New Intelligent Heuristic Algorithm to Mitigate Security Vulnerabilities in IPv6

Abdulrahman Salih[†], Dr. Xiaoqi Ma[††], Dr Evtim Peytchev[†††]
FB104480@ntu.ac.uk , xiaoqi.ma@ntu.ac.uk , peytchev.evtim@ntu.ac.uk
School of Science & Technology, Nottingham Trent University, NG11 8NS, Clifton, Nottingham, United Kingdom

Abstract
Zero day Cyber-attacks created potential impacts on the way information is held and protected, however one of the vital priorities for governments, agencies and organizations is to secure their network businesses, transactions and communications, simultaneously to avoid security policy and privacy violations under any circumstances. Covert Channel is used to in/ex-filtrate classified data secretly, whereas encryption is used merely to protect communication from being decoded by unauthorized access. In this paper, we propose a new Security Model to mitigate security attacks on legitimate targets misusing IPv6 vulnerabilities. The approach analyses, detects and classifies hidden communication channels through implementing an enhanced feature selection algorithm with a coherent Naive Bayesian Classifier. NBC is one of the most prominent classification algorithm defining the highest probability in data mining area. The proposed framework uses Intelligent Heuristic Algorithm (IHA) to analyse and create a novel primary training data, furthermore a modified Decision Tree C4.5 technique is suggested to classify the richest attribute presenting hidden channels in IPv6 network. The results evaluation showed better detection performance, high accuracy in True Positive Rate (TPR) and a low False Negative Rate (FNR) and a clear attribute ranking.

Key words:
Covert Channel, Cyber Security, IPv6, ICMPv6, Decision Trees C4.5. Naïve Bayes.

1. Introduction

Internet Protocol version 6 (IPv6) as shown in Figure 1, expressly designed as a successor for IPv4. While the protocol itself is already over a decade old but currently its adoption's infancy reaching 7% in the world. The low acceptance of IPv6 results in an insufficient understanding of its security properties as mentioned in [1], despite of the security improvements, IPv6 had no cryptographic protection when it was deployed and even the successful deployment of IPsec within IPv6 would not give any guarantee or additional security against hidden channel attacks [9]. Covert channels have been defined in many ways; Lampson (1973) was the first that recognized them as storage channels between two monolithic systems However these channels were not meant to be used for communications [5]. Most of researchers in Network and data security defined them as enforced, illicit signalling channels that allow a user to stealthily, contravene targeted objective [2], [4].

The protocol dimension representing the changed and new fields values in pcap data according to the multi-level separation policy and unobservable requirements of any RFC 2460 as shown in Figure 2. The utmost information that an IPv6 packet could carry is distributed into 40 octets (320 bits) called mandatory fixed headers and sub optional extension headers. IPv6 consists of eight main header fields, these fields have potential to carry covert channels depending on each fields modified values in the packet transmission over the net as indicated in [1], [2] and [4].

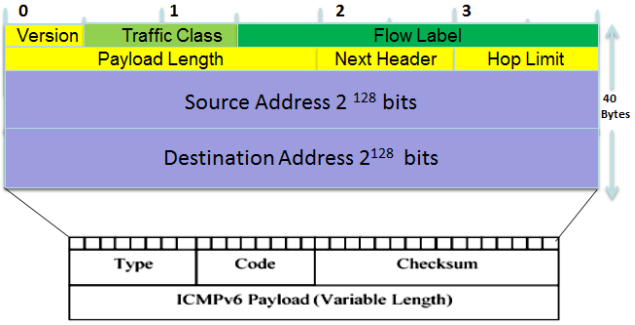

Fig. 1: IPv6 Header Format

There are two types of covert channels: storage and timing in which categorized under two types of taxonomies: variable and predictable according to our performed analysis and the protocol's RFCs standard values in [3], [4]. A predictable cover means there is no variation, whereas a variable cover means there is a limited variation. Internet Control Message Protocol version 6 (ICMPv6) as shown in Figure 3 is a vital component and an integral part of IPv6 and must be fully implemented by every IPv6 node according to RFC 4443 protocol specifications and IANA assigned values [1]. Table I shows examples of possible IPv6 covert channels characteristics.

DOI : 04.IJIS.2015.1.3

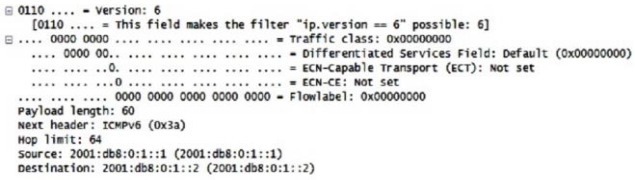

Fig. 2: IPv6 PCAP Data in Header Fields

ICMPv6 reports errors encountered in processing packets [6], [8,] and it does other internet-layer functions such as diagnostics. It produces two types of messages: Information Notification and Error Notification using type and code fields to differentiate services, in which both are vulnerable to variety of attacks including; denial of Service (DoS), Man-in-the-Middle (MITM) and spoofing attacks [3],[10].

Each of these messages carries a next header value of 58, which includes a Type value for message specification. The Type ranges (1-127) are for error messages and from (128-255) are for information messages. Having said that and the arbitrary content of the ICMPv6 payload may carry different types of data according to the messages types and ranges mentioned earlier, besides the Operating Systems type used too [12].

Table 1: Identified Covert Channels in IPv6 Header

ID	Field	Covert Channel	Bandwidth
1	Traffic Class	Set a false traffic class	8bits/packet
2	Flow Label	Set a false flow label	20 bits/packet
3	Payload Length	Increase value to insert extra data	Various
4	Next Header	Set a valid value to add an extra extension header	Various
5	Hop limit	Increase/decrease value	≈ 1 bit/packet
6	Source Address	Set a false source address	16bytes/packet

However, sometimes ICMPv6 packet contains insignificant or null values which indicate that potential covert channels could be existed, although ICMPv6 cannot do anything if the protocol itself commits an error [4].

The research question that motivated this project was: Is it possible to detect hidden communication channels in IPv6? If yes, then another sub question could be asked. Why this Internet Protocol has not been investigated by many researchers and what possible different method can be implemented to tackle and mitigate the security vulnerabilities in this New Generation Protocol?

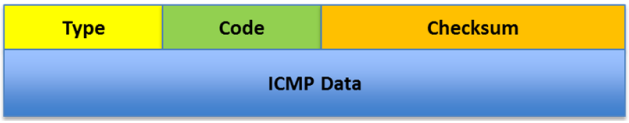

Fig. 3: ICMPv6 Header Format

2. Related Work

2.1 Network Covert Channels and Detection Techniques

Previous researchers in network covert channels focused on IPv4 [2], [4], [6], however fewer researchers concerned in the security vulnerabilities of the new generation IPv6 due to its incomplete implementation. Hidden information could be transferred very easy in the data section of the packet due to the large size and relatively unstructured in comparison to headers fields. Covert channels could be encoded in the unused or reserved bits in the packet header frame, these unused header fields are designed for future protocol improvements, and mostly they are dismissed by IDS and Firewalls [14],[17] furthermore this exception caused by the in-existence of specific values in protocol standards [8],[9].

