Improving Long-term Accuracy of DNS Backscatter for Monitoring of Internet-Wide Malicious Activity (poster)

Poster, March 2016; USC/ISI Technical Report ISI-TR-707, April 2016

Abdul Qadeer[1] John Heidemann[1] Kensuke Fukuda[2]
1: USC/ISI Los Angeles 2: NII Tokyo

ABSTRACT

Internet-wide malicious activities are prevalent on the Internet. Such activities include the malicious, like spamming and scanning, and the benign, like large e-mailing lists and content delivery networks.

We've previously shown that they can be detected centrally with DNS backscatter, and developed a classifier using supervised learning [1]. However, long-term detection is difficult because activities rapidly change with time to evade detection or as they naturally evolve, and manual training is expensive.

Our solution: we extend backscatter-based detection by identifying: how behavior evolves, how often we need to retrain, and how to retrain without human supervision.

Details are in the attached poster.

Acknowledgments

The work of Abdul Qadeer and John Heidemann here is partially sponsored by the Department of Homeland Security (DHS) Science and Technology Directorate, HSARPA, Cyber Security Division, via SPAWAR Systems Center Pacific under Contract No. N66001-13-C-3001 (Retro-Future), and via BAA 11-01-RIKA and Air Force Research Laboratory, Information Directorate under agreement number FA8750-12-2-0344 (LACREND). The U.S. Government is authorized to make reprints for Governmental purposes notwithstanding any copyright.The views contained herein are those of the authors and do not necessarily represent those of DHS or the U.S. Government.

Kensuke Fukuda's work here is partially funded by Young Researcher Overseas Visit Program by Sokendai, JSPS KAKENHI Grant Number 15H02699, and the Strategic International Collaborative R&D Promotion Project of the Ministry of Internal Affairs and Communication in Japan (MIC) and by the European Union Seventh Framework Programme (FP7/2007- 2013) under grant agreement No. 608533 (NECOMA). The opinions expressed in this paper are those of the authors and do not necessarily reflect the views of the MIC or of the European Commission.

REFERENCES

[1] Kensuke Fukuda and John Heidemann. Detecting malicious activity with DNS backscatter. In *Proceedings of the ACM Internet Measurement Conference*, pages 197–210, Tokyo, Japan, October 2015. ACM.

Improving Long-term Accuracy of DNS Backscatter for Monitoring of Internet-Wide Malicious Activity

Abdul Qadeer
University of Southern California / ISI
aqadeer@isi.edu

John Heidemann
University of Southern California / ISI
johnh@isi.edu

Kensuke Fukuda
National Institute of Informatics / Sokendai
kensuke@nii.ac.jp

INFORMATION SCIENCES INSTITUTE

USC Viterbi School of Engineering

Problem: Accurately Detect & Track Malicious Activity

Internet-wide malicious activities are prevalent on the Internet. Such activities are:

- **Malicious**: Spamming, Scanning
- **Benign** : Large e-mailing lists, Content delivery network

We've previously shown that they can be detected centrally with DNS backscatter, and developed a classifier using supervised learning[1]. However, long-term detection is difficult because **activities rapidly change with time** to evade detection or as they naturally evolve, and **manual training is expensive**.

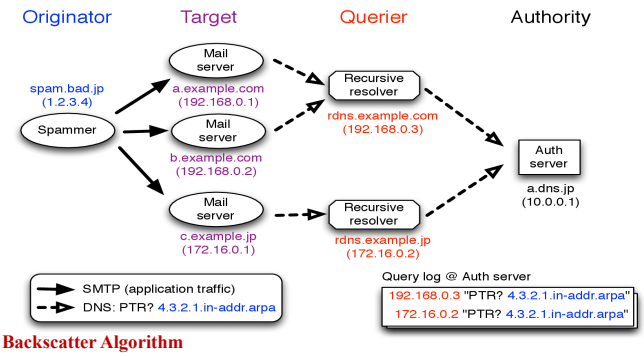
Our solution: we extend backscatter-based detection by identifying:

- how behavior evolves,
- how often we need to retrain,
- and how to retrain without human supervision.

Background: [1] Detecting Malicious Activity with DNS Backscatter, Kensuke Fukuda and John Heidemann, Proceedings of the ACM Internet Measurement Conference (Tokyo, Japan, Oct. 2015), 197–210.

When an originator sends network traffic to a target, some entity in the target organization does a 'originator IP to human readable name' query using DNS PTR facility. Such DNS PTR queries arrive at some DNS servers and these are called **DNS backscatter**.

If some originator sends traffic to many different targets, that results in amplified DNS backscatter traffic and becomes a **detectable signal of a network wide activity**.

