Whitepaper Comments:
Time Series Analysis of Recursive Queries

September 19, 2016

1 Introduction

This document provides independent comments on a whitepaper entitled *White Paper #1 - Auditable V3* (hereinafter, "the whitepaper"). The whitepaper details various domains owned by the Trump Organization, and identifies patterns of DNS queries demonstrating network connections between a Trump owned domain, Spectrum Health (a medical service provider in Michigan), and Alfa Bank (a Russian bank).

This document attempts to use alternative data sources to verify the conclusions of the whitepaper. In relevant parts, this study agrees with the whitepaper's core findings: there is a significant interaction between a domain operated by a presidential candidate, and a Russian bank. The interactions appear to be related to email delivery, using a secured server, whose messages are evidently only accessible to Spectrum Health in Michigan, Alfa Bank, and a VPN provider in Utah.

This report also shows how others can verify these conclusions themselves, using various public data sources.

2 SPF Analysis

Public online passive DNS databases, such as the Chinese DNS informational site dnsdb.io [1], show several RRsets in the `trump-email.com` zone. In relevant part these include:

```
mail1.trump-email.com A 66.216.133.29

trump-email.com TXT |v=spf1 ip4:198.91.42.0/23 ip4:64.135.26.0/24
 ip4:64.95.241.0/24 ip4:206.191.130.0/24 ip4:63.251.151.0/24
 ip4:69.25.15.0/24 mx ~all
```

The last line is a Sender Policy Framework (or "SPF") record [4], and identifies domains and address ranges used in outbound email. While complex, the SPF record essentially lists the machines authorized to send outbound email on behalf of the `trump-email.com` domain. The listed IP address ranges have mail servers that send

emails, such as office correspondence. Using SPF, recipient mail servers can verify and discard fake messages (e.g., spam from 3d party networks claiming to come from `trump-email.com`). If the sending host claims to come from `trump-email.com`, but is not inside one of the listed SPF ranges, the message delivery may "soft fail", under the SPF protocol.

Significantly, the SPF CIDR ranges ***do not*** encompass `66.216.133.29`, the address for `mail1.trump-email.com` (hereinafter the "`mail1` host"). Thus, it is unlikely that `mail1` is used for sending messages on behalf of the parent zone, `trump-email.com`, since many recipients might discard them, or score them as likely spam [1]. Instead, if the `mail1.trump-email.com` host sends mail, it likely is on behalf of the child zone `mail1`. (As noted below, the host may instead be a outbound server or forwarder.)

In effect, `mail1` is a "punch out" domain in the larger Trump zone. It has its own mail policy, separate from the parent zone which appears designed for more robust, scaled messaging. While emails do originate through `trump-email.com`, and the six listed IP address ranges, the host `mail1.trump-email.com` has its own sending policy.

These DNS records can be found in most any passive DNS data source, and there are no contrary records in any online passive DNS database, from 2010 forward. And a current DNS lookup for these records yields the same results.

3 mail1 Host Operation

One can determine the purpose of `mail1.trump-email.com` by contacting the mailserver directly, and checking if it operates an SMTP server. Novice users can even check using several online mail services themselves. For example, Pingability [6] lets one interact with the mail server through a web interface:

```
DEBUG: getProvider() returning
 javax.mail.Provider[TRANSPORT,smtp,
 com.sun.mail.smtp.SMTPTransport,Oracle]
DEBUG SMTP: useEhlo true, useAuth false
DEBUG SMTP: trying to connect to host
 "mail1.trump-email.com", port 25, is SSL false
521 lvpmta14.lstrk.net does not accept
  mail from you (72.249.37.67)
DEBUG SMTP: could not connect to host
  "mail1.trump-email.com", port: 25, response: 521
```

[1] In more detail: While there is a "soft fail" flag in the SPF record, there is no wildcard SPF record at the zone apex. Thus the child zone `mail1.trump-email.com` would need its own SPF record to facilitate delivery, and its address must be included in an appropriate ip4 stanza. Despite not having a covering SPF record, the host `mail1.trump-email.com` could still send small-scale direct messages, on behalf of `mail1.trump-email.com` (but not the parent zone), if recipients whitelist it or are configured to specifically accept the mail.. But the `mail1` domain would prove problematic for mass mailings such as newsletters, hotel customer contact, vendor communications, and such.

