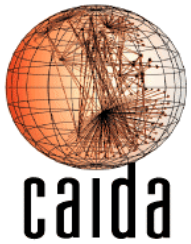
Bulk DNS Lookup Service

Josh Polterock josh@caida.org

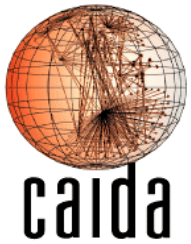
CAIDA/WIDE Workshop
January 19, 2008
Honolulu, HI
USA

http://www.caida.org/



Overview

- Motivation
- Bulk DNS Lookup Service
- Lookups of Topology Data
- Conclusions

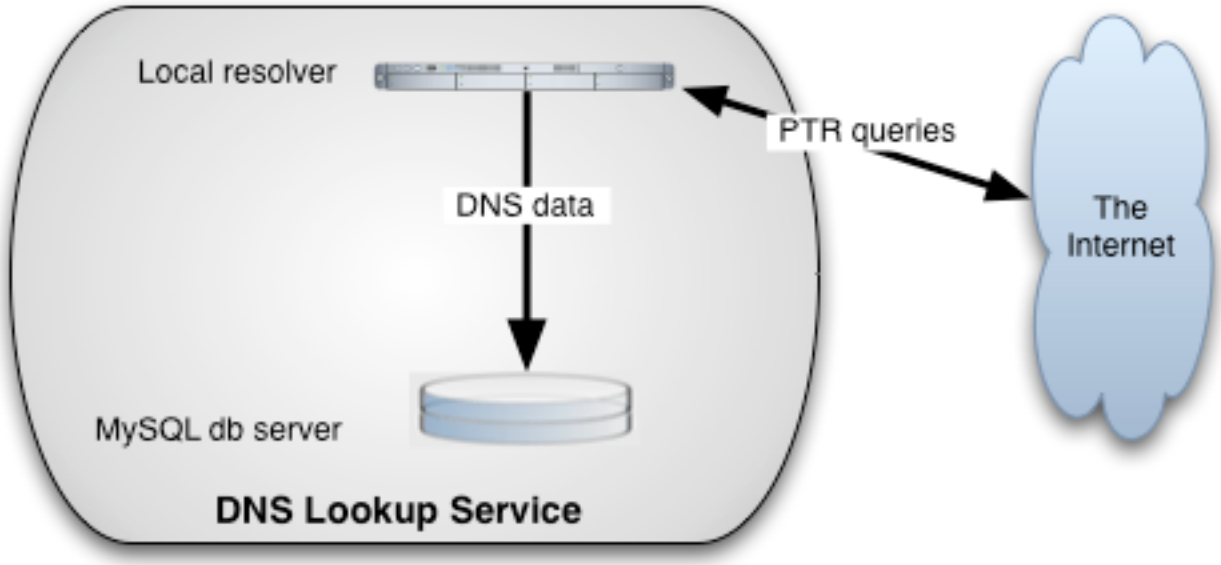


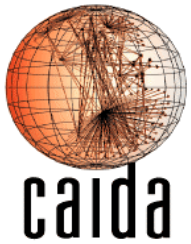
Motivation

- DNS information is valuable for many passive and active data analyses
- DNS information helps answer questions:
  - Is an IP address a router, home box, or web server?
  - Where is this host geographically?
  - Is the host at a corporate or an academic site?
  - What is the likely link speed (e.g., home broadband)?

DNS Lookup Service

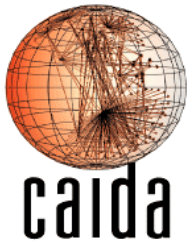
- CAIDA has an internal bulk DNS lookup service
  - Currently only PTR queries





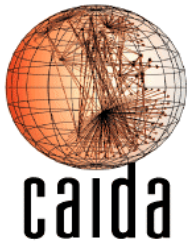
Goals

- Obtain DNS information in a timely manner
- Archive DNS lookup results
- Support querying of archived results
- Be scalable to large numbers of lookups
- Be considerate of remote nameservers



