DOMAINTOOLS

THE DOMAINTOOLS REPORT

SUMMER SUPPLEMENT 2016
MALICIOUS DOMAIN AFFIX PATTERNS

EXECUTIVE SUMMARY

In our previous reports, we profiled malicious domains by describing patterns in their registration details such as top level domain (TLD), or attributes such as domain age. In this supplementary edition, we compared the distributions of malicious domains vs neutral domains across a set of affixes (prefixes and suffixes) appearing in malicious domains, in order to see whether they occurred at higher rates than in neutral domains.

WHAT IS AN AFFIX?

noun
af-iks
1. An additional element placed at the beginning or end of a root, stem, or word, or in the body of a word, to modify its meaning.

Examples of affixes as discussed in this report:
www--apple[.]com
chaseonline-login[.]com

KEY FINDINGS

>> **Affixes convey intent to deceive:** It is common for registrants of malicious domains to use affixes such as “www” or “login” to lure victims to click on links that are controlled by the attacker.

>> **Affixes allow malicious registrants to (ab)use properly-spelled brand names:** A domain such as “www--apple[.]com” can look, at a glance, like the legitimate “www.apple.com,” which also can help the attacker lure victims who are not paying close attention.

>> **Abuse abounds:** Brand owners such as Microsoft, Google, and countless others, do not as a practice “defensively” register domains with these affixes. This leaves such domains available for malicious use. Nor do domain registrars, as a rule, place any restrictions on such registrations.

>> **Different affixes for different activities:** The top affixes varied somewhat, depending on whether the domains in question were blacklisted for spam, phishing, or malware (the three categories we examined for this report).

INTRODUCTION

Each edition of the DomainTools Report examines patterns of malicious and suspicious activity across the Internet, identifying "hotspots" of activity which can help analysts or researchers better understand threat actors and their networks of malicious infrastructure. For this supplement, we looked at patterns in domain names themselves to calculate their "signal strength" as an indication of nefarious activity.

The term "affix" encompasses prefixes, suffixes, and infixes (where the string occurs in the middle of a word). We analyzed a corpus of active domains across the Internet—that is, out of the approximately 300 million domain names that are currently registered, we examined approximately 255 million that are actively resolving in DNS—to explore whether certain patterns in prefixes or suffixes were correlated with higher rates of malicious or suspicious activity.

Affixes can serve a number of purposes. Familiar affixes such as the prefixes "www" or "account," and the suffixes "online" and "update," can convey the purpose of a domain; an example is "login-<domain>.com" for the case where an organization dedicates a specific domain to account logins. But another common purpose for these affixes is less wholesome: a malicious actor can spoof a legitimate domain by registering a new domain consisting of the target domain plus one or more affixes (as in "<domain>-account-update.com"). Because of the huge numbers of affixes, domain name variations, and top level domains (TLDs), it is hard for even large companies such as Microsoft and Google to prevent abuse of their names in this way, and to compound the problem, domain registrars generally take a laissez-faire approach to such registrations.

Most security practitioners have observed this type of malicious domain, but we wanted to investigate these affix patterns at large scale to see if they appeared disproportionately in pools of blacklisted domains, as contrasted with the general population of neutral domains. This supplement is the result of our findings.

METHODOLOGY

First, we amassed a list of affixes that appeared frequently in an initial corpus of domains used in phishing attacks and other nefarious activity. Then we queried the entire DomainTools database of extant domains—over 330 million—to assess the rates of appearance of the affixes. Then, using well-known blacklist providers, we compared the rates of occurrence of these affixes in any domains that had been identified as spam, phishing, or malware on the blacklists.

We sought answers to questions such as these:

>> Do certain affixes appear in malicious domains at higher rates than they appear in neutral domains?

>> Do the malicious activity types (malware, phishing, spam) have different constellations of affixes?

>> Does the presence of a given affix provide a meaningful signal that the domain is more likely to be malicious?

ACTIVE VS. DORMANT DOMAINS

Of all domains with current registration, only a subset are active, as observed in passive DNS sources. For neutral domains, we ran our calculations against both active and dormant neutral domains. We assume that essentially all blacklisted domains are active, since a domain has to be observed in some kind of nefarious activity to be blacklisted, so the active vs. dormant distinction is really only meaningful for the neutral domains. We did this to see whether this might signal that suspicious affixes are more prevalent in active domains; it could suggest that malicious actors don't leave domains "on the shelf." Conversely, if the signal were stronger in dormant domains, it could be indicative that there is a body of such domains awaiting later weaponization.

As in our February 2016 report, we used the concept of *signal strength* to characterize domain features. In this supplement, we apply the idea of signal strength to affixes. Signal strength is a function representation in a class of domains, where "class" means neutral domains, or domains blacklisted as spam, malware, or phishing. Thus, for all domains classified as phishing, the signal strength of a given affix, such as the prefix "app", tells us how representative that prefix is among phishing domains. We compare this to how representative it is across neutral domains.

For example: of all blacklisted domains, approximately 4.7% have the prefix "app." Of all neutral domains, just over 1/10 of 1% have this affix. Therefore, if we compare the two percentages, the rate at which malicious domains use that prefix is about 42 times the rate at which neutral domains do. We call this a "signal strength" of 42.

WHY DO WE MEASURE SIGNAL STRENGTH ACROSS EACH VALUE IN THE DISTRIBUTION?

While our report can simply compare the distributions through standard statistical measures, we wanted our research to help inform our risk scoring and reputation scoring algorithms as to which affixes indicate maliciousness, and the relative strength of the signals.

