*International Journal of Engineering & Technology*, 7 (3.20) (2018) 1-5

International Journal of Engineering & Technology

*Website: www.sciencepubco.com/index.php/IJET*

***Research paper***

# DNS Tunneling: a Review on Features

**Mahmoud Sammour [1], Burairah Hussin [2] , Mohd Fairuz Iskandar Othman [3] , Mohamed Doheir [4]**
**Basel AlShaikhdeeb [5] , Mohammed Saad Talib [6]**

[1,2,3,4,6] *Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Malaysia*
[5] *Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, 43600, Bangi, Selangor, Malaysi*a
*Corresponding Author Email:* mahmoud.samour@gmail.com

**Abstract**

One of the significant threats that faces the web nowadays is the DNS tunneling which is an attack that exploit the domain name protocol in order to bypass security gateways. This would lead to lose critical information which is a disastrous situation for many organizations. Recently, researchers have pay more attention in the machine learning techniques regarding the process of DNS tunneling. Machine learning is significantly impacted by the utilized features. However, the lack of benchmarking standard dataset for DNS tunneling, researchers have captured the features of DNS tunneling using different techniques. This paper aims to present a review on the features used for the DNS tunneling.

**Keywords:** *DNS tunneling, payload analysis, traffic analysis, feature extraction*

## 1. Introduction

The World Wide Web (WWW) is playing a crucial role in the daily lies of individuals and corporations. It provides the paradigm of accessing, surfing and searching numerous things in the education, business and entertaining domains [1]. In specific, business nowadays is mainly relying on the internet not only by making profit but also to restore data, accommodate transactions, conducting communications and other activities. Even though, this would facilitate the business procedures however, there are many potential threats could be emerged which would have a disastrous impact on the corporations and the organizations [2]. One of these threats is the DNS tunneling attacks which exploit the trivial nature of DNS protocol [3]. DNS protocol aims to provide an easy access for servers via domain's name rather than the IP address. In this manner, DNS protocol can be used to make a tunnel in order to conduct several types of attacks such as stealing critical information, fully access to the operating system particularly file transfer system and obstructing the operations of the business [4].
Domain Name System (DNS) is one of the significant protocol that has been used for the web browsing and emailing purposes [5]. The key characteristic behind the DNS protocol lies on its capability to provide accessible mechanism through the names rather than the IP address. For example, it is easy to access the Google server via the domain name 'google.com' instead of recalling the IP address '8.8.8.8' or '8.8.4.4'. Due to DNS protocols was not designated for transferring data, DNS protocol is more vulnerable for attacks or malicious communications. The majority of large corporations are possibly subjected to serious attacks vis DNS protocol [6] . This because most of the organizations are paying little or no attention for DNS but rather they are more concerned to monitor the traffic within the web or email which is commonly undergo attacks.

Several available tools for tunneling over DNS has been depicted in the internet, the common purpose of such tools is to provide a free access to the internet using WiFi networks. However, such tools would be utilized for serious attacks [1], [2]. For instance, the DNS tunneling could be used for a full remote control via a compromised internet host in which the capabilities of the operating system can be fully controlled either by file transfer or a full IP tunnel [7]. In addition, DNS tunneling can be used as a malware such as the Feederbot [8] and Moto [9] which are known to use DNS as a communication method.

Several techniques have been proposed for detecting the DNS tunneling, such techniques have been inspired by traditional methods [10]. Firewalls are one of the traditional methods that have been utilized to detect the DNS tunneling. Generally, firewalls aim at generating rules that prompt IP spoofing by denying DNS queries that came from an IP address that is not listed in a predefined numbers space. This has the ability to prevent the name resolver to be used in possible threat. On the other hand, Intrusion Detection Systems (IDS) have been used also for identifying DNS tunneling by inspecting the DNS requests from an unauthorized client. In addition, Traffic Analyzers (TA) has also been used for detecting DNS tunneling by its exploiting its ability to observe the traffic without causing any impacts on the traffic flow within the network. Finally, Passive DNS Replication method has been used for the process of detecting DNS tunneling too. This can be represented by exploiting the ability of replicating every DNS query in order to prevent the use of such queries by a domain generation algorithm.

