IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, August 2022
Impact Factor: 6.252

A Detection Method Against DNS Cache Poisoning Attacks using Machine Learning Techniques

Shashank Biradar[1], Shramik S Shetty[2], Pradeep Nayak[3], Prajakta Shetty[4], Shwetha R Sharma[5]
Students, Department of Information Science and Engineering[1,2,4,5]
Assistant Professor, Department of Computer Science and Engineering[3]
Alvas Institute of Engineering and Technology, Mijar, Karnataka, India
4al20is044@gmail.com,4al20is046@gmail.com, 4al20is045@gmail.com,
4al20is047@gmail.com, pradeepnayak@aiet.org.in

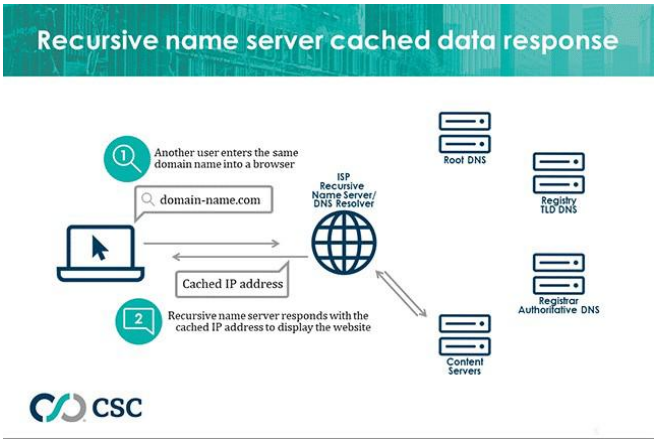
Abstract: *In this paper, we offer a machine learning-based enhanced detection strategy for DNS cache poisoning attacks. In addition to the standard DNS packet's five basic tuples, we plan to include numerous specific features that were extracted based on The heuristic components, such as the common DNS protocols "trigger," "time related features," and "GeoIP related features" of DNS cached data," etc.[1] By mapping IP and domain name, DNS's principal job is to lead users to the right computers, programmes, and data. Due to some DNS security weaknesses, attackers frequently use DNS-based malware, DNS-amplification, false-positive triggering, DNS tunnelling, etc. as a means of attack.[2] The upcoming effort comprises training with DNS traffic data and evaluations in both a small-scale experimental network and a large-scale real network environment.*

Keywords: DNS, Machine Learning.

I. INTRODUCTION

The majority of name resolution issues on the Internet have been resolved using DNS (Domain Name System) infrastructure. When there are security problems involving the DNS infrastructure, the situation is made significantly worse by the Internet's reliance on DNS-based name resolving services. DNS cache poisoning attack is one of the serious risks. The Kaminsky attack is depicted in broad strokes and has been recognised for more than ten years . All computers using a local DNS cache server's name resolution service will be subject to security threats due to the possibility of being directed to erroneous or malicious servers once that server has fallen victim to cache poisoning attacks.[2]

The three fundamental servers for TLD servers are the root DNS server, the TLD, and authoritative DNS. gov, edu, com, and org are examples of generic domain names. TLD server enables the matching authoritative DNS to receive the record. The IP address of the webpage is returned by the DNS server. DNS is susceptible to numerous threats because of the vital roles it plays. Various attack types are frequently observed on DNS. The most significant DNS amplification, DNS cache poisoning, and DNS tunnelling, NXDomain assault, and hijacking



IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, August 2022
Impact Factor: 6.252

II. ADVANCED DETECTION METHOD USING MACHINE LEARNING TECHNIQUES

In this paper, we present a machine learning-based enhanced detection strategy for both Kaminsky assaults and attacks from hacked authoritative DNS servers. As far as we are aware, outside the strategy It is the initial tactic to categorise against Kaminsky attackers .The DNS replies from a hacked reliable DNS server into both real and fake data