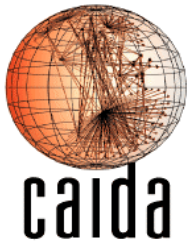
Scalability

- Enables timely data collection through scalability
  - quickly performs a large number of lookups while the data is still fresh
- Achieves scalability with multiple hosts
  - run dedicated local resolver (BIND), one per host
  - distribute lookups to hosts in a pool (up to 5 hosts)
- Sustained an average of 2 million lookups/day over a month
- In the past three months, we looked up 31 million addresses



Archiving Data

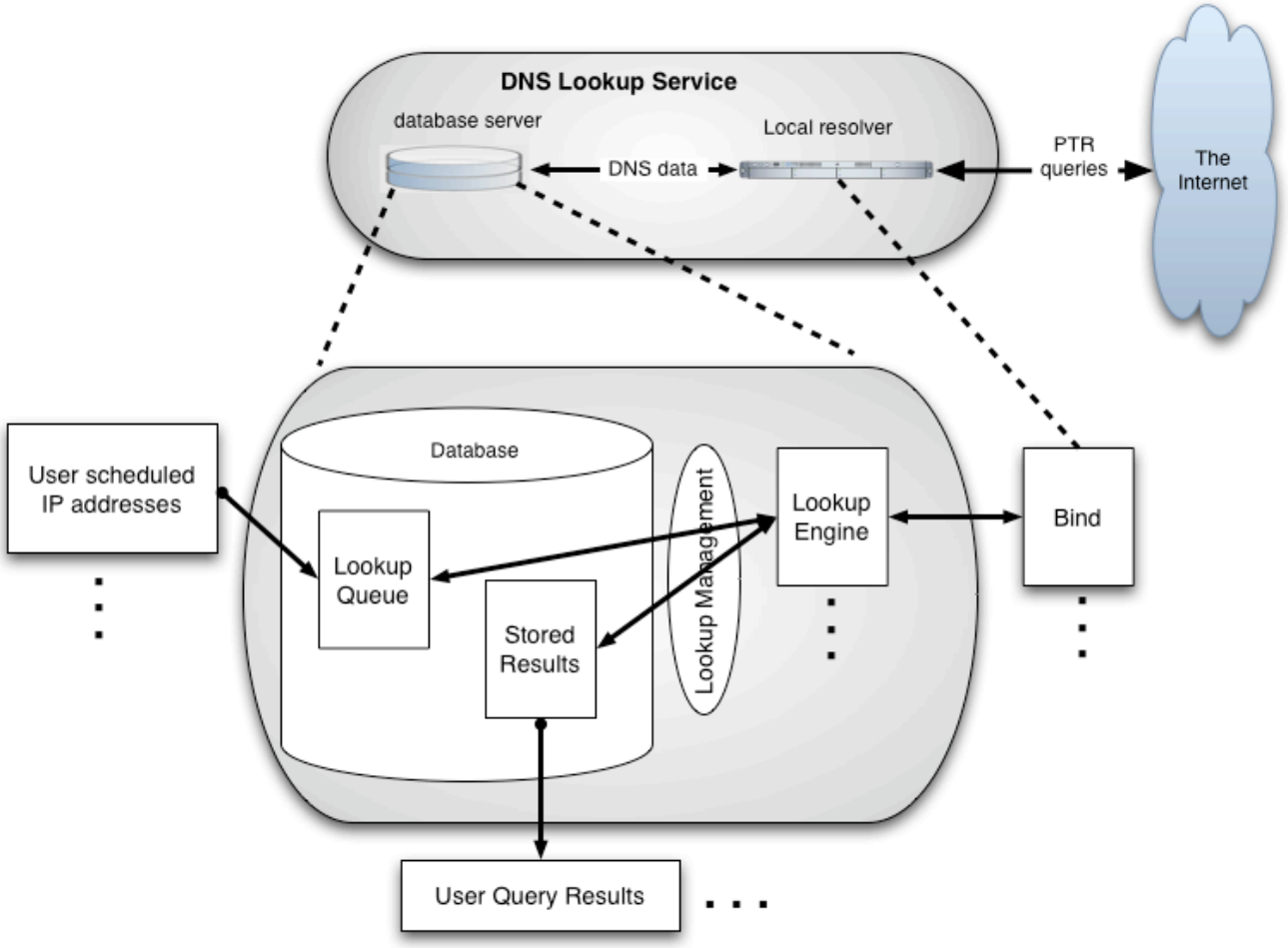
- Lookup results are archived in a database
  - columns: timestamp, address, hostname, result code
  - timestamp column allows the same address to be looked up multiple times over time
- Query by (timestamp, address) and get lookup performed nearest to the requested timestamp
- In the past six months, we stored over 42 million lookup entries

