)^2\pi^2/(8x^2)} \quad (5)$$

From Eqs. (4) and (5), K-S test achieves $F_n(x)$ fits $f(x)$ by maximizing the absolute difference as follows:

$$D_n = max_x|f(x) - F_n(x)| \quad (6)$$

In the case of the critical value ($D_{n,\propto}$) (i.e., $\propto$ denotes significance level) falls into the Kolmogorov-Smirnov table (KDDCUP1999, 2007), do $P(D_n \leq D_{n,\propto}) = 1 - \propto .D_n$ that can be used to test $\forall f$ within $f(x)$. Hence, the suitable fitting is accomplished in the case of $max_x|f(x) - F_n(x)| \leq D_{n,\propto}.F_n(x) \pm D_{n,\propto}$ affords a confidence interval to $f(x)$ to present the proper distribution of $\forall f$.

4.3. Multivariate skewness

The skewness method (Mardia, 1970) is an asymmetry measure of the probability distribution of $\forall f$ that has $x_1, x_2, \ldots, x_n$ from its M, skewness function can be defined as:

$$ske = \frac{\sum_{i=1}^{n}(x_i - M)^3}{n\delta^3} \quad (7)$$

In Eq. (7), if result is positive, the distribution with an asymmetric tail spreads toward major positive values. On the other hand, a negative value indicates that the distribution with an asymmetric tail extends toward more negative values.

4.4. Multivariate kurtosis

The kurtosis technique (Mardia, 1970) is a peakiness measure of the probability distribution of $\forall f$ that has $x_1, x_2, \ldots, x_n$, kurtosis is denoted as:

$$Kur = \frac{\sum_{i=1}^{n}(x_i - M)^4}{n\delta^4} \quad (8)$$

In Eq. (8), if the outcome is positive, the distribution is more peaked than the normal distribution. Nevertheless, a negative value indicates a flatter distribution.

It is acknowledged that if skewness and kurtosis values tend to be 0, then the distribution approximates a normal distribution.

4.5. The statistical functions utilisation on the TR_{I_N} and TS_{I_N}

The K-S test, Multivariate skewness and kurtosis functions are customised on the TR_{I_N} *and* TS_{I_N} to estimate the compatibility of them, as declared in Eqs. (9)–(11).

$$\forall f \epsilon_{TR_{I_N}} D_{n,TR_{I_N}} \overset{\rho}{\Leftrightarrow} \forall f \epsilon_{TS_{I_N}} D_{n,TS_{I_N}} \leq D_{n,\propto,\forall f} \quad (9)$$

$$\forall f \epsilon_{TR_{I_N}} ske_{TR_{I_N}} \overset{\rho}{\Leftrightarrow} \forall f \epsilon_{TS_{I_N}} ske_{TS_{I_N}} \quad (10)$$

$$\forall f \epsilon_{TR_{I_N}} kur_{TR_{I_N}} \overset{\rho}{\Leftrightarrow} \forall f \epsilon_{TS_{I_N}} kur_{TS_{I_N}} \quad (11)$$

Eq. (9) estimates the best fitting of the distribution to TR_{I_N} and TS_{I_N} of $\forall f$, achieving the two sides are less than or equal K-S test (i.e., $D_{n,\propto,\forall f}$), while Eqs. (10) and (11) calculate the skewness and the kurtosis to the TR_{I_N} *and* TS_{I_N} of $\forall f$, respectively. It is observed that ρ assigns to a threshold operator (e.g., = , < or >) which compares the results between the two sides of the equations.

Based on the above explanation, the TR_{I_N} *and* TS_{I_N} *of* $\forall f$ is analysed to evaluate the statistical relationship of them as in the following algorithm:

Algorithm 1: The statistical relationship between TR_{I_N} *and* TS_{I_N} *of* $\forall f$

Input ← TR_{I_N} *and* TS_{I_N} *of* $\forall f$

1: **for** $\forall f$ **do** // *for each feature contains in the training and testing sets do*

2: convert the values of the nominal features to numerical as described in proposition 1.

3: apply z-score from Eq. (3), subject to Eq. (2) as discussed in section 4.1.

4: apply Kolmogorov-Smirnov and multivariate skewness and kurtosis measures as declared in sections 4.2, 4.3 and 4.4 respectively).

Downloaded by [UNSW Library] at 15:45 12 January 2016

Downloaded by [UNSW Library] at 15:45 12 January 2016

5: compare the results of step 4 to the TR_{I_N} and TS_{I_N} as in formulated equations 9, 10 and 11 respectively.

6: **end for**

Output → the relationship of the TR_{I_N} and TS_{I_N} *of* $\forall f$ via the results of step 5

5. Feature correlations of the TR_{I_N} and TS_{I_N}

Correlation analysis is another aspect to identify the relationship of the TR_{I_N} *and* TS_{I_N} features. Two correlation analysis mechanisms are used. First, a Pearson's correlation coefficient technique (PCC) (Bland & Altman, 1995) measures the relevance between features without a label. Second, a Gain Ratio method (GR) (Hall & Smith, 1998) is applied to rank the correlation between features and the label. The major goal of these techniques is to recognise the correlation scores between the features either with or without the class label on theTR_{I_N} and TS_{I_N}, to estimate the efficiency of the features for discriminating the normal and attack observations.