Handel and Sandford in [1] proposed a covert channel exploiting the unused bits of the type of service (TOS) IP header or the Flags field in TCP header. Ahsan and Kundur in [2] suggested five hidden channels approaches manipulating the headers in TCP, IGMP and ICMP and one of them in packet sorting within the IPsec protocol. Hintz in [1], [9] proposed to use the Urgent Pointer in TCP to transmit covert data. Lucena et al suggested a number of covert channels in IPv6 header fields i.e. Traffic Class and Flow Label, Hop-by-Hop, Fragment, Authentication and Encapsulating Security Payload extension headers [1]. Time consumption and the complexity of detection process were noticed in her attempt. The same attempt was performed in [19] sending nonzero octets in data part of PadN option in Destination option extension header.

Rowland proposed to multiply each byte of the hidden data by 256 and use it directly as IP ID [2] meanwhile the IP identification header field is used for reassembling fragmented IP packets. The main requirement from RFC 0791 for the IP standard is that IP packet is uniquely identified by IP ID for a certain temporary time [8, 9]. Rutkowska proposed a developed covert channel using TCP ISNs for Linux using encryption [2, 8]. Furthermore Murdoch and Lewis [1, 2] proposed different idea about ISN covert channels techniques that they might produce different outcome than the real operating system implementation. Rutkowska proposed a developed covert channel using TCP ISNs for Linux deploying encryption [2], [9].
Qu et al [6] suggested a technique for covert information to be embedded into the Time to Live (TTL) and the Hop Limit field so as Lucena in [1]. Zander et al [2] analysed both proposed initial TTL values by Qu and Lucena, and suggested an encoded covert channels, which is harder to detect.

DOI : 04.IJIS.2015.1.3

Sohn et al in [13] mentioned the Support Vector Machine in passive warden to detect TCP covert channels within the IP ID and TCP ISN. This method is not preferable for well understood and explicit features in his proposed IP IDs and ISNs steganography covert channels, furthermore Support Vector Machine (SVM) can only identify simple aspects of tested data and seems unlikely to detect complex structure deployed in TCP/IP fields and their interdependencies [8].

Project Loki suggested exploring the concept of ICMP tunnelling in [8, 16] by using covert channels through the data portions of the ICMP_ECHO and ICMP_ECHOREPLY packets. The attacker wraps the commands and transmits them in the ICMP payloads, created a server Lokid once received the commands, unwrap them and execute them, then transferring the result back again wrapped in ICMP packets. Frikha and Trabelsi in [4] suggested a complex theory in triple processes within one security system, theoretically the approach was effective but it was not fully implemented.

2.2 Naïve Bayes Algorithm

NBA is a simple probabilistic classifier applying Bayes theorem but with a strong independence assumptions, which called class conditional independence because it assumes that an effect of an attributes value on a given class is independent. It allows the representation of dependencies among subsets of attributes; therefore, NBA is the fastest learning algorithm examining all its training inputs [12], [14, [18]. Let say C_k, C representing a class type with subset k as an attribute in which needs to be classified. Each class should have a probability denoted $P(C_k)$ that represents the prior probability of classifying an attribute into C_k, meanwhile the value that C_k has, will be estimated from the training dataset. Let say that an attribute such as n values, Xn, so the objective of classification is quite clearly to estimate and find the conditional probability of P (C_k| X_1, X_2, X_3...X_n) therefore the probability is calculated according to Bayes rule:

$$P(C_k|X) = \frac{P(X|C_k)P(C_k)}{P(X)} \qquad (1)$$

We can write this rule as below:

$P(C_k|X_n) = (X_1, X_2 \ldots Xn|C_k)\, P(C_k) = P(X_1, X_2 \ldots Xn)$

Key data:
P= Probability (of effective existence of the likelihood)
C= Class {Normal, Covert}
X= Data {attributes value}
k= tuple given the class {the subset values of each attributes}

3. Proposed Security Model

New attempts required to detect storage covert channel in IPv6 using different approaches, methodologies advanced MLA in respond to the novel vulnerabilities in this protocol. This approach could act as a countermeasure restrain against sophisticated attack tools used by hackers.

Using supervised Machine Learning to tackle such network threats in IPv6 will add a new rout of cutting-edge solutions for security systems. Most of the existing methods in [1]-[4] dealing with IPv6 covert channels have the following issues [8]:

- Approaches are complicated using complex algorithms to detect encrypted covert channels.
- Creating traffic congestion while processing.
- Time consumption in online detection
- Few parameters are considerable while dealing with covert channels of the specified style “paragraph” from the drop-down menu of style categories

Various approaches currently exist for anomaly detection: signature, behaviour and protocol based detection; few researchers used machine-learning technique to tackle covert channels in IPv6 and ICMPv6 due to the complexity of the protocol inherited oversight design. Our approach as shown in figure 4 uses pattern behaviour analysis of the header value to determine the identification that covert data has been transferred without affecting the normal communication and by passing the firewall too. In the first step of the proposed framework, we designed and configured a separate LAN as shown in figure 5 for IPv6 according to the network system environment. A Security script tool written in Python programming language was created as client and server running on the sender and the receiver hosts on the designed LAN, in addition to The Hacker Choice (THC) tool which is written in C programming language to simulate different attacks using ten fields to embed covert channels in both protocols IPv6 and ICMPv6 [3], [8], [16]. The framework consists of five modules as shown in Figure 4:

1) Capture Raw Data Module: Jpcap library packet sniffer is a Java API used to capture packets for 3 minutes.

2) Data Pre-processing Module: Input pcap data go through field selection and the following sub steps:

a) Packet Transformation: data needs to be decoded into numeric values in order to be compatible input for the next step analysis.
b) Packet Normalization: data need to be normalized in order to enhance the performance.
c) Packet Discretization: data needs to be discretized

DOI : 04.IJIS.2015.1.3

to create a consistency value type of the fields to facilitate feature selection.

3) Covert Channel Analyser:

a) Detection Algorithm: the input of this module

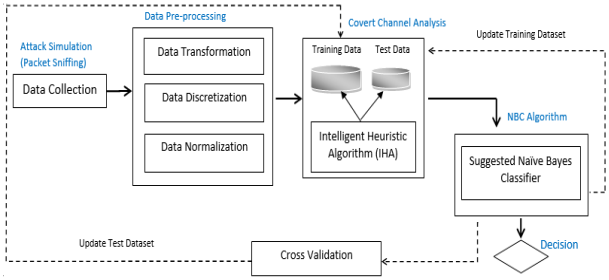

Fig. 4: Proposed Covert Channel Detection Framework in IPv6

is the selected fields with their values, here we run an Intelligent Heuristic Algorithm (IHA) to create training dataset and to detect the covert channels referring to the Request for Comments (RFC) and Internet Assigned Numbers Authority (IANA) rules. The output is the clear formatted data derived into two classes numerically 1 and 0, 1 is covert (anomaly) and 0 is normal. Each individual attribute consist of sub-set various values depending on its attributes holding type.

