International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 6, Issue 12, ISSN No. 2455-2143, Pages 82-92
Published Online April 2022 in IJEAST (http://www.ijeast.com)

DETECTION OF VPNS USING MACHINE LEARNING

Shipra Srivastava, Harigovind H, Devvrat Modi, Mohammad Bassam Salim, Yashraj Mathur
Department of IT
GNIOT, Greater Noida, Uttar Pradesh, India

***Abstract*— The nature of the tools used by criminal actors to hide their identities has made it is very challenging to detect unauthorized users. Network security is any move an association makes to forestall vindictive use or coincidental harm to the organization's private information, its clients, or their gadgets. The objective of organization security is to keep the organization running and valid for every authentic client. Since there are such countless ways that an organization's security can be compromised, network security includes a wide scope of rehearses. Most of normal assaults against networks are intended to acquire admittance to data, by keeping an eye on the correspondences and information of clients, rather than to harm the actual organization. However, assailants can do more than taking the information. They might have the option to harm clients' gadgets or control frameworks to acquire actual admittance to offices. This leaves the association's property and individuals in danger of damage. Network security extensively comprises of approaches, cycles and practices embraced to forestall, recognize and screen unapproved access, abuse, change, or disavowal of organization available assets. This work presents computational model that can detect the use of virtual private networks to gain unauthorized access.**

***Keywords*—Security, Virtual Private Network, Information, Authentic.**

I. INTRODUCTION

Virtual Private Networks (VPNs) are a typical strategy for crooks and other troublemakers to camouflage their identity on the web [1, 2]. This is helped along by the increment in simplicity of utilization of VPNs; they are at this point that they are not simply an instrument for, from a distance, getting to big business assets while going for work or while telecommuting. Indeed, this could be a utilization case for a lawbreaker. Assuming that they wish to remotely get to an undertaking network to take organization and proprietary innovations, they can utilize a VPN to conceal their own area or to cause it to show up as if another person was penetrating the organization [3]. There have been a couple of outstanding instances of this occurrence lately, for example, the Sony Pictures occurrence from 2014, where classified information including individual data about representatives was taken [4, 5]. Different assaults of note are the different information breaches which have been happening for the last number of years, for example, the LinkedIn breach [6]. Around 167 million record subtleties including messages furthermore passwords were taken. It isn't known whether the attacker(s) were utilizing a VPN administration to conceal their area.

Numerous obscurity innovations with most are being in view of organizations called "blend" organizations. These 'Blend organizations' course bundles so as to make it very troublesome a connection between the wellspring of the solicitation. This works by means of through delegates and 'blending' bundles from member. This makes it undeniably challenging for snoops to follow start to finish interchanges [7, 8]. Low inactivity frameworks incorporate the well-known unknown correspondence framework Tor as well as HTTP/SOCKS intermediary administrations and Virtual Private Networks (VPNs) [9]. Frameworks, for example, Tor fall under the class of multi-jump unknown correspondences models, while HTTP/SOCKS intermediaries and VPNs for the most part fall under the classification of single-bounce mysterious correspondence models. Intermediary servers that are utilized to give anonymisation depend on one more sort of intermediary known as an "open" proxy. Open intermediaries are an intermediary that is accessible to any client on the Web. They are generally used to set up mysterious intermediary sites and arranged as a solitary bounce mysterious correspondence model. There are a few unique executions of VPNs for giving mysterious correspondences [10, 11, 12]. The planned use for VPN executions was to permit an association's labourers to safely access inner organization assets from outside of the inner organization for example remote access. This is accomplished through setting up an association called a passage between the client's PC and the association servers. VPNs anyway can additionally be utilized as a mysterious correspondence framework in a comparable way to a mysterious intermediary server. The principle contrast between the two strategies is in the VPN's burrowed association. The burrowed association between the client and the VPN server is encoded.

II. VIRTUAL PRIVATE NETWORKS (VPNS)

A Virtual Private Network (VPN) gives private organizations of assets and data over any open network [21, 22]. It empowers a remote machine on network X to burrow traffic

that could not ordinarily have the option to be sent across the Web, to a door machine on network Y and appear to be sitting, with an inward IP address, on network Y. The passage machine gets traffic to this inward IP address, what's more sends it back to the remote machine on network X [23]. This itself doesn't give a lot of safety. Blocking these burrowed bundles would in any case take into account the substance of the private parcels to be caught and uncovered by an outsider. To conquer this, the private bundles should be encoded or more that, some type of confirmation should be utilized. VPN conventions change in their help for encryption and confirmation plans. Every one of the accompanying areas will examine a few model calculations and plans upheld by each VPN convention.

II.i. PPTP:

The Point-to-Point Tunneling Protocol (PPTP) is a connection layer VPN convention that is intended to burrow Point-to-Point Convention (PPP) associations through an IP organization, making a VPN association [10, 23]. The last bundles are sent over IP from the client to the passage PPTP server and back once more. PPTP doesn't give any strategies to keeping information classified or for giving solid verification. The Microsoft execution that was incorporated with Windows NT gives a system to arranging verification and encryption calculations among server and client which depends after existing exchanges held inside augmentations and improvements of PPP [25]. Some model verification calculations are the Password Authentication Protocol (PAP), the Challenge-Handshake Authentication Protocol (CHAP), MS-CHAPv1/v2, Microsoft's executions of CHAP, what's more Extensible Authentication Protocol (EAP). CHAP and MS-CHAPv1/v2 have confronted broad investigation throughout the long term [22, 23, 26, 27]. PAP communicates the username and secret phrase from the client through a decoded channel which leaves it defenceless against eavesdropping attacks. This leaves it in the position where it must be utilized if all else fails. Because of the weaknesses that have been found in the confirmation and encryption calculations it utilizes, PPTP doesn't see boundless use any longer.

II.ii. L2TP:

The Layer 2 Tunnelling Protocol (L2TP) is additionally a connection layer VPN that broadens the PPP model by consolidating elements of PPTP with elements of the Layer 2 Forwarding (L2F) convention [28]. L2TP works in much the same way to PPTP. More elevated level conventions, usually PPP associations, are embodied inside a L2TP burrow by setting up a L2TP meeting. The L2TP bundles thus, including both the payload and the L2TP header are shipped inside a UDP parcel. L2TP is additionally like PPTP in that it doesn't give some techniques for secrecy or validation and on second thought acquires existing securities from PPP. A convention suite called IPsec was acquainted with give further developed confirmation and privacy over the PPP strategies [12]. The first PPP techniques utilized by L2TP were viewed as helpless against a Refusal of Service (Dos) assault which included sending a solicitation to stop the association utilizing the right ID to end the VPN session [20]. This was a weakness that was addressed in a refreshed adaptation of L2TP called L2TP form 3 (L2TPv3). The new form included a discretionary validation and uprightness check that invalidated the weakness. L2TP is regularly joined with another validation and encryption convention suite called Internet Protocol security (IPSec) [29].

