Packet Length Covert Channel Detection: An Ensemble Machine Learning Approach

K.G.S.Venkatesan

1Department of Computer Science and Engineering, MEGHA Institute of Engineering and Technology for Women, Edulabad, Hyderabad, Telangana, India.

Received: 08.09.2023 ● Accepted: 20.09.2023 ● Published Online: 30.09.2023

Abstract: New undetectable attack methods made possible by advances in covert channel techniques. Due to their nonstandard data transfer methods, common countermeasures are ineffective. The detection, mitigation, and elimination of hidden channels that operate on the packet length level are extremely challenging problems. Differences in packet size can be used as part of a covert communication strategy for the delivery of sensitive information across networks. The use of machine learning techniques for detecting covert channel attacks has been praised in recent academic research. Researchers in this work developed an effective ensemble classification method for spotting such breaches. Our approach relies on an ensemble model, which is a combination of three different machine learning methods. There are several applications for classifiers including the Support Vector Machine (SVM), Random Forest (RF), and Naive Bayes (NB). The logistic regression (LR) classifier served as a meta classifier to aggregate the results of the component classifiers into an ensemble classifier. According to the findings, the proposed ensemble model is effective. Among single-classification algorithms, it is unrivalled in its ability to detect such covert channel attacks.

Key words: Network Security, Cybersecurity, Data Leakage Prevention, Threat Detection, Feature Engineering.

*Correspondence: Professor, Department of Computer Science and Engineering, MEGHA Institute of Engineering and Technology for Women, Edulabad, Hyderabad, Telangana, India. Email:drkgsvenkatcse@meghaengg.ac.in
https://doi.org/10.58599/IJSMIEN.2023.1903
Volume-1, Issue-9, PP:20-26 (2023)

This work is licensed under a Creative Commons Attribution 4.0 International License CC BY-NC-ND 4.0.
1. INTRODUCTION

Covert channel is a term used to describe a method of transmitting sensitive data outside of an authorized channel [1, 2]. Lampson first recognized the unauthorised movement of information in 1973 [3]. Data is sent over conduits that were never meant to carry such information [4]. In 1987, more work was done on the idea to make use of the power of computers and networks [5]. In order to conceal information in transit via a network, covert channels are often used within the infrastructure itself [6]. Covert channels of all kinds can flourish in the modern infrastructure of online networks [7]. Even in the most cutting-edge innovations, not enough consideration is paid to the weaknesses that can be exploited to set up covert communication channels. References [8] provide more depth on the topic of hidden channels that make use of the IPv6 protocol, which is the next iteration of Internet Protocol.

In [9], the authors outlined several characteristics that are vital to the launch of networks. Attacks against communication systems that are intended to remain undetected. As network infrastructure improves, more opportunities arise for employing covert channel methods. This is possible because switching techniques can move a covert message from one field in one protocol to another field in a completely different protocol. As a result, it becomes increasingly difficult to identify these kinds of attacks. The detection of such covert channel attacks is further complicated by the use of internal control protocol approaches that incorporate micro protocol applications. The duration of a data packet is sometimes used as a secret channel to communicate confidential information. Using the variation in data packet lengths, sensitive information can be sent securely to its destination. Many examples have been found in the current literature that show how varying packet lengths in a network can be used to launch a variety of attacks. The use of encryption has not rendered this type of assault impossible. Protocols like IPv6 datagrams can benefit greatly from packet length hiding. Creating universal detectors requires first developing high-level indicators. This can be done with the help of commonly-used machine learning classification algorithms or frameworks [10] that can spot statistical outliers. All covert channel techniques based on packet length cannot be detected by a single method. Due to the constant transfer of large amounts of data in VoIP traffic, it provides an ideal setting for the research and development of various covert channels based on packet length. Mobile Voice over Internet Protocol (VoIP) is one application that can benefit from covert channels’ high data transfer rate, as VoIP is increasingly used to transmit large amounts of data. As time goes on, it becomes clearer that novel approaches are appearing for building covert channel attacks. The widespread use of covert channels of communication poses serious risks to the confidentiality of our personal information. As a result, there should be more work done to set up adequate protections against such attacks. In this research project, an ensemble classification model was created to analyze packet data in search of evidence of covert channels. Stacking is a machine learning strategy that improves accuracy by combining the outputs of numerous classifiers into a single prediction. In this model, the results from the two underlying base classifiers are combined using the logistic regression (LR) classifier, which acts as a meta-classifier.
2. RELATED WORKS