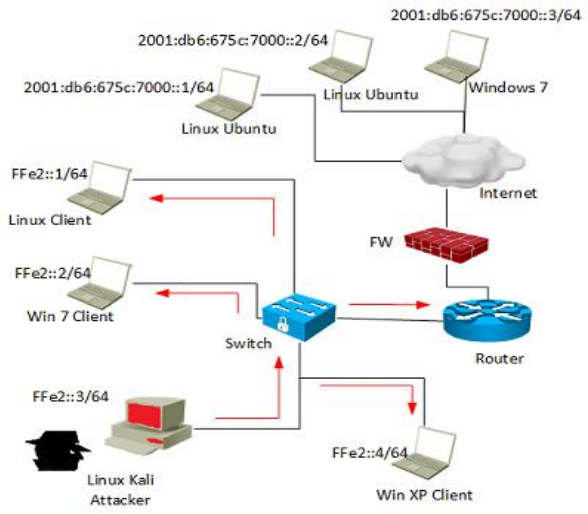

Fig. 5: Suggested Network Topology for Covert Channel attacks Simulation

4) Classification Module: formatted data pruning is vital in this stage in order to create a compatible data format to be enabled for classification process, however feature selection is performed within the classes to remove the unwanted and replicative data types. Hybrid feature selection techniques are used consisted of C4.5 with ratio gain to enable the training data as an input into the suggested NBC. Two classes will be achieved in the output data, covert in which classified as an attack, and a normal class.

Furthermore ranking of the covert channel attributes will be performed in this stage too, this is in order to weight the features gathered through data collection, and this is purely to classify the highest valued attributes in both labelled classes. More explanation will be given in sections 3.2 and 3.3.

5) Decision stage: the passive warden can take different actions such as: drop off the packet, block, and audit or limit the bandwidth of the connection as a part of mitigation security process.

3.1 Building the Classifier

The proposed Nave Bayes Classifier (NBC) is to improve the performance of the classification process by eliminating the irrelevant or the monotonous attributes from the captured dataset, then only tackling the most informative sub-values in the classification task. For the classification process, the denominator is irrelevant, since it will have the same value when for attribute values of the X_j as it is the same regardless of the value of C_k, The central assumption of Nave Bayesian classification is that every value in Xj within each class is independent from each other. Next, we get by applying the independent probability rule:

$$P(X_1 \mid \{\text{all left values of } X_j\}, C_k) = P(X_1 \mid C_k) \quad (2)$$

And therefore:

$$P(X_1, X_2 \ldots X_n \mid C_k) = P(X_1 \mid C_k)\, P(X_2 \mid C_k), P(X_n \mid C_k) \quad (3)$$

So each factor of the right hand of the equation possible to be determined from the training data because Then we can say from equation 2 we get:

$$P(Xi \mid Ck) \approx \frac{[(\#\, Xi \wedge Ck)]}{[(\#Ck)]} \quad (4)$$

Where # represents the number of such rates in the training set data. Therefore, we can classify the test dataset through calculating $P(C_k \mid X_1, X_2 \ldots X_n)$ as this is relevant to:

$$P(C_k)\, P(X_1 \mid Ck)\, P(X_2 \mid Ck)\, P(X_3 \mid C_k)\, P(X_n \mid C_k) \quad (5)$$

Let's apply this to our existing data, first we have categorized our training data characteristics into 10 main attributes (6 attributes as shown in Table I plus 4 additional attributes as shown in Table II). Let's assume Xi represents an attribute with its subset here subset i is the value held by each attribute X_1, X_2 … X_n, each group of attributes have been given a class C_k in which has a prior probability of classifying the attribute into Xi which, represent the value created by training data set.

DOI : 04.IJIS.2015.1.3

Bayes classifier will predict the class according to the higher probability (likelihood) which taken by an attribute to find the conditional probability of P (C_k| given X_1 and X_2 and X_3…and Xn).

3.2 Data Pre-Processing

1) Data Collection: The explicit unreachability of benchmark data on covert channels attacks calls for creating new models for IPv6 Intrusion detection systems. In our approach, we create primary data through simulation of different known and unknown attacks on the suggested IPv6 LAN topology (See Figure 6) suing a security tool to perform these attacks. Different attacks were simulated using covert data in IPv6 header fields. Table II shows the pre-processed output data format used into NB classification and Figure 6 shows the output dataset format with two classified classes.

Table 2: Covert Channels Data Format and Values

ID	Header Format	Value Type	Class
1	Traffic_Class	Numeric	Normal or Covert
2	Flow_Label	Numeric	Normal or Covert
3	Hop_Limit	High,Low, Moderate	Normal or Covert
4	Payload_Length	increased,decreased, Low	Normal or Covert
5	Source_Address	Numeric	Normal or Covert
6	Next_Header	Numeric	Normal or Covert
7	ICMPv6_Type	Numeric	Normal or Covert
8	ICMPv6_Code	Numeric	Normal or Covert
9	Reserve_Bit	Numeric	Normal or Covert
10	ICMPv6_ Payload	Numeric	Normal or Covert

We performed different simulations of various attacks, and then captured the raw data processed through field selection. We used two processes of selection: field selection prior to the data pre-processing phase, and feature selection post pre- processing phase.
The input here will be the captured pcap packets and should be filtered, transformed and discretized then pre-processed to create the needed training dataset; this is done by applying the Intelligent Heuristic Algorithm (IHA).The output is formatted according to the suggested classification technique, in our case; we need an Attribute Relation File Format (ARFF) containing three headers; attribute, value and class as shows in Table II.

```
00,00,High,increased,00,00,00,11,00,11,Covert
11,11,Moderate,Descreased,11,11,00,00,00,00,Normal
11,11,Low,Low,11,11,11,00,11,11,Covert
11,11,Moderate,Decreased,11,11,00,11,00,00,Normal
```

Fig 6: Output Dataset in ARFF Format before Classification

3.3 Feature Selection Algorithm

1) Decision Trees C4.5: Feature selection is the most critical step in building security system models it reduces data complexity and computational time and efforts. There are two methods to perform feature selection in [12], filter method and wrapper method, the filter method uses measures such as information, consistency or distance to compute the relevance of set of features while the wrapper predicts the accuracy of a classified as a mean to evaluate and assess the goodness of a feature set.
In our approach, we use a modified C4.5 technique. C4.5 is a popular method for inductive inference as it tolerate noisy data and has the capability to learn disjunctive expressions. It is a greedy algorithm and constructs the decision trees in a top-down recursive divide-and-conquer manner. Decision Trees considered as non-parametric estimator that reasonably approximate any function according to the increase size of the training or testing dataset, so using Nave Bayes Classifier would improve the performance in a better result.

