DOMAINTOOLS WINTER SUPPLEMENT 2017-18:

NEW PATTERNS IN PHISHY AFFIXES

EXECUTIVE SUMMARY

In a DomainTools Report from the summer of 2016, we compared the distributions of malicious domains against neutral domains across a set of affixes (prefixes, suffixes, and infixes) appearing in domain names, in order to see whether certain affixes were overrepresented in nefarious domain names, and thus presented a meaningful signal of risk. Our findings confirmed that certain affixes do portend higher risk, and we published data demonstrating which affixes were most represented in domains blacklisted for malware, spam, or phishing.

Because threat actors continually evolve their tactics, the DomainTools reports periodically update earlier findings; for this edition, we went back to the data for a new study of affix patterns. We wanted to identify what had changed, what stayed the same, and what inferences could be drawn from the data. In the interest of continuously evolving and improving our methodology, we also introduced a new method of finding affixes, and this contributed some interesting new data.

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 EVOLVE:

some new affixes have made our Top 10 lists as compared to our previous report. These include the suffix “upgrade” for phishing domains, and the prefix “support” for malware.

THREAT ACTORS CONTINUE TO BE CREATIVE WHEN CREATING NEW AFFIXES:

Our new method of hunting for affixes surfaced some surprises, such as “pineapple” and “plan-cul” (a rather salty French term for “one night stand”). In both cases, these prefixes may have represented specific campaigns rather than widespread use.

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 updated a previous study that examined domain name affixes as possible signals of nefarious activity.

The term “affix” encompasses prefixes, suffixes, and infixes (where the string occurs in the middle of a word). For each of these affix studies, we analyzed a corpus of active domains across the Internet—that is, out of the approximately 346 million domain names that are currently registered, we examined approximately 261 million that are actively resolving in DNS—to explore whether certain patterns in prefixes, infixes, or suffixes were correlated with higher rates of malicious or suspicious activity.

Affixes can serve a number of purposes. Familiar affixes such as the prefix “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” or “www<domain>.com”). Because of the huge numbers of affixes, domain name variations, and top level domains (TLDs), it is challenging for even large companies such as Microsoft and Google to prevent abuse of their brand names in this way. To make matters worse, domain registrars generally take a laissez-faire approach to such registrations, with few, if any, checks in place to ensure that the registrant has a legitimate reason to register a domain that contains a well-known brand or organization name.

Most security practitioners have observed this type of malicious domain for years in the course of incident response, threat hunting, or enforcement actions, 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. Having done an initial study in the summer of 2016, we wanted to follow up to see if, and how, patterns of affix usage are evolving over time.

METHODOLOGY

We developed a new method of hunting for suspicious affixes for this edition of the report, but some aspects of our work remained the same as in the previous study. 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 our database of over 261 million domains to assess the rates of appearance of the affixes. Next, 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.

OUR NEW TECHNIQUE OF IDENTIFYING INTERESTING AFFIXES WORKED LIKE THIS:

1. We split every existing domain name into sets of three contiguous letters, a process called tri-gramming. (Example: the word "affix" contains the tri-grams aff, ffi, and fix) We then used the signal strength algorithm (described next) to identify tri-grams that are overrepresented in the three threat categories of spam, phishing, and malware.

2. We combined overlapping high-signal tri-grams into larger word fragments. This provided hints for the most likely malicious patterns (since the tri-grams themselves were generally not words).

3. We then generated a new list of affixes based on these patterns and re-ran our affix processing.

Armed with the new data set, we compared how these new affixes stacked up against our previous list. Some of the questions we wanted to answer were these:

>> Do certain affixes still carry a strong signal of risk?

>> Have the specific affixes favored by threat actors changed over the last 12-18 months?

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

SIGNAL STRENGTH

Since the publication of our February 2016 report, we have used the concept of signal strength to characterize domain features. 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", is that affix's representation, as a percentage, in the malicious category (spam/malware/phishing) compared to its representation, as a percentage, among neutral domains.

SCORE

In order to capture both the representation in the class (signal strength) and the impact in the wild, we developed the concept of a score. Our Top 10 lists represent the affixes that have the highest combination of absolute numbers and signal strength. While some 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.

TOP 10 LISTS

While there were many similarities to the 2016 report, we found some new thematic trends, especially affixes related to various Apple products. Our malware and spam Top 10 lists include the words "apple," "iCloud," and "iPhone."

There are also some more surprising findings: the word "pineapple" made our Malware Top 10 list, and the term "plan-cul-sur" (which incorporates a rather salty French slang term which we won't translate here) reached the #1 spot on our Spam Top 10 affix list.

In each of these lists, the affixes in **bold** are those found with our original methodology that did not make the Top 10 in the previous study. Affixes in *italics* are those found by our new affix hunting methodology.