However, the latter methods have a major drawback that can be shown by the restricted capability of detecting the DNS tunneling. Handling DNS tunneling that is unknown using such approached will not be enough. Researcher of late seem to use the machine learning techniques with its ability of coming up with a statistical model that is able to detect the tunneling of the DNS from the previous experience. Yet, machine learning techniques are significantly impacted by the utilized features. Features can be defined as discriminative characteristics that facilitate the classifier to learn and identify regularities [11]. With the lack of available benchmark dataset for DNS tunneling, researchers tend to capture the features by simulating a network traffic. Hence, there are variety of features that have been utilized in the literature. This paper aims to review the common features utilized for the process of detecting DNS tunneling, as well as, the new trends that have been presented in the latest related work.

Generally, there are two main classes of DNS tunneling features which are *payload analysis* and *traffic analysis*. The payload analysis technique basically aims at monitoring and observing the DNS payload with respect to the requests and responses that use the particular indicators. On the other hand, traffic analysis is focused on concentrating on the traffic over time by the analysis of some statistical features such as the frequency as well as the count. The following sections will discuss every category and its sub-category mechanism individually.

## 2. Payload Analysis

This analysis techniques focus at identifying particular features of the payload, and this requires monitoring the capacity of the requests as well as the responses plus other indicators. Also, a light will be shed on a good and topic that is relevant that is the Domain Generation Algorithms (DGA). DGA offer the domain similarly suspicious to the names that are encoded. Basically, there are several features that could be categorized under the payload analysis. These features will be illustrated in the following sub-sections.

### A. Size of request and response

In terms of finding out the suspicious DNA tunneling traffic analyzing of the size of request as well as the response is one of the responses that have been in discussion. The main feature of this technique is based on computing the ratio of the source as well the destination bytes [14]. Such a ratio will then be compared to a particular limit value so as to find out if the DNS appears suspicious or normal.

Researchers like Skoudis [1] have argued that the length of the request along with the response is a good sign to find out if the DNS is suspicious or not suspicious. Majority of the DNS tools for tunneling try to put data as much as they can in the request and response, this it is possible the tunneling can have labels that are longer.

### B. Entropy of hostnames

This is an approach that checks in terms of the entropy, here the basic DNS names seem to contain a dictionary or words that have meaning [12]. Ironically, the encoded DNS names have a greater entropy by the use of extensive characters. There are numerous exceptions in this scenario, however, measuring the entropy is viewed as one of the important factors that can indicate the activity of DNS tunneling.

When examining the traffic features, the measurements could be conducted at different levels of abstraction including IP packet level, transport level or application level. In addition, a distinguishing could be made between the protocol client requests and server responses. In fact, both HTTP and FTP protocols contain relatively small vocabulary of commands and content in terms of the request. However, the responses of both protocols may contain large amounts of data with great variation. In this manner, analysing the requests of these protocols would be better in terms of predicting and estimating the content and variation. For this purpose, information entropy has been used in order to address the variation of the components that make up a message. Homem et al. [13] have calculated the information entropy based on the bytes that make up a packet layer or field value. In this vein, the entropy is being computed as the probability of a specific byte occurrence $p(x_i)$ multiplied by the logarithm of the probability of that occurrence, this would be summed up for all byte occurrence as in the following equation:

$$H(X) = -\sum_{i=1}^{n} p(x_i) \times log\ p(x_i) \tag{1}$$

The main idea of computing the entropy lies on the entropy distribution that would be generated for each protocol flow. In this manner, such entropy can be analysed in order to identify the trends of such distribution.

### C. Statistical analysis

Another signal of finding out the DNS tunneling is observing the frequency of the occurrence of the characters that are specific [3]. It is obvious that the normal DNS names that seem to have lesser digits unlike the names that are encoded where the digits are most likely to occur. Thus, checking the degree of occurrence of the digit in the name of DNS would be a vital element[14]. Plus, analysis of the length of the longest substring that is meaningful can another factor that will be important. Ultimately, watching the statistics of how the unique characters occur will facilitate the finding of the tunneling. For example, a DNS name that has more than 27 characters that are unique will be suspicious.

In this vein, analyzing of the DNS names with respect of the common language in which the consonants and the numbers rarely occur. A tool of DNS tunneling would possibly use consonants that are repeated so as to encode the name of the DNS.