SCORE

In order to capture both the representation in the class (signal strength) and the prevalence in the wild, we developed scores for each category. To make the top 10, an affix has to have a strong combination of absolute numbers and signal strength. While such affixes could have strong signal strengths, the very low numbers of domains suggest that users will only very rarely encounter such domains "in the wild" and so, while the domains may be malicious, the affixes tied to them are not reliable large-scale indicators of danger.

AFFIXES IN PHISHING DOMAINS

The first table shows the rates of occurrence of the top-scoring affixes found across all domains in our sample that had been blacklisted for phishing. In the first column, the parenthetical (p) (s) or (i) indicates whether the string was a prefix, suffix, or infix.

First column: Affix (p|s|i)
Second column: Absolute number of blacklisted phishing domains with the affix
Third column: Percentage of phishing domains with the affix
Fourth column: Signal strength of the affix

These affixes confirm what one might naturally surmise about phishing domains: if their purpose is to lure a victim into taking some action, the affixes related to "login," "account," and "update" may very well be tied to credential-harvesting look-alike sites.

TOP 10 PHISHING AFFIXES BY "PHISH SCORE"

AFFIX	NUMBER	PERCENT	SIGNAL STRENGTH
com- (p)	3850	1.89%	279.00
wap- (p)	933	0.46%	702.05
app (p)	9637	4.74%	42.79
account (s)	588	0.29%	242.37
account- (p)	411	0.20%	300.79
update- (p)	348	0.17%	306.75
pay (p)	3083	1.52%	31.78
update (s)	442	0.22%	184.92
login- (p)	243	0.12%	270.46
login (s)	295	0.15%	179.45

AFFIXES IN MALWARE DOMAINS

This table corresponds to the above, except that it shows the highest-scoring affixes in malware rather than phishing domains.

First column: Affix (p|s|i)
Second column: Absolute number of blacklisted malware domains with the affix
Third column: Percentage of malware domains with the affix
Fourth column: Signal strength of the affix

These affixes, too, make sense given the classification of malware. "Download" is certainly an action that the threat actor might want the victim to take.

TOP 10 MALWARE AFFIXES BY "MALWARE SCORE"

AFFIX	NUMBER	PERCENT	SIGNAL STRENGTH
app (p)	4146	0.64%	5.80
com- (p)	911	0.14%	20.79
-download (s)	531	0.08%	30.52
download (p)	983	0.15%	14.39
api- (p)	178	0.03%	36.45
update (p)	311	0.05%	18.32
vv (i)	1535	0.24%	3.66
-com (s)	441	0.07%	8.96
com (s)	1606	0.25%	2.73
download (s)	402	0.06%	6.66

AFFIXES IN SPAM DOMAINS

These tables show affix statistics for domains blacklisted as spam domains.

First column: Affix (p|s|i)
Second column: Absolute number of blacklisted spam domains with the affix
Third column: Percentage of spam domains with the affix
Fourth column: Signal strength of the affix

The spam affixes show a little more variety than the other categories. This is expected as well, since "spam" is a broad term that encompasses many types of unwanted and potentially harmful emails.

TOP 10 SPAM AFFIXES BY "SPAM SCORE"

AFFIX	NUMBER	PERCENT	SIGNAL STRENGTH
com- (p)	4225	0.41%	73.83
www (p)	14403	1.39%	32.13
vv (i)	2445	0.24%	2.27
db (p)	1303	0.13%	2.98
updates (s)	200	0.02%	9.71
-install (s)	55	0.01%	16.01
mail (p)	468	0.05%	2.40
app (p)	1423	0.14%	1.41
account- (p)	52	0.01%	5.39
www- (p)	175	0.02%	1.99

BUILDING A COMPOSITE PICTURE

As with several of the other dimensions we have studied in previous editions of the DomainTools Report, the affixes contained some expected items and some surprises. We expected to see prefixes such as "www" and suffixes such as "com." There were also some surprises along the way, such as the prefix "wap," whose semantic meaning is not immediately clear, but which was obviously present in some large campaigns.

These signals may prove extremely valuable in combination with other features we have examined. An ongoing DomainTools project seeks to use machine learning and other techniques to analyze various composites of attribute signals to develop high-confidence domain risk assessment.

In the meantime, we hope that these analyses are helpful to security professionals, researchers, and anyone else interested in better understanding large-scale patterns in domain registration data with respect to nefarious activities.

ABOUT DOMAINTOOLS

DomainTools is the leader in domain name, DNS and Internet OSINT-based cyber threat intelligence and cybercrime forensics products and data. With over 14 years of domain name, DNS and related 'cyber fingerprint' data across the Internet, DomainTools helps companies assess security threat risks, profile attackers, investigate online fraud and crimes, and map cyber activity in order to stop attacks.

Our goal is to stop security threats to your organization before they happen, using domain/DNS data, predictive analysis, and monitoring of trends on the Internet. We collect and retain Open Source Intelligence (OSINT) data from many sources and we index and analyze the data based on various connection algorithms to deliver actionable intelligence, including domain scoring and forensic mapping.

DomainTools uses over 10 billion related DNS data points to build a map of 'who's doing what' on the Internet. Government agencies, Fortune 500 companies and leading security firms use our data as a critical ingredient in their threat investigation and cybercrime forensics work.

For more information about DomainTools' data and products, please visit our website at **www.domaintools.com**.

WORLD'S LARGEST DNS FORENSICS DATABASE**

>> Over 300 Million known domains in DNS
>> 10 Billion+ current and historical Whois records
>> 4.5 Billion+ IP address change events
>> 1.8 Billion+ Registrar change events
>> 3 billion+ name server change events

** These figures are from Q1 2016, but they are inherently out of date, as we add about 5M records a day.