Course: Reseaux mobiles et avances
Academic year 2020/2021

Supervisor:
Maciej Korczyński
maciej.korczynski@univ-grenoble-alpes.fr

Analysis of CT Logs to Detect Phishing Websites

Description

Phishing

(Based on Wiki and https://www.csoonline.com/article/3290417/csos-guide-to-phishing-and-phishing-kits.html)

Phishing is a social attack, directly related to social engineering. Commonly centered around email, criminals use phishing to obtain access or information. Phishing attacks can be basic or customized toward the victim and their organization.

Most types of phishing use some form of technical deception designed to make a link in an email appear to belong to the organization (e.g. bank). Misspelled URLs or the use of subdomains are common tricks used by phishers. In the following example URL, www.bnpparibas.example.com, it appears as the URL will take you to the example section of the bnpparibas website; actually this URL points to the " bnpparibas" (i.e. phishing) section of the example website.

Certificate Transparency Logs

Certificate Transparency (CT) is an Internet security standard and open source framework for monitoring and auditing digital certificates. The standard creates a system of public logs that seek to eventually record all certificates issued by publicly trusted certificate authorities, allowing efficient identification of mistakenly or maliciously issued certificates (http://www.certificate-transparency.org).

You will use the API proposed by https://certstream.calidog.io to see newly observed domains in CT Logs in close to real-time. You will search for keywords such as paypal, societegenerale or bankofamerica, etc. in the domain names. Those domains will be candidate phishing domains. You

may also use dnstwister (https://dnstwister.report/ as well as https://github.com/elceef/dnstwist) to generate typosquating domains that look similar to the original versions (e.g. paypalaa). Attackers sometimes use misspelled names to trick their victims.

The second step of the project consist in developing a method (heuristics or a machine learning method) to categorize observed domains as phishing pages. You could also crosscheck with Google Safe Browsing, Virus Total and other blacklists if the domains have been detected as malicious.

You can find background information and a similar method in the following research paper:

https://pure.tugraz.at/ws/portalfiles/portal/25394076/156259641564590.pdf