Packet Length Covert Channel: A Detection Scheme

Muawia A. Elsadig
College of Computer Science and Technology, SUST.
Khartoum, Sudan.
Muawiasadig66@gmail.com

Yahia A. Fadlalla
Lead Consultant/Researcher, InfoSec Consulting,
Hamilton, Ontario,
Canada.

***Abstract*— A covert channel is a communication channel that is subjugated for illegal flow of information in a way that violates system security policies. It is a dangerous, invisible, undetectable, and developed security attack. Recently, Packet length covert channel has motivated many researchers as it is a one of the most undetectable network covert channels. Packet length covert channel generates a covert traffic that is very similar to normal terrific which complicates the detection of such type of covert channels. This motivates us to introduce a machine learning based detection scheme. Recently, a machine learning approach has proved its capability in many different fields especially in security field as it usually brings up a reliable and realistic results. Based in our developed content and frequency-based features, the developed detection scheme has been fully trained and tested. Our detection scheme has gained an excellent degree of detection accuracy which reaches 98% (zero false negative rate and 0.02 false positive rate).**

***Keywords*— covert channel, security; detection, prevention, elimination, packet length covert channel, packet size covert channel, network protocols, machine learning.**

INTRODUCTION

A covert channel is a communication channel that is somehow established to pass a secret information covertly. This illegal flow of information was initially defined by Lampson in 1973 [1]. At which two processes with different security policy convey a secret message in a way that violates system security policy such as a high security level process leak secret information to another process with low security level which is prohibited by system policy. In 1987. Grilling extended this concept to suit computer network environment [2, 3]. And then, the advanced development in computer network technology has enriched developing different covert channel techniques that pose many security challenges [4-7].

In covert channels literature, three types of covert channel are defined. Storage, timing, and hybrid covert channel. In storage covert channels, a storage location is exploited as a channel to convey secret information such as a sender write a secret message in a shared location and the intended receiver picks that secret message by reading the shared location. In timing covert channels, a sender exploits a system rescore aspect to signal a secret message and then, the receiver observes and decodes the secret message. The third type of covert channels is a combination of storage and trimming covert channels. It is a hybrid covert channel which poses real challenge as it benefits from the advantages of storage and timing channels. Storage channels provide high bandwidth and timing channels are hard to detect.

A packet length-based covert channel is a network-based covert channel that exploits the network packets' lengths to convey secret information. There are many techniques to establish packet length covert channels, however in this paper we consider one type of them, as to this end no detection method is presented to deal with. So, our work is considered the first try to present a detection method to deal with the aforementioned type of packet length covert channel.

In this paper, a machine learning detection scheme is presented and evaluated to achieve a high classification accuracy degree. As it commonly known that machine learning approach has a lot of contributions in different scientific areas [8-11] especially in information security filed [12-15]. This section, give a brief introduction to covert channel definitions and types while the rest of paper is organized as follows: the next section sheds lights on covert channel development. A thorough investigation of packet length-based covert channel related work and literature is presented in Section III. This section also gives details on packet length covert channels development and their countermeasures. Our proposed scheme is demonstrated in section IV and subsequently Section V presents our scheme implementation and results. Then the paper is concluded in section VI, while the future work is given in section VII.

COVERT CHANNEL DEVELOPMENT

A covert channel is a communication channel that is exploited to convey secret information in a way that breaches system security policy [16, 17]. A covet channel can cause massive risk when it is exploited to pass malicious activities. Moreover, a covert channel can be established even if data encryption is applied.

Recent survey on covert channel countermeasures has reflected that only few works have been presented to counter the rapid development in covert channel techniques [18]. Accordingly, the survey authors recommended that more work in covert channel prevention and detection is highly required as covert techniques are rapidly developed. In addition, each presented solution to counter covert channels targets single type of covert channels instead of taking into consideration the common behaviors of multiple covert channels which results in obtaining solutions that capable to deal with more than one type of covert channels. A new trend to categorize covert channels techniques is presented in [5]. Defiantly this trend contributes positively in developing efficient countermeasures that targeting multiple covert channels.

