Detecting Malicious Activity with DNS Backscatter

By Allison McDonald and Taeju Park

Paper Reference

Kensuke Fukuda, *et al.*, "Detecting Malicious Activity with DNS Backscatter," *Proc. of IMC '15*, October 28-30, 2015, Tokyo, Japan.

Summary

This paper has introduced a new way to detect anomalies for network-wide activities. In order to detect the anomalies, this paper introduced "DNS backscatter", a new source of information on network-wide activity. Network-wide activity is defined as a single source computer touching many target computers. When the activity contacts the target computers, they will sometimes perform a reverse DNS query of the source IP to collect more information about the reason for the connection. The main goal of this paper is to classify the network-wide activities based on machine-learning classification algorithms using static (specific words in the reverse domain name) and dynamic features (temporal and spatial aspects of queries) of the reverse DNS queries. The classification is done at the DNS nameserver (the authority) by gathering the multiple reverse DNS queries generated by the targets.

This paper evaluated the proposed method with the DNS datasets from three authorities and they have labeled the groundtruth of the dataset. With the 60% of the labeled data, they train the classification model using three classification algorithms and measure the accuracy of the model with the remaining 40% of the data. The random forest classification whose accuracy is about 70-80% outperforms the other algorithms. Even though the accuracy is not high enough to use solely, it can complement the prior solutions such as darknets.

Review

Strengths

1. Introduce a new method to analyze the network activities based on DNS backscatter

Since the Internet is highly distributed and decentralized, analyzing the widespread network activity is challenge because aggregating enough information from similar sources to analyze the network activity is really difficult. In this paper, the authors have introduced a new method to analyze the network activities based on DNS backscatter, reverse DNS queries made by targets of widespread network activity and collected by DNS authorities, to overcome the challenge. Even though DNS backscatter cannot capture the entire network activity due to DNS caching and target's configuration, this paper shows that careful interpretation of collective behavior of DNS backscatter can identify originator's application with the reasonable accuracy.

2. Verify the application class of originator in manual

The authors of this paper have labeled the groundtruth on the given dataset to train their classifier and evaluate the accuracy of the trained model. If the labeled groundtruth is wrong, their classifier would make a wrong decision and the measured accuracy becomes unreliable. To avoid this tragedy, the authors of this paper have verified the labeled groundtruth manually. In the paper, they describe how to manually classify the originators into twelve classes such as ad-tracker, cdn, p2p, spam and etc. This manual verification process enhances the credibility of their work.

3. Classification and Feature Selection

While building the machine learning system, the authors defined a group of static features, dynamic features, and classes of application. The static features were derived from the domain names of the querier servers, which often contain *mail*, *ns*, *www*, etc, and shed light on the type of server conducting the reverse query. These features are not particularly innovative. However, the authors also define a group of novel dynamic features that look at the spatial and temporal qualities of the reverse queries. This includes things like the volume and persistence of queries relating to a particular originator and the spatial diversity of queriers. These features, while difficult to obtain data for because they require large and lengthy data sets, are useful for conceptualizing the qualities and problems of DNS backscatter and could be useful for future work.

Similarly, the set of application classes that the originators are sorted into are also a novel addition to network-wide activity analysis. Presumably the authors put considerable thought and analysis into identifying these 12 categories of application, which inform a larger picture of batch Internet activity. On the other hand, as we will discuss below, the manual definition of these categories may be limited by the imaginations of the authors -- for example, if they had failed to identify a particular type of event -- and may not have protracted usefulness with the speed at which Internet usage changes.

Weaknesses and Extensions

1. Random choice of training data

To evaluate their work, the authors of this paper pick 60% of the labeled data randomly as the training data. However, this random choice can cause overfitting because we cannot assure that the randomly chosen data is well-balanced across the all classes with the insufficient data. Machine-learning community usually do ‘k-cross validation’ to avoid the overfitting problem. They partition the given dataset into k-subset and do k-round test. For each round, they select one subset in order to train a model and the remaining k-1 subsets are tested with the trained model. Consequently, they can estimate accuracy of the model properly by using the average the results of k-round tests.

2. Generalizability

It could also be called into question how much this analysis can be generalized. Not all of the targets of a network-wide event generate reverse DNS requests, and many of the ones that do could be responded to by lower-level DNS servers or hidden by caches. Additionally, the requests we do see are made not by the targets, but by the queriers, who sit in between the targets and the DNS servers (eg, middleboxes, firewalls, and spam filters). This can be informative but can also obscure information about the target computers. However, the authors did do a good job of recognizing the constraints of their data and their analysis by recognizing these limitations in section 3.5.

3. Usefulness of DNS Backscatter

The authors cite one motivation of the paper to be the ability to contribute to the detection and response to new or abnormal network-wide activity, such as a new attack. We are quite skeptical of the plausibility of this claim.

First of all, the trend in large-scale attacks has been towards decentralization and distribution. For example, in October 2016 we witnessed the highest bandwidth distributed denial of service (DDoS) attack that has ever occurred, which effectively prevented large portions of the United States from accessing content from CDNs serving Twitter and Facebook, among many other smaller sites. DNS backscatter (disregarding the fact that this particular attack was on a DNS service itself) is far from the most useful tool in detecting an attack stemming from many different IPs. And with the rise of the Internet of Things and the relative ease of renting time and bandwidth from existing botnets, monitoring these network-wide activities (remembering that they're defined as stemming from a single computer) may not be useful for much other than tracking normal, benign traffic.

Additionally, in order to run the analysis on this data at a useful scale, DNS providers would need to coordinate and cooperate to share and analyze data, which is a huge bureaucratic undertaking and unlikely to happen. If one top-tier DNS provider decided to do the monitoring alone, the diversity of the data possibly be enough to yield useful information, but the benefits may still outweigh the cost, for example of collecting and retraining data over time to continue having up-to-date information. However, we are skeptical that there is incentive for any DNS provider to do this type of monitoring, especially considering other ways to monitor network-wide activity.

Finally, the authors claim that a 70-80% precision is sufficient for this type of analysis. Considering the other factors discussed in this section that make DNS backscatter a less appealing method for watching this activity, we were not impressed with this accuracy.

Discussion

In this paper, the authors considers twelve classes of originator. We think that the twelve classes may not cover the originator's future behavior, both because an attacker could find new attack vectors to cause malicious behavior of target and because the general shape of even benign network traffic is apt to change as the Internet continues to expand. Thus, these pre-defined classes might not be sufficient in the future due to lack of scalability. We think that unsupervised clustering algorithms such as k-means clustering can be a good alternatives to define classes.

Even though the unsupervised clustering can reflect the future diversity well, we think that the proposed method will lose its benefits because the accuracy of classifier depends on the number of classes that the classifier considers. The authors mention that if the classifier considers fewer classes, the accuracy of the classifier increases. In other words, if the classifier considers more classes, the accuracy of the classifier could decrease. We therefore question how long this method will yield interesting and useful information.

In all, the authors propose a novel and interesting method for identifying network-wide activity through backscatter DNS requests. Although we are skeptical of the current and future usefulness of the analysis method and resulting data, the authors have conducted a worthwhile experiment and have explored the information available from DNS backscatter.