Pinning & Binning: Real Time Classification of Certificates *

Zheng Dong
School of Informatics and Computing
Indiana University
Bloomington, IN
zhdong@indiana.edu

Apu Kapadia
School of Informatics and Computing
Indiana University
Bloomington, IN
kapadia@indiana.edu

L. Jean Camp
School of Informatics and Computing
Indiana University
Bloomington, IN
ljcamp@indiana.edu

ABSTRACT

The creation of a PKI with trusted roots on a X.509 infrastructure has solved the problem of key exchange and enabled widespread use of encryption between individuals with no previous contact. However, these certificates are inadequate for making a "trust or do not trust" decision in web interactions as exemplified by MITM attacks, phishing attacks, and rogue but technically valid certificates. Thus, end users today often rely on constantly updated blacklists and whitelists. While these approaches offer a simple security solution to the end users, it is often a challenge to construct a whitelist or blacklist that simultaneously satisfies three requirements: correctness, timeliness and completeness. To complement current approaches, we propose a machine learning based approach using features from TLS certificates that addresses the inherent limitations of whitelists and blacklists. We illustrate improvements in timeliness for blacklist updates and completeness for the whitelists, and offer a correctness check for both.

Categories and Subject Descriptors

K.6.m [**Miscellaneous**]: [Security]

General Terms

Security

1 Introduction

Rogue certificates and phishing are two closely related but different security problems. Rogue certificates refer to certificates that were issued by a trusted Certificate Authority (CA) but to a different entity than the one indicated in the *subject* field of the public key certificate. While phishing is an attack that an adversary masquerades another trusted online entity to steal private information from a victim. As indicated by the online database of PhishTank, the intersection of these two attacks (using rogue certificates to conduct phishing) is constantly growing. As a valid solution for both problems, blacklists may be created for malicious websites, CAs and certificates. These lists are usually maintained by the browser manufacturer, a trusted third party, or self-organized by a group of homogeneous users. However, blacklists inherently suffer from a lag, as malicious sites are identified and promulgated. For example, Tyler and Moore compared commercial and social network approaches to blacklists and found that PhishTank was slower than other data sources on reaching decisions about phishing websites [4].

With a complete whitelist, genuine websites in important categories can be identified. For example, a financial institution list informs online users when to enter their bank credentials. Employer whitelists can identify sites which can be trusted with institutional information. While it is crucial to distinguish these websites from the general web, it is often challenging to maintain a complete and accurate whitelist. In this work, we propose that in addition to the existing black/whitelist approaches (**pinning**), we can effectively categorize the good and bad websites only from their public key certificates (**binning**). This approach is a complement to the existing lists and works primarily before a white/blacklist gets official updates.

To defeat phishing and rogue certificates, blacklists and whitelists need to fulfill the following requirements: **1) Correctness:** Both the whitelists and blacklists must distinguish the corresponding websites with high accuracy. This is a fundamental requirement for almost all security mechanisms. **2) Timeliness:** The whitelists and blacklists need to be updated quickly and consistently to reflect the latest situation online. For example, the phishing blacklist requires a prompt process from the reporting of phishing, to verification, to client-side updates. **3) Completeness:** While it remains impossible to comprehensively predict tomorrow's malicious contents, a relatively complete whitelist of yesterday's is a reasonable goal. A list of today's valid websites in a specific category is feasible. A list of all websites hosted by a legitimate financial institution would be an example: a given bank would know its own domains. Previous empirical studies were conducted to investigate the timeliness and accuracy of phishing blacklists [2]. In 2006, Ludl et al. tested 10K phishing URLs against the Microsoft and Google blacklists and received true positive rates of 65% for Google and 56% for Microsoft.

A number of complementary approaches have been proposed to identify phishing sites real time and to prevent efficient

*A Poster Proposal for ACSAC 2013

leveraging of lag time, such as analyzing the pattern of URLs [1]. Mishari et al. investigated the feasibility of using the public key certificate fields to distinguish phishing websites from regular sites [3]. There are also several approaches to defeat the TLS MITM attack. EFF actively scans web domains on the Internet for certificates and built a Certificate Observatory. With Perspective [5] and *Convergence*, end users can submit the hash value of their observed certificates and the associated URL. A centralized server then compares the submission with observations from a number of geographically distributed notary servers. While very few blacklists and whitelists can achieve the timeliness, completeness and correctness requirements simultaneously, we propose an innovative approach to identify good and bad sites through the classification of TLS public key certificates. Our work is based on reliable classifications of trusted (work, bank) sites, regular and malicious sites based on 42 features extracted from certificate fields.