978-1-5386-4427-0/18/$31.00 ©2018 IEEE

Some important factors that paly essential role in developing covert channel techniques have been presented in [19]. These factors are summarized by the following points:

- The rapid development in computer network technologies, communication protocols, cloud computing, virtualization techniques, and data centers. This advanced development represents a rich area to develop different covert channel techniques.
- Switching techniques which give a covert channel the capability to switch its appearance from one filed to another in a given protocol or from one protocol to another. The switching techniques assist in developing a covert channel that hard to detect.
- Internal control protocol technique which provides reliable and trusted communication channel for a covert message. This technique uses a micro protocol to provide reliable communication and dynamic routing to a covert message.

Basing on the above three factors that enrich the development of covert channel techniques, a new concept has been coined by Elsadig and Fadlalla. The concept is known as network covert channel triangle (DSM, the rapid Development of network techniques, Switching techniques, and Micro protocol techniques). This concept is coined based on the three factors that stated above which have the most impact in create, develop, and secure covert communications.

Recently, a new trend that focuses on defining a unified taxonomy of data hiding techniques is proposed in [6]. It seeks to establish a classification mechanism to sort out, compare, and evaluate covet techniques. This work is highly apricated because it assists in defining and categorizing covert channel techniques in specific patterns that result in developing sufficient countermeasures.

Generally, the ongoing development of covert channel techniques is clearly noticeable. Therefore, more works in developing sufficient solutions to counter these types of threat is highly needed. It is noteworthy to mention that, there are some covert channels have been presented for legal uses [20-24]. however, this doesn't affect the fact that most covert channels are developed to breach system security.

Related Work

Secret communication over network has attracted attackers to leak and discover confidential information. Network covert channel is a modern way to leak information and it is hard to detect [25]. Actually, the common detection techniques -that are used to detect network anomalies or malware activities - are not capable to detect covert channels [26], especially packet length based covert channel which generates covert traffic that very similar to normal traffic. In packet length based covert channels, network packets' lengths are exploited to encode secret message. The attackers benefit from the packets' lengths variation to exchange secret message covertly.

Initially, packet length-based covert channel concept was proposed by Padlipsky (1978) and Girling (1987)[27]. Later this type of covert channel has rapidly developed so many techniques have been presented.

Basing their work in modulating the length of link layer frames to convey secret messages, Padlipsky [28] and Girling [3], developed packet-length covert channels. In these channels, sender and receiver should share some rules before the start of transmission. Due to the random selection of the frame lengths that are used to convey a secret message from a predefined group of frame lengths, the generated covert traffic can easily be distinguished by detection methods. Moreover, these channels use an static predefined dictionary for decoding and deducing a covet message which is vulnerable to statistic detection methods [29]. Therefore, these covert channels are vulnerable to detect.

A packet length covert channel that is based on exchanging of a shared secret matrix was proposed by Yao et al in 2008. The shared matrix contains selected unique packets' lengths. Both sender and receiver must exchange this matrix prior the starting of their transmission session [30]. However, this covert channel is vulnerable to discover [31].

In 2009, Ji et al. constructed a packet length-based covert channel capable to generate covert traffic that imitates normal traffic. They use real packets' lengths as a reference to construct a covert traffic [31]. However, an approach to detect this type of covert channel is presented in [32]. Also in 2009, Ji et al. proposed another packet length based covert channel that utilizes enough packets' lengths as a reference trying to imitate normal traffic and thus to be difficult to detect. [29]. However, the regular distribution of their covert message bits leads to discover this covert channel [33].

In 2011, a high bandwidth covert channel that uses the data payload and network packet lengths to convey secret message was developed [34]. The authors claimed that this covert channel capable to provide high bandwidth channel to send covert messages. However, this technique also uses the data payload to encode secret messages which leads to make it vulnerable to detect. Moreover, this covert channel technique is more complicated compared with other techniques [26].

In 2013, a packet length covert channel that doesn't require a shared key to be exchanged between the communication parties was developed [35]. This feature reflects the important of this covert channel technique compared with the previous packet length covert channels that are required a shared key\rules. This feature makes this this type of covert channel more secure and reduces its overheads. As per the authors claim, the existing detection techniques fail to detect their covert channel. Their covert channel sufficiently imitates normal traffic behavior. Based on our recent survey, to this end, there is no detection method is presented to detect such type of covert channel.

The Proposed Detection Scheme

Our scheme focuses to present a detection method that capable to detect the existence of a packet length-based covert channel that described in [35].