II.iii. IPSec:

IPsec incorporates an assortment of normalized conventions for shared validation between two hosts toward the start of a VPN meeting and for the arrangement of cryptographic keys used to empower encryption for the session [29]. Information is kept secure by verifying organization bundles to ensure the trustworthiness of the parcel and that epitome has been carried out accurately. There are two modes in which IPsec can give this usefulness: transport mode and passage mode. In transport mode, the first bundle is altered to incorporate another IPsec header in the first IP header. This extra header contains the data expected to perform verification and uprightness checking. In correlation, burrow mode gives greater adaptability. In burrow mode, the total of every unique IP bundle is epitomized inside another IP bundle comprising of another IP header and the IPsec header [29]. This adds a layer of deliberation from the first IP bundle's substance hence giving privacy to the payload. To figure out which mode is to be utilized during an association, security data characterizing the modes that each end point upholds should be traded. This is alluded to as a security affiliation. It contains data on the mode of IPsec to be utilized, the encryption calculations to be utilized and the encryption keys used to set up the encryption. Trade of this data is finished utilizing the Internet Key Exchange (IKE) convention.

II.iv. OpenVPN:

OpenVPN [30] has a straightforward arrangement and the combination of big business level security, ease of use and different elements, in addition to its help for the vast majority of the working frameworks that are accessible, it is generally viewed as among the best VPN arrangements [31]. OpenVPN utilizes Hash-based message validation codes (HMAC) in blend with the SHA1 hashing calculation for guaranteeing parcel respectability. OpenVPN has two validation modes. In mode one, a pre-shared static key is utilized to give verification and encryption. In mode two, SSL/TLS mechanisms are utilized for authentication and key exchange [31]. In static key mode, a pre-shared key is divided among the two hosts before the passage is set up. This static key contains four autonomous sub-keys: HMAC send, HMAC get, encode and decode. The favoured method of activity is mode

International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 6, Issue 12, ISSN No. 2455-2143, Pages 82-92
Published Online April 2022 in IJEAST (http://www.ijeast.com)

two which utilizes SSL/TLS. In this mode an SSL meeting is laid out requiring the two hosts to introduce their own verification endorsement. Assuming the validation of the hosts succeeds arrangement and trade of the encryption /decoding and HMAC keys starts. Rather than the keys being static as in mode 1, in mode 2 the keys are haphazardly created either by OpenSSL's RAND_bytes work or by utilizing the TLS pseudorandom work (PRF) close by irregular source material from the two hosts. The keys are then traded over the SSL/TLS association and the burrow sending process starts. The information to be encoded what's more moved in the passage incorporates a 64-bit sequence number and the payload information comprising of an IP parcel or Ethernet outline. Encryption of the passage parcels is conveyed out utilizing the Blowfish secret key Block Cipher [32]. OpenVPN then multiplexes the SSL/TLS meeting that is utilized for confirmation and key trade with the encoded burrow information. SSL/TLS is intended to work utilizing a dependable transport convention so OpenVPN gives a dependable vehicle layer on top of UDP. The genuine IP bundles are burrowed over UDP without an additional dependability layer after they have been verified with a HMAC as the IP parcel forwarder has been intended to work over a questionable vehicle layer.

III. SECURITY AND COMMUNICATION NETWORKS:

To empower the correspondence between the PCs, TCP/IP stack was carried out. The stack was executed without the thought of safety of data being moved in the correspondence [48]. This issue raised a great deal of safety concerns which are continually overseen by different security administrations [49]. Secure Sockets Layer (SSL) is has colossally expanded for some, security reasons. One of the reasons might be to just approve parties associated with the communication [52]. Straightforward firewalls are by and large not outfitted with SSL assessment or off-stacking which permits scrambled traffic to pass with no investigation [53]. This permits pernicious traffic inside the organization over clandestine channels that are not assessed by the firewall [54]. There is a desperate need to recognize genuine and ill-conceived traffic with negligible organization upward and generally framework cost. This will permit any scale association to more readily oversee their authoritative strategies. Virtual private organization (VPN) administration might be utilized to conceal the genuine traffic in the organization which might be in any case not permitted or might be checked [55]. A client utilizing VPN administration interfaces with a VPN server utilizing ordinary Transport Layer Security (TLS) association outside the organization. Once associated, it demands the site or administration from the server [56, 57]. The VPN server starts the solicitation for the benefit of the client to the server mentioned. The encoded reaction is shipped off the client on currently settled channel; subsequently, the entire movement passes any channel on the organization firewall. Such methods might be utilized by the clients who mean to stow away from or mislead the association of their Internet action [56]. This paper proposes an original strategy to distinguish VPN traffic inside an organization. The proposed method extricates the organization traffic elements and characterizes the traffic to demonstrate in the event that the traffic is authentic or not. Key elements are removed from the organization traffic and are looked at against the generally distinguished highlights of traffic viewed as ill-conceived or VPN traffic. The framework is additionally ready to order the traffic which isn't following the example of ordinary traffic or typical client action and banners that specific traffic stream to be invalid. We tried our framework against five notable uninhibitedly accessible electronic VPN specialist co-ops; the proposed framework had the option to group every one of them accurately. More traffic-portraying highlights might be added to distinguish more applications.

IV. RELATED WORK AND COMPARISON:

Different VPN administrations like TOR [58], Hotspot Shield, and different administrations have exceptional fingerprints, and not all the administrations can be recognized utilizing a comparable measure. Yamada et al. talked about a procedure that utilizes measurable examination on the encoded traffic [59]. The plan talked about, utilizes information size of organization bundles and performs timing investigation on the got parcels to distinguish malevolent traffic inside an encoded channel. This procedure is exceptionally valuable for Web specialist co-ops to dissect the traffic coming to their servers and distinguish any malevolent movement coming from outside the organization.

