Detection and Classification of Covert Channels in IPv6 Using Enhanced Machine Learning

Abdulrahman Salih
School of Science And Technology
Nottingham Trent University
Nottingham, United Kingdom
Email:fb104480@ntu.ac.uk

Xiaoqi Ma
School of Science And Technology
Nottingham Trent University
Nottingham,United Kingdom
Email:xiaoqi.ma@ntu.ac.uk

Evtim Peytchev
School of Science And Technology
Nottingham Trent University
Nottingham,United Kingdom
Email: evtim.peytchev@ntu.ac.uk

***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 Machine Learning approach to detect covert channel implementing an enhanced feature selection algorithm supporting 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 create novel primary training data, in addition to a modified Decision Tree C4.5 technique to detect and classify hidden channels in IPv6 network. The results showed better detection performance and high accuracy in True Positive Rate (TPR) and a low false negative rate (FNR) in comparison to other previous techniques.**

***Keywords*—CyberSecurity,Covert Channel, ICMPv6, IPv6,Naive Bayes, Dicision Trees C4.5, MLA ,DARPA, NSL-KDD.**

I. 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 6.69% 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 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) in [5] in was the first that recognized them as storage channels between two systems, however these channels were not meant to be used for communication. Then commonly researchers defined them as enforced, illicit signalling channels that allow a user to stealthily, contravene targeted objective in [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. 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],[4].

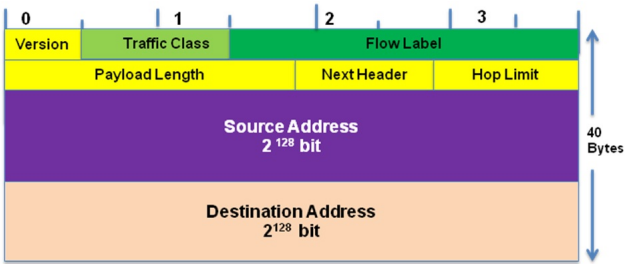

Fig. 1: IPv6 Header Format

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.

```
0110 .... = Version: 6
   [0110 .... = This field makes the filter "ip.version == 6" possible: 6]
.... 0000 0000 .... .... .... .... .... = Traffic class: 0x00000000
   .... 0000 00.. .... .... .... .... .... = Differentiated Services Field: Default (0x00000000)
   .... .... ..0. .... .... .... .... .... = ECN-Capable Transport (ECT): Not set
   .... .... ...0 .... .... .... .... .... = ECN-CE: Not set
.... .... .... 0000 0000 0000 0000 0000 = Flowlabel: 0x00000000
Payload length: 60
Next header: ICMPv6 (0x3a)
Hop limit: 64
Source: 2001:db8:0:1::1 (2001:db8:0:1::1)
Destination: 2001:db8:0:1::2 (2001:db8:0:1::2)
```

Fig. 2: IPv6 PCAP Data in Header Fields

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 in [1]. Table I shows examples of possible IPv6 covert channels characteristics in ICMPv6. 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; 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 between (1-127) are for error messages and from (128-255) are for information messages [2]. Having said that and the arbitrary content of the ICMPv6 payload may carry

DOI: 07. ICCTIS.2015.1.1

different types of data according to the messages types and ranges mentioned earlier, besides the Operating Systems type used too [12].
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].

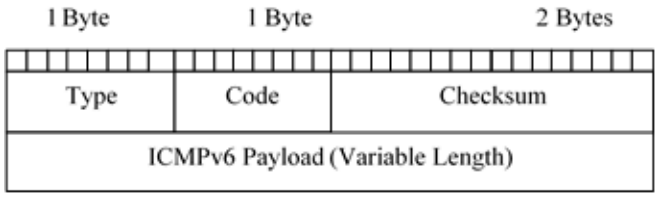

Fig. 3: ICMPv6 Header Format

TABLE I: IDENTIFIED COVERT CHANNELS IN IPV6 HEADER FIELDS

ID	Field)	Covert Channel	Bandwidth
1	Traffic Class	Set a false traffic class	8 bits/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	$\approx$ 1 bit/packet
6	Source Address	Set a false source address	16 bits/packet

The rest of the paper is organized as follow: section II discusses the related work in two subsections; covert channels detection in IPv6 including ICMPv6, and Nave Bayes Algorithm. Section III discusses the proposed framework, section IV discusses the experiments and initial results obtained from the testing phases, and finally section V discusses Conclusion and Future work. Our proposed system offers a better performance in high accuracy and prediction of the future unknown attacks against legitimate targets.