2) Information Gain Algorithm: In order to select the best test attributes we need to work out the entropy measurement to calculate the purity in an arbitrary collection of examples. Let S be a set of consisting of s data samples. Suppose that the class label attributes has m distinct values defining m distinct classes C_k. Moreover, let Si be the number of samples of S in class C_k, so we need to classify the expected information as follow:

$$I(S_1, S_2, \ldots S_m) = -\sum_{k=1}^{m} P_k \log(P_k) \qquad (6)$$

Where P_k is the probability that an arbitrary sample belongs to class C_k and estimated by S_k = S. Let attribute A obtains x as distinct values, a1, a2 … ax. We can use attribute A to split S into x subsets S1, S2 …Sx, where Si contains the samples in S which have the value of aj of A. Then let S_{kj} be the sample numbers of class C_k in a subset Sj. So the entropy in which the expected information in the splitting subsets by A will give:

$$E(A) = \sum_{j=1}^{v} \frac{S_{1j} + \cdots + S_{mj}}{S} I(S_{1j} + \cdots .. + S_{mj}) \qquad (7)$$

In order to work out the weight we assume the term$(\frac{S_{1j+\cdots+S_{mj}}}{S})$ to be the j[th] subset and is the number of samples in the divided subset by total number of samples in S in equation (5). For a given subset S_j,

$$I(S_{1j}, S_{2j}, \ldots, S_{mj}) = -\sum_{k=1}^{m} P_{kj} \log_2(P_{kj}) \qquad (8)$$

Where $P_{kj} = S_{kj} / S_j$ and it is the probability in which any sample of S_j would belong to class C_k.

DOI : 04.IJIS.2015.1.3

This will make the entropy value zero if the sample is pure as all samples S should belong to one class, and the entropy has a maximum positive value such as 1. When the sample occasionally is impure and it could contain some negative and positive sub value examples also. Finally, the information gain expression would be achieved by:

$$InformationGain\ (A) = I(S1, S2, \ldots Sn) - E(A) \quad (9)$$

Finally, we work out the gain ratio and calculate as below:

$$GainRatio\ (A) = \frac{InformationGain(A)}{SplitInformation\ (A)} \quad (10)$$

4. Experiments and Results

The primary dataset in this experiment was obtained from a generated script simulating attacks on a separated IPv6 LAN network environment from the internet. Due to ethical issues concerning Data Protection Act 1998 realistic attacks are illegal. However, the IPv6 simulated topology as shown in Figure 5 configured successfully. After summing the proposed processes step 1-3. Packet were captured using Wireshark 1.12.1 The training dataset was created and streamed into the classification model using Weka 3.7 java built database system. We performed two segments of experiments using two types of training dataset: in phase one we used our primary dataset to elaborate the accuracy and the performance improvement of the suggested model implementing the suggested detection algorithm.

4.1 Detection Algorithm

In Figure 7 and 8 we see an example of five other implemented logic algorithms of an Intelligent Heuristic Algorithm (IHA) to obtain and extract the targeted data type from the following fields: flow label, traffic class, hop limit, payload length, next header, source address, icmpv6 type, icmpv6 code, reserve bit, icmpv6 payload. The process in this stage starts with checking the collected packets, targeted header fields values after being decoded, discretized and normalized, the verification is done through the validating against the original values of each field assigned by to references; IANA and the RFC's 2401, 2406, 2675, 4443 and more than 30 RFC's [22].

The algorithm will give an output of pure data format ready for classification along to raise a flag whether the packet contains covert data or to process normal if it is not affected.

The overall Algorithm is shown in figure 9 describes the model functionalities and tasks. The process starts with shuffling training dataset samples captured and

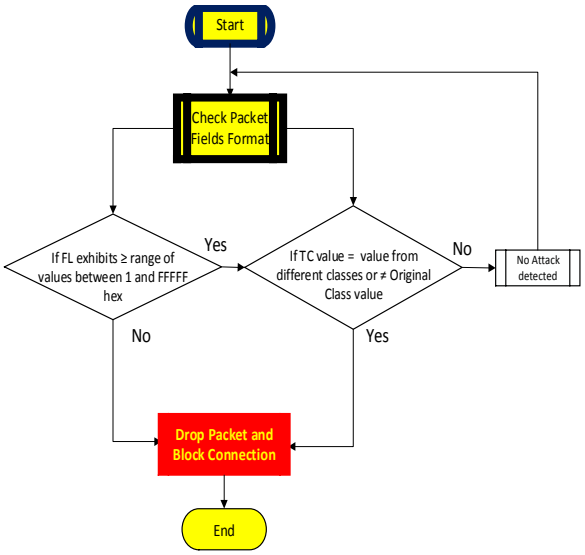

Fig. 7: Intelligent Heuristic Detection Algorithm for Flow Label and Traffic Class fields in IPv6

pre- processed, then we select 20% of the whole packets and run the IHA to create the compatible data format specify the covert channel characteristics.

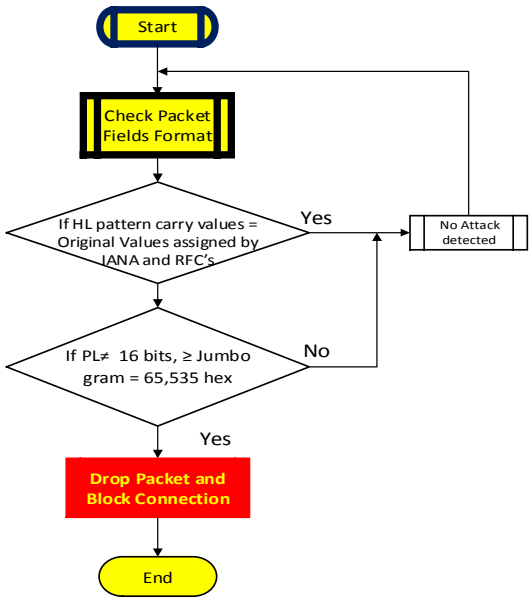

Fig. 8: Intelligent Detection Algorithm for Hop Limit and Payload Length fields in IPv6

However we used two different training datasets to examine the detection and performance accuracy of the proposed algorithm. Then we extract the attributes and each subset values up down to level 5 of each instance. We run the modified C4.5 along with Info Gain feature selection to obtain the decision tree and to measure the entropy of the pre pruned classes and the ranking of the most weighted detected attribute signalling an alarm.

DOI : 04.IJIS.2015.1.3

The IHA depends solemnly on the exact specifications given by those two protocol standards details IANA and RFCs which demand all network protocol fields "**should**" or "**must**" be used in their implementation over the network. This means the possibility of covert channels [24] as the firewalls don't depict these modified values. Finally running the Naïve Bayes classifier on the input data to classify statistically the covert channels and expose the accuracy percentage of the depicted and classified targeted data format.

