Institutional Privacy Risks in Sharing DNS Data

Basileal Imana, Aleksandra Korolova and John Heidemann

Applied Networking Research Workshop

July 26-28, 2021

USC Viterbi School of Engineering

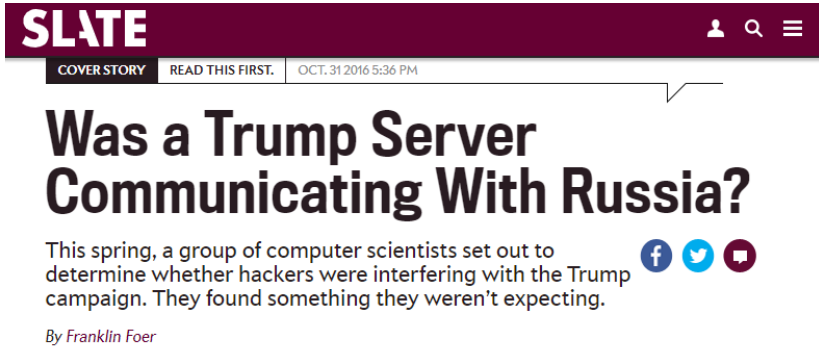
Information Sciences Institute

DNS Queries Leak Data About End-users’ Online Activities

What about Institutional Privacy in DNS?

- Institutional privacy
  - The behavior of an institutions traffic
  - Not closely studied before
  - Vs. individual privacy

- Institutions' internal activities can leave a digital trail in DNS
  - Sending/receiving an email
  - Accessing sensitive websites
  - …



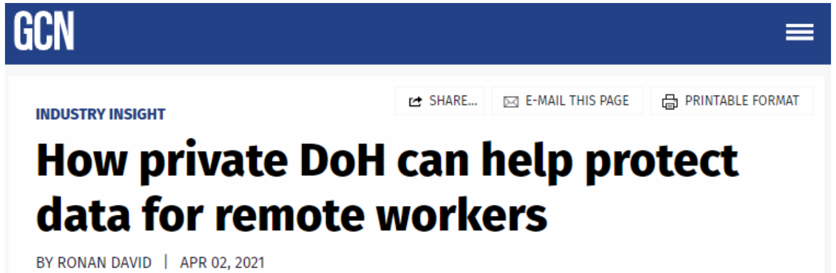
SLATE

COVER STORY | READ THIS FIRST. | OCT. 31 2016 5:36 PM

Was a Trump Server Communicating With Russia?

This spring, a group of computer scientists set out to determine whether hackers were interfering with the Trump campaign. They found something they weren't expecting.

By Franklin Foer



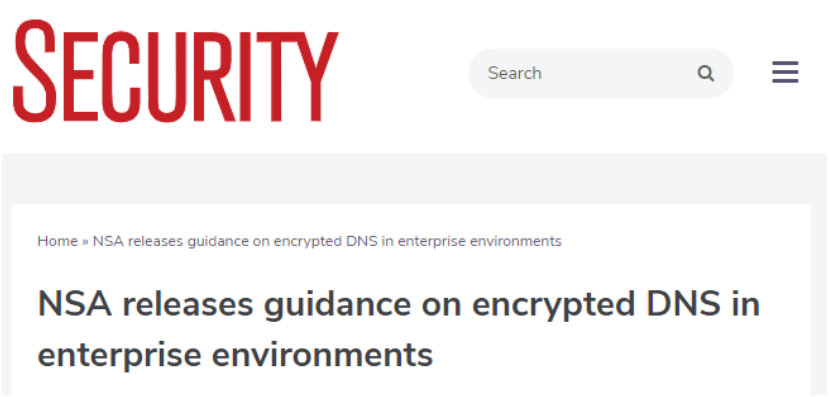
GCN

SHARE... | E-MAIL THIS PAGE | PRINTABLE FORMAT

INDUSTRY INSIGHT

How private DoH can help protect data for remote workers

BY RONAN DAVID | APR 02, 2021



SECURITY

Search

Home » NSA releases guidance on encrypted DNS in enterprise environments

NSA releases guidance on encrypted DNS in enterprise environments

Our Contributions

- We define institutional privacy as a new privacy risk in DNS
- Give a methodology for finding institutional privacy leaks
- Demonstrate the privacy risks using anonymized real-world data
  - Prefix-preserving anonymization not sufficient to prevent institutional leaks