II. RELATED WORK

A. 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 in [1] suggested a number of covert channels in IPv6 header fields hence time consumption and the complexity of the process was noticed in attempted approach.

Rowland proposed in [2] a method to multiply each byte of the hidden data by 256 and use it directly as IP ID 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 [9],[10]. Rutkowska proposed a developed covert channel using TCP ISNs for Linux using encryption [2],[9]. Furthermore, Murdoch and Lewis [1],[2] proposed different idea about ISN covert channels.

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 in [2] analysed both proposed initial TTL values, which is out of our research scope.

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 SVM can only identify simple aspects as its unlikely to detect complex structure deployed in TCP/IP fields and their interdependencies [8].

Project Loki suggested exploring the concept of ICMP tunneling in [9],[17] by using covert channels through the data portions of the ICMP_ECHO and ICMP_ECHOREPLY 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.

B. Naive 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, X_n, 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:

$$P(C_k \mid X) = \frac{P(X \mid C_k)\, P(X_k)}{P(X)} \tag{1}$$

We can write this rule as below:

$$P(C_k|X_n) = (X_1, X_2, X_n|C_k)P(C_i)/P(X_1, X_2, X_n)$$

III. PROPOSED FRAMEWORK

New attempts required to detect storage covert channel in IPv6 using advanced MLA in respond to the novel vulnerabilities in this protocol. This approach could act as a countermeasure restrain to 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.

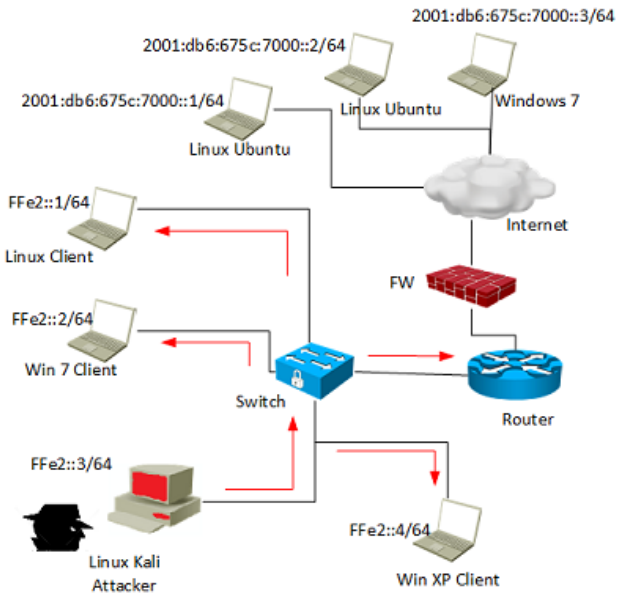

Fig. 4: Simulation of Covert Channels in LAN Topology

Various approaches existed 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. Our approach uses pattern behaviour of the header value to determine the identification that covert data has been transferred without affecting the normal communication. In the first step of the proposed framework, we designed and configured a separate LAN as shown in Figure 4 for IPv6 according to the network system environment. A Security tool was created along with The Hacker Choice (THC) in [16] to simulate different attacks using ten fields to embed covert channels in both protocols IPv6 and ICMPv6 [3],[8]. The framework consists of five modules as shown in Figure 5:

1) Capture Raw Data: Jpcap library packet sniffer is a Java API used to capture packets for 3 minutes.
2) Covert channel Analysis: Input pcap data go through field selection and the following sub steps:
   a) Packet Transformation:data needs to be transformed into numeric values in order to be compatible input for NBC.
   b) Packet Normalization: data need to be normalized in order to enhance the performance of the detection and create a consistency for the values.
   c) Packet Discretization: data needs to be discretized to create a consistency value type of the fields to facilitate feature selection.