The line starting with "521" is from the Trump server, and lines starting with "DEBUG" are generated by the Pingability testing service. Other online services may show slightly different output, but the message from the `mail1` host is always the same. The 521 reply code means the host is refusing to accept incoming email, per RFC 1846 and subsequent revisions [2].

In detail, the response indicates the Trump host refused mail, identifying itself as a "Listrack" virtual mail transfer agent (or "lvpmta"). The host is likely configured as an "outbound server", where users of in the Trump organization send emails, while using another email server to receive messages. The Listrak service is powered by "Port25 powerMTA", a commercial vendor of high quality SMTP software, which offers an "access control list" (ACL) capability. This permits the `mail1` host to filter based on user IP. In this case, it appears the Trump host is private, and was configured to only permit connections from specific hosts.

4 Query Rates

The whitepaper provides query logs for the `mail1` DNS record, in an ancillary file called "Log-Of-DNS-Lookups-For-mail1.trump-email.com-851.txt". The data is plotted in the whitepaper, which appears to show similar query frequencies originating from Alfa Bank and Spectrum Health networks. We note that, using old, well-known cache inspection techniques [3], it might be possible for users to iteratively query for such data themselves, to construct a similar pattern of usage. This would be time consuming, and was not attempted.

So using just the log files in the whitepaper, we consider whether the DNS lookups were for email delivery. This appears to be the case, because of these established facts:

- The host `mail1.trump-email.com` has two domain labels often associated with SMTP operation.
- The `mail1` host is colocated in a commercial network often used for managed email handling.
- The `mail1` host listens on port 25, and responds with an error message found only in Listrak mail servers.

The lack of MX lookups may be associated with the configuration of the Listrak host (e.g., being used for secure relay to a specific host, using a `relay-domain` option in the Port25 software), though this is not certain.

Because the log file has time stamps with precision down to the second, it is possible spot patterns of DNS lookups from Alfa Bank and Spectrum networks. Significantly, we note that there are very few source networks resolving the `mail1` host: just Alfa Bank, Spectrum Health, and (at a distant third) the VPN provider in Utah. The whitepaper analysis rightly dismisses the handful of other DNS lookups as noise, e.g., originating from infected hosts that just do a single lookup, and never interact further. A check of various online DNSBL sources [5] confirms this diagnosis.

We can therefore look for patterns among the three major resolver networks (Alfa, Spectrum and the VPN), and create a time series analysis. Consider:

1. If these DNS lookups are human-driven email message delivery attempts, one would see a pattern often associated with normal email threads: quick replies in some cases, and slight delays in some replies.