Problem Statement

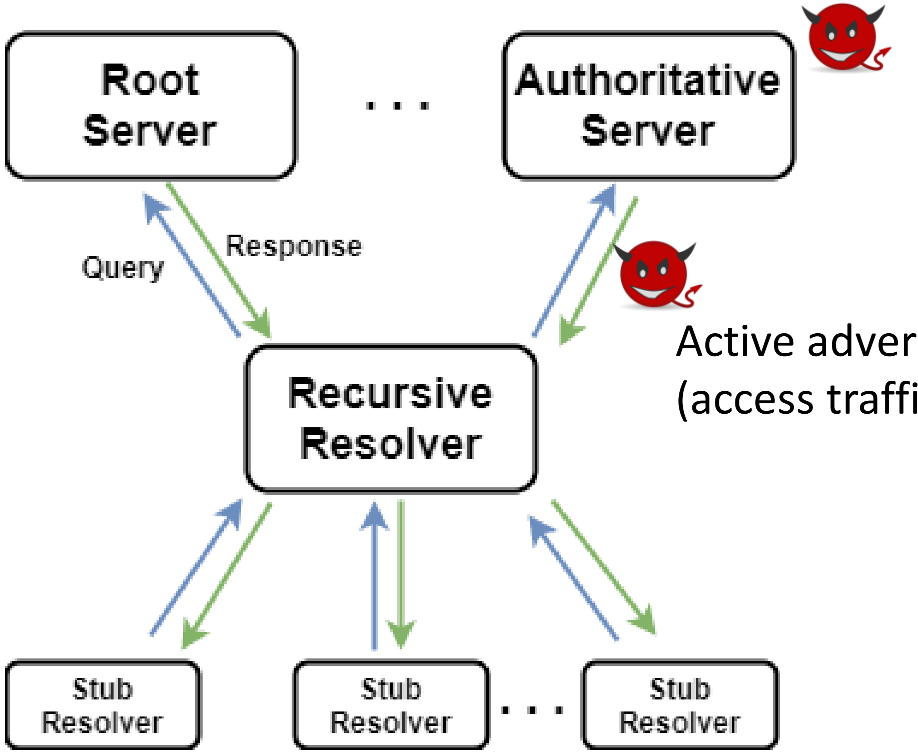
Defining Institutional Privacy in DNS

- Definition: Confidentiality of digital footprints of an institution's internal activities

- Specific activities we look at that may leak information through DNS:
  - Sending/receiving an email
    - May reveal relationships between institutions
  - Accessing privacy sensitive or embarrassing websites
    - May be considered sensitive from a company's PR perspective
    - Example: illegal or adult websites

Threat Model: Who is the Adversary?

- Adversary is at authoritative server
- The adversary sees:
  - Source IP of DNS query
  - Domain looked up
  - Query type
- Goal: associate source IPs and domains to institutions