TOP 10 PHISHING AFFIXES BY "PHISH SCORE"

AFFIX p|s|i designate prefix, suffix, or infix	COUNT Absolute # of blacklisted phishing domains with the affix	PERCENTAGE % of phishing domains with the affix	SIGNAL Signal strength of the affix
upgrading (i)	7790	2.27	6738.88
upgrading (s)	5813	1.69	8138.95
updating (s)	7964	2.32	5828.98
updating (i)	10661	3.10	4328.93
warnlng- (i)	508	0.15	64013.92
lmportant- (p)	519	0.15	56057.18
warnlng (i)	508	0.15	48010.44
update (i)	30218	8.79	487.95
upgrade (i)	17807	5.18	713.63
upgrades (s)	6061	**1.76**	**1982.93**

Our new methodology populated most of the Top 10 list for phishing, though variations on "update" did also make our Top 10 list in 2016. Certainly the themes of upgrading/updating are working well for phishers—the strong inference is that victims must be falling for domains containing these words. Typo-ed attention-getters "lmportant" and "warnlng" (note the substitution of L for I) made this list, as well as the malware list below. Phishers (and other cybervcriminals) are nothing if not results-driven. Another change worth noting is the absence, this time around, of the terms "account" and "login." One possible conjecture about this change is that would-be victims are becoming more savvy about credential-harvesting phishing pages.

TOP 10 MALWARE AFFIXES BY “MALWARE SCORE”

AFFIX	COUNT	PERCENTAGE	SIGNAL
p|s|i designate prefix, suffix, or infix	Absolute # of blacklisted phishing domains with the affix	% of phishing domains with the affix	Signal strength of the affix
warnlng- (i)	417	0.09	40183.63
Important- (p)	346	0.07	28578.70
apple (i)	8703	1.94	50.10
error- (i)	1156	0.26	158.87
-server (i)	2469	0.55	73.88
virus- (p)	1480	0.33	111.95
iphone (i)	3795	0.84	43.78
icloud (i)	1513	0.34	79.12
pineapple (i)	1038	0.23	106.90
-alert (i)	1220	0.27	83.66

The entire Top 10 list for malware affixes was surfaced by our new affix hunting methodology. While the first two affixes clearly urge the victim to update something, three of the other affixes in this list are more specific: “apple,” “iPhone,” and “iCloud” are all apparently paying off for criminals. In an irony that will not surprise anyone who has been watching informwation security trends, many malware domains purport to have something to do with security (“virus” as well as “warnlng-” and “Important.”. The most unusual entry on this list is “pineapple,” which may have made its way into the Top 10 on the strength of a single campaign or actor.

TOP 10 SPAM AFFIXES BY “SPAM SCORE”

AFFIX	COUNT	PERCENTAGE	SIGNAL
p|s|i designate prefix, suffix, or infix	Absolute # of blacklisted phishing domains with the affix	% of phishing domains with the affix	Signal strength of the affix
plan-cul-sur (p)	223	0.02	1850.27
warn (s)	901	0.08	167.87
prize (i)	1504	0.14	17.11
apple (s)	1233	0.12	17.26
new (p)	5463	**0.52**	2.19
vv (i)	2578	0.25	2.92
prize (p)	270	0.03	13.33
www (p)	3350	0.32	1.91
db (p)	955	0.09	2.48
1111 (s)	244	0.02	4.55

The prefix “new” is the lone representative from our original affix-hunting methodology, but in 2016 it did not make our previous Top 10. The bawdy French prefix “plan-cul-sur” had a very strong signal, though its absolute numbers of domains were low. Like the malware list, “apple” makes an appearance in spam domains. Variants of “prize” as both a prefix and infix suggest that spammers are having some success in enticing victims with promises of winning something; the rest of the list is something of a smattering of different kinds of affixes. As in our previous report, 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.

THE BIG PICTURE

The changes in favored affixes are one small insight into the many ways that attackers are evolving. Cybercriminals are nothing if not pragmatic; it is likely that the new affixes in our Top 10 lists are connected to a demonstrable ROI for the criminals. As in the 2016 affix report, the data contained some expected items and some surprises. We expected to see prefixes such as "www" and suffixes such as "www" and "com," and these seem likely to remain in our Top 10 lists for some time, but we didn't expect to see "pineapple" or "plan-cul." A slightly discouraging trend is the apparent payoff to criminals for including variations on the words "upgrade," "update," and "security;" it appears that victims are being successfully lured by those terms.

These signals may prove extremely valuable in combination with other features we have examined. Our Threat Profile project uses various attributes of domains to develop predictive risk models for domains. These affixes are examples of the kind of attributes that can, at statistically relevant scale, classify dangerous or unsavory domains.

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 15 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 345 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 Q4 2017, but they are inherently out of date, as we add about 5M records a day.