A review on android-based applications which use VPN administration [60] to show that these VPN administrations might utilize outsider trackers to follow client conduct, and some might be utilized to sidestep android sandbox climate. Once a malware is conveyed to the gadget inside the organization, the entire organization is vulnerable against attacks [61].

VPN clients inside the organization go about as an intermediary, which interface with the separate VPN server. When the association is laid out, the VPN specialist organization can change or listen in on the data and network traffic as required [62, 63]. This draws in some outsider commercial or following elements [64, 65]. Any pernicious element can peruse, save, or potentially alter our solicitation and the connected data to and from the foreordained assistance. VPN administrations can change the information as they are in charge of approaching and active traffic from organization to gadget. VPN administrations are likewise ready to perform TLS interception [66] attempt by utilizing their own authentications which is trusted locally by the framework, for VPN administration to work appropriately. This prompts an all the more possibly hazardous circumstance when the gadget associated contains touchy information [59,

68]. One of the countermeasures to this issue is certificate pinning [59, 69]. In this way, recognizing such VPN administrations inside your organization can save you from immense misfortunes as far as the data lost.

Goh et al. [70] proposes a man-in-the-center methodology to identify VPN traffic in the organization. The article advances an answer that utilizes secret-sharing plan which includes an enormous key administration upward utilizing public key framework (PKI) method. The paper expects to be simply the traffic coming to the framework is decoded and the information are accessible in plain structure for the framework to break down and recognize VPN traffic. This is accomplished by utilizing application layer intermediary which creates the duplicate of decoded traffic against every association which is then shipped off the framework for additional investigation. This procedure roughly copies the organization traffic and computational assets of existing framework while expanding the memory prerequisites to unscramble and yet again scramble the web traffic. Another arrangement that utilizes Deep Packet Inspection strategy [71] utilizes different sensors all through the organization to get the decoded traffic from the end has and send it back to grunt based IDS [72] to distinguish uncommon conduct in rush hour gridlock. It expands the general organization traffic in light of the fact that a sensor is to be introduced on each organization machine to have the option to distinguish any uncommon movement. Another method is to duplicate the whole association traffic and utilize pre-shared mystery to break down any malevolent traffic [74].

To recognize applications being run inside the organization, network examination is utilized widely. The work examined by He et al [75]. utilizes fundamental yet one of the best and involved strategies in network traffic examination for traffic grouping. In light of five-tuple association grouping, the procedure utilizes association qualities like parcel size, their inter-arrival time, and the course and request of the bundles to recognize the organization mark of any android application. The plan gives fundamental comprehension of traffic characterization. Be that as it may, network traffic produced by online VPN administrations will have no significant contrast or distinguishing qualities, different to a standard HTTPS association.

The utilization of decoded traffic to make due, investigate, and arrange encoded traffic is an interesting idea, examined by Niu et al [76]. The plans utilize named DNS-based dataset to recognize vindictive order and control traffic and mark the traffic as dubious or ordinary. The idea gives a remarkable forthcoming to dissect the organization traffic past five-tuple/current association procedure examined 2 Security and Communication Networks beforehand [75].

Our proposed framework investigates DNS records to distinguish vindictive or ill-conceived VPN server names. Association highlights are separated utilizing five-tuple approach. Five-tuple approach characterizes each new association by five credits recorded beneath:

(I) Source IP
(ii) Destination IP
(iii) Protocol (TCP/UDP)
(iv) Source port
(v) Destination port

DNS-based traffic investigation and association the board were finished utilizing five-tuple methods; our proposed framework goes above and beyond to break down HTTPS handshake. This is finished to confirm the server name utilized in the association with the DNS movement which the client has produced by his organization action. Utilizing this original methodology of dealing with an association by utilizing the action going before the current association, we can distinguish and recognize VPN traffic inside the organization.

V. VPN CLASSIFICATION:

A dataset comprising of TCP bundles caught utilizing the parcel examination apparatus Wireshark from an OpenVPN association was made and tried utilizing precisely the same Azure machine learning apparatuses. The outcomes for this showed that the organization was over fitting the issue as it was accomplishing 100% grouping exactness for both VPN traffic and non-VPN traffic. In outer approval tests, the organization was basically speculating, as it was grouping each example as having come from a VPN. To conquer this issue, it was speculated that a new dataset comprising of TCP stream records/measurements would be more fitting for examination. Stream measurements give an undeniable level perspective on network interchanges by detailing the addresses, ports and byte what's more bundle includes contained in those communications [42]. This information can be particularly important when organization traffic is being scrambled which can be the situation with VPN traffic. Wireshark shaped the premise of the bundle catch for this fresher dataset as was additionally the situation for the first dataset. The PC framework used to catch the traffic was an Ubuntu 16.04 put together virtual machine running with respect to a Windows 10 host. The organization association utilized in the analysis is a virtualised Intel PRO gigabit Ethernet card. Linux was utilized as it takes into account a better level of command over a portion of the inside frameworks included, for example, the systems administration stack. Utilizing a few inherent instruments, it is not difficult to robotize associations and detachments to various organizations and different organization interfaces. This was an especially supportive element when managing the catch of VPN based parcels. In ordinary activity, an association with a VPN begins with a common TCP "hello" arrangement and key trade. When the association is arrangement, it is just brought down at whatever point the client quits utilizing the VPN. The association is one long TCP association between the client's machine and the VPN server.

International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 6, Issue 12, ISSN No. 2455-2143, Pages 82-92
Published Online April 2022 in IJEAST (http://www.ijeast.com)

V.i. OpenVPN using Stunnel:

Stunnel is an open source, multiplatform application that is intended to add SSL/TLS encryption capacity to clients and servers that don't locally uphold the SSL/TLS conventions. While OpenVPN itself has support for SSL/TLS, procedures like Deep Packet Inspection (DPI) can possibly distinguish OpenVPN while utilizing SSL/TLs [47]. Stunnel can be used to conquer this and present the traffic to DPI structures as ordinary SSL web traffic running on port 443. This brought about whether or not a comparable strategy for characterization that was utilized to order OpenVPN traffic utilizing a brain organization could likewise be prepared to perceive OpenVPN traffic that was utilizing Stunnel. To utilize Stunnel, the client should introduce and arrange the application on both the OpenVPN server and on anything that OpenVPN client they are utilizing to associate with the VPN. On Linux this includes introducing the application by downloading the stunnel4 bundle, making also sharing another OpenSSL authentication between the client and the server, making and altering Stunnel config files and configuring the firewalls of both the server and client to permit the Stunnel traffic to be shipped.