The term "covert channel" is used to describe a method of sending secret information through a line of communication that is not sanctioned by law [1, 2]. Lampson in 1973 [3] was the first to recognize the phenomenon of unauthorised transfer of information. It’s true that a lot of data transmission nowadays happens across connections that weren’t meant for it [4]. 1987 saw the start of dedicated research and development into the best ways to utilize computers and networks [5]. The use of covert channels within the core network infrastructure is common practice for protecting data in transit [6]. A number of hidden pathways have flourished in the modern online network infrastructure [7, 8]. Even with cutting-edge technology, it is not always given enough consideration to address the weaknesses that could be exploited to set up covert channels of communication [11]. To learn more about hidden channels and how they work with IPv6, the next version of the Internet Protocol, check out the aforementioned references [9]- [12]. In their research, the authors presented a thorough review of the most important characteristics that should already be in place before a network is created. attacks on communication networks that are meant to remain undetected for as long as possible. There is a growing possibility of employing covert channel techniques due to the development of network infrastructure. Switching methods make it possible for a hidden message to be moved from one field in one protocol to another field in a completely different protocol, making this phenomenon possible. That’s why it’s getting harder and harder to spot these kinds of assaults. Integrating micro protocol applications into internal control protocol methodology increases the complexity of detecting covert channel attacks. Sometimes, someone will use the length of a data packet to secretly send overly personal or confidential information. Secure communication of sensitive information to the intended recipient is made possible via the use of changes in data packet lengths. In the existing corpus of research, numerous examples have been found to demonstrate the use of varying packet lengths inside a network to begin a variety of assaults. While encryption has reduced the likelihood of this type of assault, it is still possible. Hiding the actual length of a packet, like in IPv6 datagrams, can have several practical benefits. In order to begin working on universal detectors, it is necessary to first establish high-level indications. In order to identify statistical outliers, many researchers have turned to widely used machine learning classification algorithms or frameworks. While many covert channel techniques depend on packet length, there is no foolproof mechanism for discovering them all. Voice over IP (VoIP) traffic’s constant, large-volume data transfers make it an ideal testing ground for the development and study of various covert channels dependent on packet length. Given the increasing need to send large amounts of data, Mobile Voice over Internet Protocol (VoIP) may benefit from the use of covert channels due to their high data transfer rate. Covert channel attack development has been showing signs of progress toward more novel approaches over time. There are serious risks to people’s privacy because of the widespread use of underground communication channels. That’s why it’s crucial to put in more work creating adequate protections against attacks like this. The goal of this study was to create an ensemble classification model for analyzing packet data in order to spot clues to hidden channels. Combining the results of various classifiers into one cohesive prediction, stacking is a machine learning strategy that improves
predictive accuracy. For the purposes of combining the results of the two basic classifiers, the present model makes use of the logistic regression (LR) classifier as a meta-classifier.

3. METHODOLOGY
This research presents a new ensemble classification approach for identifying and protecting against stealth channel attacks, which use differences in packet sizes inside a network to secretly distribute sensitive data. The aforementioned attack method entails adjusting the packet length to send a "0" bit of the secret message when the value is odd and a "1" bit when the value is even, or vice versa. This suggests that a malicious actor is using packet timing manipulation to send a secret message over a network. In order to decode the modulated message, the receiver maintains track of the individual data lengths that it receives. A dataset has been compiled for the purpose of studying the covert channel as defined before. The network traffic of the Skype application was collected with the Wireshark tool. The collected data was divided into two distinct portions. Half of the data was left unaltered to symbolize benign traffic, which refers to lawful traffic. Conversely, the remaining half was manipulated utilizing the Scapy tool in the Python programming language to produce malicious traffic, also known as covert traffic. Hence, a dataset including 200 observations. The number of cases manufactured. There exist two distinct categories of traffic, each manifesting itself on 100 separate occasions: overt traffic and covert traffic.

3.1. Ensemble Classifier Model
The ensemble approach, which is widely recommended, employs a stacking mechanism to aggregate the outputs of many classifiers. It is widely acknowledged in academic literature that ensemble classifiers exhibit superior accuracy compared to individual classification models. Considerable work was invested in the meticulous selection of foundation classifiers for our ensemble technique, which have demonstrated a consistent ability to accurately represent performance. Two often used classifiers in machine learning are the Naive Bayes (NB) and Random Forest (RF) methods. The ensemble strategy we suggest involves utilizing the Logistic Regression (LR) classifier as a meta classifier to amalgamate the outcomes of the basic classifiers.