1. Scramble the training dataset samples
2. Select and Create 20% from the shuffled packets.
3. Run IHA on data from step 2
4. Extract a set of preassigned attributes with its subset level 5 values instances.
5. Run C4.5 and Info Gain

$$SplitInformation(A) = -\sum_{j=1}^{x} \frac{S_j}{S} \log_2 \frac{S_j}{S} \quad (11)$$

6. Repeat 10 times (step1-5)
7. Collect all attributes from 10 rounds.
8. Run Naïve Bayes Classifier on the training data with test data using only the latest features collected in step 7.

Fig. 9: The Suggested Hybrid Feature Selection Algorithm

The cross validation is done through evaluation process of the experimental results given from the classification module. Running each data training and data testing set 10 folds with different feature selection and with a combination with the suggested NBC and feature selection, then calculate the True Positive Rate, False Positive Rate and the precision of the targeted attributes.

4.2 Primary Dataset Testing (Phase 1)

The primary data set used as a training dataset in the first attempt consisted of 700.000 instances which were collected from the captured packets, then we performed detection process through the covert channel's analyser as mentioned in step 2 by running the IHA in order to detect examine the input filtered data and to detect covert types anomaly) and normal data type as the final version prior to the classification process is shown in Figure 6.

The simulated attacks and commands as shown in Table 3 and should fall in one of the following four types of attacks [23]:

1) Probe.
2) Denial of Service (DoS).
3) Covert Channel: Echo_request, echo_reply, false traffic class, false flow label, increase or decrease in the hop-by-hop values, set false address, increase of Payload length value.
4) Root to Local (R2L).

Table 3: A sample of Covert Channel Test Cases Commands

Attack Test Case	Performed Commands
Payload fields covert channels	Scapy: send(IPv6(dst= "2001:db6:675c:7000::1") /IPv6DestOpt(options= [PadN(optdata=("22222222"))]+[PadN(optdata= ("3333333333333333"))]) / ICMPv6EchoRequest(id=1))
Covert channel Using PadN Option	IPv6DestOpt(type=02data="YYYYY")/ icmpechorequest

There are two types of Machine Learning; Supervised and unsupervised. When there is a training dataset and testing dataset processed to build a model and create a pattern it called supervised, if there is no training dataset it called unsupervised machine learning.

For the supervised learning algorithm, we used the learning dataset created by IHA carrying the characteristic described in Table 1 and 2. The training dataset as shown in Table 5 contained 10 attributes or features with one target value or labelled class either normal or covert (attack). Then we performed 10 fold cross validation to test the efficiency of the built model through the training phases. We performed all experiments in Windows 7 OS platform, CPU Core i5 processor, with 8 GB RAM. The experimental result of the created primary data in phase one testing shown in Table 4.

Table 4: Accuracy of NBC in comparison to other techniques

Classifier	Acc (%)	TPR	FPR	Precision	Time
Naïve Bayes	76.65	0.766	0.23	1.000	0.22
NB+InfoGain	65.89	0.833	0.38	0.827	0.18
NBC	**94.47**	**0.985**	**0.02**	**0.960**	0.15
NB+SubSetVal	55.32	0.81	0.32	0.274	0.25

4.3 Classification Metrics

The normal classification machine process normally intakes the unknown input data as Yu Liu mentioned in [20]. This is to identify the targeted class according to the assigned perimeter 'Covert' if the packet contains anomaly, and 'normal' if the packet belongs to the source type. The accuracy percentage is the actual measurement for a liable

DOI : 04.IJIS.2015.1.3

classification technique in which consists of the following metrics as shown in Table 4:

Table 5: Sample outcome of Covert Channel Characteristics from Analysis process

R #	TC	FL	HL	PL	NH	SA	Type	Code	RB	PYL	Class
1	0	0	High	Increased	0	0	0	1	0	1	Covert
2	1	1	Low	Unchanged	1	1	1	0	1	1	Covert
3	1	1	Moderate	Descreased	1	1	0	0	0	0	Normal
4	1	1	Moderate	Decreased	1	1	0	1	0	0	Normal
5	1	1	Low	Unchanged	1	1	1	0	1	1	Covert
6	1	1	Moderate	Decreased	1	1	0	0	0	0	Normal
7	0	1	Moderate	Unchanged	1	1	1	0	1	1	Covert
8	1	1	Low	Unchanged	1	1	0	0	1	1	Covert
9	0	1	Moderate	Unchanged	1	1	1	1	0	0	Covert
10	1	1	Low	Unchanged	1	1	1	0	1	1	Covert

true positive, false positive, true negative, false negative, precision. Below is a brief explanation of the mentioned metrics [21]:

1) True Positives: it is the correctly classified packets according to the assigned type.
2) False Positives: it is the incorrectly classified packets according to the assigned type.
3) Precision: it shows the correct instances that accurately holds the targeted type, in which picked up among the classified assigned type.

4.4 Secondary Dataset Testing (Phase 2)

In order to extend the proficiency of the pro- posed model, we used the DARPA 1999 IDS dataset [15]. This dataset was collected at the Massachusetts Institute of Technology (MIT) in Lincoln Lab to evaluate intrusion detection systems, however it lacks instances of IPv6 attack types except the ICMPv4, and IP ID covert channels [14] that has similar techniques principles manipulating such attacks. McHugh and Mahoney in [11], [15] claimed that DARPA dataset does not containing some background noise i.e. packet storms, strange packets, etc.

This dataset has binary class attribute as shown in Table 8 along with numerous realistic numbers of training and test instances that simplifies our experiment in this paper. Each connection record consists of 41 features and labelled in order sequences such as: 1,2,3,4,5,6,7... 41 and falls into four main categories below and their details are in Table 6.

1. Category 1 (1-9): Contains features of individual TCP connections.
2. Category 2 (10-22): Contains features within a connection suggested by domain knowledge.
3. Category 3 (23-31): Contains traffic features computed using two-second time windows.

Table 6 Attack Categories in NSL DARPA Dataset.

Category	Type
Normal	Normal
DoS	smurf, neptune, back, teardrop, pod, land, apacha2 Back, Mailbomb, Netpune, Pod, Processtable, Udpstorm, Buffer_overflow
Probe	satan, ipsweep, portsweep, nmap, saint
R2L,	warezmaster, ftp_write, multihop, phf, warezclient, guess_passwd spy, imap, worm, xlock, Xsnoop, Named, snmpguess, snmpgetattack, sendmail,
U2R	buffer_overflow,rootkit, loadmodule, perl, Httprunnel, PS, Sqlattack, Xterm

4. Category 4 (32-41): Contains traffic features computed using a two-second time window from destination to host.

Ciza in [15] described the features and values of NSL-KDD99 cup including a version of DARPA 1999 dataset attacks types. The suggested NBC in testing DARPA training dataset gave a slightly higher detection rate than our primary captured data in comparison to other techniques used in the process as shown in Table 7. The second phase with 10 folds resulting a lower false rate and a higher detection rate so far.

