Exploring Domain Name-Based Features on the Effectiveness of DNS Caching

> Shuai Hao, Haining Wang. In ACM SIGCOMM Computer Communication Review 2017.

Background

DNS cache: the acquired mapping results will be cached locally to answer the following queries in a specific duration.

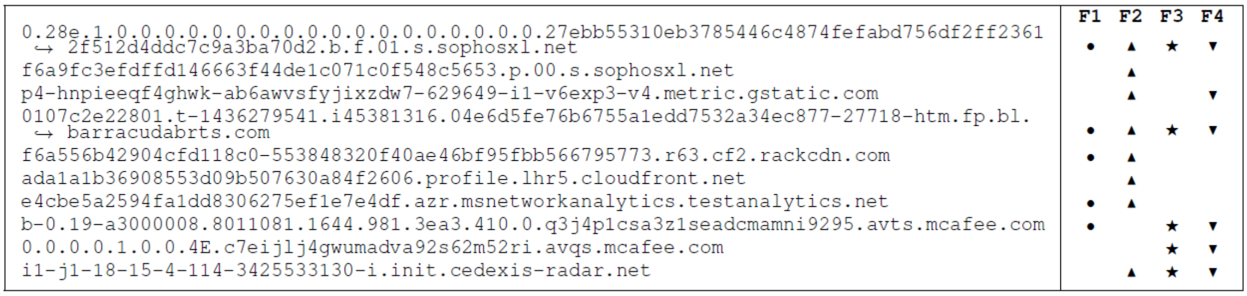
RRs: the DNS resource records

Target

Ensure that the cached RRs would be likely to be accessed again.

Motivation

Most repeatedly appeared domains have a short name and limited subdomain depth, and a significant portion of domains have a long query name and a large number of subdomains.



	F1	F2	F3	F4
0.28e.1.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.27ebb55310eb3785446c4874fefabd756df2ff2361 ↪ 2f512d4ddc7c9a3ba70d2.b.f.01.s.sophosxl.net	●	▲	★	▼
f6a9fc3efdffd146663f44de1c071c0f548c5653.p.00.s.sophosxl.net		▲		
p4-hnpieeqf4ghwk-ab6awvsfyjixzdw7-629649-i1-v6exp3-v4.metric.gstatic.com		▲		▼
0107c2e22801.t-1436279541.i45381316.04e6d5fe76b6755a1edd7532a34ec877-27718-htm.fp.bl. ↪ barracudabrts.com	●	▲	★	▼
f6a556b42904cfd118c0-553848320f40ae46bf95fbb566795773.r63.cf2.rackcdn.com	●	▲		
ada1a1b36908553d09b507630a84f2606.profile.lhr5.cloudfront.net		▲		
e4cbe5a2594fa1dd8306275ef1e7e4df.azr.msnetworkanalytics.testanalytics.net	●	▲		
b-0.19-a3000008.8011081.1644.981.3ea3.410.0.q3j4p1csa3z1seadcmamni9295.avts.mcafee.com	●		★	▼
0.0.0.0.1.0.0.4E.c7eijlj4gwumadva92s62m52ri.avqs.mcafee.com			★	▼
i1-j1-18-15-4-114-3425533130-i.init.cedexis-radar.net		▲	★	▼

Figure 2: Sample of domains with the domain name-based features.

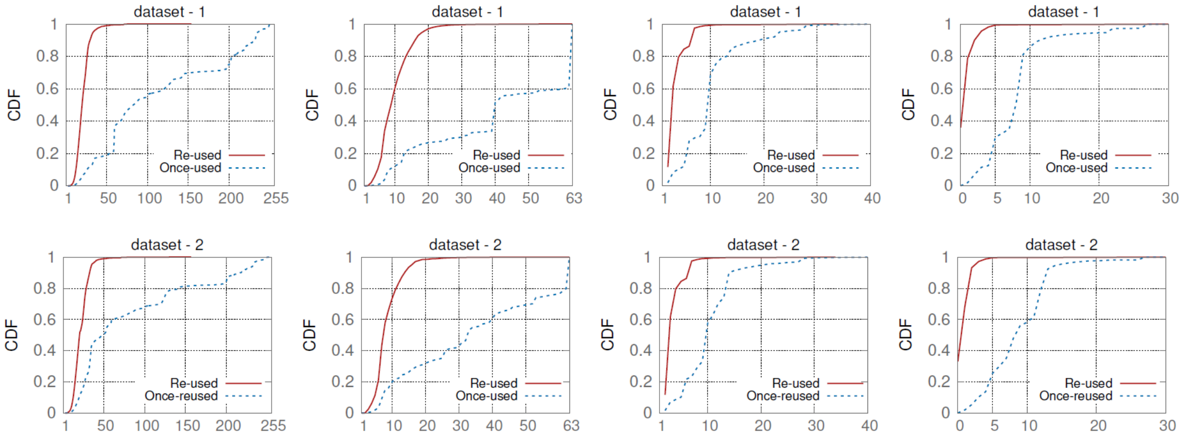
Contributions

- Characterize the properties of re-used and once-used domains;
- Train a classifier to classify the entries;
- Conduct a trace-driven simulation to validate their efficacy in caching. (LRU>FIFO)

Features

- $F1$: Length of Query Name.
- $F2$: Length of the Longest Subdomain Name.
- $F3$: Number of Format Fields.
- $F4$: Total number of L-FF and S-FF.

Validation



(a) DNS Query-Name Length (b) Longest-Subdomain Length (c) Number of Format Field (d) Total Number of L-FF and S-FF

Figure 3: Distribution of domain name-based features for `re-used` **and** `once-used` **domains.**

Detailed descriptions are shown in Part 4.2.

Experiments

Dataset (manual, disclosed)

The trace logs of outgoing DNS queries captured at local DNS servers at the College of William andMary (WM) and the University of Delaware (UD) over a period of two weeks.

Model

decision tree + random forest

> Types of RRs: A AAAA TXT PTR SRV SOA NS other

Useful info

- distribution of types of malicious RRs

- features of RRs

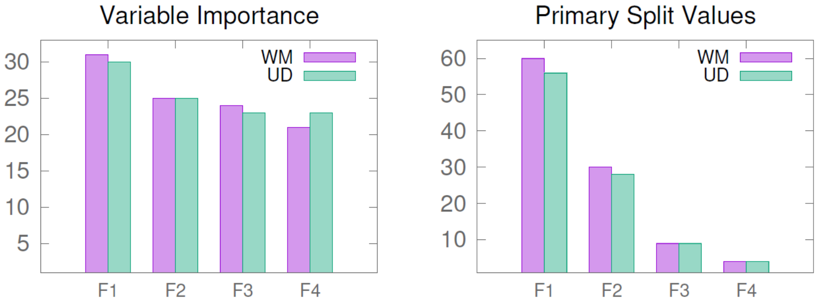


Figure 4: Training Results (with Decision Tree).

- Why not use TTL: in part 5.3

Related Work (Part 6)

- DNS Caching and TTL characterization.
- Cache modifications.
- Malicious domain detection.