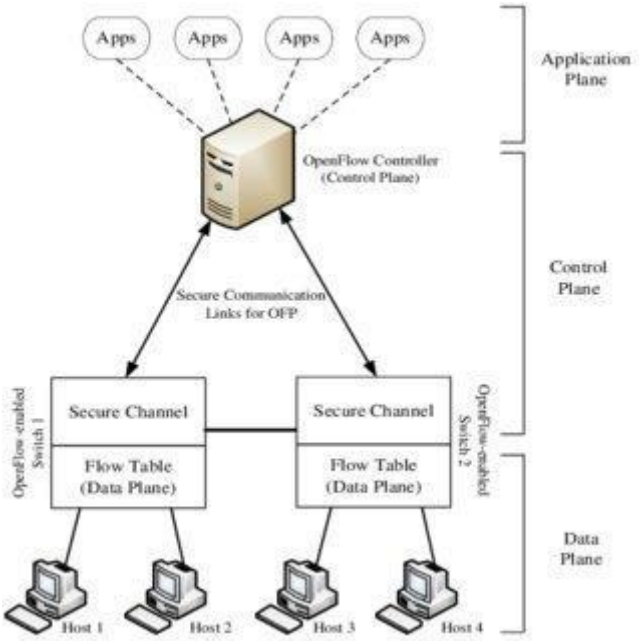
2.1 Objective and basic idea

For Kaminsky attacks, the proposed method takes into account a heuristic approach in addition to the detection of source IP address spoofing on a significant volume of DNS queries and responses because we cannot simply set a threshold on the number of DNS responses from an authoritative DNS server in a given timeframe (as described in section I). In particular, the likelihood of being subject to Kaminsky attacks grows for a DNS cache server as it receives more DNS answers with the "NXDOMAIN" (Non-Existing Domain) code [5] in response to DNS queries for non-existent sub-level domain names of a target Top Level Domain name. Consequently, a DNS traffic training model may be developed based on the feature by classifying normal and abnormal DNS query-response pairs in order to identify Kaminsky.[6]

The core concept is to heuristically compare newly received DNS data from the same authoritative DNS server to cached DNS data in order to detect attacks coming from a compromised authoritative DNS server, which is the major goal of this effort. In essence, the majority of DNS cache servers are designed to only submit queries to authoritative DNS servers and receive responses from those servers. It should be noted that some DNS cache servers have the option of being set up as "forwarders," which simply send all incoming DNS requests to another DNS cache server and only provide the final DNS results back to the clients[3]. Since we only need to identify DNS cache poisoning attacks on the DNS cache servers, we exclude these occurrences.

2.2 Main Objective

The core concept is to heuristically compare newly received DNS data from the same authoritative DNS server to cached DNS data in order to detect attacks coming from a compromised authoritative DNS server, which is the major goal of this effort.[2] In essence, the majority of DNS cache servers are designed to only submit queries to authoritative DNS servers and receive responses from those servers. It should be noted that some DNS cache servers have the option of being set up as "forwarders," which simply send all incoming DNS requests to another DNS cache server and only provide the final DNS results back to the clients. Since we only need to identify DNS cache poisoning attacks on the DNS cache servers, we exclude these occurrences.

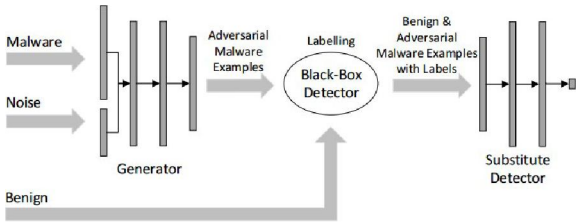


IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, August 2022
Impact Factor: 6.252

2.3 Methodology

The main objective of this project is to detect assaults emanating from a compromised authoritative DNS server, and the key notion is to heuristically compare newly received DNS data from the same authoritative DNS server to cached DNS data. In essence, the bulk of DNS cache servers are made to send queries to and receive responses only from authoritative DNS servers. It should be noted that some DNS cache servers offer the choice to be configured as "forwarders," which only transfer all incoming DNS requests to another DNS cache server and only return the final DNS results back to the clients. We ignore these events since we only need to identify DNS cache poisoning attacks on the DNS cache servers.

A few specific features that will be included specifically to each of the two training phases are listed above, however the exact amount of features will be changed depending on the network topologies when building the model, it should be noted. SVM (Support Vector Machine) is an option for the training method's two stages, classification type and regression type, respectively, but it is not the only algorithm that can be used; it is possible to combine two or more algorithms.

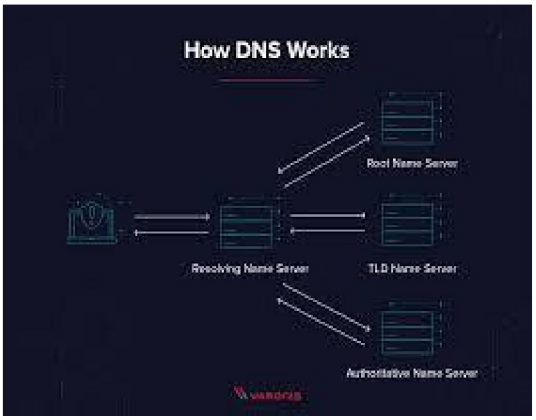


2.4 Proposed Method

A query-response pair that is likely related to any compromised authoritative DNS servers will be detected and prevented from being cached on the DNS cache server as well as replied to by the end client using the proposed method, which uses machine learning techniques for heuristic analysis and DNS data classification. According to the three categories listed below, we want to categorise the physical location of
an authoritative DNS server and the application server that responds to it.[2]

1. In the organisation network (owned by the owner themselves), both the authoritative DNS server and the replied application server are situated, allowing their IP addresses to be assigned from the same space that the organisation
2. Half outsourcing: One of them—the application server or the authoritative DNS server—is situated somewhere other than the company's internal network, like a for-profit cloud network. Their IP addresses will therefore be assigned from two distinct spaces, one of which belongs to the organisation and the other to the cloud service provider.
3. Complete outsourcing: The application server and the authoritative DNS server are both situated on an external network.