Table 7: Accuracy Detection of NBC Using Primary Data

Classifier	Acc (%)	TPR	FPR	Precision	Time
Naïve Bayes	80.04	0.802	0.198	0.907	0.27
NB+InfoGain	93.67	0.939	0.013	0.939	0.23
NBC	**96.46**	**0.945**	**0.012**	**0.989**	**0.20**
NB+SubSetVal	96.36	0.936	0.016	0.977	0.25

4.5 Discussion

The results of both experiments confirm the accuracy of initial hypothesis in which our NBCs performance is impressive with regards to the significant accuracy of each classifier detection rate in separate testing phases so far. In Table 4 and Figure 10, we observe the distinguished

DOI : 04.IJIS.2015.1.3

correctness and low false positive of the suggested classifier by **94.46%** after working out 10 folds using testing data and the cross validation True Positive detection Rate was **96.46%.** Let's discuss the experimental and evaluation results in more details.

Table 8: Attack Attribute Types in NSL DARPA Dataset

Nr	Feature Type	Nr	Feature Type
1	Duration	21	Is_host_login
2	Protocol_type	22	Is_guest_login
3	Service	23	Count
4	Flag	24	Srv_count
5	Src_bytes	25	Serror_rate
6	Dst_bytes	26	Srv_error_rate
7	Land	27	Rerror_rate
8	Wrong_fragment	28	Srv_error_rate
9	Urgent	29	Same_srv_rate
10	Hot	30	Diff_srv_rate
11	Num_failed_logins	31	Srv_diff_host_rate
12	Logged_in	32	Dst_host_count
13	Num_compromised	33	Dst_host_srv_count
14	Root_shell	34	Dst_host_same_srv_rate
15	Su_attempted	35	Dst_host_diff_srv_rate
16	Num_root	36	Dst_host_same_src_port_rate
17	Num_file_creation	37	Dst_host_srv_diff_host_rate
18	Num_shells	38	Dst_host_serror_rate
19	Num_access_files	39	Dst_host_srv_rerror_rate
20	Num_outbound_cmds	40	Dsthost_rerror_rate
		41	Class

Running the IHA to process the selected characteristics of covert channels in IPv6 and processing the classification module resulted better performance with amazing outcomes. We performed experiments on the original dataset which includes 10 attributes using Nave Bayes classifier once, and Subset Evaluation Technique, Naïve Bayes only, NBC with InfoGain algorithm in a second run with 10 folds each; the results were significantly obvious as shown in Table 4 and 7. The modified feature selection technique offered a higher prediction rate in detection process as shown in Figure 10 and 12, in addition to creating a better positive impact on the precision rate of the proposed method as shown in Figure 11 and 13.

The suggested decision tree C4.5 created a positive power along with Nave Bayes algorithm on the detection rate with a precision accuracy of **0.960%** as shown in Table 4. Here we observe an obvious improvement with **94.46%** in comparison to the first experimental results testing similar implemented techniques.

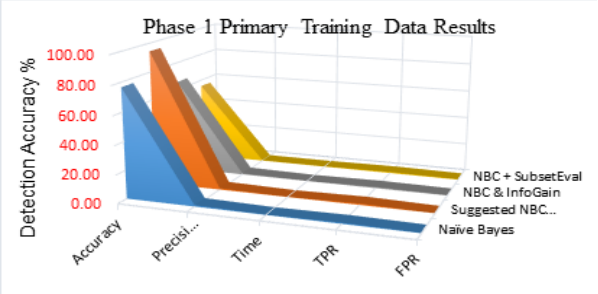

Fig. 10 Covert Channel Detection Accuracy

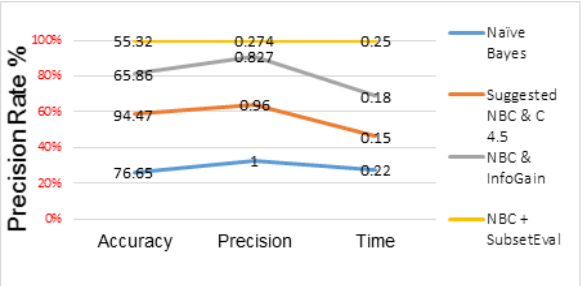

Fig. 11: Precision Rates of the Suggested Model

The NBC is fastest among other classifiers because fewer attributes are involved in learning process, however it seems to be liable for testing large size of data such as DARPA example, this is quite clear in Table 7 and Figure 12 and 13. The time that our proposed NBC spent in building the data model is **0.15** milliseconds in comparison to other elapsed time using NB alone was **0.22** milliseconds, NBC with InfoGain technique was **0.18** milliseconds, and finally the result of testing SubsetVal was **0.25** as shown in Table4.

The Amazing performance of our suggested Model has an improvements of **0.3** milliseconds time elapsed in building the model in compare to InfoGain, and **0.07** milliseconds better than NB, and finally **0.10** milliseconds in compare to SubSetVal classification technique as shown in Figure 11.

The True Positive Rate using NBC for the first detection process was **0.985%** in which is an obvious better performance of the suggested algorithm in comparison of other records of TPR; **0.152%** against InfoGain, **0.219%** against NB, **0.175%** against SubsetVal.
False Positive Rate also was tremendously better and low percentage with **0.02%** in compare to all other tested techniques rates; **0.36%** against InfoGain, **0.21%** against NB **0.30%** against SubsetVal.

DOI : 04.IJIS.2015.1.3

Table 9: Comparison of Covert Channel Attack Detection and Other Unknown Attacks

Dataset	NB	NBC without FS	NBC	C4.5	SubsetValue
Covert Channel	76.65	65.86	**94.47**	55.32	60.32
Other Attacks	80.04	**93.67**	**96.46**	72.56	**96.36**
Mean	78.34	79.76	**95.46**	63.94	78.34

The DARPA dataset has extraordinary huge amount of data so we had to cut (20%) of the whole dataset to create testing dataset with 41 attributes in order to evaluate the accuracy performance of the proposed method. NBC performance also potential in phase 2 despite of using more instances than the original dataset as shown in Figure 11. Finally Table 7 and Figure 12 show the accuracy of the detection rate in using NBC which is **(96.46%)** and potentially higher than using other techniques as well as the time elapsed in building the data model is **0.20** milliseconds. Occasionally the Mean value of the suggested NBC resulted a better among the other tested techniques too as shown in Table 9.