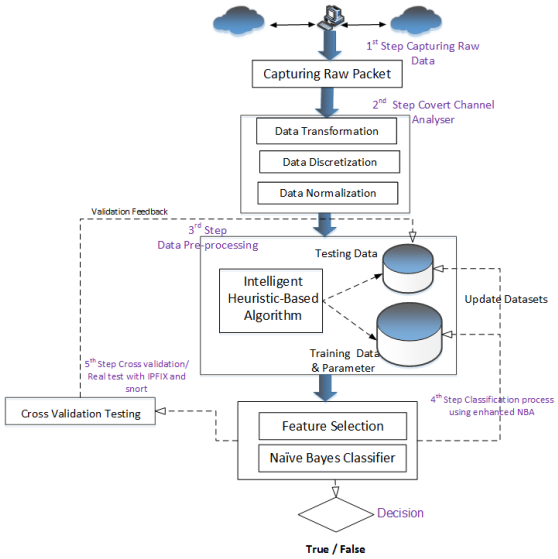

Fig. 5: Proposed Covert Channel Detection Framework in IPv6

3) Data Pre-processing: the input of this module is the selected fields with their values. Here we run an Intelligent Heuristic Algorithm (IHA) to create training dataset referring to the Request for Comments (RFC) and Internet Assigned Numbers Authority (IANA) rules. The output is the clear formatted data into two classes numerically 1 and 0, 1 is covert and 0 is normal. The classes consist of attributes with sub-set various values depending on each attributes holding type.
4) Classification and Detection step: feature selection is needed to prune classes and remove the unwanted repetitive data types, a modified C4.5 is used with ratio gain to enable the training data as an input into the NBC. Two classes will be achieved in the output data, covert in which classified as an attack, and a normal class.
5) Decision step: the passive warden can take different actions such as: block, audit or limit the bandwidth of the connection as a part of mitigation security process.

A. *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 X_j within each class is independent from each other. Next, we get by applying the independent probability rule:

$$P(X_1 \textbf{ all left values of } X_j = P(X_1 \mid C_k) \qquad (2)$$

And therefore:

$$P(X_1, X_2, X_n|Ck) = P(X_1|C_k)P(X_2|C_k), .P(X_n|C_k) \quad (3)$$

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

$$P(X_i|C_k) \approx [(\#X_i \wedge C_k)/[(\#C_k)] \quad (4)$$

Where # represents the number of such occurrences in the training set data. Therefore, the classification of the test set can now estimate by: $P(C_k|X_1)X_2.X_n)$ so this will be proportional to:

$$P(C_k)P(X_1|C_k)P(X_2|C_k)P(X_3|C_k)P(X_n|C_k) \quad (5)$$

Let's apply this to our existing data, first we have categorized our training data characteristics into 6 main attributes as shown in Table I plus 4 additional attributes as shown in Table II. Let's assume X_i 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. Bayes classifier will predict the class according to the higher probability (likelihood) which taken by an attribute to find the conditional probability of $P(C_k|givenX_1andX_2andX_3...andX_n)$.

B. 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 suing a security tool to perform these attacks. Different attacks were simulated using covert data in the IPv6 header. Table II shows the pre-processed output data format used into NB classification and Figure 6 shows the data set format with classified class.

TABLE II: COVERT CHANNELS DATA FORMAT AND VALUES

ID	Header Attribte	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	Increase, decrease, 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: Data Set Format Before Classification

C. Feature Selection Algorithm

1) Data Trees C.45: 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 [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 this 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:

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

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 $S_1, S_2, .S_x$, where S_i contains the samples in S which have the value of a_j of A. Then let S_kj be the sample numbers of class C_k in a subset S_j. So the entropy in which the expected information in the splitting subsets by A will give:

$$E(A) = \sum_{j=1}^{v} \frac{(S_1j + ... + S_mj)}{S} \quad (7)$$

In order to work out the weight, we assume the term $\frac{(S_1j+...+S_mj)}{S}$ to be the (j^th) subset and it the number of samples in the devided subset by total number of samples in S as in equation (5). For a given subset S_j,

$$I(S_1j, S_2j, ..., S_mj) = -\sum_{k=1}^{m} P_kj \log_2(P_kj) \quad (8)$$

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. 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:

$$InformationGain(A) = I(S_1, S_2, ..., S_M)E(A) \quad (9)$$

So here, we choose the attribute with the highest information gain to test the current node. In order to avoid focusing only on attributes with many values rather than attributes with few values we need to modify the C4.5 techniques with another alternative measurement called Information Gain Ratio (IGR) which, maximizes the probabilities of considering each value of any attribute no matter how many values can have. This split of the information takes into account that an attribute having many values like:

$$SplitInformation(A) = -\sum_{j=1}^{x} \frac{S_j}{S} log_2 \frac{S_j}{S} \quad (10)$$