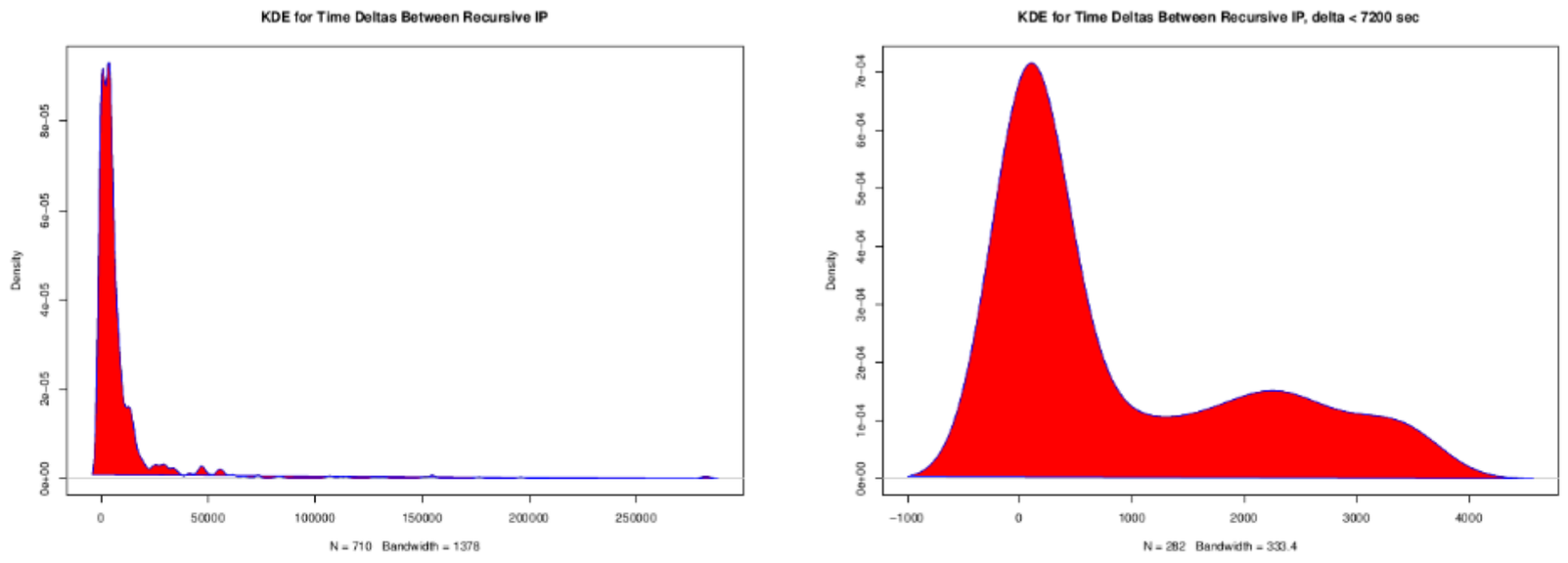
2. If these DNS lookups are instead associated with malware or some infection vector, one would expect to see a more automated lookup pattern. (Indeed, one would likely see the resolution of a 3d party command-and-control domain, instead of just a Michigan hospital and a Russian bank.) The lookup volume and paucity of qname diversity likely rules out this theory.

3. If these DNS lookups were associated with bulk email delivery newsletters, or customer contact, one would expect to see a more distributed period of lookups, with an exponential rate. (That is, regular high volume resolutions, followed by low volume periods.) As noted above, the SPF records significantly complicate the use of `mail1` for anything like this. But we can investigate this alternative theory.

We first calculate the inter-arrival time between DNS lookups from the Spectrum and Alfa Bank resolvers using a simple time series analysis. Using a stateful window, we note which network resolved the `mail1` host last, and at what time. When a different network (AlfaBank, Spectrum, or the Utah VPN) resolves the `mail1` host, we note the length of time, δ, between the recursive change, and update the state window. Informally, this measures the speed or pace of any message exchanges, or the "tick-tock" of humans sending messages back and forth. That is, when both Alfa Bank and Spectrum's recursives no longer have `mail1` in local cache, we can measure the speed of the "reply" to the first message putting `mail1` back in cache on the other network. Since the data spans a lengthy period, and the `mail1` TTL cache period is short, we have many such observations.

In traditional network analysis, spam, viruses or scheduled bulk newsletter deliveries exhibit less "back-and-forth", where a sending network is contacted by the recipient. Indeed users seldom reply to spam, viruses or even newsletters. And if both Spectrum and Afla Bank were automating their lookups, then the time delta distribution would peak around the greatest common divisor for both lookup periods, and with minimal heteroscedasticity (informally, with little dispersion or "flatness"), due only to nework lag.

After processing the time stamps in the log file, we plot the inter-arrival of queries from different recursives. Figure 1(a) and (b) show the kernel density estimate (KDE) for the distribution of these time deltas. We use a KDE instead of a traditional bin distribution plot, because the latter easily skew results based on bin size. Here, the optimal bin size is calculated algorithmically, with the smoothing parameter, h, reported as "bandwidth" in the plot. Figure 1(a) shows a wide distribution of times, suggesting these are unlikely to be "cron'd" or automated lookups. Figure 1(b) zooms in on the distribution (where $\delta < 7200$ seconds), showing the distribution of paired DNS lookups just seconds and minutes apart. (I.e., short episodes, when another network put `mail1` back in cache, perhaps in reply to a message.) In other words, there are many instances where the conversations are active, rapid-fire, and other instances where the cache refresh changes appear to take hours.