In this kind of packet length covert channels, network packets' lengths are exploited to encode a secret message by

a sender while a receiver decodes the secret message. Each packet length is exploited to encode one bit of a secret message and that is based on one of two scenarios, either (i) the even number of a packet length encodes 1 and the odd number encodes 0, or (ii) the even number of a packet length encodes 0 and the odd number encodes 1. Let us give an example for constructing a covert channel that is based on the first scenario (even length encodes 0 and odd length encodes 1).

Example: after capturing a real network traffic, assume a sender wants to send a secret message of three bits (100), which means only three packets' lengths are needed to do that. The sender examines the first packet length which is required to encode 1, if its length is even number; then, one byte should be padded to this packet in order to change its length to be odd; otherwise, its length is kept as it is. The sender applies the same scenario for the next two packets' lengths in order to encode the rest of the secret message.

A. Dataset

As there is no available dataset of such type of covert channels, we have generated our own dataset to train and test our proposed detection scheme classifier.

The dataset contains two types of network traffics, normal traffic and covert traffic. 100 records of covert and normal traffic are used to train and test the scheme classifier. 70% is used for training phase and 30% for testing phase. These training and testing phases are repeated 20 times using sampling random validation technique to reflect agreeable and reliable classification results.

The normal traffic is captured using the Wireshark tool, which is a well-known capturing tool that is used to capture real network traffic. A part of this normal traffic is used to construct our covert traffic that is based on network packet length covet channel technique that described above. Python language and Scapy library are used for the purpose of modifying and padding network packet lengths according to the different secret message that are encoded to form our covert traffic.

The modulated messages on both covert traffic and normal traffic have been calculated and accordingly a dataset that continues two types of covert messages and normal messages has been created. We notice that, mostly the covert messages are meaningful sentences and take the normal sentence format. In contracts, the messages that obtained from normal network packets do not form meaningful sentences as well they are not taking the normal sentence format and mostly they contain special symbols that are not usually used in forming sentences. Therefore, effective features have been generated to assist in our classification process between covert messages and normal message and thus the traffic that contains covet message is a covert traffic and the other is normal traffic. Our classifiers are fully trained and tested based on out data set and our obtained feathers.

B. Data preprocessing and features extraction

A preprocessing operation is done to enhance our input data and to remove unwanted features that are selected automatically by a classifier. Then a number of effective features are selected to enhance the classifier prediction. These features either content based or frequency-based features that include:

1- By examine covert and normal traffic, it clearly noticed that most time, a number of symbols are usually included at normal traffic but were not appeared at a covert traffic. This obtained feature has highly impact on enhancing our detection scheme capability. This finding is approved practically by comparing our classifier results before and after applying this feature.

2- The repetition of characters more than two times – one after another - can happen in normal traffic.

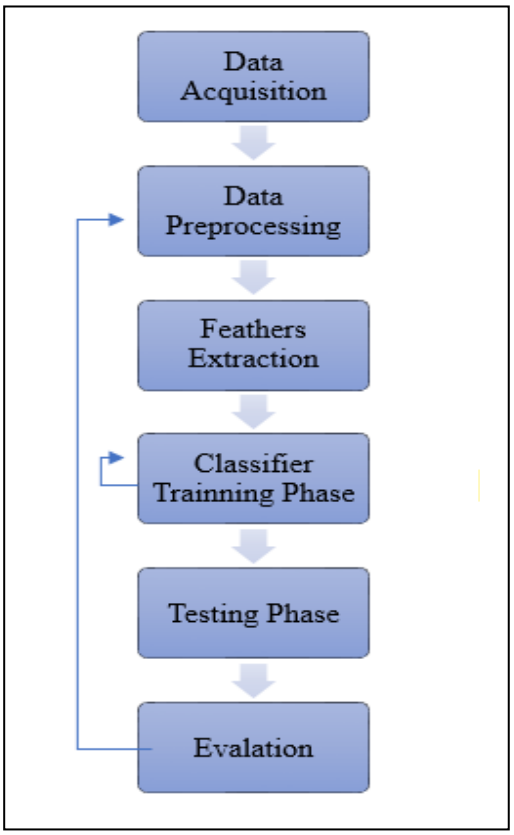

Figure 1. Detection Scheme Diagram

C. Our scheme classifier:

To obtain a proper classier that attains high accuracy to detect the existence of packet length covert channels, five different classifier approaches are fully trained and tested as illustrated in the next section. These classifies are Neural Network, Logistic Regression, Naïve Bayes, Support Vector Machine (SVM) and Random Forest. Based on our experimental results and evaluation, the best classifier that archives higher accuracy is obtained and recommended.