Definition 1.1 (the extension of Definition 1): Let data set (T) has multiple features $f_1, f_2, \ldots, f_d \in I_{1:N}$ and class (C), each feature and the class have multiple values. For instance, $f_1 = \{x_1, x_2, \ldots, x_N\}$, $f_2 = \{y_1, y_2, \ldots, y_N\}$ and $C = \{c_1, c_2, \ldots, c_w\}$, where d is the number of features, N denotes the number of instances and w is the number of classes.

5.1. Feature correlations without labeled

Pearson's correlation coefficient (PCC) (Bland & Altman, 1995) is one of the simplest linear correlation methods to measure the dependency between features. From Definition1.1, the PCC of the features $f_1 and f_2$ is formulated as:

$$PCC(f_1, f_2) = \frac{cov(f_1, f_2)}{\sigma_{f_1} . \sigma_{f_2}} = \frac{\sum_{i=1}^{N} (x_i - M_{f_1})(y_i - M_{f_2})}{\sqrt{\sum_{i=1}^{N} (x_i - M_{f_1})^2} . \sqrt{\sum_{i=1}^{N} (y_i - M_{f_2})^2}} \quad (12)$$

where $cov()$ is the covariance and σ is the standard deviation, $M_{f_1} = {}^1/_N \sum_i^N x_i$ and $M_{f_2} = N \sum_i^n y_i$ indicate the means of f_1 and f_2 respectively.

Based on Eq. (12), the output of the PCC is in a range [−1, 1]. If the value is near to −1 or 1, it indicates a strong correlation between the two features. However, if the value is close to 0, it shows that there is no correlation between the features. A positive sign means that the two features are in the same direction, while a negative sign indicates that the two features are in the opposite trend.

To rank the strongest features, the mean for each PCC feature (i.e., $M_{pcc_{f_i}}$)is calculated, such that $M_{pcc_{f_i}} = {}^1/_N \sum_{i=1}^{N} PCC_{f_i}$, then the means are ordered descendingly to define the closest dependency features.

5.2. Feature correlations with labeled

The Gain Ratio technique (Hall & Smith, 1998) estimates the ratio between an information Gain method (IG) (Jain et al., 2005) and feature values. The GA is prepared to solve the problem of the IG when a feature has a large number of values compared with other features values. The IG is a feature selection method depends on an entropy function which is a measure of uncertainty of any feature. Let I be set of samples with w distinct classes. The entropy or the expected IG that is required to classify the observations is given by:

$$G(I) = -\sum_{i=1}^{m} p_i \log_2(p_i) \quad (13)$$

where p_i is a likely of the instance I that belongs to class C_i and is calculated by ${}^{I_i}/_{I_N}$.

For partitioning the I_i into subsets for each feature f, the expected IG is denoted as:

$$E(f) = -\sum_{i=1}^{m} G(I) \frac{I_{1i} + I_{2i} + \ldots\ldots + I_{mi}}{I_N} \quad (14)$$

From Eqs. (13) and (14), the encoding information that is gained to F is:

$$Gain(f) = G(I) - E(f) \quad (15)$$

The splitting value between the subsets I_{ir} is indicated as:

$$Split(I) = -\sum_{i=1}^{r}\left(\left|\frac{I_i}{I_N}\right|\right)\log_2\left(\left|\frac{I_i}{I_N}\right|\right) \quad (16)$$

In Eq. (16), the split value of I expresses the information generated by dividing I into r parts conforming to r on the features. From Eqs. (14) and (15), the GA can be defined as: $GR(f) = Gain(f)/Split(I)$, where the feature with the highest Gain ratio is selected as the splitting feature. Thus, the strongest features with the class label are evaluated and ranked by utilising the GR, where the scores of the features are in the descending order.

6. Techniques for evaluating the complexity

This section discusses the techniques that are used to evaluate the complexity in terms of accuracy and false alarm rates (FAR) on the UNSW-NB15 data set. The five techniques used are Naïve Bayes (NB) (Panda & Patra, 2007), Decision Tree (DT) (Bouzida & Cuppens, 2006), Artificial Neural Network (ANN) (Bouzida & Cuppens, 2006; Mukkamala et al., 2005), Logistic Regression (LR) (Mukkamala et al., 2005), and Expectation-Maximization (EM) Clustering (Sharif et al., 2012). Each technique has its own characteristics to learn and evaluate data points of the TR_{I_N} and TS_{I_N} which are described respectively in the following section. First, the NB classifier is a conditional probability model which constructs the classification of the two classes (i.e., normal (0) or anomaly (1)). It is applied by the maximum a posterior (MAP) function which is denoted as:

$$P(C|I) = \underset{w\in\{1,2,..,N\}}{\operatorname{argmax}}\ P(C_w)\prod_{j=1}^{N}P(I_j|C_w) \quad (17)$$