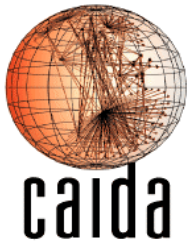


Scheduling Lookups

- To avoid high load on remote nameservers, the service skips requests for addresses already queried in the past 7 days
  - a trade-off between reducing load and obtaining timely information
  - however, can force immediate lookups of addresses
    - useful for security events
- Supports prioritization of lookups on per address basis
  - user can reduce priority of frequently looked up addresses

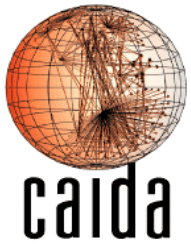
Database Engine





Our uses of DNS Lookup Service

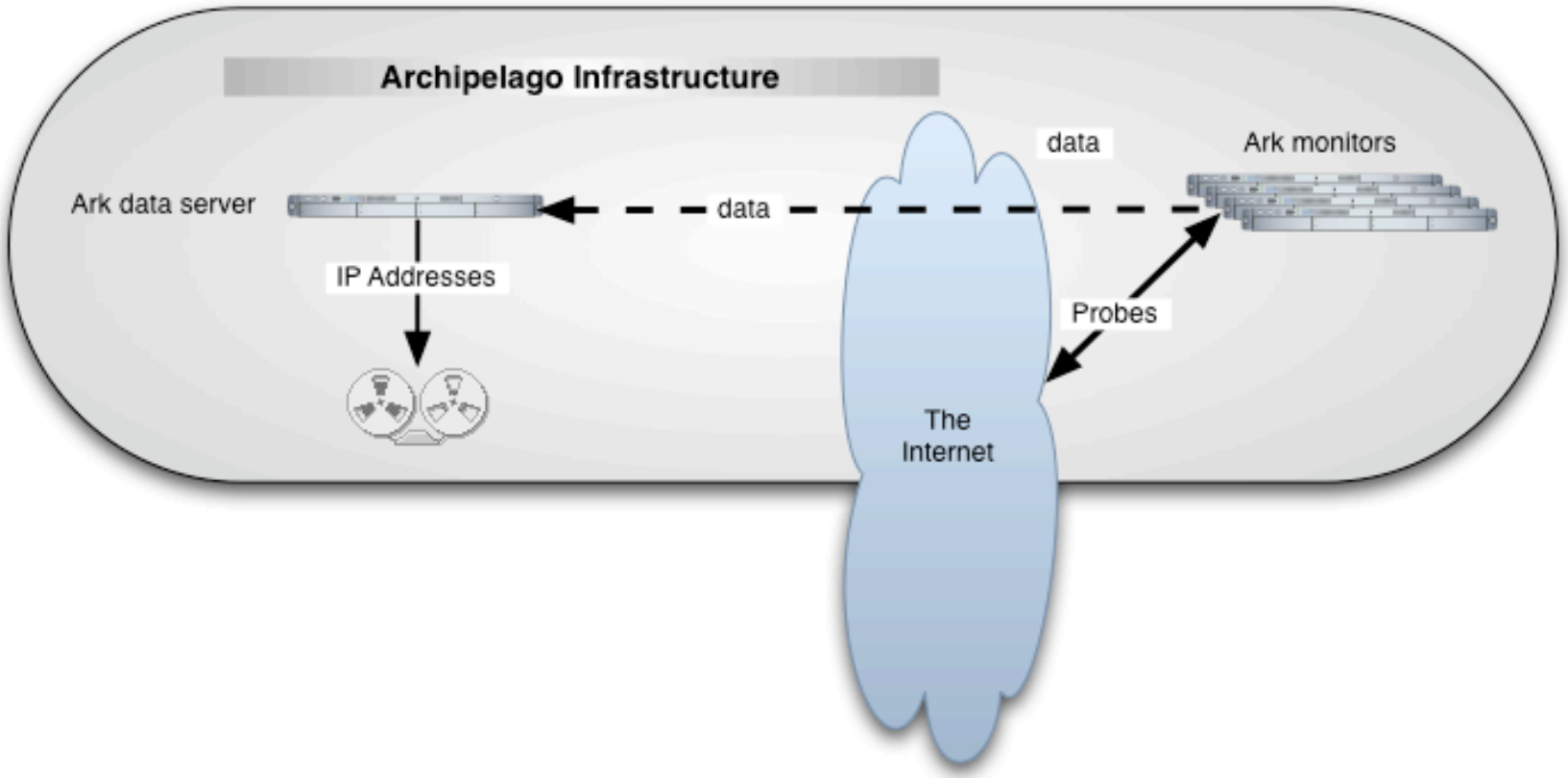
- Security data
  - Backscatter
  - UCSD Network Telescope
  - Worm data
- Network traffic traces
- Topology data



