

## Distributed Denial-of-Service: What Datasets Can Help?


John Heidemann<sup>1,2</sup>

joint work with Terry Benzel<sup>1</sup>, Wes Hardaker<sup>1</sup>, Christian Hesse<sup>1</sup>, Zi He<sup>1,2,7</sup>, Allison Mainkin<sup>4,8</sup>,  
Urbashi Mitra<sup>3</sup>, Giovanni Mura<sup>3</sup>, Moritz Müller<sup>1</sup>, Ricardo de O. Schmidt<sup>1</sup>, Nikita Somaiya<sup>2,7</sup>,  
Gautam Iyengar<sup>3</sup>, Wouter de Vries<sup>3</sup>, Lan Wei<sup>1</sup>, Diane Wessels<sup>1</sup>, Liang Zhu<sup>1</sup>

<sup>1</sup>USC/ISI <sup>2</sup>USC/CS <sup>3</sup>SIDN Labs <sup>4</sup>Verisign Labs <sup>5</sup>USC/EE <sup>6</sup>U. Twente  
(and now at <sup>7</sup>Amazon, <sup>8</sup>Salesforce)

ACM ACSAC, Los Angeles, 2016-12-07

Copyright © 2016, by John Heidemann  
Release terms: CC-BY-NC 4.0 international



USC Viterbi School of Engineering

## Network Security Needs Data

- DDoS attacks
- DNS privacy leaks
- DNS filtering and censorship
- experimenting to test solutions



[After J. M. Flagg's poster, with apologies to Schneier]

**Security Needs Data!**

USC Viterbi School of Engineering

DDoS Datasets / 2016-12-07

## DDoS: Bad and Getting Worse


- big and **getting bigger**
- easy and **getting easier**
- frequent and **getting frequent-er**

USC Viterbi School of Engineering

DDoS Datasets / 2016-12-07

## Network Security Needs Data

- DDoS attacks
  - like the Oct. 2016 Dyn attack
- DNS privacy leaks
  - what does DNS say about you?
- DNS filtering and censorship
  - multiple countries (Turkey, China filter DNS)
- experimenting to test solutions
  - does my fix help?



[After J. M. Flagg's poster, with apologies to Schneier]

**Security Needs Data!**

USC Viterbi School of Engineering

DDoS Datasets / 2016-12-07

## DDoS: Bad and Getting Worse

- big and **getting bigger**
  - 2012: first 100Gb/s [Arbor12a]
  - 2016: 100Gb/s common; Oct.: 1Tb/s vs. Dyn
- easy and **getting easier**
  - 2012: several 1000+-node botnets
  - 2016: 10k+ nodes and DDoS-as-a-service: \$1/attack
- frequent and **getting frequent-er**
  - 2002: the October 21 DNS root event
  - 2016: 3 recent big attacks (2015-11-30, 2015-12-01, 2016-06-25)

USC Viterbi School of Engineering

DDoS Datasets / 2016-12-07

## USC Has DDoS-Relevant Data


- detecting Distributed Denial-of-Service
- understanding effects of DDoS
- evolving DNS to prevent DDoS and improve privacy
- DNS as a data source and as a target platform

USC Viterbi School of Engineering

DDoS Datasets / 2016-12-07


## USC Has DDoS-Relevant Data

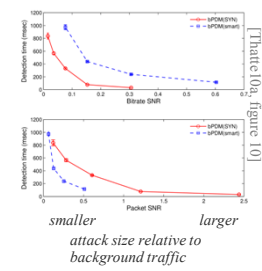
- **detecting Distributed Denial-of-Service**
- understanding effects of DDoS
- evolving DNS to prevent DDoS and improve privacy
- DNS as a data source and as a target platform


DDoS Datasets / 2016-12-07
7


## Judging Effectiveness via Controlled Attacks

- goal: detecting low-rate attacks
- judging sensitivity? *need test data*
- we generated *synthetic* test data
  - mix controlled attack traffic
  - in with real-world traffic
  - at *controlled* rates (the SNR)
- dataset: UniformAttack\_Traces\_Generated20070821-20041202

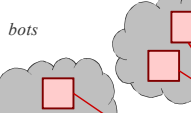

DDoS Datasets / 2016-12-07
10




## DDoS' Cumulative Power



bot master



bots




victim

DDoS is obvious at the victim: too much traffic


can we move detection near the bots?