V.ii. Dataset:

Similarly as with the past investigations, a dataset containing network traffic from Stunnel OpenVPN associations and non-VPN traffic is expected to prepare the brain organization. With the foundation previously finished with the arrangement of the OpenVPN server on AWS for the past analysis, this was generally basic. The Streisand VPN bundle additionally contained all things needed to arrangement Stunnel for use with OpenVPN, just requiring a couple of setup documents to be adjusted. Once the VPN was arrangement and the association stable, catch of the network traffic started involving a similar strategy as utilized for the OpenVPN information catch. Wireshark was utilized to catch network bundles; the VPN was set to separate and reconnect at regular intervals and programmed perusing script was utilized to create traffic from a similar determination of sites. When the bundles were caught, they were handled utilizing the TCP stream send out instrument NetMate all together to acquire stream measurements of the new information. The aftereffect of this information catch was a complete dataset of 3,952 examples, of which 1,931 were Stunnel OpenVPN and 2,021 were non-VPN. This dataset was then stacked into Weka.

V.iii. Feature Selection:

Feature selection was applied to the catch information to diminish the quantity of elements created by NetMate. Once more, a similar Weka strategy utilized for the OpenVPN try was utilized. This was the CorrelationAttributeEval model which was additionally working under a similar limit of 0.5. The subsequent highlights are shown in Table 1. The element determination for the Stunnel information seems, by all accounts, to be to a great extent unique to the elements chose for the first VPN dataset. A few ascribes make a return, like span, yet with an alternate relationship coefficient. A portion of the qualities chose this time have not been seen before which would appear to demonstrate that there is a distinction in how Stunnel alters the OpenVPN association.

Attribute Name	Correlation Coefficient
min_fpktl	0.992
duration	0.937
max_fpktl	0.913
max_idle	0.78
max_biat	0.763
std_idle	0.719
max_fiat	0.673
mean_idle	0.575
min_idle	0.562
mean_fpktl	0.561
mean_active	0.512
max_active	0.511
std_fpktl	0.506

Table 1: Correlation Coefficients for Stunnel attributes

Following similar advances utilized in the past analysis, the dataset was resampled into discrete preparation, testing and approval sets. The preparation set contains 3160 examples, the testing set contains 633 examples and the approval set contains 127 examples subsequent to resampling.

V.iv. Neural Network Setup:

For this test the objective was to look at how well the model created in the past trial could likewise play out something similar with network traffic from an alternate source. Hence, the brain network model utilized in the past try was reused with practically no alteration. Weka was taught to make a completely associated network with a covered up layer which aggregates together the quantity of traits with the number of classes and separation the outcome by 2. In this case there are 13 credits and 2 classes which brings about 15 partitioned by 2 which is 7.5. Weka adjusts down to the closest entire number so the quantity of stowed away hubs is set to 7. Once at this stage, the model is fit to be prepared utilizing the dataset. In the past investigation, the model was prepared, tried and approved utilizing three resampled sets of information. A similar technique was utilized for this model with extra tests being run utilizing 10-crease cross-approval and Leave One Out Cross Validation (LOOCV). On beginning testing utilizing these approval techniques, the outcomes accumulated showed that the model was getting ridiculously high exactness, conceivably giving indications of over fitting of the model to the issue. To cure this, the learning rate and afterward the force of the model were brought from 0.1 down to 0.01.

International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 6, Issue 12, ISSN No. 2455-2143, Pages 82-92
Published Online April 2022 in IJEAST (http://www.ijeast.com)

VI. RESULTS:

Table 2, Table 3 and Table 4 show the results of each validation method used once the neural network had been finally trained using the updated configuration. Table 5, Table 6 and Table 7 show the confusion matrices for each of the tests.

Correctly Classified Instances	98.4252%
Incorrectly Classified Instances	1.5748%
Average True Positive Rate	0.968
Average False Positive Rate	0.000
Average Precision	1.000
Average Recall	0.968
Average F-Measure	0.984

Table 2: 80/20 split Validation test results

Table 2 shows the outcomes accumulated from Weka for the test that utilized a 80/20 rate split on the dataset to make separate preparation, testing and approval sets. The outcomes shown are taken from the last approval set test, which utilizes information that was kept separate from the preparation and tuning of the model to mimic as close as conceivable this present reality execution of the model. The general precision of the model was demonstrated to be 98.42%.

Correctly Classified Instances	97.8998%
Incorrectly Classified Instances	2.1002%
Average True Positive Rate	0.969
Average False Positive Rate	0.012
Average Precision	0.987
Average Recall	0.969
Average F-Measure	0.978

Table 3: 10 fold Cross Validation test results

Table 3 shows the outcomes accumulated from the test that pre-owned 10-overlay cross approval to approve the model. For approval of this model the dataset was parted into 10 similarly measured subsamples or folds. Of these 10 subsamples, one is held as the approval information for testing of the model and the leftover 9 subsamples are utilized as preparing information. This cycle is then rehashed multiple times so every one of the folds is actually once as the approval information. These outcomes are then arrived at the midpoint of to give a solitary assessment of the exhibition of the model. The general exactness as shown by this approval is demonstrated to be 97.89%.

Correctly Classified Instances	97.8239%
Incorrectly Classified Instances	2.1761%
Average True Positive Rate	0.968
Average False Positive Rate	0.012
Average Precision	0.987
Average Recall	0.968
Average F-Measure	0.978

Table 4: Leave One Out CrossValidation test results

Table 4 shows the outcomes assembled from the test that pre-owned Leave One Out cross approval to approve the model. LOOCV includes a comparative interaction to 10-overlap Cross Validation where, rather than dividing the information into equivalent estimated folds, just a single example is held as the approval information, with the rest being utilized as preparing information. This cycle is rehashed however many times as there are tests in the dataset for example until each and every example has been utilized as the approval information once. The general exactness accomplished utilizing this approval strategy was viewed as 97.82%.

Table 5: Confusion Matrix for 80/20 split Validation test

Classified as	VPN	Normal
VPN	60	2
Normal	0	65

Table 5 shows the confusion matrix for the test that utilized a 80/20 rate split on the dataset. It shows 60 examples were accurately recognized as VPN, 65 examples were accurately distinguished as non-VPN and 2 were mistakenly recognized as non-VPN. Intriguing is the absence of tests that were mistakenly recognized as VPN.