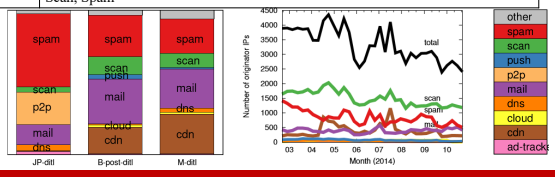


Backscatter Algorithm

- Accumulate DNS backscatter from a DNS server (e.g. a root server or an authoritative DNS server)
- Set a threshold that how much backscatter per originator will trigger the alarm for network wide activity
- Curate labelled examples from available data
- Make feature vectors using static and dynamic features for the originators
- Use supervised machine learning algorithms (e.g. random forest) to classify network wide activity into application classes

Static Features: Extracted from DNS PTR Replies Conducted on Querier IPs			
Keyword Based	Home, Mail, NS, FW, Antispam, WWW, NTP, AWS		
Well Known Organization	CDN (Akamai, Edgecast, CDNetworks etc.) AWS, MS, Google		
Others	Nxdoman , unreachable and other non-classifiable		
Dynamic Features: Extracted from Originator's DNS Backscatter Data			
Queries per querier	Query persistence	Local entropy	Global entropy
Unique ASes	Unique countries	Queries per country	Queries per AS
Labels (Application Classes)			
Benign	Ad-tracker, CDN, Cloud, P2P, Push, DNS, Mail, NTP		
Malicious	Scan, Spam		

Backscatter Algorithm In Action



Challenge

How is Machine Learning (ML) classifier performing over time in terms of accuracy, precision, recall and f-score?

Malicious traffic changes its characteristics often with time to evade detection. It causes the ML feature vectors to change over time. If the feature vectors associated with human supervised labelled examples are not updated according to **changing world**, the ability of classifier to correctly work diminishes with time. Additionally it is desirable to automate the process of updating of labelled examples feature vectors to **minimize human intervention** so that algorithm could work on continuous basis.

Approach

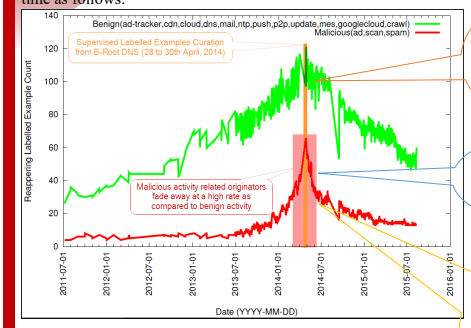
Our **insight** is that feature vectors of human supervised labelled examples can be updated by an algorithm, as long as sufficient number of originators present in labelled examples remain active in current data. We have extended base backscatter algorithm by:

- Measuring ML performance on daily basis
- Updating feature vectors on daily basis to keep track of world changes
- Retraining classifier with updated feature vectors
- Precisely declaring when human intervention is needed

We use **experimental approach** to validate our extended backscatter algorithm on **multi-year B-Root's DNS data**.

Extended Backscatter Algorithm Results

Research question: Does supervised labelled examples (originators) continue their activities beyond curation day? If yes for how long?
Answer: Count re-appearance of curation day originators of malicious and benign activities over time as follows:

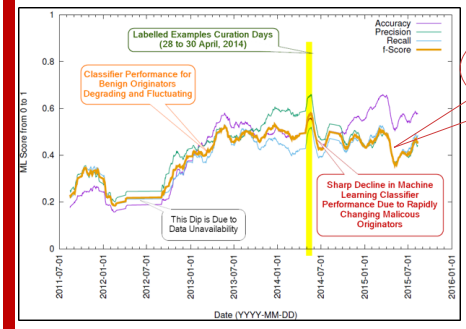


Benign originators mostly use well publicized IPs, which does not change over months

Malicious originators cycle their IPs within a month to evade detection

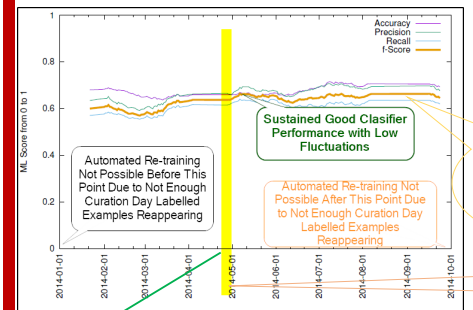
We precisely know when some application class labelled examples go below a certain threshold. At that point new supervised labels should be added by a human expert

Research question: How backscatter algorithm accuracy changes over time with one curation day training?
Answer: Find which curation day originators are present on day n, use 2/3 of them for training and 1/3 for validation. Validation gives us ML performance for day n as follows:



Changing feature vectors over time cause the classifier to perform poorly

Research question: How does ML algorithm performance change after utilizing our proposed solution?
Answer: Find which curation day originators are active on day 'n', extract their new feature vectors from the day, retain the classifier and validate to measure ML algorithm performance for day n. Run these steps over time. The graph is on the left:



For each new day, labelled examples feature vectors are re-evaluated afresh and classifier retrained

Original supervised label examples curation days

- Malicious originator are tracked with good sustained performance for 15 days beyond the curation day on either side
- We cannot measure malicious originator related classification performance beyond 15 days on either side of curation day due to insufficient number of curation day originators reappearing
- Accurately tracking benign activity over about 8 months is feasible, without the need for fresh labelling

Conclusion

- Malicious activities on the Internet are highly volatile over time.
- Supervised learning algorithms need to adapt according to changing world so that we could track malicious activity correctly.
- Our extended backscatter algorithm regularly updates feature vectors and retrains the classifier to sustain good accuracy.
- Our extended algorithm precisely informs when fresh labelling effort is required. In our experimental data its after 30 days for malicious activities and 8 months for benign activities.