(a) (b)

Figure 1: Distribution of inter-arrival of networks querying `mail1` host. (a) KDE of time between all changes in source networks. (b) Distribution of short-period changes, $\delta < 7200$ seconds.

These results are consistent with human-driven email interactions with the `mail1` domain. As noted, if this were instead automated, or driven by an infectious or spamming process, the rate, volume and frequencies would be more periodic for the former, or random for the latter.

One can further constrast this result for `mail1` with the volumetric network graphs for the domain `trump-mail.com`. **Nota Bene:** here, we refer to `trump-mail.com`, not `trump-`**`e`**`mail.com`. We note that the `trump-mail.com` domain has distinct whois data, and even may have separate owners. The `trump-mail.com` domain is hosted in a colocation facility often abused by spammers. In contrast, the `mail1.trump-`**`e`**`mail.com` domain (note the 'e' in email) has the correct contact information for the Trump Organization, and is hosted in a facility commonly used for legitimate enterprise mail handling.

Since the MikroTik router on `trump-mail.com` is public facing, anyone can look at the volume of traffic on the two interfaces. Figure 2 shows the network graphs for `trump-mail.com` external interface, available from its public web page. It shows a periodic spike for inbound traffic (green peaks), with no spikes on the weekends. Further, the spikes always occur at the same hour, every weekday. This is classically found in automated network use, such as backups, newsletter delivery, and the like. This automated, periodic pattern provides a useful contrast for the human-driven `mail1` trump server interactions.

5 Conclusions

This paper verified some statements in a whitepaper describing DNS interactions with `mail1.trump-email.com`, offering other sources of data were possible. This

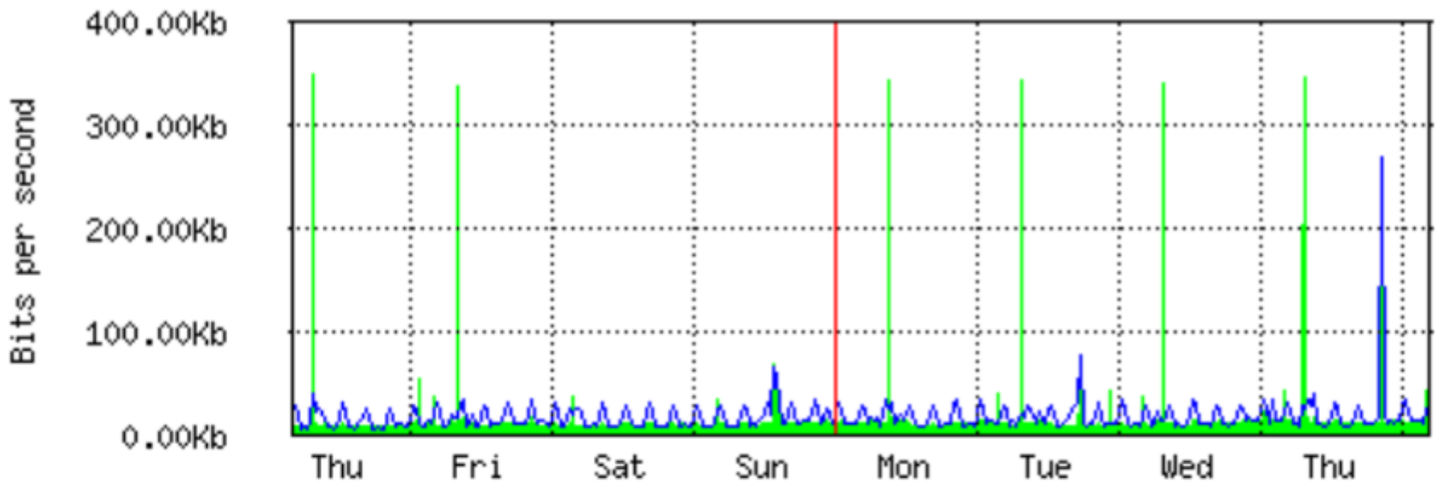


Figure 2: Periodic network traffic on `trump-mail.com`, consistent with bulk receiving and cron backup. Green lines indicates inbound, and blue represent outbound traffic.