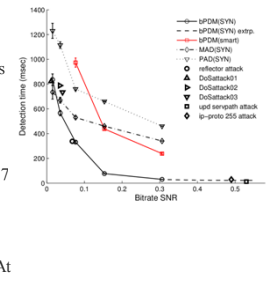
DDoS' strength: thousands of small attackers (in Mirai: IOT); there it is *not obvious*


DDoS Datasets / 2016-12-07
8

## Judging Effectiveness via Real-World Attacks


- goal: confirming results in practice
- replay and try detection:
  - synthetic attacks
  - *and* collection of real-world attacks
- datasets:
  - DoS\_traces-20020629
  - DoS\_80-20110715
  - DoS\_DNS\_amplification-20130617
  - DARPA\_2009\_DDoS\_attack-20091105
  - DARPA\_2009\_malware-DDoS\_attack-20091104
  - FRGP\_SSDP\_Reflection\_DDoS\_Attack\_Traffic-20140930


DDoS Datasets / 2016-12-07
11




## Challenge: Detecting Low-Rate DDoS

- catching bots is part of stopping DDoS
- DDoS traffic is low-rate at the bots
  - can detection be sensitive enough?
  - and can we do it in *aggregate* traffic? (to avoid expensive flow separation)
- approach:
  - model background traffic as Poisson (not correct, but sufficient)
  - apply Sequential Probability Ratio Test
  - result: rapid and sensitive detection
- details:
  - Thatte, Mitra, and Heidemann. Parametric Methods for Anomaly Detection in Aggregate Traffic. *ACM/IEEE Transactions on Networking*, V. 19 (N. 2), August, 2010. <http://dx.doi.org/10.1109/TNET.2010.2070845>


DDoS Datasets / 2016-12-07
9

## Data for DDoS Replay

- these datasets can test *your* DDoS detection algorithms
- paper about our approach and datasets
  - Thatte, Mitra, and Heidemann. Parametric Methods for Anomaly Detection in Aggregate Traffic. *ACM/IEEE Transactions on Networking*, V. 19 (N. 2), August, 2010. <http://dx.doi.org/10.1109/TNET.2010.2070845>
- our datasets
  - <https://impactcybertrust.org>
  - <https://ant.isi.edu/datasets/all/>
  - look for anything with “DoS” in the title


DDoS Datasets / 2016-12-07
12

## USC Has DDoS-Relevant Data

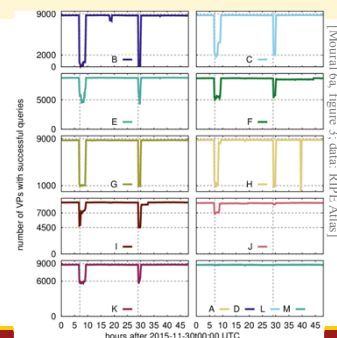
- detecting Distributed Denial-of-Service
- **understanding effects of DDoS**
- evolving DNS to prevent DDoS and improve privacy
- DNS as a data source and as a target platform

## How About the Root Letters?

**some did great:**  
D, L, M: not attacked  
A: no visible loss

**most suffered:**  
a bit (E, F, I, J, K)  
or a lot (B, C, G, H)

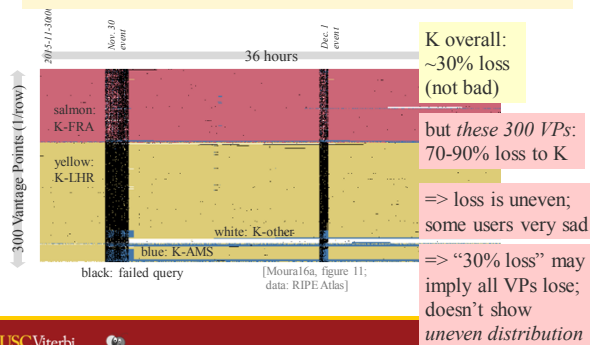
but does “x%”  
measure what  
*users actually see?*



## What are the Results of DDoS?

- we studied the effects of two large DDoS attacks on the DNS root
- goals were to understand
  - what responses *do* happen
  - how should we *quantify* the effects
  - what responses *should* or *could* happen
- details:
  - Moura, Schmidt, Heidemann, de Vries, Müller, Wei, and Hesselman. Anycast vs. DDoS: Evaluating the November 2015 Root DNS Event. ACM IMC 2016. <<http://dx.doi.org/10.1145/2987443.2987446>>

## View from Atlas Vantage Points



## Analysis of DNS DDoS Data

- re-analysis of RIPE Atlas probes
  - queries to each DNS Root letter every 4 minutes
  - from about 9000 places around the world
  - queries return *CHAOS* strings, showing what anycast site the vantage point connects to
  - provided by RIPE: <https://atlas.ripe.net>
- mapping from CHAOS replies to sites
  - we did it by hand; now we’re automating it
  - (work in progress)

## Data for Root DNS DDoS

- reanalysis of RIPE Atlas tells about DDoS response
  - some users will see persistent loss
  - “x% loss” is not complete picture
- paper examines response strategies
  - Moura et al, ACM IMC 2016 <http://www.isi.edu/~johnh/PAPERS/Moura16b>
- can you use the data?
  - <https://impactcybertrust.org>
  - <https://ant.isi.edu/datasets/anycast/>
  - Root\_DNS\_Event-20151130
  - contact me for in-progress CHAOS mapping