Figure 1 shows our detection scheme diagram which consists six phases: data acquisition, data prepressing, features extraction, classifier training, classifier testing and evaluation. Implementation

IMPLEMENTATION & EVALUATION

In order to come up with a realistic decision which a classifier is better in terms of detection accuracy. Five common classifiers are selected to be under test. These

classifiers are Neural Network, Logistic Regression, Naïve Bayes, Support Vector Machine (SVM) and Random Forest. Based on our own-made dataset and the obtained features, these classifiers were sufficiently trained and tested.

The performance of these classifiers is estimated by using confusion matrix that reflects the classification accuracy as well as the misclassed cases. Figure 2 shows the percentage of classification accuracy for each classifier along with the percentage of misclassified instances. It is clearly noticeable that neural network classifier has outperformed the other classifiers by obtaining a higher degree of detection accuracy and fewer misclassified instances. Moreover, Table 1 shows a performance comparison for all investigated classifiers. The common evaluation measures that are used to evaluate the performance of binary classifiers are computed. These performance evaluation measures include classification accuracy, sensitivity (recall), precision and specificity.

Classification Accuracy = (TP+TN) / (TP+TN+FP+FN).

Specificity = TN / (FP+TN).

Sensitivity (Recall) = TP / (TP+FN).

Precision = TP / TP + FP.

Table. 1 Classification performance Comparison

Classifier	Confusion matrix		Classification Performance			
	TP	FN	Sensitivity	Specificity	Precision	Accuracy
	FP	TN				
Neural Network	300	0	100%	96%	96.2%	98%
	12	288				
Naïve Bayes	300	0	100%	95.3%	95.5%	97.7%
	14	286				
Logistic Regression	300	0	100%	93%	93.5%	96.5%
	21	279				
Random Forest	285	15	95%	92.6%	92.8%	93.8%
	22	278				
SVM	300	0	100%	86.6%	88.2%	93.3%
	40	260				

Where TN: True Negative which is the number of normal traffic instances that classified as normal traffic; TP: True Positive indicates the number of covet traffic instances that classified as covert traffic; FP: False Positive refers to the number of normal traffic instances that classified as covert traffic; and FN: False Negative shows the number of covert traffic instances that classified as normal traffic.

Sensitivity (Recall) is a measure of the true positive rate, which is the percentage of positive cases that are identified correctly. In other words, the percentage of true positive cases among all positive cases, e.g. the number of covert traffic instances that correctly classified among all covert traffic instances.

Specificity is a measure of the true negative rate which means the percentage of negative cases that are identified correctly.

Classification accuracy is the percentage of cases that are classified correctly [36].

Precision is the percentage of true positive cases among cases that are classified as positive [37], e.g. the percentage of covet traffic cases that are identified correctly as covert traffic.

To evaluate the performance of the investigated classifiers, random sampling validation technique is performed on our dataset which includes 50 covert traffic instances and 50 normal traffic instances. This random sampling technique is repeated 20 times. Each time, 70 % of our dataset is used for training and 30% is used for testing. Accordingly, True positive and True negative which indicate the prediction of target class and the prediction of none target class respectively. Classification errors (FP and FN) are evaluated using confusion matrices.

It is clearly noticed that based on our conducted comparison results that is shown in Table 1, the neural network classifier outperforms the other classifier and Naïve Bayes classifier comes next and then Logistic Regression; while SVM and Random Forest are lagged behind.