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

$$GainRatio(A) = \frac{InformationGain(A)}{SplitInformation(A)} \quad (11)$$

IV. Experiment and Result Discussion

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 4 configured successfully. After sunning the proposed processes step 1-3 The training dataset was created and streamed into the classification model using Weka 3.7 java built database system. We performed two phases 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:

- *Phase 1*: The primary data set used as a training dataset in the first attempt consisted of 10.000 instances and we performed classification types: covert (anomaly) and normal as shown in Figure 6. The simulated attacks should fall in one of the following four types of attacks:
  1) Probe
  2) Denail of Service (DoS)
  3) Covert Channel: ICMPv6 covert channel, Flow label, hop-by-hop, and Traffic class covert channels.
  4) Root to Local (R2L).

TABLE III: ACCURACY OF DIFFERENT DETECTION TECHNIQUES

Classifier	Accuracy(%)	TPR	FPR	Precision	Model Built
Nave+Bayes	76.650	0.766	0.234	1.000	0.22
NBC+InfoGain	65.86	0.833	0.380	0.827	0.18
Suggested NBC	**94.47**	0.985	0.015	0.960	0.15
NBC+SubsetEval	55.32	0.810	0.319	0.274	0.25

For the supervised learning algorithm, we used the learning dataset created by IHA with the characteristic described in Table I and table II. The training dataset contained 11 attributes or features including the one target value or labelled class

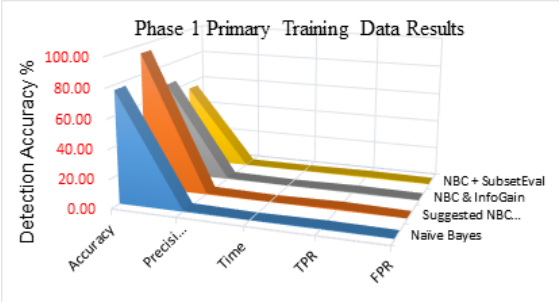

Fig. 7: Performance of The Proposed Model

either normal or covert (attack) in order to build the detection system. Then we performed 10 fold cross validation to test the efficiency of the built model through the training phase. We performed all experiments in Windows 7 OS platform, CPU Core i5 processor, with 8 GB RAM. The initial result of the phase one testing shown in Table III.

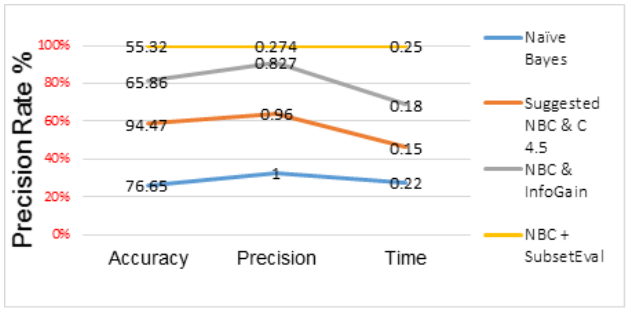

Fig. 8: Precision Rates of The Suggested Model

- *Phase 2*: In order to extend the proficiency of the proposed 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] criticized the DARPA dataset for not containing some background noise i.e. packet storms, strange packets, etc.

This dataset has binary class attribute 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:

1) *Cat 1 (1-9)*: Contains features of individual TCP connections.
2) *Cat 2 (10-22)*: Contains features within a connection suggested by domain knowledge.
3) *Cat 3 (23-31)*: Contains traffic features computed using two-second time windows.
4) *Cat 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 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 IV. The second phase with 10 folds resulting a lower false rate and a higher detection rate so far.