where C is the class label, I is the observation of each class, w is the class number, $P(C|I)$ denotes the probability of the class given a specified observation and $\prod_{j=1}^{N}P(I_j|C_w)$ indicates to multiply all the probabilities of the instances conditionally to their classes to achieve the maximum outcome. Second, the DT classifier is a structure similar to a flowchart which consists of root, nodes and branches to represent the classification rules. Each node denotes rules or procedures on a feature, each branch contains the results of the rules; and each leaf node expresses a class label. Third, the ANN learning is used to approximate an activation function that depends on a large number of input observations I. The basic ANN function can be defined as:

$$f(I) = \tau\left(\sum_j W_j.I_j\right) \quad (18)$$

where $f(I)$ represents a predicted output of the class label, τ is an activation function (i.e., Sigmoid), W_j is a weight of each input instance I_j. Fourth, Logistic Regression algorithm establishes the correlation between a dependent variable (C) and independent variables (F). It uses the maximum likelihood function to estimate the regression parameters. Fifth, Expectation-Maximization (EM) clustering technique depends on maximizing the probability density function of a Gaussian distribution to calculate the mean and the covariance of each instance I in T. The EM clustering algorithm encompasses into two steps (i.e., Expectation (E-step) and Maximization (M)). In the E-step, it estimates the likelihood for each instance I in T. whilst, the M-step is utilised to re-estimate the parameter values from the E-step to achieve the best expected output.

Two parameters (accuracy and false alarm rate) are calculated from the outcomes of these techniques to measure the complexity of the UNSW-NB15 data set. Let the factors of the classification are $FC = \{TP, TN, FP, FN\}$ where TP (i.e., true positive) denotes a number of the correctly attack classified, (i.e., true negative) expresses a number of the correctly normal classified, FP (i.e., false positive) is a number of the misclassified attacks and FN (i.e., false negative) is a number of the misclassified normal records (Sokolova, Japkowicz, & Szpakowicz, 2006). The accuracy (So-In et al., 2014; Sokolova et al., 2006) is the rate of the correctly classified records to all the records, whether correctly or incorrectly classified, which is denoted as

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (19)$$

Downloaded by [UNSW Library] at 15:45 12 January 2016

Downloaded by [UNSW Library] at 15:45 12 January 2016

The false alarm rate (FAR) is the average ratio of the misclassified to classified records either normal or abnormal as denoted in Eq. (22). It is designed from Eqs. (20) and (21) to calculate the false positive rate (FPR) and the false negative rate (FNR), respectively.

$$FPR = \frac{FP}{FP + TN} \tag{20}$$

$$FNR = \frac{FN}{FN + TP} \tag{21}$$

$$FAR = \frac{FPR + FNR}{2} \tag{22}$$

7. Results and discussion

This paper examines analytical approaches to measure the complexity of the UNSW-NB15 data set which was developed to evaluate NIDSs. The study uses three approaches: (7.1) the statistical explanation (i.e., K-T test, multivariate skewness and kurtosis measures), (7.2) the features correlations (i.e., PCC and GR), and (7.3) and the complexity evaluation using the five classifiers. To measure the complexity of the UNSW-NB15 data set within the adopted part of the training and the testing sets, as presented in Table 6. The features that are selected to execute these aspects are reflected in Table 7.

7.1. The statistical explanation

The SPSS tool (SPSS tool, 2014) is utilised to analyse the statistical explanation and to determine the distribution nature of the training and the testing sets. Figure 1 shows the probability of the features on the training (TR) and the testing (TS) sets. The results demonstrate that the features distribution is a nonlinearity and nonnormality representation. The fitting percentage of these features in the two sets is 78%, when the two lines of the TR and the TS are identical. On the other hand, the nonfitting percentage is 22%, which is caused of the TR records are excessively more than the TS records.

To measure the asymmetry of the *TR* and the *TS* sets features (*F*); a multivariate skewness (*skw*) function is executed. In the *TR*, the empirical outcomes show that all the features, except 7, are positive. Therefore, the majority of the features are on the right side of the probability density function distribution which is longer or fatter than the left side.

Table 7. The features of the analysis.

Id	Names	Id	Names
1	dur	22	synack
2	spkts	23	ackdat
3	dpkts	24	smean
4	sbytes	25	dmean
5	dbytes	26	trans_depth
6	rate	27	response_body_len
7	sttl	28	ct_srv_src
8	dttl	29	ct_state_ttl
9	sload	30	ct_dst_ltm
10	dload	31	ct_src_dport_ltm
11	sloss	32	ct_dst_sport_ltm
12	dloss	33	ct_dst_src_ltm
13	sinpkt	34	is_ftp_logn
14	dinpkt	35	ct_ftp_cmd
15	sjit	36	ct_flw_http_mthd
16	djit	37	ct_src_ltm
17	swin	38	ct_srv_dst
18	stcpb	39	is_sm_ips_ports
19	dtcpb	40	proto
20	dwin	41	service
21	tcprtt	42	state