## USC Has DDoS-Relevant Data

- detecting Distributed Denial-of-Service
- understanding effects of DDoS
- **evolving DNS to prevent DDoS and improve privacy**
- DNS as a data source and as a target platform

USC Viterbi School of Engineering  
DDoS Datasets / 2016-12-07 19

## Confront Tradition: Connection-Oriented DNS

DNS over UDP\* **Vulnerable!**  
(\*except for zone transfers and fallback)

Amplification **TCP => prevent spoofing** - Service (as victim)

spoofed queries appear from victim

Victim (too many replies)

Victim (too many requests)

**TLS => reduce eavesdropping**

**No Privacy!** mail.sensitive.org ?

USC Viterbi School of Engineering  
DDoS Datasets / 2016-12-07 22

## DNS is Essential

Example.com  
192.0.2.5

DNS is simple request-response

www.example.com ?

DNS

192.0.2.5

Perfect for UDP  
(TCP is supported too, but as fallback and zone transfers)

USC Viterbi School of Engineering  
DDoS Datasets / 2016-12-07 20

## The Challenge

- but won't DNS over TCP and TLS be horrible?
  - everyone knows it won't work
  - DNS has to be UDP
  - you have to use DTLS and not TLS

USC Viterbi School of Engineering  
DDoS Datasets / 2016-12-07 23

## Pitfall of Current DNS

DNS over UDP **Vulnerable!**  
(\*except for zone transfers and fallback)

Amplification

DNS servers

spoofed queries appear from victim

Victim (too many replies)

Denial-of-Service (as victim)

DNS server

Victim (too many requests)

**No Privacy!** mail.sensitive.org ?

USC Viterbi School of Engineering  
DDoS Datasets / 2016-12-07 21

## Data Answers The Challenge

- but won't DNS over TCP and TLS be horrible?
  - everyone knows it won't work
  - DNS has to be UDP
  - you have to use DTLS and not TLS
- no, no, no! (if you're careful)
  - caching works well
  - careful TCP optimizations matter
  - DTLS is exactly the same as TLS (by design!)

**Only Data-driven Experiments can refute incorrect common wisdom**

USC Viterbi School of Engineering  
DDoS Datasets / 2016-12-07 24

## Threat Model

- Denial of Service
- eavesdropping
- weak crypto choices

USC Viterbi School of Engineering  
DDoS Datasets / 2016-12-07 25

## Cost of Connection Reuse? (ok!)

Connections => Memory

method: replay same 3 traces (here we show 2 biggest)  
Assumes Google-style TLS optimizations to 10kB/conn [2]  
(experimental estimate of memory: 360kB/connection, very conservative estimate)  
(graph shows medians and quartiles)

conclusion: connection reuse is often helpful and not too costly

[2] <https://www.imperialviolet.org/2010/06/25/overlocking-ssl.html>

USC Viterbi School of Engineering  
DDoS Datasets / 2016-12-07 28

## DNS-over-TCP: Protocol Optimizations

- Connection reuse
  - Persistent connections
  - TCP fast open
  - TLS resumption
- Query Pipelining
  - Send queries as fast as possible
- Out-of-order processing (OOOP)
  - Server processing in parallel

IETF Internet-Draft draft-ietf-dnscop-5966bis-01  
DDoS Datasets / 2016-12-07 26

## Data for Protocol Design

- trace-driven experimentation shows
  - how much optimizations matter
  - how critical caching can be
  - necessary to correct common wisdom
- paper with details
  - Zhu, Hu, Heidemann, Wessels, Mankin, and Somaiya. Connection-Oriented DNS to Improve Privacy and Security. IEEE S&P, 2015. <http://dx.doi.org/10.1109/SP.2015.18>
- can you use the data?
  - <https://imactcybertrust.com> and [https://ant.isi.edu/datasets/Root\\_DNS\\_Event-20151130](https://ant.isi.edu/datasets/Root_DNS_Event-20151130)

USC Viterbi School of Engineering  
DDoS Datasets / 2016-12-07 29

## Connection Reuse Helps? (YES!)

Suggested connection time-out :  
20 s authoritative servers and 60 s elsewhere

what fraction of queries find open TCP connections?