### D. Uncommon record types

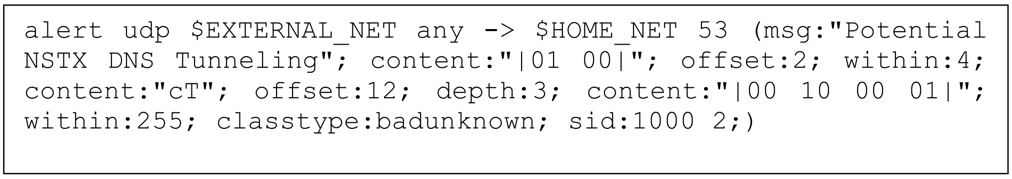
Analysis of the record types that are uncommon is another signal that can be used in detecting the DNS tunneling. There are some types of record types that are rarely made use of a regular customer like the TXT record[15]. Thus, being able to identify the type of record will be an important factor to check for in the tunneling of DNS.

### E. Policy violation

At times there will be a particular policy that is needed by all DNS like it needs all DNS to pass through a DNA server which is internal. Therefore, being in vi0lation of this policy is a good signal of DNS tunneling occurring [17]. It is vital to note that majority of DNS tunneling utilities are planned in way that they can work even as they forward request through a DNS server that is internal.

### F. Specific signatures

Other authors have suggested using particular signatures of the DNS tunneling names. Majority of the tools of DNS result in DNS names that have particular signatures or some particular patterns. This particular patterns can be located in the DNS header and content that is found in the payload. For example, there is an approach that has been proposed for the purpose of detecting tunneling that relies on the signatures, it is known as the snort signature[16]. Figure 2.6 shows an example of detected DNS tunneling by the Snort signature.



```
alert udp $EXTERNAL_NET any -> $HOME_NET 53 (msg:"Potential
NSTX DNS Tunneling"; content:"|01 00|"; offset:2; within:4;
content:"cT"; offset:12; depth:3; content:"|00 10 00 01|";
within:255; classtype:badunknown; sid:1000 2;)
```

**Figure** 1: Example of detected DNS tunneling by Snort signature

As shown in the example depicted by Figure 2.6, there is multiple correspondences content regarding the signature rule. First content which is '01 00' contains the ID of the DNS header that is represented by the first two bytes, while the other two bytes represent multiple header attributes. In particular, the bits from 0 to 6 are being represented via 0, whereas the 7 bit is being represented by 1. This reveals that the first content is matching a standard DNS query in which the bit '1' refers to the demand of recursion process. The second content which is 'ct' indicate a request for a QNAME field as part of the first 3 bytes of the domain name. The final content which contains hexadecimal values as '00 10 00 01' within the first 255 bytes of the payload. This reveals that this content is matching QTYPE and QCLASS which correspond to a 'TXT' request. To sum up, this signature is seeking a regular DNS query for a record type of 'TXT' but with a text 'ct' in the beginning of the domain name [17].

## 3. Traffic Analysis

Analysis of traffic aims on focusing the two requests and responding in a manner of time in which the amount and the frequency of this elements can be utilized as a signal for tunneling. Similar to the payload analysis, there are several features that are related to the network traffic. These features will be illustrated in the following sub-sections.

### G. Volume of DNS traffic per IP address

After discussing the among of traffic, the first thing that was ready for grabs first was analyzing the DNS traffic amount which is produced by a specific IP address[15, 16]. The number of requests that are needed in conducting the tunneling as the data of tunneling is restricted is what makes this type of analysis to be feasible. The tunneling is restricted to 512 bytes per each request, more so there are cases that the client can poll the server which then leads to many requests being sent.

### H. Volume of DNS traffic per domain

This is the same as the first method, it focuses on the quantity of the DNS traffic although the main distinguishing feature is that it focuses on the traffic of the DNS which is generated by a specific domain name[12]. This technique originates from the fact that majority if the tunneling tools of DNS are adjust the tunneling of data while utilizing a particular domain name. Thus, all the traffic that is tunneled will be limited to a domain name. Nevertheless, some DNS tunneling tools are capable of adjusting the data if the tunnel by use of various domain names so as to reduce the quantity of the traffic in each domain.

### I. Number of hostname per domain

Analyzing the amount of hostname for a specific domain is another method for analyzing the traffic. This is because majority of the tunneling tools for DNA usually request particular hostname for each request, thus the number of requests for such a particular hostname becomes bigger than the normal DNS name of the domain. This method can be considered as an example of the technique for traffic analysis.