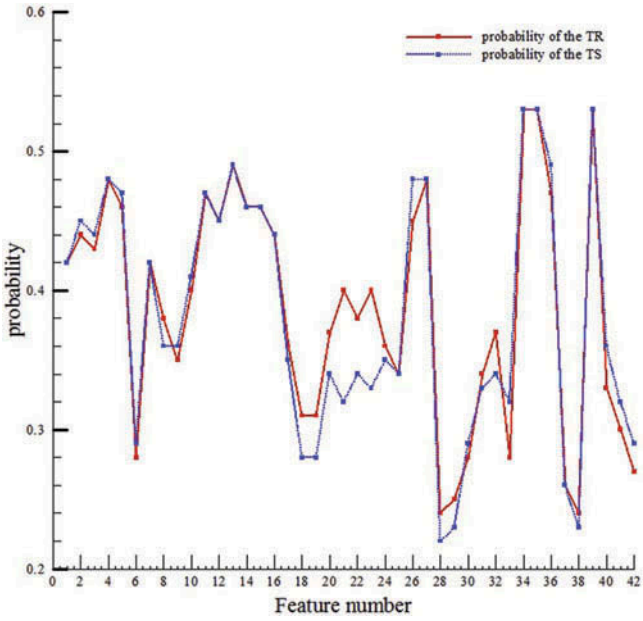

Figure 1. The probability distribution of the features on the training and testing sets.

The *TS* is approximately similar to the *TR*, almost majority of the features are positive, except the features 7, 17 and 20. Figure 2 demonstrates the skewness of the *TR* and *TS* features, where the relationship percentage is 82% when the two lines are similar. The most skewed features are 25, 26, 27 and 28, conversely, the lowest skewness are 6, 7, 8, 17, 18, 19, 29, 30, 31, 32, 37, 38 and 42.

To estimate the Preakness of the *TR* and *TS* features, a multivariate kurtosis function is

Downloaded by [UNSW Library] at 15:45 12 January 2016

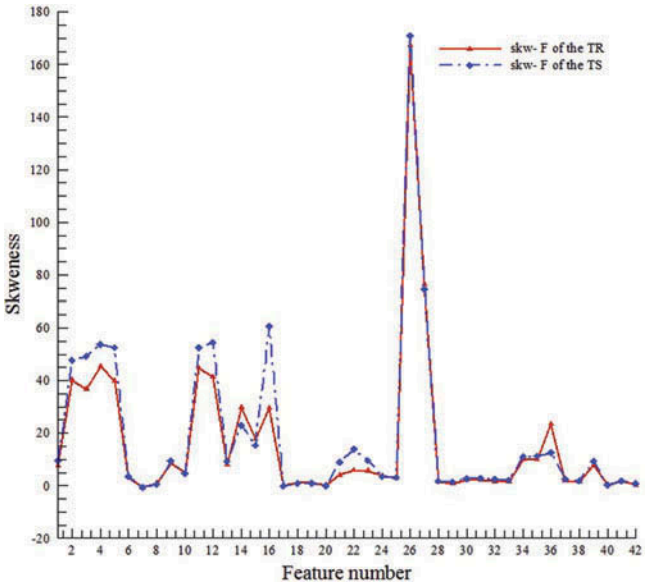

Figure 2. The skewness of the features on the training and testing sets.

implemented. In the *TR*, the experimental results indicate that the 7 features, 7, 8, 17, 18, 19, 20 and 40, have a negative value which is a flatter distribution. However, the rest of the features are positive; this leads to the distribution are higher than a normal distribution. In Figure 3, the kurtosis comparative of the TR and TS features (F) illustrates that the two lines fit 76% of the features. The kurtosis of the TS is higher than the TR in the features 2–6, 10–12, 15–16, 25–28 and 35–37. Conversely, the other features are almost in the close proximity.

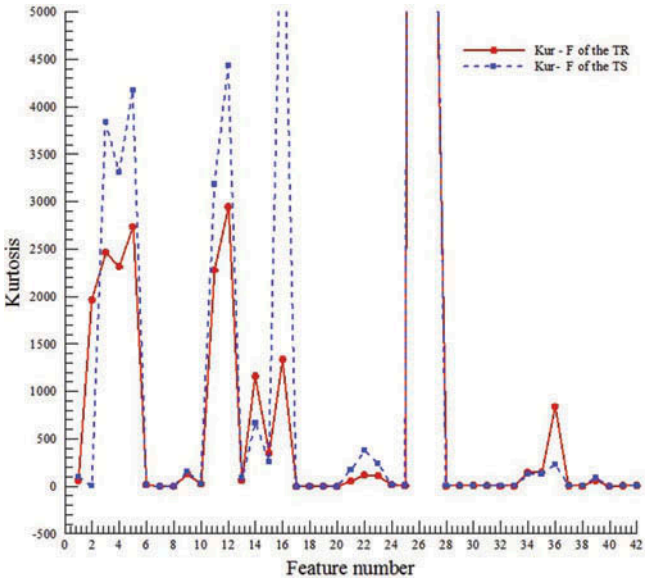

Figure 3. The kurtosis of the features on the training and testing sets.