method: replay 3 traces: recursive (DNSChanger, Level3) and authoritative (B-Root)  
(graph shows medians, quartiles are tiny)

conclusion: connection reuse is often helpful

USC Viterbi School of Engineering  
DDoS Datasets / 2016-12-07 27

## USC Has DDoS-Relevant Data

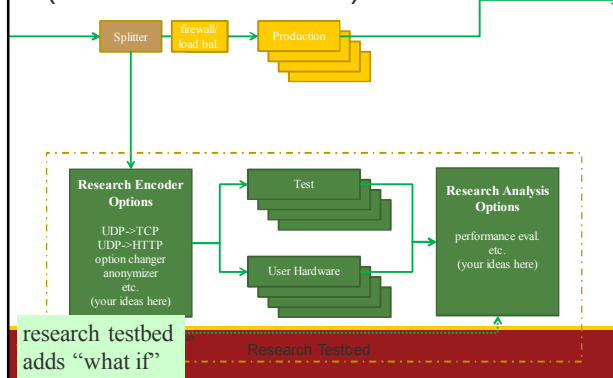
- detecting Distributed Denial-of-Service
- understanding effects of DDoS
- evolving DNS to prevent DDoS and improve privacy
- DNS as a data source and as a target platform**

USC Viterbi School of Engineering  
DDoS Datasets / 2016-12-07 30

### Challenge: Testing Your Ideas

- how do you test *your* ideas?
- where can you get real-world data?
  - that reflect real DDoS events
  - and a *real traffic mix* (good, bad, and ugly)
- experimental test platforms?
  - that run at *operational scales*

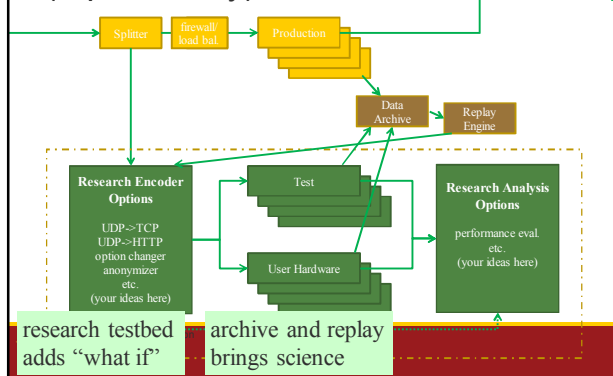
### Parallel Experiments (similar but different)



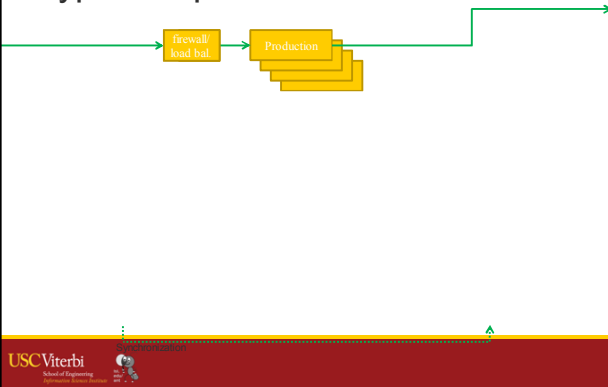
### Our Solution: A Testbed Married With Operations

- B-Root Operations
- ISI background in net measurements and research
- together, they can fill in:
  - sharable long-term data collection, archive and sharing
  - experimentation on a real platform
  - path to deployment for new ideas
  - community built around these ideas

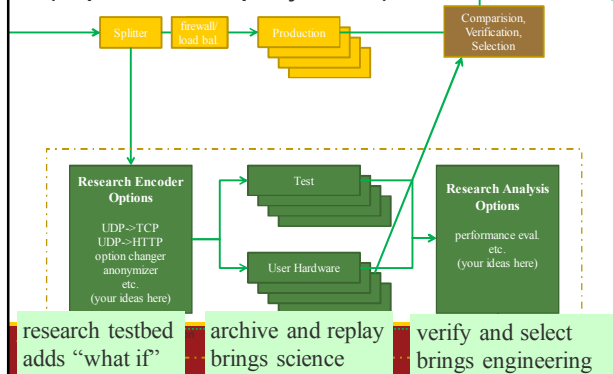
### Long-term, Replayable History (repeatability)



### Typical Operational DNS



### Compare and Validate (a path to deployment)



## Our NIPET DNS Testbed

- **looking for feedback on testbed**
  - <https://ant.isi.edu/nipet/>
  - Join our mailing list (on that page)
  - Send us ideas, suggestions, feedback
- what are *your* use cases?
- **some data available today:**
  - <https://imactybertrust.com> and <https://ant.isi.edu/datasets/>
  - DITL\_B\_Root-20130528, DITL\_B\_Root-20140428, DITL\_B\_Root-20150413, DITL\_B\_Root-20160405

## Conclusions

- lots of DDoS-related data is available
  - <https://imactybertrust.com>
  - <https://ant.isi.edu/datasets/>
- we've used it many ways
  - detecting DDoS
  - evaluating DDoS effects
  - improving protocols to counter DDoS
- and we're planning a testbed for live experiments
- *does this data apply to your work?*