Atal Bihari Vajpayee Indian
Institute of Information Technology
and Management, Gwalior

Report On

Discovering HTTPSified Phishing Websites Using the TLS Certificates Footprints

Shubham Kumar(2020IS-09)

M.Tech - Information Security

Table of Contents

Chapter 1

Introduction

In the latest report of "Google Transparency Report", it was found that the percentage of traffic that has been encrypted is over 90% as in fig1.

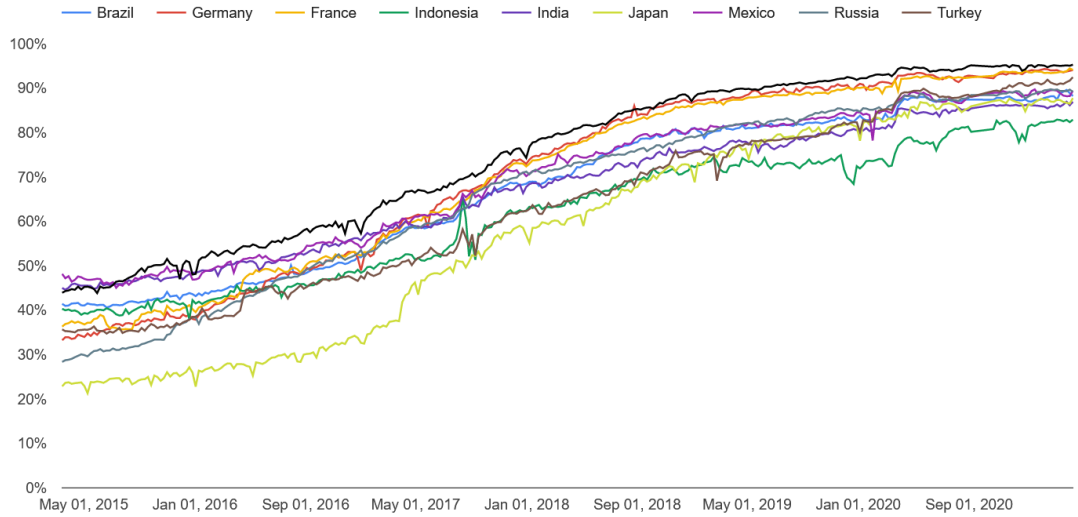

FIG:1

In the year 2015, it was reported that the percentage of usage of https was bout 50% for most of the countries. But by the end of 2020, it is approximately more than 90%. And this trend shows that the httpsified websites are increasing rapidly. And this same trend has also been discovered for phishing websites also. With the increase of httpsified phishing websites, it is important to find out such websites for security purposes.

While HTTPSifying a phishing website may bring several advantages for the attacker, but it can also contribute in generating intrinsic footprints, which in turn be used to systematically detecting the HTTPSified phishing website. The key insight behind this assumption is that byHTTPSifying a website, an attacker must register a valid public-key certificate.

Chapter 2

Background

In this research, the researcher has focused on the digital certificates for identifying phishing websites. The main reason to do so was that digital certificates are necessary for phishing websites, an attacker must register a valid public-key certificate (i.e., TLS certificate) that contains intrinsic features such as issued date, issuer name (CA), and common name (CN). The CN in a certificate is equivalent to the fully qualified domain name (FQDN)of a server.
So basically, the researcher classified this into two phases. In the first phase preparation phase and second one is attack phase as we can see in fig2.

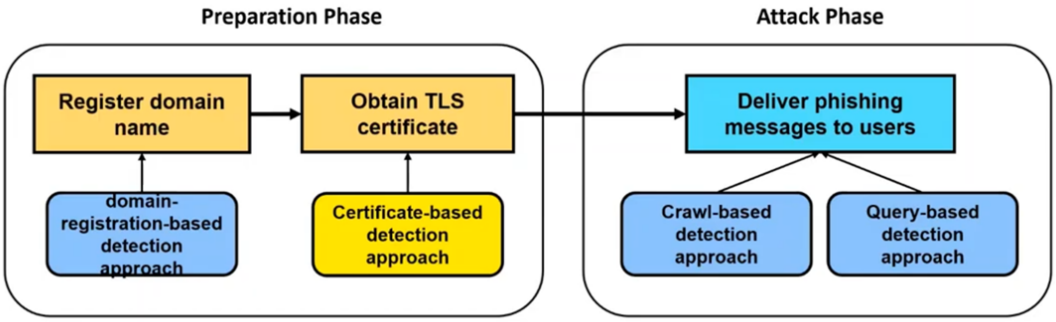

FIG : 2

So what happens here is that during the attack phase, we can find websites only when the attack is done, the reason for this is that these attacks get triggered only when the phishing is delivered. So this basically the drawback of this approach. In order to overcome this drawback what the researcher did was, they came up with the idea of using digital certificates. Using digital certificates we can stop the attack during the initial phase only. This will even help the users for phishing attacks done by the attackers using DDNS or hosting services.