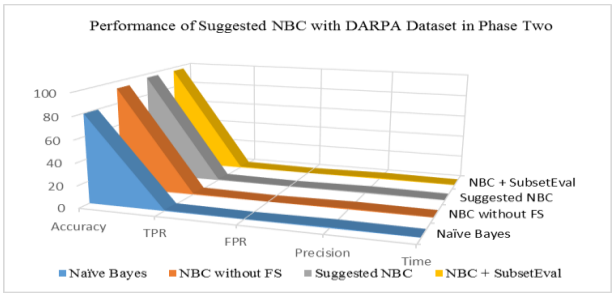

Fig. 12: Precision Rates of the Suggested NBC Testing DARPA Dataset

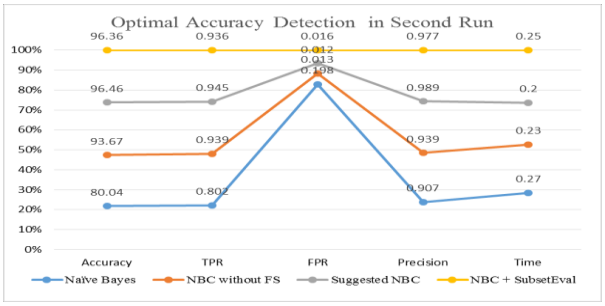

Fig. 13: The Optimal Performance of NBC in Testing DARPA Dataset

The cross validation result was impressive in testing DARPA dataset which consisted of 4.900.000 instances confirming the high accuracy of the suggested NBC. As we see in Table 7. The detection rate between the NBC and NB was **16.42%**, against NBC with InfoGain was **2.79%,** and against Subset Val techniques was **0.1%.**

The evaluation results of the True Positive Rate (TPR) also was distinguish and higher than the compared techniques with the following differences; NB has **0.802%** so the difference was **0.143%** against it, InfoGain has **0.939%** the difference was **0.006%** against it, Subset Val was **0.936%** so the difference was **0.009%** against it. In order to observe the error detection rate we examined the False Positive Rate (FPR) for NBC against the other techniques as shown in Table 7 and Figure 12 the following statistical analysis were given; NB has a higher rate of FTP with **0.186%** than the suggested model, this means that Naïve Bayes alone could not flag and detect the utmost real security threats as well as InfoGain error was **0.01%,** and Subset Val has **0.04%** false positive rate in detection process.

The above analysis of evaluation results and figures during the cross validation indicated a better and higher accuracy in NBC in comparison to the other feature selection techniques testing huge size of data, furthermore even testing the whole nearly five million instances or cutting 20% of the whole data size. We can consider that this approach is one of the effective methods to depict the unknown attacks in the future. We perceive the differences and the better performance in the rest of the statistical figures given in Precision rates and Data model creation time elapsed testing DARPA dataset as shown in Table 7, Figure 12 and 13.

The mean value of the overall testing techniques as shown in Table 9 give an indication of the high performance too for the suggested NBC with the highest value of **95.46%** against all other figures in the same table.

To see the best attributes given in the ranking calculation as shown in Table 10, we observe the highest attribute classified in which the ICMPv6_PYL is obtaining level 1 as it has a sequence number 8 in the covert channel characteristics testing list. However Hop_Limit attribute was the most splitting features among the other tested features with three subset values as shown in Figure 14 and 15. And finally Figure 16 shows as the weighting values of each attributes holding the best relevant attributes and in corresponds to Table 10.

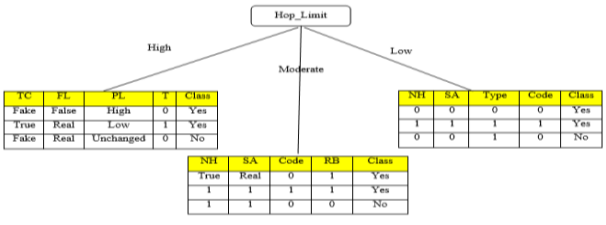

Fig. 14: Hop Limit has the highest Information and the Splitting Attribute at the root of the Decision Tree

DOI : 04.IJIS.2015.1.3

Table 10: Classification of Sequence Fields Used Covert Channel

Ranking	Att Sequence	Attribute Name
1	8	ICMPv6_PYL
0.55	3	HOP_Limit
0.55	4	Payload_L
0.335	5	Next_Header
0.335	2	Flow_Label
0.335	1	Traffic_Class
0.335	7	ICMPv6_Code
0.258	6	ICMPv6_Type

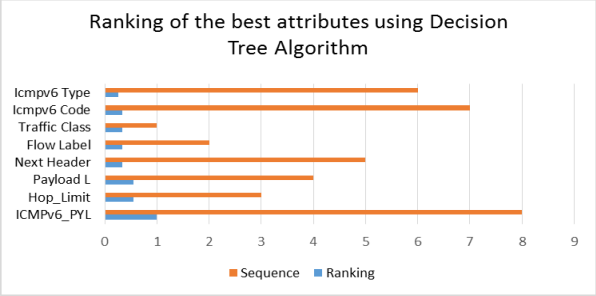

Fig. 15: The Best Attributes Ranking Percentage Using C4.5

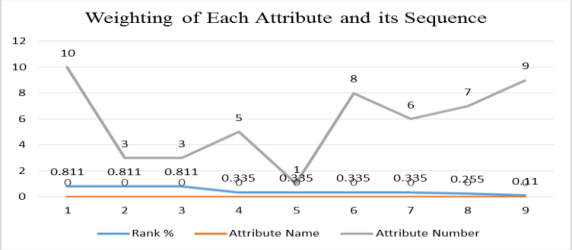

Fig. 16: A Parallel weighting of Each Classified Attributed

5 Conclusion

A new hybrid method in feature selection that uses C4.5 decision trees with Information gain technique is presented. This method is used to classify and detect covert channels in IPv6. The suggested Model aimed to use advanced Data Mining Techniques in such complex designed network protocol vulnerabilities, however the suggested NBC was improved by Nave Bayes learning algorithm.

This proposed approach implementing an enhanced feature selection technique i.e. C4.5 decision trees with Information Gain in a heterogeneous form, reduces the probabilistic stimulation, which leads to higher accuracy in detection and classification process, consequently led to lower false negative rate (FNR) and higher true positive rate (TPR).

The reason behind this result is that we reduced the entropy and the noisy data in both training datasets: Our Original primary data and the secondary data DARPA 1999 dataset led to pure data pruning and significant compatible data as shown in Figure 6. Our approach examined and depicted the hidden covert channel features selected in the primary dataset of the IPv6 and its attacks in captured packets

We have answered the research question about the possibilities of existence of such complex hidden storage channels in IPv6 and explained the security threats impact created by attackers on network security policy and privacy, furthermore we also justified our findings by practical evidence and evaluation of the suggested new model to mitigate such vulnerabilities in IPv6.

Future work is further planned to examine the TCP and UDP covert channels weighting and classifying their attributes implementing different approach and techniques in order to obtain more accurate and efficient results. In addition to the ranking trees process, we aim to use more different advanced feature selection algorithms to elaborate the vulnerabilities in the oversight design of the TCP/IP Suite Protocol.

References

[1] Lucena N, Grzegorz Lewandowski, Steve J. Chapin, "Covert Channels in IPv6," Privacy Enhancing Technologies, Springer Berlin / Heidel- berg., vol.3856, pp. 147–166, May. 2005.

