Characterization of the Dynamics and Interactions of Domain Names and Name Server

Mansurul Bhuiyan[1], Aziz Mohaisen[2], Yannis Labrou[2] and Mohammad Hasan[1]
[1]Dept. of Computer Science, Indiana University—Purdue University, Indianapolis, [2]Verisign Labs
[1]{mbhuiyan,alhasan}@cs.iupui.edu, [2]{amohaisen,ylabrou}@verisign.com

I. Introduction

In the domain name system (DNS), the life cycle of a domain starts by its registration under one of the TLDs (Top Level Domains) registries followed by a pairing with a name server that serve as a gate keeper of the domain name. Name servers, specially the authoritative, are a significant entity of a domain's operation as they tell users where to look for the domain. It is desirable to maintain multiple name servers for a domain name for many reasons, including fault tolerance, load balancing, and geographical load distribution.

There exists a large number of domains that do a frequent name server switchings, even across several service providers. This phenomena of name server switching is not usual because it demands interference. In [1], the author interprets name servers switching as a hiding mechanism of a domain, and show that domains with such behavior tend to display unsavory behavior i.e. hosting malware, pornography, and un-authorized pharmacy, among others. In [2], researchers used the number of name servers of a domain within a period of time as a feature to develop a classifier for detecting malicious domain names. In [3], researchers developed an inference system to build proactive blacklist of domains where they used the name servers information of already blacklisted domains.

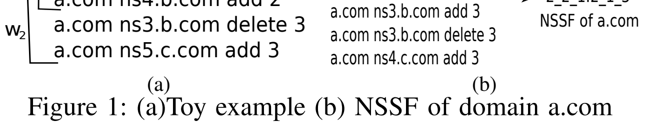
In this work, we build on [1], [3], [2] and look deeper into the subject matter by considering the evolution of the name servers association with a domain name. We use this evolution information to build an identifier called "NS-Switching Footprint" (NSSF). Following the same motivation as in [3], we propose to use the "NSSF" as a vital artifact for building an inference-based security system that evaluates domain names based on their name-servers dynamics. This footprint is robust in comparing NS behavior of blacklisted and potential future bad domains to be considered for blacklisting. To demonstrate the efficiency of "NSSF", we show that domains of similar types tend to have the same footprint: we identify two special types of domains: takeover domains and domains that work for increasing page-rank of some third party domains. Finally, we also developed a prediction model, that can predict how many name servers that a domain might interact with given the history of name server interaction of the respective domain. This prediction model will allow us to get an idea of what type of name server switching behavior that the domains of special interest will exhibit and allow us to plan actions based on that.

II. Dataset

Upon activation of a domain, the registry stores the DNS information of the domain in the corresponding DNS **zone file**. Any operation in the zone file can be considered as a transaction. The addition, removal or updates of domains, as well as the corresponding name servers, are considered zone impacting transactions. Verisign Inc maintains a dataset called Domain Name Zone Alerts (DNZA) that stores all .com and .net zone impacting transactions ordered by their occurrence. In this work, we use DNZA data for the period of March 28, 2013 to June 27, 2013. We only consider domains that are registered within these 90 days. We have approximately 31 million transactions, 350K transactions per day, 7.9 million domains and 480K unique name servers.

III. Proposed Measure of Switching Behavior

NS Switching. Let a window (w) be defined as a succession of at least one add operation followed by at least one delete operation followed by at least one add operation. Let a transition be defined as the set-theoretic difference of the set of name servers of a window (NSw) and the set of name servers of the previous window. A NS switching occurs when the transition set is non empty. The intuition is that by ignoring individual successive additions/deletions of NSs and instead focusing on aggregate changes, we will capture significant changes in the state of a domain's NS services provider. A possible, future refinement of this metric, might be to take into account changes in the namespace of NS's in the transitions. Figure 1(a) illustrates a simple example of NS switching behavior. According to the definition of window, in figure 1(a) there are 2 windows, w_1 and w_2. We can assume that $|NSw_0| = 0$. At the end of window 1, we have $NSw_1 - NSw_0 \; != \; \emptyset$, so we record one NS switching of domain "a.com". Similar conclusion can be made by comparing NSw_1 and NSw_2.



```
a.com ns1.b.com add 1
a.com ns3.b.com add 1
a.com ns1.b.com delete 2
a.com ns4.b.com add 2
a.com ns3.b.com delete 3
a.com ns5.c.com add 3
```

(a)

```
a.com ns1.b.com add 1
a.com ns3.b.com add 1
a.com ns1.b.com delete 1
a.com ns3.b.com delete 1
a.com ns3.b.com add 3
a.com ns3.b.com delete 3
a.com ns4.c.com add 3
```

(b)

Figure 1: (a)Toy example (b) NSSF of domain a.com

NSSF: NS Switching Footprint. The NSSF is a domain name's unique identifier for characterizing the pattern of name servers switching over time. In Figure 2, we present

domain	Switching count	Type of domains
amazingweb007.com	164	Adult Dating
teknotigr.com	151	Empty Blog
climate13.com	148	Fake Conference
zqbifen8.com	84	Advertisement
dxsmalvn.com	81	Page Not found