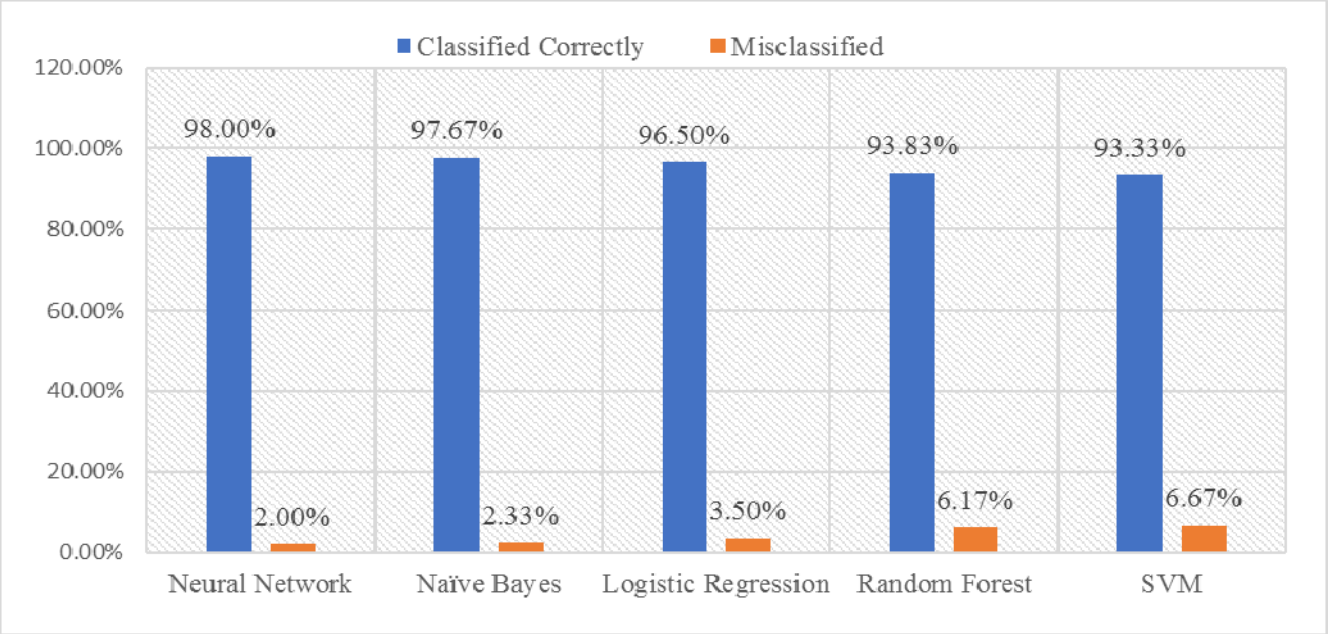

Figure 2. True Classification vs. False Classifcation.

Figures 3, 4, 5, 6, and 7 show the Receiver Operating Characteristic curve (ROC) for Neural Network, Naïve Bayes, Logistic Regression, Support Vector Machine (SVM) and Random Forest classifiers respectively. Figure 8 shows the ROC curves for all classifiers together in one diagram. These figures reflect that, in term of specificity and sensitivity, the neural network is the best. So, our proposed detection scheme which is built on neural network classifier and our selected features attains an excellent classification accuracy in detection of this type of network covert channel.

CONCULSION

Since a packet length covert channel generates a covert traffic that is typically close to normal traffic, this kind of covert channels is considered a one of the most undetectable network covert channels. This research focus on one type of packet length covert channel that is, to this end, no detection solution is presented to deal with. Therefore, this motivates us to introduce a machine learning detection scheme that capable to detect the existence of such type of covert channels. The investigated packet length covert channel in this work is exploited the variations of network packet lengths to send a secret message. It uses the even and odd value to modulate a secret message. This paper is not only purposed a detection scheme, it introduces a comparison study between different classifiers. These classifies are fully trained and tested based on our self-made data set and a number of content and frequency-based features. Accordingly, one classifier is recommended as it gains the highest classification accuracy compared with the others.

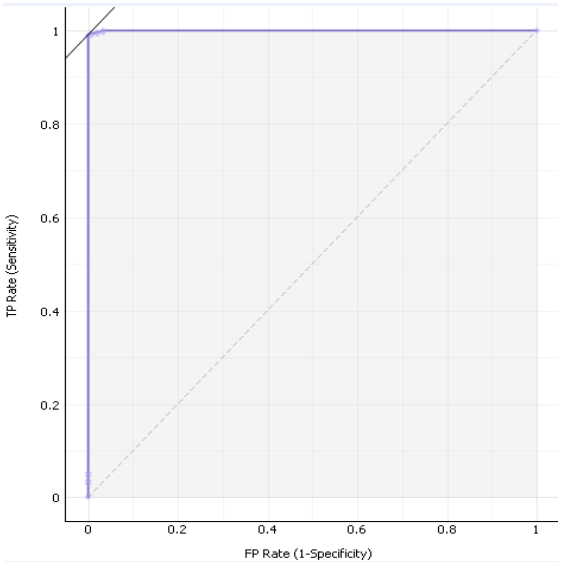

Figure 3. Neural Network ROC curve.