Based on the above observation, the features of the training and testing sets are a highly statistical correlation. Consequently, this part of the UNSW-NB15 data set is reliable to evaluate classifier techniques because the training and the testing sets have the same characteristics of the non-linearity and non-normality, and the compatibility of the skewness and kurtosis values is acceptable. Overall, this shows that the data set can be used to evaluate the existing and novel methods of NIDSs.

7.2. Feature correlations

The correlation of features is evaluated based on two aspects; without the label using the PCC and with the label using the GR using the MTLAB tool (Matlab Tool, 2014) to identify the percentage of the closeness or the remoteness of the features.

The PCC estimates the score of each feature in the TR and the TS. As can be seen in Figure 4, in the TR and the TS, the features have the same correlated score. These features are ranked into a specified range [−0.01, 0.11]. The highest related features are *5, 3, 6, 12, 13, 29, 31, 32, 33, 34, 38* and *39*. On the contrary, the lowest correlated features are *7, 8, 10, 11, 14, 40* and *41*. Otherwise, the features almost fall into the middle of the range. The correlated features are classified into high, middle and low, which get probability values $\frac{12}{42}$, $\frac{23}{42}$ and $\frac{7}{42}$. This means that 87.5% are acceptable correlations of the high and middle correlated features.

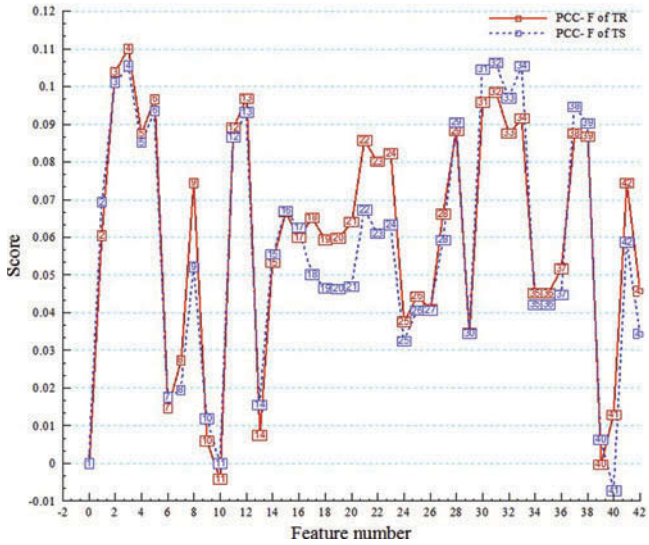

Figure 4. The PCC of the features on the training and testing sets.

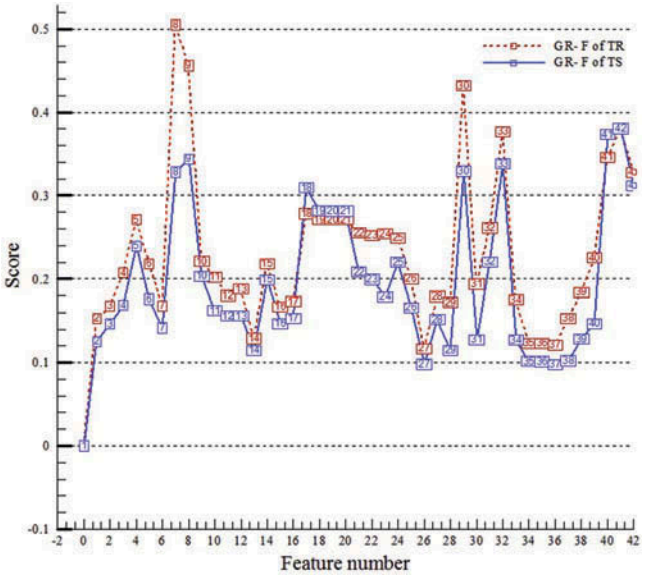

Figure 5. The Gain Ratio of the features on the training and the testing set.

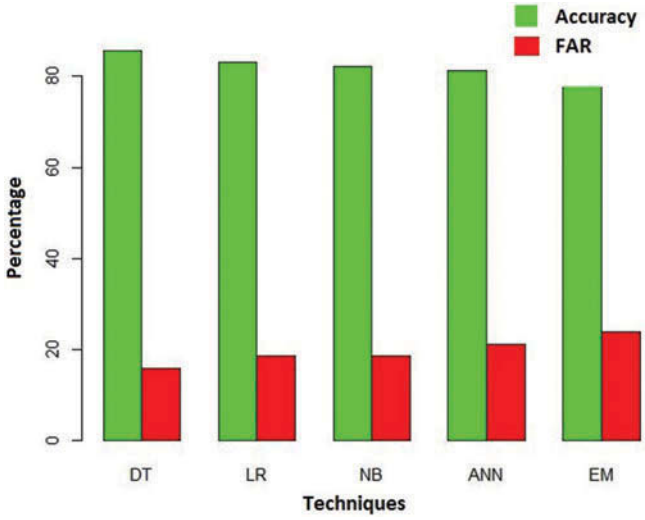

Figure 6. The comparison of the fifth techniques on the UNSW-NB15 data set.