TABLE IV: PERFORMNCE OF NBC IN COMPARISON TO OTHER TECHNIQUES

Classifier	Accuracy(%)	TPR	FPR	Precision	Model Built
Nave+Bayes	80.04	0.802	0.198	0.907	0.27
NBC without FS	**93.67**	0.939	0.013	0.939	0.23
Suggested NBC	**96.46**	0.945	0.012	0.989	0.20
NBC+SubsetEval	**96.55**	0.948	0.014	0.987	0.25

A. *Discussion*

The results of both experiments confirm the initial hypothesis in which our NBCs performance is impressive with regards to the significant accuracy of each classifier in separate testing phases so far. In Table III and Figure 7, we observe the distinguished correctness and low false positive of the suggested classifier. The suggested decision tree C4.5 created a positive impact along with Nave Bayes algorithm on the detection rate as shown in Table III.

The NBC is fastest among other classifiers because fewer attributes are involved in learning, furthermore and the time that our proposed NBC spent in building the data model is 0.15 milliseconds, is obviously less than other used techniques.

To see a better performance of NBC, we performed another experiment on the original dataset with 10 attributes using Nave Bayes classifier once, and with Subset Evaluation Technique in a second run; the results were significantly obvious as shown in Table III. The modified feature selection technique offered a higher prediction rate in detection process as shown in Figure 7, in addition to creating an impact on the precision rate of the proposed method as shown in Figure 8.

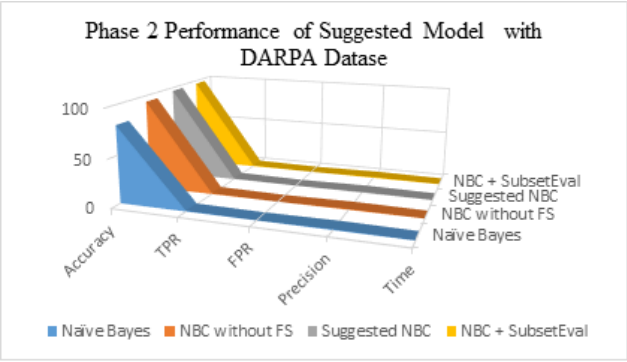

Fig. 9: Performance of the proposed Model Using DARPA

The DARPA dataset has extraordinary huge amount of data so we had to cut 10(%) of the whole dataset to create testing dataset with 41 attributes in order to see 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 9.
Finally Table IV and Figure 10 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.

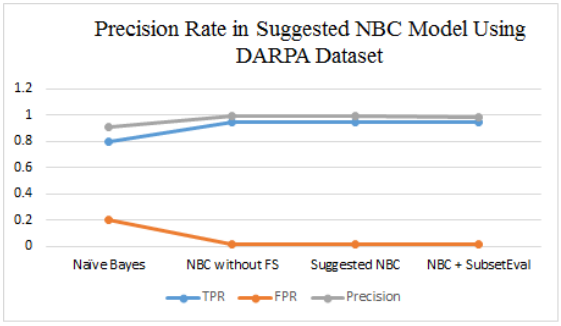

Fig. 10: Precision Rates of the Suggested Model Using DARPA

V. CONCLUSION AND FUTURE WORK

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 improving the Nave Bayes learning algorithm.

This proposed approach implementing an enhanced feature selection technique.i.e C4.5 decision trees with Information Gain heterogeneously reduces the probabilistic stimulation, which leads to higher accuracy in detection and classification process, consequently leads 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 DARPA 1999 dataset which led to pure data pruning and significant compatible data as shown in Figure 6.

Future work is further planned to examine the weighting and ranking of the features selected in the primary dataset of the IPv6 and its attacks captured packets. In addition to the ranking trees process, using more different advanced feature selection algorithms is planned to be investigated too.

REFERENCES

[1] Lucena N, Grzegorz Lewandowski, Steve J. Chapin, "Covert Channels in IPv6," *Privacy Enhancing Technologies,Springer Berlin / Heidelberg.*, 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.

[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 Laboratory 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 Defense and Security*, Proceedings of SPIE,vol 6973,no 15, 2008.

[16] ,Marc Hauser,"IPv6 Security Vulnerabilities" ,2014,"https://www.thc.org/thc-ipv6/."

[17] [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.

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 rejoining NTU. He is the founder and CEO of KNCIS in 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 Networking; Enterprise Computing; and Computer Architecture.