Summary

As the domain traffic is usually allowed to pass through enterprise firewalls without deep inspection the attackers exploit this to steal sensitive information. Therefore, in this paper, they developed a mechanism for real-time detecting exfiltration and tunneling of data over DNS. they worked with an operational network large University and a mid-sized Government Research Institute. They had done two experiments:

1. They built, optimized, and trained a machine-learning system to identify suspicious DNS queries using a known dataset of benign domains as ground truth.
2. They used Data Extrafilteration Toolkit (DET) to generate the data which will be used in their schema to identify malicious activity.

As their goal is to maximize the detection of anomalous queries while reducing the rate of false alarms. They used "Isolation Forest (iForest)" which is an effective algorithm in detecting anomalous instances in high-dimensional datasets with minimal memory and time complexities. And they fined-tune this algorithm they found that the detection performance rises by increasing the maximum limit of trees and gets stabilized at the value of 18 with the best accuracy of more than 90% and 98% for ground-truth benign and malicious instances respectively. To summarize, their approach proved to be a promising solution when it was tested on real 10 Gbps traffic streams from the two firms' networks by injecting over a million malicious DNS requests using an exfiltration tool.

Critical Review

- **Research Goal**

  The author's objective is to create a mechanism that can identify data exfiltration and tunnelling through DNS in real time. As with all previous solutions, this one will be used off-line or in the network core.

- **Clarity**

  The general idea of the paper is quite clear but there were some details didn’t mention clearly. And the authors used some concepts in cybersecurity and machine learning that were not expected from the reader of low background to know it. Like (ground truth, Feederbot and botmaster, Morto worm, and Wekbypisloader).

  The paper was very interesting at first when reading the abstract. but when you go deeper the sequence isn’t fully organized. And they are repeating themself as their flow is clear.I think I can’t fully trust the paper result as they are using open-source synthetics tools to generate malicious data and their target is real-time detection. The attacker can understand the trick of this tool and how is it works and create a new attack. As the attackers are very intelligent.

- **Related Work**

The authors well addressed the related work in a wise sequence and mention their concerns after each work. However, this is good, but they didn’t mention evidence to their concerns.

•Methods

The authors do two experiments:

1. They built, optimized, and trained a machine-learning system to identify suspicious DNS queries using a known dataset of benign domains as ground truth.
2. They used Data Extrafilteration Toolkit (DET) to generate the data which will be used in their schema to identify malicious activity.

Although authors mention their contribution several times in the paper and mention they design their scheme without clearly explaining it in detail all they mention is that their scheme is worked on 10Gpbs stream data.

• Results and Claims

They obtained a high accuracy of more than 91 percent for benign and 63 percent for malicious cases using "Isolation Forest (iForest)" with high limit of 2 trees in experiment 1 of algorithm fine-tuning, and they claimed that increasing this parameter does not improve the accuracy but increases the model size and prediction time.

• Support of Results and Claims

To back up their claims in experiment 1, they conduct experiment 2 in which they work with two operational networks and a live 10Gbps stream, dividing the data into four days for training their anomaly detection machine with benign data and keeping the remaining three days for testing and evaluation:

1. Cross-validating and assessing the trained model's accuracy for benign cases.
2. Using our technology, we tested the detection rate for fraudulent DNS requests they create.
3. Measuring the performance of the two firms' live 10 Gbps traffic streams in real time.

• Missing Claims and Results

I think they cover the part of results quite well.

• Discussion

Generally, the authors made a new benchmark to DNS extrafilteration as their solution works in real-time detecting their results were very promising. And they illustrated their solution, but they have some limitations to clearly describe their full work.

• Future Work

They didn’t mention their next step. But I believe this research will open another insight to other research.

• Reference

J. Ahmed, H. H. Gharakheili, Q. Raza, C. Russell and V. Sivaraman, "Real-Time Detection of DNS Exfiltration and Tunneling from Enterprise Networks," 2019 IFIP/IEEE Symposium on Integrated Network and Service Management (IM), 2019, pp. 649-653.