Archipelago (Ark) Data Collection

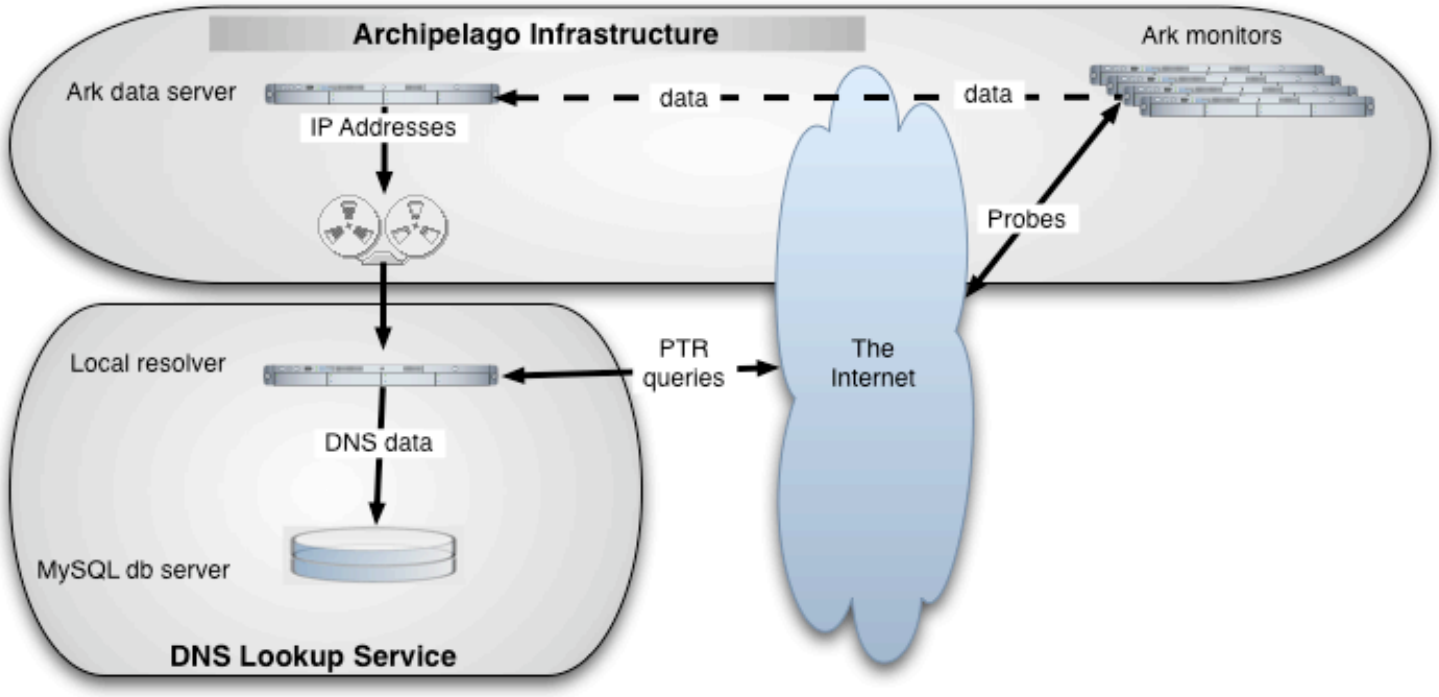
- Ark is our next generation infrastructure supporting:
  - long-running, large scale experiments,
  - coordination via local and global tuple spaces.
- We probe a random destination in every routed /24 (IPv4) each cycle
  - about 7M /24s in RouteViews BGP table
- 13 monitors
- 2-3 days/cycle
  - Collected 41 cycles since 12 Sept 2007

Ark Data


Archipelago Infrastructure
data
Ark monitors
Ark data server
data
IP Addresses
Probes
The
Internet

Lookups of Topology Data Diagram

- We lookup IP addresses found in Ark traces
  - routers and responding destinations



Topology Lookup Results

- We have automated daily lookups of over 600K addresses/day.
- Analysis: TLD breakdown for six cycles (one month) of addresses.