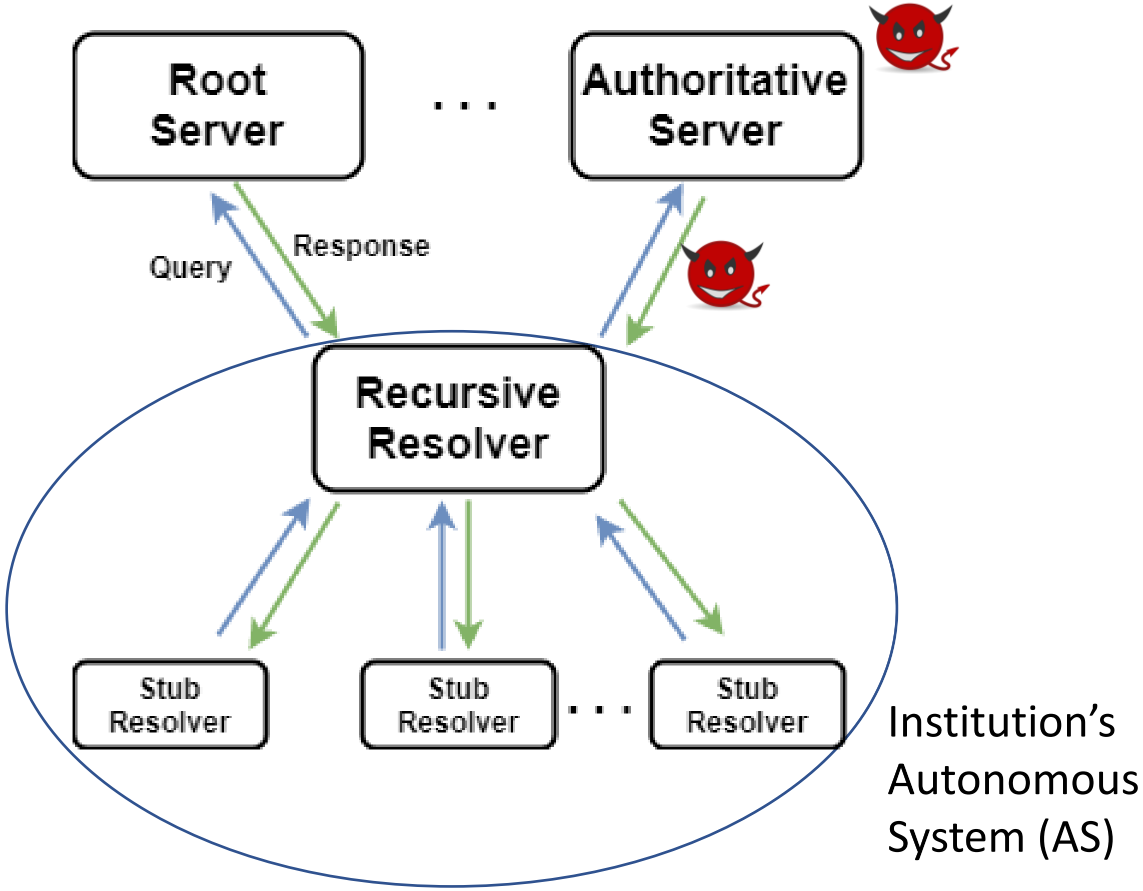
Passive adversary
(access server logs)

Active adversary
(access traffic)

Threat Model: Who is the Target?

An institution that:

1. Runs its own recursive resolver
   - Resolver’s IP can be used to identify the institution’s traffic
2. Routes traffic from its own Autonomous System
   - Resolver's IP can be mapped to the AS the IP belongs to



Many Institutions and Adversaries Fit The Threat Model

- We pick 66 institutions that represent diverse sectors
  - S&P 500 companies, Government institutions, UC Schools, Airlines, …
  - Exclude institutions that have apparent deniability (E.g., ISPs)

- Example of potential real-world adversaries
  - DNS service providers (E.g., Public DNS resolvers)
  - Researchers with access to DNS data (E.g., DITL initiative)
  - Government or state-level actors

Methodology

1. Associating Queries with an Institution

2. Finding Queries Related to Email Exchange

3. [Paper S4.3] Finding Queries to Sensitive Sites

Associating Queries with an Institution

Goal is to find which institutions are associated with a query’s:

1. Source IP
2. Domain name

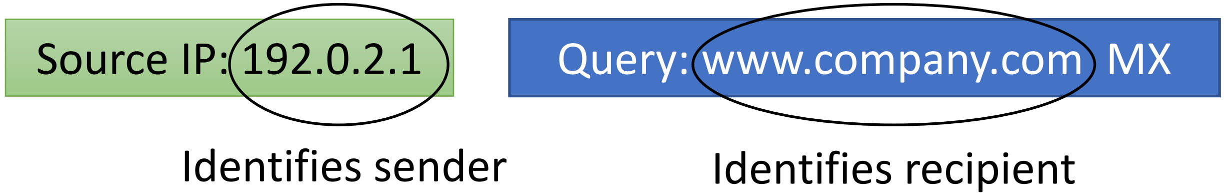
1. Source IP --> Autonomous System Number --> Institution
   - Using public IP to ASN mapping data
   - Works even if partial (host-only) prefix-preserving anonymization is used

2. Domain name --> Domain Owner --> Institution
   - Using public WHOIS data
   - Assumes Qname minimization (QMIN) is not used

Finding Queries Related to Email Exchange

Goal: Find out when an email is sent or received

- Sent: Watch outgoing MX queries



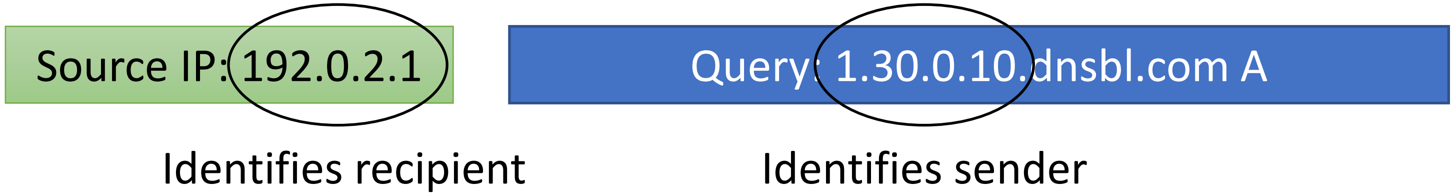
Source IP: 192.0.2.1

Query: www.company.com MX

Identifies sender

Identifies recipient

- Received: Watch DNSBL queries made by anti-spam services



Source IP: 192.0.2.1

Query: 1.30.0.10.dnsbl.com A

Identifies recipient

Identifies sender

Experiment Results

Dataset

- 1 week of b-root data from Jan 9-15, 2019
    - Similar results on a second week

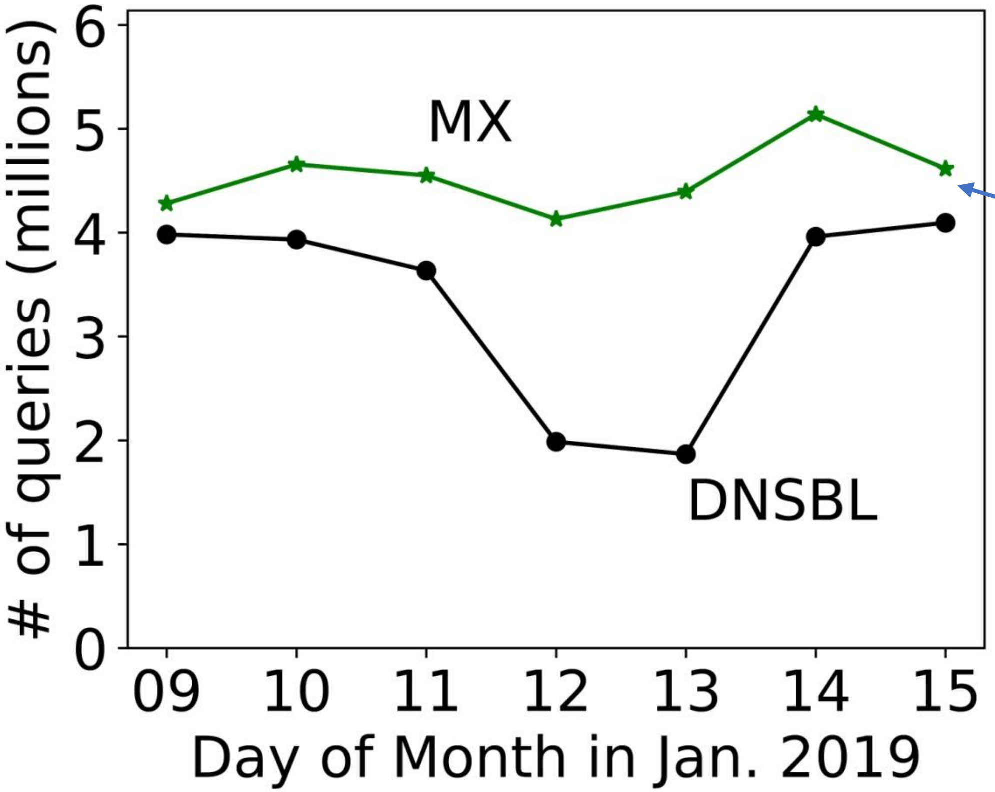
- Source IP addresses are anonymized using prefix-preserving method
    - Bottom-8 bits are anonymized