### J. Geographic location of DNS server

The location of the DNS geographically is also another indicator which is linked to the analysis of traffic mechanism. A large quantity of the DNS traffic is aimed at targeting the particular locations that are irrelevant which can be a signal of the activities of tunneling [18]. For the organizations that lack branches across the globe, the method does not appear to be feasible.

### K. Domain history

Taking a look at the history of domain of another method that is associate to the analysis of traffic, domain history has been used in detecting the activities that are malicious. This technique has originated from the idea that some domain names are intended to fulfill particular record types. Therefore, detecting a request for the various techniques in the history of the domain will be a signal for the tunneling activities.

### L. Volume of NXDomain responses

This is another traffic analysis technique that analyzes the NXDomain responses that are excessive. The method is good for identification of the Heyoke DNS utility that results in big amounts of NXDomain responses [19].

### M. Visualization

The visualization of the DNS traffic would be a technique that is useful with regards of identifying the activities of tunneling [20]. In this way, the traffic is analyzed by professionals whereby the traffic that is tunneled will be greatly emerged.

### N. Orphan DNS server

This technique will be centered on things that can be expected and not the things that can be seen. For example, a web page request is normally linked with the application of https, therefore, identifying the DNS orphan requests which lack a request that is corresponding with another program that will be a useful method in detecting tunneling [23]. Even though there exist some exceptions such as the security devices may accommodate the reverse lookups on the addresses, that is the IP addresses. Nonetheless, these can be easily sorted out.

### O. General covert channel detection

This is an analysis method that converts the detection channel in which a channel that is hidden which is an independent protocol will be utilized in detecting the tunneling. The main function that is behind this method is based on examining the time or day of the traffic and associating it against a fingerprint that is statistical [24].

## 4. State of the Art

Numerous research studies have addressed the task of feature extraction for DNS tunneling. However, one of the significant approaches is the one that introduced by [21]. The authors have presented an approach for profiling DNS tunneling using feature transformation method. The proposed method is intended to utilize feature extraction mechanism by exploiting specific features such as time and size of the DNS messages. The authors have used a simulated network traffic with a tunneling tool (i.e. dns2tcp) in order to capture these features. Based on the statistics collected from the network traffic, a Principle Component Analysis is being adopted to synthesize the features. As well as, the Mutual Information is also being utilized to identify discriminated separations between regular and normal connections. Results showed that the proposed method has the ability to profile some malicious scenarios but there are still some cases that need to be investigated.

Dusi et al. [22] have proposed a fingerprint statistical approach that exploit the machine learning technique in order to detect tunneling activities. The authors have utilized statistical features related to traffic aspect such as regular remote interactive logins or secure copying activities. Using a rule-based classification method, the authors have demonstrated the effectiveness of their proposed method regarding the detection of tunneling.

Similarly, Dusi et al. [10] have proposed a machine learning technique that is intended to identify the tunnels within the application-layer. The proposed technique aims to exploit specific features that discriminate the legitimate and tunneling behavior. Such features are being constructed statistically from the TCP session and it is related to the traffic including packet size and the inter-arrival time. Consequentially, a Decision Tree classifier is being used to train of the latter features. After the training, the classifier has been tested in a set of DNS queries to evaluate its performance regarding the process of detecting tunneling. The proposed method has showed superior performance compared with the traditional techniques.

For instance, Allard et al. [23] suggested a solution that is based on the machine learning mechanisms whereby two types of classifiers have been utilized that is inclusive of the Decision Tree as well as the Random Forest so as to find out the DNS tunneling. The suggested classifiers have been trained on the flows that have been ciphered by being able to accommodate an analysis of a statistics model on the inner protocol [23]. Each flow has is analyzed with respect to the specific features, the size as well the inter-arrival delays of the packets that are found in the flow.

Aiello et al. [24] suggested the idea of a traditional support vector machine classifier that would be used for identifying the DNS tunneling. They used the statistical characters of the DNS queries as well as the answers [24]. They focused on the content of the queries of DNS along with the answers so as to get the data that is malicious that is hidden by the normal DNS.

In the same way, Aiello et al. [25] put forward a machine learning mechanism that would detect the DNS tunneling. The technique that they proposed examines the normal statistical features of the protocol messages like the statistics of the packets at the inter-arrival times along with the sizes of the packets [25]. These features look forward to differentiating the traffic that is legit from the DNS tunneling.