Classified as	VPN	Normal
VPN	1872	59
Normal	24	1997

Table 6: Confusion Matrix for 10 fold Cross Validation test

Table 6 shows the confusion matrix for the test that pre-owned 10-overlay cross approval. It shows 1872 examples were accurately recognized as VPN, 1997 examples were accurately distinguished as non-VPN, 24 examples were mistakenly recognized as VPN and 59 examples were mistakenly recognized as non-VPN.

Classified as	VPN	Normal
VPN	1870	61
Normal	25	1996

Table 7: Confusion Matrix for Leave One Out Cross Validation test

Table 7 shows the confusion matrix for the test that involved LOOCV for approving the model. It shows 1870 examples were accurately distinguished as VPN, 1996 examples were accurately recognized as non-VPN, 25 examples were

International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 6, Issue 12, ISSN No. 2455-2143, Pages 82-92
Published Online April 2022 in IJEAST (http://www.ijeast.com)

erroneously recognized as VPN and 61 examples were mistakenly recognized as non-VPN.

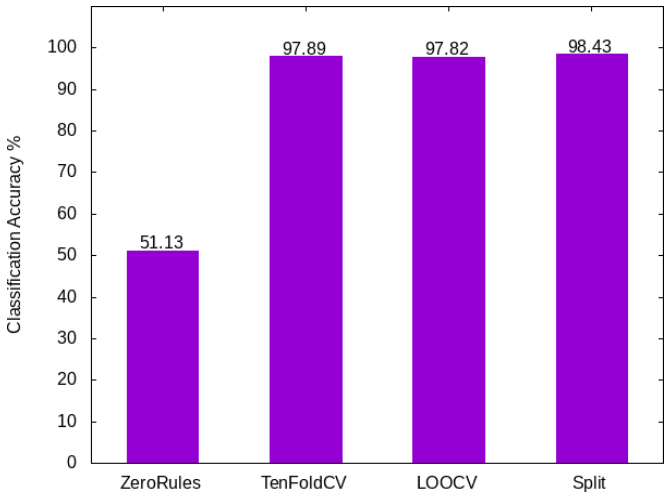

Figure 1: Graph comparing accuracies of different validation techniques against Zero Rules

The 80/20 split approval technique had the option to accomplish a precision pace of 98.43%. At first this would propose that the 80/20 preparation and test split gives the best model, on the grounds that the general number of tests in the approval set is nearly low, the outcomes may not be solid. This passes on the two kinds of cross approval to be contrasted with one another. 10-crease cross approval is one of the more famous types of cross approval and is generally utilized. LOOCV is basically cross approval where the quantity of folds that the information is sub-isolated into is equivalent to the absolute number of tests in the dataset, for this situation that would be 3952 folds. In the outcomes the general exactnesses of the two strategies are exceptionally near each other. In any case, LOOCV has a lot higher calculation time when contrasted with 10-overlay cross approval in spite of the singular crease calculation time being lower. Whenever 10-overlay approval is utilized the model just must be prepared and tried once for every one of the 10 overlap, the model for this situation should be prepared and tried multiple times while utilizing LOOCV. Since the consequences of the two approval strategies are so near each other, this implies the advantages of LOOCV are conceivably useless. Along these lines, assuming we take the consequence of the 10-overlay cross approval of 97.89% as the best pointer, one might say that the brain organization can precisely recognize an OpenVPN association utilizing Stunnel and typical non-VPN traffic. Notwithstanding, as seen with the past OpenVPN analyze, the disarray networks for all of the approval strategies utilized this time round show that the model is somewhat excessively tolerant, with a bigger number of misleading negatives than bogus up-sides. Figure 1 shows the general exactnesses of each test to a trial with practically no guidelines applied. The Zero Rules technique in Wek a shows what the outcomes would be in the occasion where everything is delegated one of the classes, for this situation that was the ordinary class. Contrasted with the zero standards result, the brain network performs quite well.

VII. CONCLUSION:

The aim was to investigate methods that would support the discovery of VPN advances that are being utilized to stow away an aggressor's character. While VPNs have genuine purposes, for example, interfacing with a business network from a far off area, they are as yet manhandled by lawbreakers who use them to carry out wrongdoings while staying undetected and unidentified. Without a strategy to distinguish when a VPN is interfacing with a web confronting server, organizations could be helpless against having their network penetrated and having information taken while being impeded in their capacity to certainly say who took it. This can be especially impeding to sites who manage client subtleties and monetary records. There are techniques accessible for assessing network traffic at the place of entrance and departure. An illustration of one of these strategies is Deep Packet Inspection (DPI). It is firmly connected with another strategy called Shallow Packet Inspection (SPI), but SPI just can examine the headers of organization bundles that are utilized to move the parcels to their objective. DPI goes above and beyond and reviews those headers and the real satisfied of the bundle, which on account of a HTTP parcel could be a solicitation for information from a site. A counter to DPI is the utilization of start to finish encryption on the substance of parcels to conceal those substance from inquisitive eyes. This is done honestly enough with the objective being to stop possible man in the center assaults from taking delicate information, for example, usernames and passwords or monetary subtleties as they are being sent. In any case, intermediary and VPN innovations additionally can utilize encryption advancements with the utilization of IPSec and SSL/TLS. This expands the requirement for a strategy to recognize these kinds of organization traffic. AI strategies are one manner by which to achieve this.

The examinations directed to arrange OpenVPN use observed that the Neural Network had the option to accurately recognize the VPN traffic with a general exactness of 93.71%. The further work done to order Stunnel OpenVPN utilization observed that the Neural Network had the option to accurately distinguish VPN traffic with a general precision of 97.82% exactness while utilizing 10-overlay cross approval. This last examination likewise gave a perception of 3 distinct approval methods and the different precision results acquired. Upon fruitful investigations led for the location of Anonymising Proxy traffic, the center was reached out to incorporate VPN traffic. The VPN innovation OpenVPN was picked as the concentration for the trials, which thus observed that the Neural Network was equipped for grouping network traffic as either VPN traffic or as non-VPN traffic. This prompted a

further arrangement of examinations which endeavored to group a type of OpenVPN traffic that utilized Stunnel to give encryption. These observed that a Neural Network prepared on the Stunnel OpenVPN information could characterize network traffic as either VPN traffic or non-VPN traffic. Once more, the tests were directed in, for example, design as to wipe out inclination where conceivable. This included keeping a part of the caught dataset away from the preparation and tuning stages, so it very well may be utilized to mimic true information that the model had never seen.