Our contributions are two-fold. **First,** we propose an alternative security solution before traditional black and whitelists receive official updates. Different from other machine learning based approaches, we moved beyond binary categorization of malicious and non-malicious websites. This multiclass categorization is needed for two reasons. Supported by the certificate data we compiled, phishing sites increasingly use legitimate sites as hosting platforms. Thus, the certificate is valid and correct for the organization with a subverted server. For example, many phishing sites use Google Docs. The sites need to be identified as *Not financial* or *Not employer* as phishing has moved beyond the simple attempts to steal bank passwords. While common phishing is still apparently profitable, spear-phishing targets specific organizations or datasets. Spear-phishing is a more difficult challenge for blacklists, as there are fewer potential reporters, and knowing only that a site is not a bank may be inadequate.

Second, we identified attributes from certificate fields and built machine learning models for categorizing websites. We applied five machine learning algorithms to our certificate dataset. Ten-fold cross validation was applied in our model building process to avoid bias on building the classifiers. Our experiment showed that the detection accuracy could achieve 99% for rogue certificates, banks, and work sites.

2 Data & Attributes

To validate the feasibility of building blacklists and whitelists from public certificates, we created an experimental dataset by downloading TLS certificates according to four public website lists: Alexa (General), FDIC (Banks), two educational institutions (Work), and PhishTank (Phishing). We also created a list of rogue certificates by collecting certificates issued from well-known CA subversion events. Bank and Work datasets were created to examine the capability of building whitelists, while the sample certificates from phishing websites and sites with rogue certificates were used to test blacklists. Overlaps between lists were handled by removing the record from the larger category.

We started with the data collection of general websites. Our script downloaded the list of the top 1 million websites daily from Alexa. Each listed website was then connected through its TCP port 443. We obtained a list of bank websites from the FDIC. Similar to general websites, our script requested server certificates from the websites once a TCP connection had been successfully established. Among the 27,599 FDIC-insured banks, 4,111 of them enabled HTTPS connections on their homepages (e.g. `www.chase.com`). Our data source PhishTank updates its list every hour for active online phishing as phishing pages usually have a short lifetime. While phishing is normally linked with web pages instead of domains, we discovered that an increasing number of phishing pages resided in websites with HTTPS. We attempted to connect to all websites on the PhishTank list through HTTPS and downloaded certificates when available. Rogue certificate is the final category we consider for categorizations. These certificates may still be valid but were issued because of a mistake, CA subversion, or an interception from other organizations. We identified 42 features from X.509 certificate fields. We recognize that a malicious attacker with a trusted root key can write any attribute; thus one of attributes we examine is change in public key.

3 Conclusions

We will present a machine-learning based mechanism to effectively augment whitelists and blacklists that addresses spear-phishing as well as phishing. As it is obviously impossible to predict the future pattern of rogue certificates and phishing websites, our mechanism offers a simple solution that is complementary to the regularly updated lists.

We conclude that it is feasible to classify websites into phishing and rogue with a high degree of accuracy using the set of classifiers we developed. We have illustrated that phishing sites can be identified using the associated certificates, even if the site is not using TLS by default. Using these decisions trees and regressions, a new certificate can be locally evaluated with a high degree of accuracy. We further conclude that it is possible to use such a classification to identify an increasing class of sites using legitimate certificates for subverted and illegitimate purposes, i.e. Google is not a bank, nor the US Government, nor the employer of any author. With this classification, users can be notified that sensitive information is being entered into an incorrect site. We also illustrate that different classifiers misidentify different certificates. Comparing these outputs for a newly-encountered certificate can classify a site, and identify uncertain classifications as such.

4 References

[1] S. Garera, N. Provos, M. Chew, and A. D. Rubin. A framework for detection and measurement of phishing attacks. In *ACM Recurring Malcode*, pages 1–8. ACM, 2007.

[2] C. Ludl, S. Mcallister, E. Kirda, and C. Kruegel. On the effectiveness of techniques to detect phishing sites. In *4th DIMVA*, pages 20–39, DE, 2007. Springer-Verlag.

[3] M. A. Mishari, E. De Cristofaro, K. E. Defrawy, and G. Tsudik. Harvesting ssl certificate data to identify web-fraud. *arXiv preprint arXiv:0909.3688*, 2009.

[4] T. Moore and R. Clayton. Financial cryptography and data security. chapter Evaluating the Wisdom of Crowds in Assessing Phishing Websites, pages 16–30. Springer-Verlag, Berlin, Heidelberg, 2008.

[5] D. Wendlandt, D. G. Andersen, and A. Perrig. Perspectives: Improving ssh-style host authentication with multi-path probing. In *ATC*, volume 8, pages 321–334, 2008.