.

analysis concludes:

- The domain `mail1.trump-email.com` has a distinct mail policy from the parent zone `trump-email.com`. It is unlikely the `mail1` could effectively send mail on behalf of `trump-email.com`, without delivery complications.
- The `mail1` host operates its own SMTP, a Listrak server. The server appears configured for secure communications or forwarding. It clearly permits connections only from a set of authorized hosts.
- Resolvers in Alfa Bank, Spectrum Health, and a Utah-based VPN provider are the only hosts resolving `mail1.trump-email.com`. A few other hosts around the Internet resolving `mail1` are low in volume (e.g., once in a month period), or exhibit infectious behaviors, and are not relevant. On the whole, only Alfa Bank (in Russia), Spectrum Health (in Michigan) and a VPN provider in Utah interact with or consume messages from the specialized Trump mail server, in any volume.
- There are many DNS lookups for `mail1` from these networks, and the timing pattern is consistent with human driven email resolutions. The resolution patterns are not consistent with automation, backups, or infectious behavior.
- Alfa Bank, Spectrum Health, and Trump's networks interact with each other regularly, evidently sending email to a secured server, `mail1.trump-email.com`.

This analysis has not addressed a few statements in the whitepaper, not considered material to the overall analysis. These include:

- The nature of the host in the Spectrum Health network and whether it was operated as a Tor exit node. The status as a Tor exit node can be verified by other sources (e.g., the Tor project), but did not appear dispositive on the core question of a messaging nexus between key networks.

- The whitepaper commented on Alfa Bank's recursive resolvers, which did not appear to "respect caching behavior". While this observation is clearly supported by the data, we speculate the explanation is quite mundane. Likely, Afla runs a caching farm, which does not share cache results among individual resolvers. (This type of resolver configuration is efficient and inexpensive, and common in enterprises the size of Alfa Bank.) While more analysis could prove this, or even estimate the number of independent cache lines behind the Alfa Bank egress IP, this did not appear material to the core question in the white paper.

- We have not reviewed the accuracy of the other data files, which generally just provide lists of domain names with "trump" substrings, or show registration information. Given that one can use public DNS sources to find the anomalous SPF records around `mail1.trump-email.com`, these steps were not necessary. If needed anyone can trivially query for the listed domains or use public passive DNS databases to verify the reported RRsets.

- We do not comment on any associated materials or analysis about Spectrum, their interest in Trump or Russian banks, Alfa Bank or its organization or operation, or its connection (beyond frequent messaging) with Trump owned networks. Other experts may look at the timing of the query volumes, in relation to other exogenous events associated within these organizations, e.g., investments of funding activities, [7]. Such details are beyond the narrow technical focus of this whitepaper.

References

[1] QQ Group 437080096. Dnsdb. https://dnsdb.io/en/search?q=trump-email.com, 2016.

[2] A. Durand and F. Dupont. Smtp 521 reply code. https://www.ietf.org/rfc/rfc1846.txt, September 1995.

[3] Luis Grangeia. Dns cache snooping or snooping the cache for fun and profit. http://cs.unc.edu/~fabian/course_papers/cache_snooping.pdf, February 2004.

[4] M. Kucherawy. Email authentication status codes. https://tools.ietf.org/html/rfc7372, September 2014.

[5] The Spamhaus Project Block List. SBL - spamhaus DNSBLs. http://www.spamhaus.org/sbl/, 2004.

[6] Pingability. Web check and alert service. https://pingability.com/smtptest.jsp, 2016.

[7] Irina Reznik. Russian oil billionaires' next big investment - american health care. http://www.bloomberg.com/news/articles/2016-07-14/russian-billionaires-plan-u-s-health-push-with-d-c-insiders, july 2016.