VIII. REFERENCES:

[1]. Miller, S., Curran, K., Lunney, T. (2018) Multilayer Perceptron Neural Network for Detection of Encrypted VPNNetwork Traffic IEEE International Conference on Cyber Situational Awareness, Data Analytics and Assessment (Cyber SA 2018), 11-12 June 2018, Scotland, UK

[2]. Miller, S., Curran, K. and Lunney, T. (2018) 'Detection of Anonymising Proxies using Machine Learning', Special issue on Machine Learning for Cyber Security in Journal of Information Science (MDPI), ISSN 2078-2489, (Accepted) 2019

[3]. Geetha, S. and Phamila, A. V. (2016) Combating Security Breaches and Criminal Activity in the Digital Sphere. First. IGI Global. doi: 10.4018/978-1-5225-0193-0.

[4]. Peterson, A. (2014) 'The Sony Pictures hack, explained.', Washington Post, 18 December. Available at: https://www.washingtonpost.com/news/the-switch/wp/2014/12/18/the-sony-pictures-hack-explained/.

[5]. Pagliery, J. (2014) What caused Sony hack: What we know now, CNN. Available at: http://money.cnn.com/2014/12/24/technology/security/sony- hack-facts/ (Accessed: 6 December 2017).

[6]. Hunt, T. (2016) Observations and thoughts on the LinkedIn data breach, troyhunt.com. Available at: https://www.troyhunt.com/observations-and-thoughts-on- the-linkedin-data-breach/ (Accessed: 6 December 2017).

[7]. Chaum, D. L. (1981) 'Untraceable electronic mail, return addresses, and digital pseudonyms', Communications of the ACM. ACM, 24(2), pp. 84–90. doi: 10.1145/358549.358563.

[8]. Yang, M. (2015) 'De-anonymizing and countermeasures in anonymous communication networks', IEEE Communications Magazine, 53(4), pp. 60–66. doi: 10.1109/MCOM.2015.7081076.

[9]. Wood, D.. (1988) 'Virtual private networks', in 1988 International Conference on Private Switching Systems andNetworks, New York, USA, pp. 132–136.

[10]. Zorn, G. (1999) 'Point-to-Point Tunneling Protocol (PPTP)', RFC 2637, pp. 1–57. Available at: https://tools.ietf.org/html/rfc2637 (Accessed: 12 January2018).

[11]. Rawat, V. et al. (2001) Layer Two Tunneling Protocol {(L2TP)} over Frame Relay. doi: 10.17487/RFC3070.

[12]. Lawas, J. B. R., Vivero, A. C. and Sharma, A. (2016) 'Network performance evaluation of VPN protocols (SSTP and IKEv2)', in 2016 Thirteenth International Conference on Wireless and Optical Communications Networks (WOCN). IEEE, pp. 1–5. doi: 10.1109/WOCN.2016.7759880.

[13]. Thomas, K. et al. (2011) 'Design and evaluation of a real-time URL spam filtering service', in Proceedings - IEEE Symposium on Security and Privacy, pp. 447–462.

[14]. Cisco (2006) Access Control Lists: Overview and Guidelines. Available at: http://www.cisco.com/c/en/us/td/docs/ios/12_2/security/conf iguration/guide/fsecur_c/scfacls.html.

[15]. Dharmapurikar, S. et al. (2003) 'Deep packet inspection using parallel Bloom filters', IEEE Micro. IEEE Comput. Soc, pp. 52–61. doi: 10.1109/CONECT.2003.1231477.

[16]. Yu, F., Chen, Z., Diao, Y., Lakshman, T., Katz, R. (2006) 'Fast and memory-efficient regular expression matching for deep packet inspection', 2006 Symposium on Architecture For Networking And Communications Systems. New York, New York, USA: ACM Press, pp. 1–10. doi: 10.1145/1185347.1185360.

[17]. Sherry, J. et al. (2015) 'BlindBox', ACM SIGCOMM Computer Communication Review. ACM, 45(5), pp. 213–226. doi: 10.1145/2829988.2787502.

[18]. Dainotti, A., Pescape, A. and Claffy, K. (2012) 'Issues and future directions in traffic classification', IEEE Network, 26(1), pp. 35–40. doi: 10.1109/MNET.2012.6135854.

[19]. Miller, S., Curran, K. and Lunney, T. (2015) 'Traffic Classification for the Detection of Anonymous Web Proxy Routing', IJISR, 5(1), pp. 538–545. doi: 10.20533/ijisr.2042.4639.2015.0061.

[20]. Miller, S., Curran, K. and Lunney, T. (2016) 'Cloud-based machine learning for the detection of anonymous web proxies', in 2016 27th Irish Signals and Systems Conference, ISSC 2016. IEEE, pp. 1–6. doi: 10.1109/ISSC.2016.7528443.

[21]. García-Teodoro, P. et al. (2009) 'Anomaly-based network intrusion detection: Techniques, systems and challenges', Computers & Security, 28(1), pp. 18–28. doi:10.1016/j.cose.2008.08.003.

[22]. Hawkes-Robinson, W. (2002) 'SANS Institute - Microsoft PPTP VPN Vulnerabilities - Exploits in Action'. SANS Institute. Available at: https://www.researchgate.net/publication/235927650_SANS_Institute_-_Microsoft_PPTP_VPN_Vulnerabilities_-_Exploits_in_Action (Accessed: 19 January 2018).

[23]. Schneier, B. and Mudge (1998) 'Cryptanalysis of Microsoft's point-to-point tunneling protocol (PPTP)', 5th ACM Conference on Computer and Communications Security, pp. 132–141. doi: 10.1145/288090.288119.

[24]. Farinacci, D. et al. (1994) 'Generic Routing Encapsulation over IPv4 networks', RFC1702, pp. 1–4. Available at: https://tools.ietf.org/html/rfc1702 (Accessed:17 January 2018).

[25]. Simpson, W. (1996) 'PPP CHAP'. Network Working Group. Available at: https://tools.ietf.org/rfc/rfc1994.txt (Accessed: 19 January 2018).

[26]. Schmidt, J. (2012) A death blow for PPTP - The H Security: News and Features. Available at: http://www.h- online.com/security/features/A-death-blow-for-PPTP- 1716768.html (Accessed: 19 January 2018).

[27]. Microsoft (2012) Microsoft Security Advisory 2743314 | Microsoft Docs, Microsoft Security Advisory. Available at: https://docs.microsoft.com/en-us/security-updates/SecurityAdvisories/2012/2743314 (Accessed: 12 January 2018).