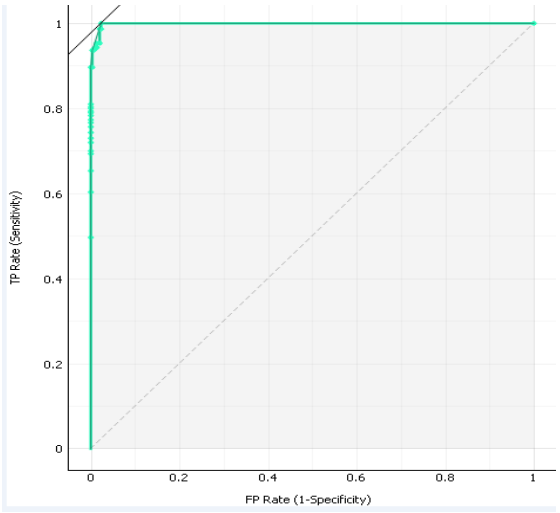

Figure 4. Naïve Bayes ROC curve.

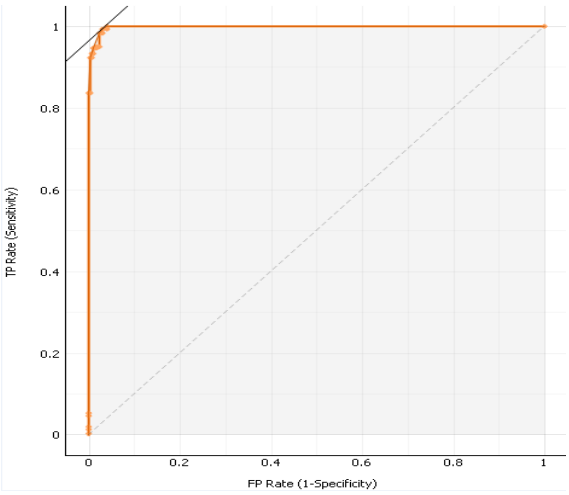

Figure 5. Logistic Regression ROC curve.

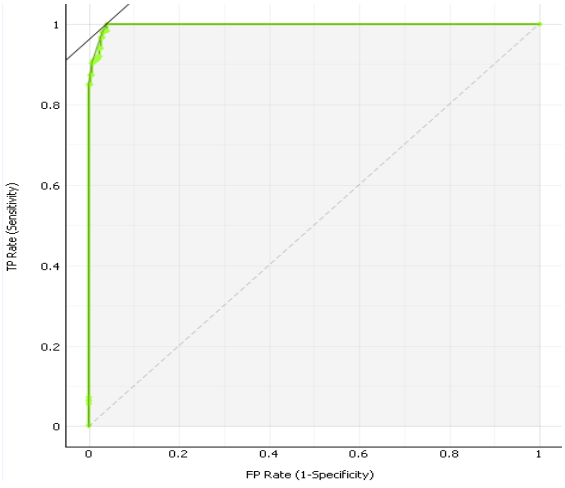

Figure 6. SVM ROC curve.

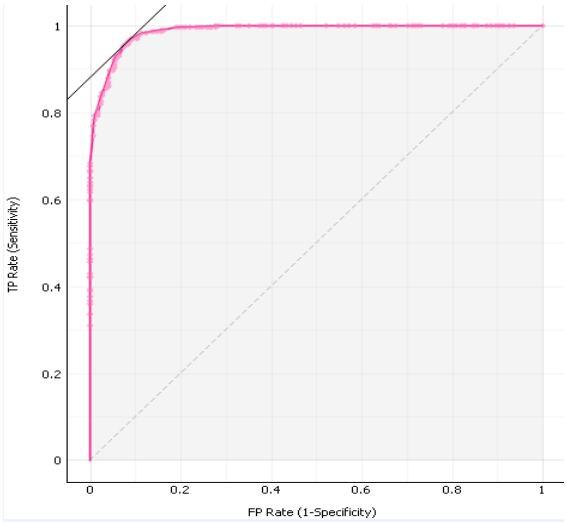

Figure 7. Random Forest ROC curve.