Buczak et al. [26] have also suggested a Random Forest Classification technique for finding the DNS tunneling. They addressed the various types of the features which are inclusive of the number answers that offered in the reply or the response, time in between two consecutive packets for a particular domain as well as the time in that is between the consecutive responses for a particular domain. The classifier that has been proposed is trained on such elements so as to classify the tunneling that is new and tunneling that

is unseen. The random forest classifier depends majorly on the rules; this is also the same for the decision tree which has a voting mechanism that selects the ultimate classes of prediction.

Aiello et al. [27] have suggested a system that is innovative and profiling for DNS tunneling that uses the Principal Component Analysis (PCA) as well as the Mutual Information (MI) element for extraction feature approaches [27]. Such approaches target at addressing the occurrence of statistics of particular features that discriminate the behavior of tunneling. In addition, the authors have also trained a KNN classifier on such characteristics. There have also been experiments that have been done on the network that are live. The approach that has been proposed shows how it can be a tool that is efficient in the profiling of traffic in the presence of the tunneling of DNS.

Homem et al. [13] proposed a detection system that is founded on the technique of entropy classification. They have examined the internal packet structure of the DNS tunneling mechanisms and characterize the entropy of information of the various protocols as well as their equal DNS equivalents [13]. Thus, the authors put forward a prediction protocol technique that utilizes the distribution averaging entropy.

Just recently, Van Thuan Do et al. [28] also proposed a SVM which is a support Vector Machine classifier that will identify the tunneling of DNS in a mobile network. The proposed SVM is capable of identifying tunneling within the mobile network using unique features like time, destination, protocol, length-up and the length down that is of the query of DNS [28].

**Table1**. Summary of related work

| Author | Technique | Features |
|---|---|---|
| Dusi et al. [22] | Rule-based classification | regular remote interactive logins and secure copying activities |
| Dusi et al. [10] | Decision Tree | packet size and the inter-arrival time |
| Allard et al. [23] | Decision Tree & Random Forest | the size and the inter-arrival delays of the packets in the flow |
| Aiello et al. [24] | Vectorization using SVM | Statistical features of the DNS answers and responses content |
| Aiello et al. [25] | Vectorization using SVM | Statistical features of packets inter-arrival times and of packets sizes |
| Buczak et al. [26] | Random Forest | number of answers providing in the response, Time between two consecutive packets for a specific domain and Time between two consecutive responses for a specific domain |
| Aiello et al. [27] | Principle Component Analysis and Mutual Information | Statistical of time and size |
| Homem et al. [13] | Entropy classification | internal packet structure of DNS tunneling techniques |
| Van Thuan Do et al. [28] | Vectorization using SVM | time, source, destination, protocol, length-up and length down of the DNS query |
| Cambiaso et al.[21] | Principle Component Analysis and Mutual Information | Statistical of time and size |

## 5. Conclusion

This paper has provided a review on the features of the DNS tunneling. Two main categories of features have been reviewed which are the payloa

.+d analysis and traffic analysis. Both categories contains numerous sub-instances features which have been discussed extensively. For future researchers, establishing a review on the classifiers used for the DNS tunneling would be a useful and challenging issue where each classifier can be assessed based on its mechanism.

## Acknowledgment

The authors would like to thank the UTeM Zamalah Scheme. This research is supported by Universiti Teknikal Malaysia Melaka (UTeM) under UTeM Zamalah Scheme.

## References

[1] Basel Alshaikhdeeb and Kamsuriah Ahmad, "Integrating correlation clustering and agglomerative hierarchical clustering for holistic schema matching," Journal of Computer Science, vol. 11, p. 484, 2015.

[2] Pure Hacking, "Reverse DNS Tunneling–Staged Loading Shellcode," Ty Miller, Blackhat, 2008.

[1] M. H. Ali, M. F. Zolkipli, M. M. Jaber, and M. A. Mohammed, “Intrusion detection system based on machine learning in cloud computing,” J. Eng. Appl. Sci., vol. 12, no. 16, 2017.

[2] M. H. Ali, M. F. Zolkipli, M. A. Mohammed, and M. M. Jaber, “Enhance of extreme learning machine-genetic algorithm hybrid based on intrusion detection system,” J. Eng. Appl. Sci., vol. 12, no. 16, 2017.

[5] Kenton Born and David Gustafson, "Detecting dns tunnels using character frequency analysis," arXiv preprint arXiv:1004.4358, 2010.