3.2. Experiments and Evaluation
The ensemble approach proposed in this paper was evaluated using a dataset consisting of 200 occurrences of both concealed and visible traffic. The dataset was utilized for both the training and testing phases of the underlying classifiers. An ensemble technique was developed by integrating the outcomes of the classifiers stated above using the RF classifier. The efficiency of the suggested model was evaluated using a cross-validation method. The approach described in this document proposes the utilization of a rotational scheme, wherein multiple subsets of the dataset are employed for the purposes of training and testing a machine learning model. The approach of machine learning assessment resampling is employed in accordance with the methodology described in reference. In the
conducted experiment, a partitioning strategy was employed whereby 20% of the dataset was allocated for testing purposes, while the remaining 80% was designated for training. In the study, the dataset was partitioned into five distinct portions, commonly referred to as folds, to facilitate the implementation of the 5-fold cross-validation technique. The model undergoes five iterations of training and testing. In each iteration of model training, the complete dataset is utilized, however in testing rounds, just a subset, often known as a “fold,” of the data is employed. Numerous modifications are implemented on the training and testing datasets, leading to a more comprehensive validation process. In order to assess the effectiveness of the proposed ensemble method, a confusion or error matrix was built. The matrix provides a rapid overview of the accuracy rates for correctly classified and wrongly identified events. The objective of this study was to assess the performance of our ensemble technique relative to the utilization of a single classifier. Additional evaluation measures, such as accuracy in classification, sensitivity, precision, and specificity, were also computed as listed in Table-1.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Specificity</th>
<th>Sensitivity</th>
<th>Precision</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>RF</td>
<td>95.1%</td>
<td>97%</td>
<td>93.3%</td>
<td>95%</td>
</tr>
<tr>
<td>SVM</td>
<td>97.6%</td>
<td>98%</td>
<td>95.3%</td>
<td>97.5%</td>
</tr>
<tr>
<td>Naïve Bayes</td>
<td>99%</td>
<td>98%</td>
<td>98.1%</td>
<td>99%</td>
</tr>
<tr>
<td>The proposed ensemble Classifier (stack)</td>
<td>99.5%</td>
<td>98%</td>
<td>99%</td>
<td>99.5%</td>
</tr>
</tbody>
</table>

4. RESULTS AND DISCUSSIONS

Table 1 illustrates that the accuracy of the ensemble model much surpassed that of the individual classifiers employed in its construction. The detection accuracy of the system for concealed channels, as determined by analyzing packet length, was found to be 98.5%. This result highlights the system’s remarkable performance in this aspect. In terms of evaluating the effectiveness of binary classifiers, the ensemble classifier demonstrated robust performance, as seen by its notable performance in commonly employed measures such as specificity, accuracy, and recall. All of these indicators achieved statistically significant values. The investigation’s outcomes provide support for our initial hypothesis that the use of an ensemble classifier has the capacity to enhance the accuracy of categorization. The results of a comparison of the classification accuracy of different classifiers, including the proposed ensemble classifier, are presented in Figure 1 to facilitate the understanding of our findings. The results of utilizing the confusion matrix to compute classification errors are presented in Table 1. In comparison to the classifiers currently available, the proposed classifier has excellent performance. The system demonstrates a notable level of accuracy with minimal mistake rates. The Naive Bayes (NB) and Support Vector Machine (SVM) classifiers, which adhere to the suggested classifier, exhibit strong performance. However, the Random Forest (RF) classifier falls short since it yields a greater number of errors. ROC curves were obtained for subsequent assessment. The receiver operating characteristic (ROC) curves for the recommended classifier, Naive Bayes, Support Vector Machine , and Random
Forest are presented below. The receiver operating characteristic (ROC) curves for all of the classifiers are depicted in Figure 1. The receiver operating characteristic (ROC) curves demonstrate that the classifier being evaluated had superior performance compared to the other classifiers utilized in the experiment.

![Figure 1. Accuracy of the Classifiers](image)

The efficacy of our suggested model in enhancing the detection accuracy of the packet length covert channel is demonstrated through experiments and evaluation tools. The classifier demonstrated superior accuracy compared to its competitors, while also exhibiting a lower rate of errors than any other classifier. Based on existing knowledge, the classifier under consideration demonstrates superior performance compared to all other existing detection algorithms employed for the prevention of this specific type of assault. To clarify, our proposed model demonstrates superior performance compared to NB and SVM classifiers, both of which have exhibited adequate levels of classification accuracy.

5. CONCLUSION

In fields as diverse as information security, machine learning has proven its utility. Many different machine learning strategies have been used with considerable success to detect and identify harmful activities. Using data packet lengths as a proxy for the presence of covert channels, a classification model was developed in this research. To transmit secret data, covert channels of this type take advantage of differences in network packet lengths. Covert traffic is a dangerous and stealthy sort of assault that can cause damage to a network. Because this communication so closely resembles "overt" network activity, traditional detection methods are generally incapable of spotting it. The proposed ensemble mode is comprised on the three widely-acknowledged base classifiers: the Support Vector Machine (SVM), the Random Forest (RF), and the Naive Bayes (NB). The LR classifier is used to merge the results of these classifiers. As indicated before, this model serves as a meta classifier. The
ensemble detection model outperformed individual classification methods by a wide margin, with a 98.5% detection accuracy.

References