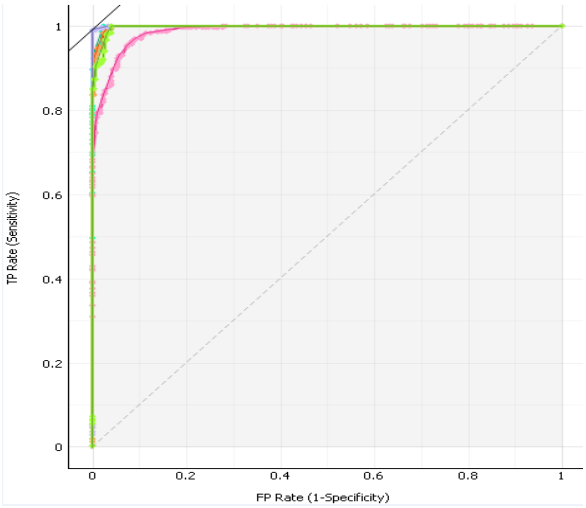

Figure 8. The ROC curves for all classifiers in one diagram.

REFERENCES

[1] B. W. Lampson, "A note on the confinement problem," *Communications of the ACM,* vol. 16, no. 10, pp. 613-615, 1973.

[2] L. Zhang, G. Liu, and Y. Dai, "Network packet length covert channel based on empirical distribution function," *Journal of Networks,* vol. 9, no. 6, pp. 1440-1446, 2014.

[3] C. G. Girling, "Covert Channels in LAN's," *IEEE Transactions on software engineering,* vol. 13, no. 2, p. 292, 1987.

[4] M. Elsadig and Y. Fadlalla, "Survey on Covert Storage Channel in Computer Network Protocols: Detection and Mitigation Techniques," *International Journal of Advances in Computer Networks and Its Security,* vol. 6, no. 3, pp. 11-17, 2016.

[5] S. Wendzel, S. Zander, B. Fechner, and C. Herdin, "Pattern-based survey and categorization of network covert channel techniques," *ACM Computing Surveys (CSUR),* vol. 47, no. 3, p. 50, 2015.

[6] S. Wendzel, W. Mazurczyk, and S. Zander, "Unified Description for Network Information Hiding Methods," *J. UCS,* vol. 22, no. 11, pp. 1456-1486, 2016.

[7] M. Wojciech, W. Steffen, Z. Sebastian, H. Amir, and S. Krzysztof, "Control Protocols for Reliable Network Steganography," in *Information Hiding in Communication Networks:Fundamentals, Mechanisms, Applications, and Countermeasures*: Wiley-IEEE Press, 2016, p. 296.

[8] P. Danaee, R. Ghaeini, and D. A. Hendrix, "A deep learning approach for cancer detection and relevant gene identification," pp. 219-229: World Scientific.

[9] M. Abdelmagid, A. Ahmed, M. Himmat, "Information Extraction Methods and Etraction Techniques in the Chemical Documents Contents: Survey", ARPN Journal of Engineering and Applied Science, vol. 10, no. 3, 2015, ISSN 1819-6608.

[10] S. Begum, S. P. Bera, D. Chakraborty, and R. Sarkar, "Breast cancer detection using feature selection and active learning," p. 43: CRC Press.

[11] G. Kyriakides, K. Talattinis, and G. Stephanides, "A Hybrid Approach to Predicting Sports Results and an AccuRATE Rating System," *International Journal of Applied and Computational Mathematics,* vol. 3, no. 1, pp. 239-254, 2017.

[12] T. Shon and J. Moon, "A hybrid machine learning approach to network anomaly detection," *Information Sciences,* vol. 177, no. 18, pp. 3799-3821, 2007.

[13] J. Sahs and L. Khan, "A machine learning approach to android malware detection," pp. 141-147: IEEE.

[14] R. B. Basnet, S. Mukkamala, and A. H. Sung, "Detection of Phishing Attacks: A Machine Learning Approach," *Soft Computing Applications in Industry,* vol. 226, pp. 373-383, 2008.

[15] R. A. R. Ashfaq, X.-Z. Wang, J. Z. Huang, H. Abbas, and Y.-L. He, "Fuzziness based semi-supervised learning approach for intrusion detection system," *Information Sciences,* vol. 378, pp. 484-497, 2017.

[16] D. C. Latham, "Department of defense trusted computer system evaluation criteria," *Department of Defense,* 1986.

[17] B. Carrara and C. Adams, "A Survey and Taxonomy Aimed at the Detection and Measurement of Covert Channels," in *Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security*, 2016, pp. 115-126: ACM.