Chapter 3

Framework

The researcher firstly collected the data from various websites. The researchers collected total of 1,634 unique certificates from OpenPhish and Censys. Then they extracted total of 69 templates. To extract the templates they performed the DBScan algorithm in the data, as we can see in the fig:3.

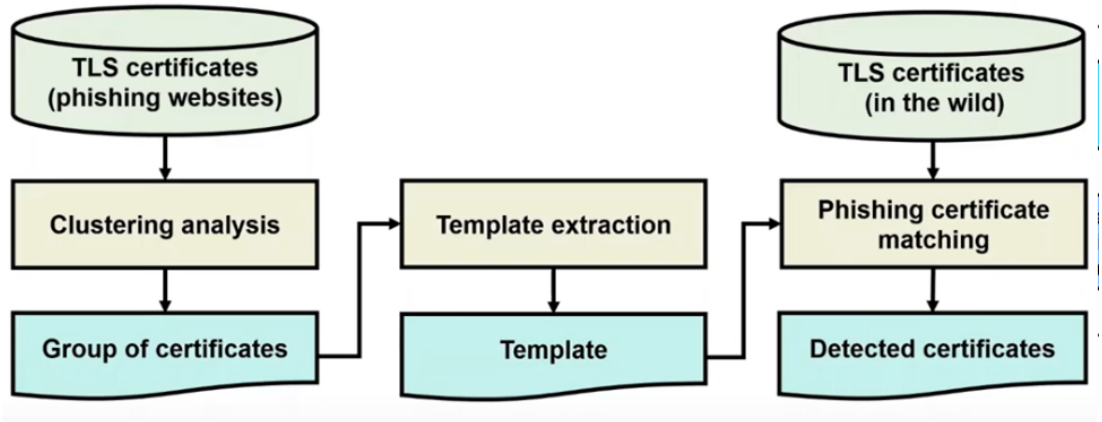

FIG : 3

They found the common names to get the group of certificates. So with the extracted templates they basically merged the templates using maximum and minimum length of the regular expressions. And lastly what they did was, the searched for the templates that matched up with templates in the wild. Here templates in the wilds are those templates which the researchers have considered to be most likely to be used for the phishing attack. The total of 38.7 million certificates were collected by the researchers from Let's Encrypt and CPanel.

Chapter 4

Results

During the result phase, it was found that a total of 1,650 certificates has common names that involve the extracted templates. It was found in the research that there are 1049 certificates has unique common names, in other words we can say that they has a unique websites. 90.8% of them were flagged as merchant which means that at least one antivirus software has raised the alerts.
In CASE STUDY it was found that:
Phishing certificates using a template that yielded the largest number of phishing websites was [a-z]6,8.runescape.com-[a-z]1,8[.TLD].Using this template, they detected the following two patterns of domain names:secure.runescape.com-[a-z]1,8[.TLD] and services.runescape.com-[a-z]1,8[.TLD].
This certificate template was detected 65% of the time, out of which 93% of them were issued by the same certifying authority. So basically the researchers speculated this and found that there are some phishing kits that target runescape. So, after depth analysis, it was found that the service offers by this is really advance and has some useful features such as mass mailer to send a large number of emails, logging, analytics, market place where they can sell or buy stolen credentials.
These results has some limitations though. The researchers has basically used the static data so this may also leads to bias inherent. Second thing is while evaluating, the researchers didn't have the direct access to the phishing websites and lastly it may be some of the websites didn't get flagged this may be due to data shortage or some other reasons which can lead to threats to validity.

Chapter 5

Conclusion and Recommendations

The idea proposed in the research paper is better than the traditional approach of identifying phishing attacks. It can detect previously unknown phishing attacks during the attack phase only. It also gives quite a good understanding of the phishing infrastructure.// As the number of phishing websites with HTTPS has been increasing. It is really important to update the dataset of the certificates. In order to over come the issue of data set, the future related studies can be done in finding a method which can overcome the static approach of identifying it and have a dynamic method for solving this problem with very much less time.