[6] R Rasmussen, "Do you know what your dns resolver is doing right now," Security Week. DOI= http://www. securityweek. com/do-you-know-what-your-dnsresolver-doing-right-now, 2012.

[7] M Haroon, "Squeeza: Sql injection without the pain of syringes," ed, 2007.

[8] CJ Dietrich, "Feederbot-a bot using DNS as carrier for its C&C," ed, 2011.

[9] C Mullaney, "Morto worm sets a (DNS) record," Symantec Official Blog, 2011.

[10] Maurizio Dusi, Manuel Crotti, Francesco Gringoli, and Luca Salgarelli, "Tunnel hunter: Detecting application-layer tunnels with statistical fingerprinting," Computer Networks, vol. 53, pp. 81-97, 2009.

[11] Basel Alshaikhdeeb and Kamsuriah Ahmad, "Biomedical Named Entity Recognition: A Review," International Journal on Advanced Science, Engineering and Information Technology, vol. 6, 2016.

[12] Patrick Butler, Kui Xu, and Danfeng Yao, "Quantitatively analyzing stealthy communication channels," in Applied Cryptography and Network Security, 2011, pp. 238-254.doi.

[13] Irvin Homem, Panagiotis Papapetrou, and Spyridon Dosis, "Entropy-based Prediction of Network Protocols in the Forensic Analysis of DNS Tunnels," 2016.

[14] Leyla Bilge, Engin Kirda, Christopher Kruegel, and Marco Balduzzi, "EXPOSURE: Finding Malicious Domains Using Passive DNS Analysis," in Ndss, 2011.

[15] T Pietraszek, "DNSCat," ed, 2004.

[16] Maarten Van Horenbeeck, "Dns tunneling," online], http://www. daemon. be/maarten/dnstunnel. html, 2006.

[17] Ron Aitchison, Pro Dns and BIND 10: Apress, 2011.

[18] Ed Skoudis, "The six most dangerous new attack techniques and what’s coming next," in RSA Conference (RSA’12), 2012.

[19] Manos Antonakakis, Jeremy Demar, Christopher Elisan, and John Jerrim, "Dgas and cyber-criminals: A case study," ed: Tech. rep., Damballa, 2012.

[20] J Guy, "DNS part ii: visualization, 13 February 2009," ed.

[21] E. Cambiaso, M. Aiello, M. Mongelli, and G. Papaleo, "Feature transformation and Mutual Information for DNS tunneling analysis," in 2016 Eighth International Conference on Ubiquitous and Future Networks (ICUFN), 2016, pp. 957-959.doi:10.1109/ICUFN.2016.7536939.

[22] Maurizio Dusi, Manuel Crotti, Francesco Gringoli, and Luca Salgarelli, "Detection of encrypted tunnels across network boundaries," in Communications, 2008. ICC'08. IEEE International Conference on, 2008, pp. 1738-1744.doi.

[23] Fabien Allard, Renaud Dubois, Paul Gompel, and Mathieu Morel, "Tunneling activities detection using machine learning techniques," DTIC Document2010.

[24] M. Aiello, M. Mongelli, and G. Papaleo, "Basic classifiers for DNS tunneling detection," in 2013 IEEE Symposium on Computers and Communications (ISCC), 2013, pp. 000880-000885.doi:10.1109/ISCC.2013.6755060.

[25] Maurizio Aiello, Maurizio Mongelli, and Gianluca Papaleo, "DNS tunneling detection through statistical fingerprints of protocol messages and machine learning," International Journal of Communication Systems, vol. 28, pp. 1987-2002, 2015.

[26] Anna L Buczak, Paul A Hanke, George J Cancro, Michael K Toma, Lanier A Watkins, and Jeffrey S Chavis, "Detection of Tunnels in PCAP Data by Random Forests," in Proceedings of the 11th Annual Cyber and Information Security Research Conference, 2016, p. 16.doi.

[27] Maurizio Aiello, Maurizio Mongelli, Enrico Cambiaso, and Gianluca Papaleo, "Profiling DNS tunneling attacks with PCA and mutual information," Logic Journal of IGPL, p. jzw056, 2016.

[28] Paal Engelstad Van Thuan Do, Boning Feng, and Thanh van Do, "Detection of DNS Tunneling in Mobile Networks Using Machine Learning," Information Science and Applications 2017: ICISA 2017, vol. 424, p. 221, 2017.