In the TR and TS, the GR measures the correlation of each feature with the class label. Figure 5 shows the features are extremely similar in the correlation. The specified range of the score is [0.01,0.56] which is classified into low, middle and high according to the ranges [0.01, 0.2], [0.21,0.3], [0.31,0.56], respectively. The lowest related features are *1, 2, 3, 7, 12, 13, 14, 16, 17, 27, 28, 29, 34, 35, 36, 37, 38* and *39*. The middle ranked features are *4, 5, 6, 10, 11, 15, 18, 19, 20, 21, 22, 23, 24, 25, 26, 31, 32* and *40*. The highest correlated features are 8, 9, 30, 33, 41, and 42. The probability of this group is $\frac{18}{42}$, $\frac{18}{42}$, and $\frac{6}{42}$ respectively, thus the acceptable correlation rate is 57.2% of the high and the middle correlated features.

7.3. Complexity evaluation of the classifier techniques

To evaluate the complexity of the UNSW-NB15 data set (i.e., the training and the testing sets) in terms of accuracy and false alarm rate (FAR), the five techniques (e.g., NB, DT, ANN, LR, and EM clustering) are executed. These techniques are built in Visual Studio Business Intelligence 2008 (Visual Studio 2008, 2014) and are implemented with the default input parameters. Figure 6 represents the comparison of the fifth techniques in the terms of the accuracy and FAR in the x-axis and in the y-axis shows the percentage. The DT technique accomplishes the highest accuracy (i.e., 85.56%) and the lowest FAR (15.78%). On the other hand, the EM-clustering achieves the lowest efficiency where the accuracy is 78.47% and the FAR is 23.79%.

In Table 8, the comparative results of the KDD99 and the UNSW-NB15 data sets are elaborated. Overall, the results of the accuracy and the FAR of these techniques using the KDD99 data set are better than the UNSW-NB15 data set. There are two perspectives demonstrate the complexity of the UNSW-NB15 data set compared to the KDD99 data set. First, from the perspective of the network traffic behavior, the UNSW-NB15 data set contains a variety of the contemporary attack and normal

Table 8. Comparison between the results of the KDD99 and UNSW-NB15 data set.

	KDD99 data set			UNSW-NB15 data set	
Techniques	Reference	Accuracy (%)	FAR (%)	Accuracy (%)	FAR (%)
DT	(Bro-IDS Tool, 2014)	92.30	11.71	85.56	15.78
LR	(Witten & Mining, 2005)	92.75	-	83.15	18.48
NB	(Shyu et al., 2005)	95	5	82.07	18.56
ANN	(Witten & Mining, 2005)	97.04	1.48	81.34	21.13
EM clustering	(Salem & Buehler, 2012)	78.06	10.37	78.47	23.79

Downloaded by [UNSW Library] at 15:45 12 January 2016

behaviors. On the contrary, the attack and normal behaviors of the KDD99 data set are outdated. Additionally, the similarities of the normal and the attack observations in majority of the features add another factor to the complexity of the UNSW-NB15 data set.

Second, from the perspective of the statistical based test, as shown in Figures 1, 2, and 3, the features of the training and the testing sets are a highly correlation because the features are almost similar in the skewness and the kurtosis indicators. Further, the training and the testing sets have the same distribution which is non-linear and non-normal. As a result, the two perspectives demonstrate the major reasons of the complexity of the UNSW-NB15 data set compared to the KDD99 data set.

8. Conclusion and future work

In this paper, the analysis and the evaluation of the UNSW-NB15 data set are discussed. A part from this data set is divided into a training set and testing set to examine this data set. The training and testing sets are analysed in three aspects of the statistical analysis phase, the feature correlation phase and the complexity evaluation phase. First, the features of the two sets are converted into numerical values to be statistically processed and normalized using the z-score transformation to prevent the change in the original distribution. The statistical results show that the two sets are of the same distribution, nonnormal and non-linear, using the Kolmogorov-Smirnov test. Further, the skewness and kurtosis indicators of the training and the testing set are statistically similar. Second, the feature correlations of the training and the testing sets are measured either with the class label (i.e., the Pearson's correlation coefficient method) or without the label (i.e., the Gain Ratio technique). The feature correlations results demonstrate that these features are highly relevant observations. Third, the five techniques of the DT, LR, NB, ANN, and EM clustering are used to measure the complexity in terms of accuracy and False Alarm Rate (FAR) of this data set, and then the results are compared using the KDD99 data set. The evaluation results of the five techniques show that the DT technique accomplishes the best efficiency compared to others. For comparing the results of the two data sets, the efficiency techniques using the KDD99 data set are better than the UNSW-NB15 data set. As a consequence, the UNSW-NB15 data set is considered complex due to the similar behaviours of the modern attack and normal network traffic. This means that this data set can be used to evaluate the existing and the novel methods of NIDSs in a reliable way.

In the future, we plan to develop a new classification technique to identify the anomalies from the nonlinearity and non-normality data representation.