- Ethics
    - USC IRB#: UP-20-00477
    - Used with permission of b-root operators
    - Agreed to not identify queries that reveal relationships not publicly known

Research Questions

- How common are sensitive email-related queries from institutions?
- Are specific relationships between institutions visible?
- [Paper S5.3] How common are queries to sensitive sites?

How Common Are Sensitive Email-related Queries?

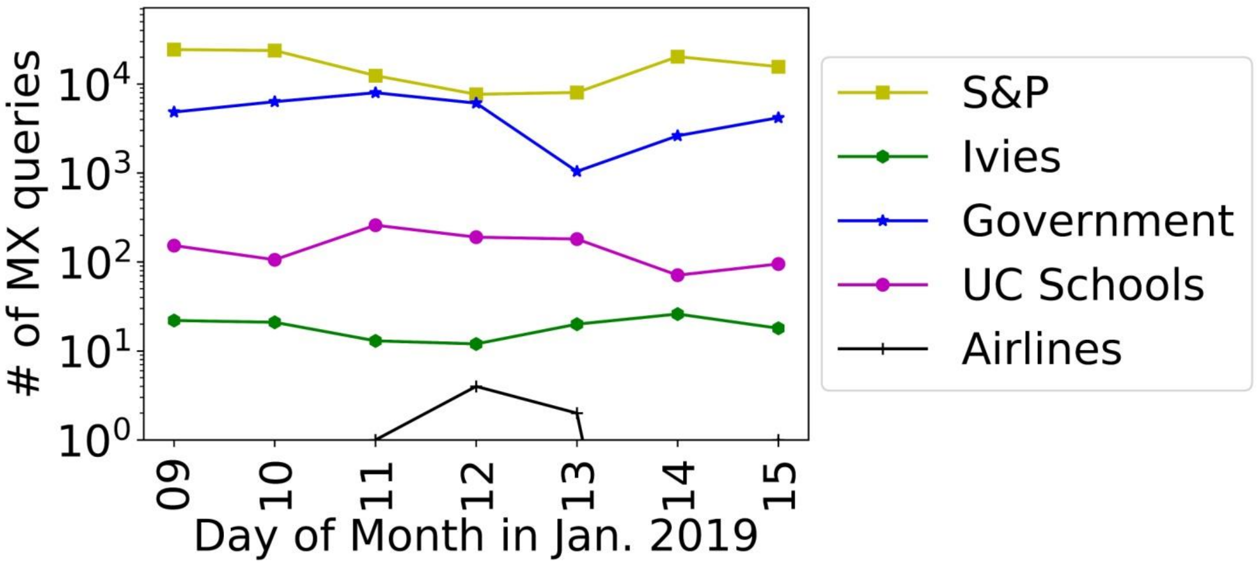


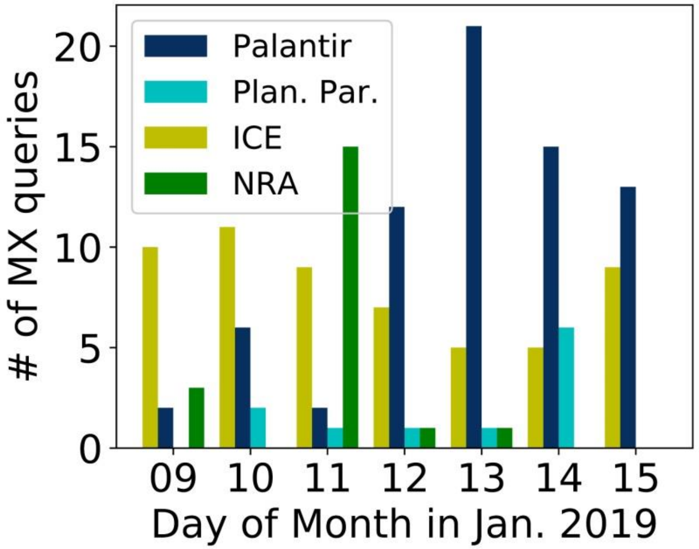
Several millions of DNSBL and MX queries made each day

→ Significant source for leakage of email-related activity

Are Specific Institutional Relationships visible?

- We can group queries by ASes/Domains to narrow down





Specific relationships are present in the data:

→ A U.S. DOJ IP address requests MX record of palantir.com

→ A school board in Jefferson Parish requests MX record of ice.dhs.gov