[18] M. A. Elsadig and Y. A. Fadlalla, "Survey on Covert Storage Channel in Computer Network Protocols: Detection and Mitigation Techniques," pp. 79-85.

[19] A. Epishkina and K. Kogos, "A Traffic Padding to Limit Packet Size Covert Channels," in *Future Internet of Things and Cloud (FiCloud), 2015 3rd International Conference on*, 2015, pp. 519-525.

[20] R. DeGraaf, J. Aycock, and M. Jacobson Jr, "Improved port knocking with strong authentication," in *Computer Security Applications Conference, 21st Annual*, 2005, pp. 10 pp.-462: IEEE.

[21] H. Qu, Q. Cheng, and E. Yaprak, "Using Covert Channel to Resist DoS attacks in WLAN," in *ICWN*, 2005, pp. 38-44.

[22] W. Mazurczyk and Z. Kotulski, "New security and control protocol for VoIP based on steganography and digital watermarking," *arXiv preprint cs/0602042,* 2006.

[23] D. D. Dhobale, V. R. Ghorpade, B. S. Patil, and S. B. Patil, "Steganography by hiding data in TCP/IP headers," in *2010 3rd International Conference on Advanced Computer Theory and Engineering(ICACTE)*, 2010, vol. 4, pp. V4-61-V4-65.

[24] H. Xie and J. Zhao, "A lightweight identity authentication method by exploiting network covert channel," *Peer-to-Peer Networking and Applications,* vol. 8, no. 6, pp. 1038-1047, 2015.

[25] S. Z. Goher, B. Javed, and N. A. Saqib, "Covert channel detection: A survey based analysis," in *High Capacity Optical Networks and Emerging/Enabling Technologies*, 2012, pp. 057-065.

[26] A. Epishkina and K. Kogos, "A random traffic padding to limit packet size covert channels," in *Computer Science and Information Systems (FedCSIS), 2015 Federated Conference on*, 2015, pp. 1107-1111: IEEE.

[27] M. A. Elsadig and Y. A. Fadlalla, "A balanced approach to eliminate packet length-based covert channels," pp. 1-7: IEEE.

[28] M. A. Padlipsky, D. W. Snow, and P. A. Karger, "Limitations of end-to-end encryption in secure computer networks: Technical report ESD-TR-78-158," *Massachusetts: The MITRE Corporation,* 1978.

[29] L. Ji, H. Liang, Y. Song, and X. Niu, "A normal-traffic network covert channel," in *Computational Intelligence and Security, 2009. CIS'09. International Conference on*, 2009, vol. 1, pp. 499-503: IEEE.

[30] Q.-z. YAO and P. ZHANG, "Coverting channel based on packet length," *Computer engineering,* vol. 34, no. 3, pp. 183-185, 2008.

[31] L. Ji, W. Jiang, B. Dai, and X. Niu, "A novel covert channel based on length of messages," in *2009 International Symposium on Information Engineering and Electronic Commerce*, 2009, pp. 551-554: IEEE.

[32] A. S. Nair, A. Sur, and S. Nandi, "Detection of Packet Length Based Network Steganography," in *2010 International Conference on Multimedia Information Networking and Security*, 2010, pp. 574-578.

[33] A. Epishkina and K. Kogos, "Covert Channels Parameters Evaluation Using the Information Theory Statements," in *IT Convergence and Security (ICITCS), 2015 5th International Conference on*, 2015, pp. 1-5.

[34] M. Hussain and M. Hussain, "A high bandwidth covert channel in network protocol," in *Information and Communication Technologies (ICICT), 2011 International Conference on*, 2011, pp. 1-6: IEEE.

[35] O. I. Abdullaziz, V. T. Goh, H. C. Ling, and K. Wong, "Network packet payload parity based steganography," in *2013 IEEE Conference on Sustainable Utilization and Development in Engineering and Technology (CSUDET)*, 2013, pp. 56-59.

[36] R. Mizoguchi and J. Slaney, *PRICAI 2000 Topics in Artificial Intelligence: 6th Pacific Rim International Conference on Artificial Intelligence Melbourne, Australia, August 28-September 1, 2000 Proceedings*. Springer Science & Business Media, 2000.

[37] M. Abdelmagid, M. Himmat, and A. Ahmed, "Survey on Information Extraction from Chemical Compound Literatures: Techniques and Challenges," Journal of Theoretical and Applied Information Technology, vol. 67, no. 2, 2014..