ORCID

Nour Moustafa http://orcid.org/0000-0001-6127-9349

References

Argus tool. (2014). Retrieved from http://qosient.com/argus/flowtools.shtml.

Australian Center for Cyber Security (ACCS). (2014). Retrieved from http://www.accs.unsw.adfa.edu.au/

Aziz, A. S. A., Azar, A. T., Hassanien, A. E., & Hanafy, S. E. (2014). Continuous features discretization for anomaly intrusion detectors generation. In *Proceedings of the 17th Online World Conference on Soft Computing in Industrial Applications* (pp. 209–221). Switzerland: Springer.

Bhuyan, M. H., Bhattacharyya, D. K., & Kalita, J. K. (2014). Network anomaly detection: Methods, systems and tools. *IEEE Communications Surveys & Tutorials*, *16* (1), 303–336. doi:10.1109/SURV.2013.052213.00046

Bland, J. M., & Altman, D. G. (1995). Statistics notes: Calculating correlation coefficients with repeated observations: Part 2—correlation between subjects. *Bmj*, *310* (6980), 633. doi:10.1136/bmj.310.6980.633

Bouzida, Y., & Cuppens, F. (2006). *Neural networks vs. decision trees for intrusion detection*. IEEE/IST Workshop on Monitoring, Attack Detection and Mitigation (MonAM), Tuebingen, Germany.

Bro-IDS Tool. (2014). Retrieved from https://www.bro.org/.

Cherkassky, V., & Mulier, F. M. (2007). *Learning from data: Concepts, theory, and methods*. Hoboken, NJ: John Wiley & Sons.

Cieslak, D. A., & Chawla, N. V. (2009). A framework for monitoring classifiers' performance: When and why failure occurs? *Knowledge and Information Systems*, *18* (1), 83–108. doi:10.1007/s10115-008-0139-1

Cunningham, R., & Lippmann, R. (2000). Detecting computer attackers: Recognizing patterns of malicious stealthy behavior. *MIT Lincoln Laboratory–Presentation to CERIAS*, *11*, 29.

Denning, D. E. (1987). An intrusion-detection model. *IEEE Transactions on Software Engineering*, *SE-13* (2), 222–232. doi:10.1109/TSE.1987.232894

Downloaded by [UNSW Library] at 15:45 12 January 2016

Downloaded by [UNSW Library] at 15:45 12 January 2016

DeWeese, S. (2009). *Capability of the People's Republic of China (PRC) to conduct cyber warfare and computer network exploitation*. Darby, PA: DIANE Publishing.

Eom, J.-H., Kim, S.-H., & Chung, T.-M. (2012). *Cyber military strategy for cyberspace superiority in cyber warfare*. in 2012 International Conference on Cyber Security, Cyber Warfare and Digital Forensic (CyberSec), IEEE.

García-Teodoro, P., Díaz-Verdejo, J., Maciá-Fernández, G., & Vázquez, E. (2009). Anomaly-based network intrusion detection: Techniques, systems and challenges. *Computers & Security*, *28* (1–2), 18–28. doi:10.1016/j.cose.2008.08.003

Ghosh, A. K., Wanken, J., & Charron, F. (1998). *Detecting anomalous and unknown intrusions against programs*. Computer Security Applications Conference, 1998. Proceedings. 14th Annual. IEEE.

Hall, M. A., & Smith, L. A. (1998). Practical feature subset selection for machine learning. In McDonald, C. (Ed.), *Computer Science '98 Proceedings of the 21st Australasian Computer Science Conference ACSC'98* (pp. 181–191). Berlin, Germany: Springer.

IXIA PerfectStormOne Tool. (2014). Retrieved from http://www.ixiacom.com/products/perfectstorm

Jain, A., Nandakumar, K., & Ross, A. (2005). Score normalization in multimodal biometric systems. *Pattern Recognition*, *38* (12), 2270–2285. doi:10.1016/j.patcog.2005.01.012

Justel, A., Peña, D., & Zamar, R. (1997). A multivariate Kolmogorov-Smirnov test of goodness of fit. *Statistics & Probability Letters*, *35* (3), 251–259. doi:10.1016/S0167-7152(97)00020-5

KDDCUP1999. (2007). Retrieved from http://kdd.ics.uci.edu/databases/kddcup99/KDDCUP99.html

Lazarevic, A., Ertoz, L., Kumar, V., Ozgur, A., & Srivastava, J. (2003). *A comparative study of anomaly detection schemes in network intrusion detection*. SDM. SIAM.