Implications

- For institutions:
  - Use Qname minimization where possible (RFC 7816)
  - Local Root (https://localroot.isi.edu/) (RFC 8806)

- For DNS service providers that share data:
  - Host-only anonymization is not sufficient for protecting institutional privacy
  - Putting legal constraints
  - More rigorous privacy-preserving data sharing approaches?

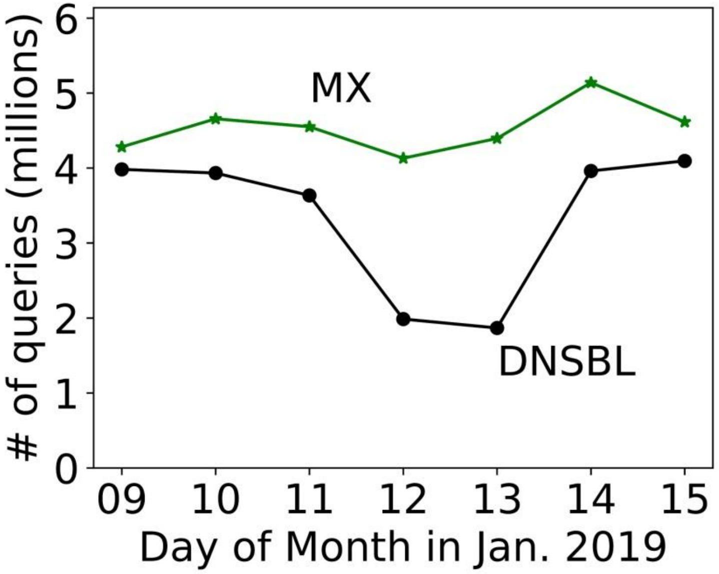
Conclusion

- DNS queries may leak significant institutional information that is private
- Institutions should deploy QMIN where possible
- Service providers should evaluate institutional privacy risks when sharing data

Contact: **Basileal Imana,** imana@usc.edu
Data: https://ant.isi.edu/datasets/dnsprivacy/

Institutional Privacy Risks in Sharing DNS Data

Basileal Imana, Aleksandra Korolova and John Heidemann


MX
DNSBL
# of queries (millions)
Day of Month in Jan. 2019