DNS data collection and analysis

Kazunori Fujiwara, JPRS
<fujiwara@jprs.co.jp>

2015/10/31

DNS data collection and analysis

- DNS traffic analysis starts with DNS query data collection
- Methods
  - Collect query logs or capture DNS packets
  - Easy to collect full capture with libpcap because DNS query rate is not high except under DoS attacks
- Issues
  - Data size problem
    - Long term dataset (A.DNS.JP dataset is 16 TB / 11 years)
    - Data set from many sites (2015 DNS-OARC DITL dataset is 9 TB / 50 hours)
  - Privacy concern
    - DNS queries may contain privacy sensitive information
    - Especially in full-service resolver data
    - Captured data may not be exported from each organization (or strict non-disclosure agreement required)

Activities of DNS data collection

- Root DNS servers : DNS-OARC DITL project
  - Access: https://www.dns-oarc.net/ditl/2011/
- Top level domains
  - Some TLDs collect and analyze their data
  - JPRS collects JP TLD servers' queries
  - ICANN requires statistics of new gTLD DNS servers
- Other authoritative DNS servers
  - Some RIRs collect reverse DNS data
  - DNS hosting providers may collect
- Full-service resolver
  - Researchers collect DNS data at their full-service resolvers and analyze it, then write papers
  - xSPs may collect and analyze for their use

Combined data analysis may results interesting outputs

- “Reduction of Root DNS Server Queries”, “ルートDNSサーバへのクエリ数の削減”, K. Fujiwara, A. Sato, K. Yoshida, The IEICE Transaction on communications (Japanese Edition), Vol. J98-B No.6, June 2015
  - DNS-OARC root dataset analysis resulted
    - More than 30,000 IP addresses sent more than 100,000 queries to Root DNS servers in 48 hours
  - Tested well used full-service resolvers with an university’s capture data and found:
    - BIND 9 full-service resolver which is widely used sends many reducible queries to Root DNS servers (88% of existing TLD name queries may be reducible)
    - Queries for non-existent TLDs from few stub resolvers cause many queries to Root (78% of non-existent TLD name queries may be reducible)
  - Part of the paper is presented at DNS-OARC workshop (in English)
    - https://indico.dns-oarc.net/event/19/material/slides/3?contribId=18