[28]. Kazemi, K. and Fanian, A. (2015) 'Tunneling protocols identification using light packet inspection', in 2015 12th International Iranian Society of Cryptology Conference on Information Security and Cryptology (ISCISC). IEEE, pp. 110–115. doi: 10.1109/ISCISC.2015.7387907.

[29]. Kent, S. and K. Seo, "Security Architecture for the Internet Protocol", RFC 4301, DOI 10.17487/RFC4301, December 2005, <https://www.rfc-editor.org/info/rfc4301>..

[30]. Feilner, M. (2006) Open VPN : building and operating virtual private networks. Packt. Publishing, ISBN: 190481185X, 2006.

[31]. Pohl, F. and Schotten, H. D. (2017) 'Secure and Scalable Remote Access Tunnels for the IIoT: An Assessment of openVPN and IPsec Performance', in, pp. 83–90. doi: 10.1007/978-3-319-67262-5_7.

[32]. Schneier, B. (1994) 'Description of a new variable-length key, 64-bit block cipher (Blowfish)', in. Springer, Berlin, Heidelberg, pp. 191–204. doi: 10.1007/3-540-58108-1_24.

[33]. Scarfone, K. and Mell, P. (2007) 'Guide to intrusion detection and prevention systems (idps)', NIST special publication, 800(2007), p. 94.

[34]. Lin, W.-C., Ke, S.-W. and Tsai, C.-F. (2015) 'CANN: An intrusion detection system based on combining cluster centers and nearest neighbors', Knowledge-Based Systems, 78, pp. 13–21. doi: 10.1016/j.knosys.2015.01.009.

[35]. Xiang, C., Yong, P. C. and Meng, L. S. (2008) 'Design of multiple-level hybrid classifier for intrusion detection system using Bayesian clustering and decision trees', Pattern Recognition Letters, 29(7), pp. 918–924. doi: 10.1016/j.patrec.2008.01.008.

[36]. Khan, L., Awad, M. and Thuraisingham, B. (2006) 'A new intrusion detection system using support vector machines and hierarchical clustering', The VLDB Journal, 16(4), pp. 507–521. doi: 10.1007/s00778-006-0002-5.

[37]. Özyer, T., Alhajj, R. and Barker, K. (2007) 'Intrusion detection by integrating boosting genetic fuzzy classifier and data mining criteria for rule pre-screening', Journal of Network and Computer Applications, 30(1), pp. 99–113. doi: 10.1016/j.jnca.2005.06.002.

[38]. Ghosh, A. K., Schwartzbard, A. and Schatz, M. (1999) 'Learning Program Behavior Profiles for Intrusion Detection.', in Workshop on Intrusion Detection and Network Monitoring.

[39]. Samuel, A. L. (1959) 'Some Studies in Machine Learning Using the Game of Checkers', IBM Journal of Research and Development, 3(3), pp. 210–229. doi: 10.1147/rd.33.0210.

[40]. Khriplovich, I. B. and Pomeransky, A. A. (1998) 'Equations of Motion of Spinning Relativistic Particle in Electromagnetic and Gravitational Fields', EUA: Prentice Hall, 178, p. 640. doi: 10.1080/01422419908228843.

[41]. Russel, S. J. and Norvig, P. (2010) Artificial intelligence: a modern approach. Third Edit, EUA: Prentice Hall. Third Edit. doi: 10.1017/S0269888900007724.

[42]. Cisco (2018) Encrypted Traffic Analytics. Available at: https://www.cisco.com/c/dam/en/us/solutions/collateral/en te rprise-networks/enterprise-network-security/nb-09-encrytd- traf-anlytcs-wp-cte-en.pdf (Accessed: 12 January 2018).

[43]. Liu, C., White, R. W. and Dumais, S. (2010) 'Understanding web browsing behaviors through Weibull analysis of dwell time', in Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval - SIGIR '10. New York, New York, USA: ACM Press, p. 379. doi: 10.1145/1835449.1835513.

[44]. Arndt, D. (2011) NetMate-flowcalc. Available at: https://dan.arndt.ca/projects/netmate-flowcalc/ (Accessed: 4October 2018).

[45]. Stibler, S., Brownlee, N. and Ruth, G. (1999) 'RTFM: New Attributes for Traffic Flow Measurement', pp. 1–18.doi: 10.17487/RFC2724.

[46]. Frank, E., Hall, M. A. and Witten, I. H. (2016) 'The WEKA Workbench Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques" Morgan Kaufmann, Fourth Edition, 2016', Morgan Kaufmann, Fourth Edition. Available at: https://www.cs.waikato.ac.nz/ml/weka/Witten_et_al_2016_appendix.pdf (Accessed: 8 November 2017).

[47]. Deri, L., Martinelli, M., Cardigliano, A. (2014) 'nDPI: Open-source high-speed deep packet inspection', in 2014 International Wireless Communications and Mobile Computing Conference (IWCMC). IEEE, pp. 617–622. doi:10.1109/IWCMC.2014.6906427.

[48]. B. Harris and R. Hunt, "Tcp/ip security threats and attack methods," Computer Communications, vol. 22, no. 10, pp. 885–897, 1999.

[49]. X. Li, M. Wang, H. Wang, Y. Ye, and C. Qian, "Toward secure and efficient communication for the internet of things," IEEE/ ACM Transactions on Networking, vol. 27, no. 2, pp. 621–634, 2019.

[50]. E. Rescorla, SSL and TLS: Designing and Building Secure Systems, vol. 1, Addison-Wesley, Boston, MA, USA, 2001.

[51]. A. P. Felt, R. Barnes, A. King, C. Palmer, C. Bentzel, and P. Tabriz, "Measuring HTTPS adoption on the web," in Proceedings of the 26th USENIX Security Symposium (USENIX Security 17), pp. 1323–1338, USENIX Association, Vancouver, BC, Canada, August 2017.

[52]. J. Clark and P. C. Van Oorschot, "SoK: SSL and HTTPS: revisiting past challenges and evaluating certificate trust model enhancements," in Proceedings of the 2013 IEEE Symposium on Security and Privacy, pp. 511–525, IEEE, Berkeley, CA, USA, May 2013.

[53]. C. Paya and O. Dubrovsky, "Inspecting encrypted communications with end-to-end integrity," US Patent 7562211, 2009.