In this scenario, their IP addresses may have been assigned from the same space or from distinct spaces, but they could both be those of cloud service providers.



IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, August 2022
Impact Factor: 6.252

III. IMPLEMENTATION

Python will be used to create the models, and SDN (Software Defined Network) technologies will be used to build the network [2]. An illustration of a network topology using OpenFlow is shown in Figure 6. OpenFlow controller, Open vSwitch, DNS cache server, DNS database, analyzer, and many end clients playing the following functions are among the components.

1. End client: An internal computer that connects to the Internet and sends DNS requests to the DNS cache server.
2. DNS cache server: Offers services for name resolution.
3. DNS database: Keep all cached DNS information and the triggering DNS queries together.
4. Analyzer: Determines whether or not to cache the received DNS records by analysing all DNS query-response pairs in conjunction with the DNS database
5. OpenFlow controller: Based on the outcomes of the analysis, it decides whether or not to permit the DNS answer packets.[3]

3.1 Achieve DNS traffic data

To collect DNS traffic data, we set up a small experimental network. The "DNS cache server" resolves names by repeatedly contacting the Internet's authoritative DNS servers. In order to log specific DNS name resolution data, we used BIND (Berkeley Internet Name Domain) for the "DNS cache server" and enabled "dnatap" model.

The impact of internet voting on voter turnout is unknown. In 2017, a study of internet voting in two Swiss cities was published.

As a result, the client sends DNS requests to the "DNS forwarder," who simply transmits them to the "DNS cache server." After receiving the final DNS responses from the "DNS cache server," the "DNS forwarder" relays them to the client. The "Analyzer" will be used to examine the DNS traffic data that has been recorded in order to categorise authoritative DNS servers and the application servers that have responded. We obtained the FQDNs of the top 500 websites according to Alexa [4] and carried out "A record queries" every weekday for five weeks. This proved that we could successfully extract all the required features from the DNS traffic data obtained using the "dnstap" model on the "DNS cache server."Analysis:

IV. CONCLUSION

DoH is primarily intended to enhance security and privacy utilising encryption techniques over conventional DNS. It is accurate to say that encryption obscures user data connected to DNS queries. The existing security measures' inadequate understanding of DoH feeds several security concerns.

We were therefore motivated to conduct an examination of encrypted DoH traffic. On do this, we used well-known ML classifiers to a brand-new benchmark dataset that is openly accessible. The performance assessment clearly distinguishes between two types of traffic—benign DoH queries and malicious ones. The RF and GB recorded a maximum of 100% accuracy and F1-measure in both the traffic, as can be shown in the results section.

The RF and GB recorded a maximum of 100% accuracy and F1-measure in both the traffic, as can be shown in the results section. For malicious DoH, KNN and LR accuracy rates are 99% and 98%, respectively, with higher performance. In comparison to the other four classifiers, the performance of the NB classifier is inferior.

It can be inferred from the performance study that ensemble learning-based classifiers like RF and GB are the best options for this kind of issue. With the use of the datasets that are currently accessible, we have attempted to categorise both benign and malicious DoH traffic. In the future, we'll try to collect some fresh data using DoH and non-Doh and attempt to forecast harmful and regular DNS requests or DoH.

REFERENCES

[1]. A Detection Method Against DNS Cache Poisoning Attacks Using Machine Learning Techniques Yong Jin∗ , Masahiko Tomoishi† , and Satoshi Matsuura‡ Tokyo Institute of Technology, 2-12-1 O-okayama, Meguroku, Tokyo, JAPAN[1]

[2]. Detecting Malicious DNS over HTTPS Traffic Using Machine Learning Sunil Kumar Singh School and

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, August 2022
Impact Factor: 6.252

Pradeep Kumar Roy

[3]. Detection of Hijacked Authoritative DNS Servers by Name Resolution Traffic Classification Yong Jin∗ , Masahiko Tomoishi† , and Satoshi Matsuura‡ Tokyo Institute of Technology, 2-12-1 O-okayama, Meguroku, Tokyo, JAPAN

[4]. Classifying DNS Servers Based on Response Message Matrix Using Machine Learning Keiichi Shima;Ryo Nakamura;Kazuya Okada;Tomohiro Ishihara;Daisuke Miyamoto;Yuji Sekiya

[5]. Recovering and Protecting against DNS Cache Poisoning Attacks Xi Yu;Xiaochen Chen;Fangqin Xu

[6]. Detection of Kaminsky DNS Cache Poisoning Attack Yasuo Musashi;Masaya Kumagai;Shinichiro Kubota;Kenichi Sugitani