Table I: Top 5 switching domains

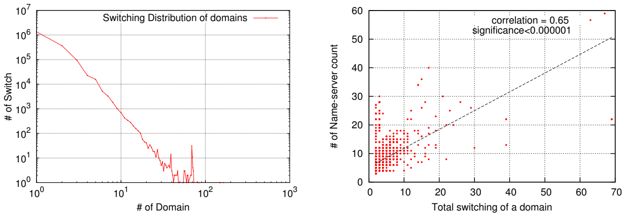
the pseudo code for building the NSSF for a domain. In the footprint we incorporate the number of name servers added and deleted along with the time period of these operations. Since we have 90 days of data, we set length of each time period to one day. Figure 1(b) illustrates a simple example on NSSF building.

```
# T = Total time period = 90
**Build_NSSF**(domain $d$):
**for** $d$ exists in $t$ time period where $t \in (1, T)$:
    NSSF = concate(NSSF, concate(#ofAdded-NS(d), #ofDeleted-NS(d) ,
                t,sep="_"), sep=":")
```

Figure 2: Footprint Building Algorithm

IV. Observations and Evaluation

Analysis of switching phenomena. We ran our algorithm to compute the total switchings of name servers for each domain in the DNZA dataset. We notice that 25% of all domains perform at least 1 NS switching. Figure 3(a) shows the switching distribution of domains, where the distribution exhibits a power law characteristic. In Figure 3(b) we plot the total number of switching versus the number of unique name servers of a domain. We also fit a straight line to demonstrate the positive correlation between the number of name servers and the count of overall switching of a domain. We also observe that most of the domains with higher switching count tend to exhibit unusual behavior/type as discussed in section I. Table I shows the findings, where we can see that "amazingweb007.com" is an adult dating website which does not follow mainstream dating website concept, "teknotigr.com" is an empty blog that has a lots of NS activity, "climate13.com" is a fake conference website mentioned in scamwarners.com, and "dxsmalvn.com" is an NS active website but does not load.



(a) (b)

Figure 3: (a) Log-log plot of switching. (b)Scatter plot of switching vs # of name server of all domain

Clusterings domains on footprint. We first compute the NSSF of each domain from its zone impacting transactions extracted from the DNZA dataset, then we cluster the domains on footprint. We pick only those clusters that have size $\geq$ 10 and have footprint of length $>$ 5. Using that, we ended up with 2604 domains in 84 clusters with median and maximum size of 23 and 99, respectively. After analyzing the clusters, we identify two special types: take over domains and domains that work to increase the page rank of third party domains. We found 31 takeover domains in 2 clusters of size 18 and 13 with NSSF of length 40. All of these domains used DNS providers that have the service of domain parking. By examining their WHOIS information, we observe that all of these domains are owned by one entity. In the second case, we found a set of domains that maintain links (urls) to the target domains to increase their page rank. After analyzing the contents hosted on these domains, we found that all of these domains are synthetically generated; we also found 343 domains work together to increase the page rank of 32 third party domains, where all of them are owned by one entity and maintain similar patterns of NS switching. In figure 4, we show the distribution of incoming links from 343 domain to the 32 third party domains.

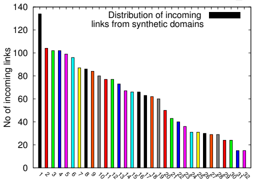


Figure 4: Distribution of incoming links

Prediction model. We build a time series based prediction model to predict the number of name servers that a domain will interact with at any time, given the history of interaction of the respective domain. We use Autoregressive Integrated Moving Average (ARIMA) [4] model for prediction. In this model, instead of time, we consider ordered events to build the time series data. We trained our model with populated data series and predicted the number of interactive name server for future events. To validate our model, we omit the data points for latest 2 events, predict them, and compute the error using the mean sum of squares (MSE). For space constraints in Table II, we present prediction result for top three switching domains mentioned in Table I.

Domain	Ture value	Predicted value	MSE
amazingweb007.com	2 , 2	2.47 , 2.23	0.136
teknotigr	2 , 1	2.4 , 1.3	0.125
climate13.com	2 , 2	2.35 , 2.15	0.072

Table II: Prediction accuracy for top three switching domains

V. Conclusion

In this work, we look at name server switching of domains, with the potential abuse and security as the motivation to understanding name server dynamics. We use the evolution of name servers to build an identifier for domains that can be used to group domains of similar behavior. We have also presented a time series based number of name servers prediction model for domains.

References

[1] T. Snoke, "Watching domains that changes dns servers frequently," CERT/CC Blog, 2013.

[2] Y. He, Z. Zhong, S. Krasser, and Y. Tang, "Mining dns for malicious domain registrations," in *CollaborateCom*, 2010.

[3] M. Felegyhazi, C. Kreibich, and V. Paxson, "On the potential of proactive domain blacklisting," in *LEET*, 2010.

[4] G. Box and G. Jenkins, *Time series analysis: Forecasting and control*, 1970.