[54]. V. Lifliand and A. Michael Ben-Menahem, "Encrypted network traffic interception and inspection," US Patent 8578486, 2013.

[55]. N. Leavitt, "Anonymization technology takes a high profile," Computer, vol. 42, no. 11, pp. 15–18, 2009.

[56]. Z. Zhang, S. Chandel, J. Sun, S. Yan, Y. Yu, and J. Zang, "VPN: a boon or trap?: a comparative study of MPLs, IPSec, and SSL virtual private networks," in Proceedings of the 2018 2nd International Conference on Computing Methodologies and Communication (ICCMC), pp. 510–515, IEEE, Erode, India, February 2018.

[57]. K. Karuna Jyothi and B. I. Reddy, "Study on virtual private network (VPN), VPN's protocols and security," International Table 3: Alerts generated for the user activity. User details Alerts classification (connection based) Total Legitimate activity IP-based VPN DNS-based VPN NO DNS User 1 178 59 4 109 6 User 2 85 50 0 35 0 User 3 250 114 0 135 1 User 4 71 24 2 41 4 User 5 145 82 0 63 0 Table 2: Forensic analysis of freely available VPN services. VPN services Classifiers for forensic analysis IP Host name Nonstandard HTTPS DNS activity TOR browser ✓ × ✓ ✓ Hotspot Shield free × ✓ ✓ ✓ Browsec VPN × ✓ × × ZenMate VPN × ✓ × × Hoxx VPN × ✓ × × 16 Security and Communication Networks Journal of Scientific Research in Computer Science, Engineering and Information Technology, vol. 3, no. 5, 2018.

[58]. D. Roger, N. Mathewson, and S. Paul, "TOR: the secondgeneration onion router," Technical report, Naval Research Laboratory, Washington, DC, USA, 2004.

[59]. A. Yamada, Y. Miyake, K. Takemori, A. Studer, and A. Perrig, "Intrusion detection for encrypted web accesses," in Proceedings of the 21st International Conference on Advanced Information Networking and Applications Workshops (AINAW'07), vol. 1, pp. 569–576, Niagara Falls, Ont., Canada, May 2007.

[60]. M. Ikram, N. Vallina-Rodriguez, S. Seneviratne, M. A. Kaafar, and V. Paxson, "An analysis of the privacy and security risks of android VPN permission-enabled apps," in Proceedings of the 2016 Internet Measurement Conference, pp. 349–364, ACM, Santa Monica, CA, USA, November 2016.

[61]. S. Sudin, R. B. Ahmad, and S. Z. Syed Idrus, "A model of virus infection dynamics in mobile personal area network," Journal of Telecommunication, Electronic and Computer Engineering (JTEC), vol. 10, no. 2–4, pp. 197–201, 2018.

[62]. N. Weaver, C. Kreibich, M. Dam, and V. Paxson, "Here be web proxies," in Proceedings of the International Conference on Passive and Active Network Measurement, pp. 183–192, Springer, Los Angeles, CA, USA, March 2014.

[63]. C. Reis, S. D. Gribble, T. Kohno, and N. C. Weaver, "Detecting in-flight page changes with web tripwires," in Proceedings of the 5th USENIX Symposium on Networked Systems Design and Implementation, vol. 8, pp. 31–44, San Francisco, CA, USA, April 2008.

[64]. N. Vallina-Rodriguez, S. Sundaresan, C. Kreibich, and V. Paxson, "Header enrichment or ISP enrichment?: emerging privacy threats in mobile networks," in Proceedings of the 2015 ACM SIGCOMM Workshop on Hot Topics in Middleboxes and Network Function Virtualization, pp. 25–30, ACM, London, UK, August 2015.

[65]. N. Weaver, C. Kreibich, and V. Paxson, "Redirecting DNS for ads and profit," in Proceedings of the UNISEX Workshop on Free and Open Communications on the Internet 2011, vol. 2, no. 2–3, San Francisco, CA, USA, August 2011.

[66]. N. Vallina-Rodriguez, J. Amann, C. Kreibich, N. Weaver, and V. Paxson, "A tangled mass: the android root certificate stores," in Proceedings of the 10th ACM International on Conference on Emerging Networking Experiments and Technologies, pp. 141–148, ACM, Sydney, Australia, December 2014.

[67]. Y. Song and U. Hengartner, "Privacyguard: a VPN-based platform to detect information leakage on android devices," in Proceedings of the 5th Annual ACM CCS

Workshop on Security and Privacy in Smartphones and Mobile Devices, pp. 15–26, ACM, Denver, CO, USA, October 2015.

[68]. S. Fahl, M. Harbach, T. Muders, L. Baumgartner, ¨ B. Freisleben, and M. Smith, "Why eve and mallory love android: an analysis of android ssl (in) security," in Proceedings of the 2012 ACM Conference on Computer and Communications Security, pp. 50–61, ACM, Raleigh, NC, USA, October 2012.

[69]. V. T. Goh, J. Zimmermann, and M. Looi, "Towards intrusion detection for encrypted networks," in Proceedings of the 2009 International Conference on Availability, Reliability and Security, pp. 540–545, IEEE, Fukuoka, Japan, March 2009.

[70]. A. A. Abimbola, J. M. Munoz, and W. J. Buchanan, "Nethostsensor: investigating the capture of end-to-end encrypted intrusive data," Computers & Security, vol. 25, no. 6, pp. 445–451, 2006.

[71]. R.. Martin, "Snort—lightweight intrusion detection for networks," in Proceedings of the 13th USENIX Conference on System Administration, LISA '99, pp. 229–238, USENIX Association, Seattle, WA, USA, November 1999.

[72]. X. Li, S. G. Karanvir, G. H. Cooper, and J. R. G., "Encrypted data inspection in a network environment," US Patent 9176838, 2013.

[73]. G. He, B. Xu, and H. Zhu, "AppFA: a novel approach to detect malicious android applications on the network," Security and Communication Networks, vol. 2018, Article ID 2854728, 15 pages, 2018.

[74]. W. Niu, X. Zhang, G. W. Yang, J. Zhu, and Z. Ren, "Identifying APT malware domain based on mobile DNS logging," Mathematical Problems in Engineering, vol. 2017, Article ID 4916953, 9 pages, 2017.

[75]. A. Nath, Packet Analysis with Wireshark, Packt Publishing Ltd., Birmingham, UK, 2015.

[76]. Netresec, Network miner.

[77]. AnchorFree, Hoptspot Shield VPN.