Lee, W., Stolfo, S. J., & Mok, K. W. (1999). *A data mining framework for building intrusion detection models*. Proceedings of the 1999 IEEE Symposium on Security and Privacy, 1999. IEEE.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, *57* (3), 519–530. doi:10.1093/biomet/57.3.519

Massey, F. J., Jr (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association*, *46* (253), 68–78. doi:10.1080/01621459.1951.10500769

Matlab Tool. (2014). Retrieved from http://au.mathworks.com/products/matlab/?refresh=true

McHugh, J. (2000). Testing intrusion detection systems: A critique of the 1998 and 1999 DARPA intrusion detection system evaluations as performed by Lincoln Laboratory. *ACM Transactions on Information and System Security*, *3* (4), 262–294. doi:10.1145/382912.382923

Moustafa, N., & Slay, J. (2014, May) *UNSW-NB15 DataSet for Network Intrusion Detection Systems*. Retrieved from http://www.cybersecurity.unsw.adfa.edu.au/ADFA%20NB15%20Datasets

Moustafa, N., & Slay, J. (2015a). *Creating novel features to anomaly network detection using DARPA-2009 data set*. 14th European Conference on Cyber Warfare and Security ECCWS-2015. The University of Hertfordshire, Hatfield, UK.

Moustafa, N., & Slay, J. (2015b). *UNSW-NB15: A comprehensive data set for network intrusion detection*. 2015 Military Communications and Information Systems Conference. Canberra, Australia: MilCIS 2015-IEEE Stream.

Mukkamala, S., Sung, A. H., & Abraham, A. (2005). Intrusion detection using an ensemble of intelligent paradigms. *Journal of Network and Computer Applications*, *28* (2), 167–182. doi:10.1016/j.jnca.2004.01.003

Panda, M., & Patra, M. R. (2007). Network intrusion detection using naive bayes. *International Journal of Computer Science and Network Security*, *7* (12), 258–263.

Salem, M., & Buehler, U. (2012). Mining techniques in network security to enhance intrusion detection systems. *International Journal of Network Security & Its Applications (IJNSA)*, *4* (6). doi:10.5121/ijnsa

Sharif, I., Prugel-Benett, A., & Wills, G. (2012). Unsupervised clustering approach for network anomaly detection, networked digital technologies. In Benlamri, R., (Ed), *Networked Digital Technologies* (Vol. 293, pp. 135–145). Communications in Computer and Information Science. Berlin, Germany: Springer Berlin Heidelberg.

Shyu, M.-L., Sarinnapakorn, K., Kuruppu-Appuhamilage, I., Chen, S.-C., Chang, L., & Goldring, T. (2005). *Handling nominal features in anomaly intrusion detection problems*. 15th International Workshop on Research Issues in Data Engineering: Stream Data Mining and Applications, 2005. RIDE-SDMA 2005. IEEE.

So-In, C., Mongkonchai, N., Aimtongkham, P., Wijitsopon, K., & Rujirakul, K. (2014). *An evaluation of data mining classification models for network intrusion detection*. 2014 Fourth International Conference on Digital Information and Communication Technology and its Applications (DICTAP), IEEE.

Sokolova, M., Japkowicz, N., & Szpakowicz, S. (2006). Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation. In *AI 2006: Advances in artificial intelligence* (Vol. 4304, pp. 1015–1021). Lecture Notes in Computer Science. Berlin, Germany: Springer.

SPSS tool. (2014). Retrieved from http://www-01.ibm.com/software/analytics/spss/

Tavallaee, M., (2009). A detailed analysis of the KDD CUP 99 data set. In *Proceedings of the Second IEEE Symposium on Computational Intelligence for Security and Defence Applications 2009* (pp. 53–58). Piscataway, NJ: IEEE.

tcpdump tool. (2014). Retrieved from http://www.tcpdump.org/

Valdes, A., & Anderson, D. (1995). Statistical methods for computer usage anomaly detection using NIDES (Next-

Generation Intrusion Detection Expert System). In *Proceedings of the Third International Workshop on Rough Sets and Soft Computing (RSSC94)*, (pp. 306–311). San Jose, CA: USW.

Vatis, M. A. (2001). *Cyber attacks during the war on terrorism: A predictive analysis*. DTIC Document. Hanover, NH: Institute for Security Technology Studies at Dartmouth College.

Vigna, G., & Kemmerer, R. A. (1999). NetSTAT: A network-based intrusion detection system. *Journal of Computer Security*, *7*, 37–71.

Visual Studio 2008. (2014). Retrieved from http://msdn.microsoft.com/en-us/library/ms175595(v=sql.100).aspx

Witten, I. H., & Mining, E. F. D. (2005). *Practical machine learning tools and techniques (The Morgan Kaufmann Series in Data Management Systems)*. San Francisco, CA: Elsevier.

Biographies

Nour Moustafa is a PhD candidate at the School of Engineering and Information Technology (SEIT) in the University of New South Wales, Canberra, Australia. He is an IEEE student member. He received his bachelor degree in 2009 and his master's degree in 2014, at the faculty of computer and Information, Helwan University, Egypt. His areas of interests include cyber security, in particular, network intrusion detection systems, data mining, and machine learning mechanisms

Professor Jill Slay is director of the Australian Centre for Cyber Security at UNSW Canberra at ADFA. She has established an international research reputation in cyber security and has worked in collaboration with many industrial partners. She has published more than 92 research outputs in information assurance and supervised 16 PhDs.

Downloaded by [UNSW Library] at 15:45 12 January 2016

View publication stats