[2] S. Zander, G. Armitage, and P. Branch, "Covert Channels in the IP Time To Live Field," in Proc. Australian Telecommunication Networks and Applications Conf, (ATNAC), December, 2006.

[3] Martin, Cynthia E. Dunn, Jeffrey H, "Internet Protocol Version 6 (IPv6) Protocol Security Assessment," in IEEE. Military Communications Conference, MILCOM 2007, IEEE USA, IEEE, 29-31 2007, pp. 1–7.

[4] Lilia Frikha, Zouheir Trabelsi, Sami Tabbane, "Simulation, Optimization and Integration of Covert Channels, Intrusion Detection and Packet Filtering Systems," in IEEE Global Information Infrastructure Symposium (GIIS 2009), IEEE,23-25 June 2009, Hammamet Tunisia.

[5] B. Lampson, "A Note on the Confinement Problem," Communication of the ACM, vol.16, no. 10 pp. 613–615, 1973.

[6] Cabuk, S., Brodley, C.E., Shields, C, "Ip covert timing channels: Design and detection," in Proceedings of the 11th ACM Conference on Computer and Communications Security, ACM Press, Washington DC, USA. 2004, pp. 178–187.

DOI : 04.IJIS.2015.1.3

[7] Zagar, D. Grgic, K, 'IPv6 Security Threats and Possible Solutions',' in Automation Congress, IEEE, WAC '6, World, 24-26 July 2006, USA, IEEE pp. 1–7.
[8] Choudhary, Abdur.Rahman, "In-depth Analysis of IPv6 Security Posture," in DOI. Collaborative Computing: Networking, Applications and Work sharing, Collaborate Com 2009. 5th International Conference, Digital Object Identifier, 11–14 Nov 2009, pp. 1–7.
[9] Supriyanto , Raja Kumar Murugesan , Sureswaran Ramadass, "Ipv6 Security Vulnerability Issues and Mitigation Methods," International Journal of Network Security and its Applications (IJNSA), vol. 4, no. 6, pp. 173–185, Nov. 2012.Malaysia.
[10] Carp, Alexandru; Soare, Andreea; Rughinis, Razvan, "Practical analysis of IPv6 security auditing methods," in IEEE. Roedunet International Conference (RoEduNet), 2010 9th, IEEE USA., 24–26 June 2010, USA pp. 36–41.
[11] M Mahoney and P Chan, "An analysis of the 1999 DARPA/Lincoln Lab- oratory evaluation data for network anomaly detection," in Proceeding of Recent Advances in Intrusion Detection (RAID), vol 2820 Pittsburgh, PA, USA.,2003,8–10 September 2003. pp. 220–237.
[12] N.B. Amor, S. Benferhat, and Z. Elouedi, "Nave Bayes vs. decision trees in intrusion detection systems," in Proc. of 2004 ACM Symposium on Applied Computing, USA. 2004,pp. 420–424.
[13] Sohn, T., Seo, J., Moon, J, "A study on the covert channel detection of TCP/IP header using support vector machine," Information and Communications Security, In Perner, P., Qing, S., Gollmann, D., Zhou, J., eds, vol. 2836, Lecture Notes in Computer Science, Springer-Verlag, pp. 313–324,2003.
[14] Yogita B. Bhavsar, Kalyani C. Wghmare," Intrusion Detection System Using Data Mining Technique: Support Vector Machine," International journal of Emerging Technology and Advanced Engineering, vol. 3, no. 3, March. 2013.
[15] Ciza Thomas, Vishwas Sharma and N. Balakrishnan, "Usefulness of DARPA Dataset for Intrusion Detection System Evaluation," in International Symposium on Defence and Security, Proceedings of SPIE,vol 6973,no 15, 2008.
[16] Marc Hauser,"IPv6 Security Vulnerabilities", 2014, https://www.thc.org/thc-ipv6/.
[17] T K Vivek, M Kalimuthu, "Improving Intrusion Detection Method for Covert Channel in TCP/IP," International Journal of Computer Science Trends and Technology (IJCST), vol. 2, no. 2, March. 2014.
[18] Kavitha, P., and M. Usha, "Anomaly Based Intrusion Detection In WLAN Using Discrimination Algorithm Combined with Naive Bayesian Classifier," Journal of Theoretical and Applied Information Technology, (JATIT and LLS), vol. 62, no. 3, pp. 646–653, 2014.
[19] M. Mavani and Leena Ragha, "Covert Channel in IPv6 Destination option Extension Header", International Conference Circuits, Systems, Communication and Information Technology Application (CSCITA), IEEE, Mumbai, 4-5 April 2014, pp 219-224, India.
[20] Yu Liu. A Survey of Machine Learning Based Packet Classification; Symposium on Computational Intelligence for Security and Defence Applications (CISDA), 2009; IEEE.
[21] Thuy T.T. Nguyen and Grenville Armitage. A survey of Techniques for Internet traffic classification using Machine Learning; Communications Surveys & Tutorials; IEEE; 2008;vol. 10; p. 56-76.
[22] Hogg, S. and E. Vyncke, IPv6 Security 2009: Cisco Press.
[23] Abdulrahman Salih., Xiaoxi Ma, and Evtim Peytchev, "Detection and Classification of Covert Channels in IPv6 Using Enhanced Machine Learning" Proc of the International Conference on Computer Technology and Information Systems. ICCTIS DUBAI, UAE, 2015, N&N Global Technology 2015.
[24] Murphy, R. P. "IPv6 / ICMPv6 Covert Channels." Las Vegas: Defcon14, Aug. 2006. PDF

Abdulrahman Salih is a PhD candidate at Nottingham Trent University. He received his MSc with Distinction in IT Security from University of Westminster, London in 2010, and his BSc (Horns) Software Engineering from Nottingham Trent University in 2007. He worked as a Network Security Engineer for Planet Solutions in London before re-joining NTU. He is the founder and CEO of KNCIS in London and Sweden, specializing in Cyber Security Analysis.

DR. XIAOQI MA is a Senior lecturer and a leader of many modules; Security Technologies, Computer Security and Advanced Security Technologies in the School of Science and Technology at Nottingham Trent University. He is a member of the Intelligent Simulation, Modelling and Networking Research Group (ISMN). He obtained PhD from Reading University in 2007 in Cryptographic Network Protocols. He contributed in more than 20 publications in International Journals, conferences and book chapters.

DR. EVTIM PEYTCHEV is a Reader in Wireless, Mobile and Pervasive Computing in the school of Science and Technology at Nottingham Trent University, UK. He is leading the Intelligent Simulation, Modelling and Networking Research Group, which consists of 5 lecturers, 3 Research Fellows and 6 research students. He is the Module Leader for Systems Software; and Wireless and Mobile Communications. He also teaches on the modules Software Design and Implementation; Mobile Net- working; Enterprise Computing; and Computer Architecture.

DOI : 04.IJIS.2015.1.3