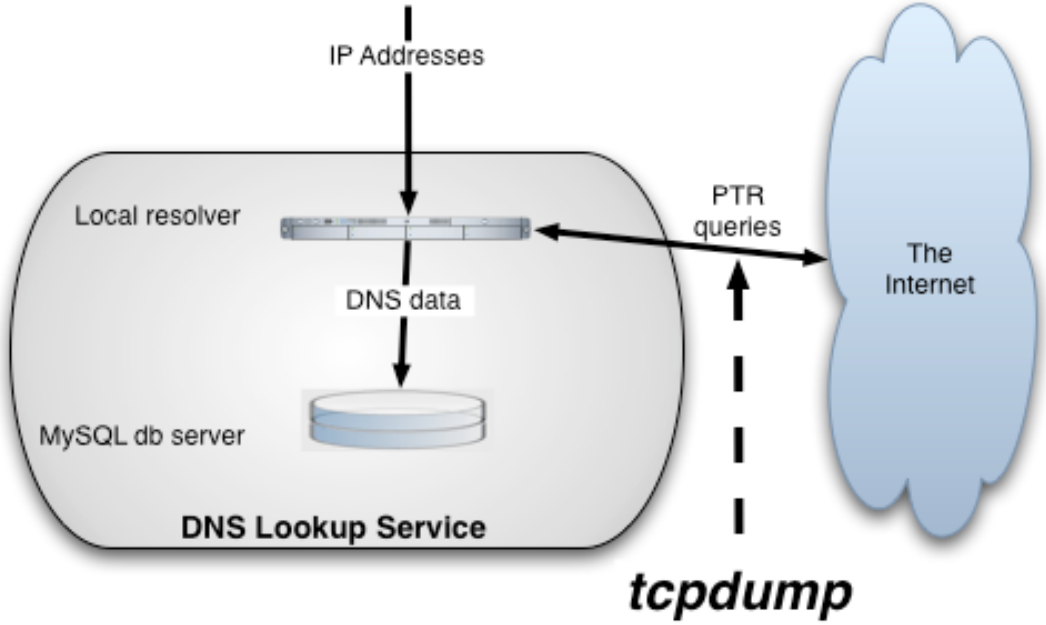
Top 10 TLDs:

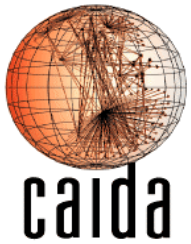
1.	net	793,407	(42.5%)
2.	com	270,259	(14.5%)
3.	jp	114,167	(6.1%)
4.	de	79,533	(4.3%)
5.	br	53,017	(2.8%)
6.	mx	45,134	(2.4%)
7.	it	43,781	(2.3%)
8.	cn	36,258	(1.9%)
9.	edu	31,581	(1.7%)
10.	pl	25,894	(1.4%)

Total Addresses: 3,176,655
success: 1,865,978 (58.7%)
failure: 1,310,677 (41.3%)

Examination of Raw DNS Queries and Responses

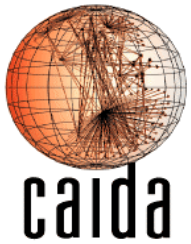
- Experiment examined the raw DNS query and response traffic between the local recursive resolver and remote nameservers





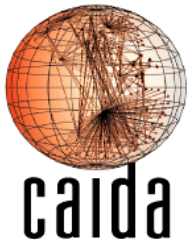
Raw DNS Statistics

- We collected 807MB of compressed pcap traces covering about 8 full days (Dec 12-20th); UDP only.
- 17M DNS packets were successfully captured.
  - 8.9M query packets
    - 1.0M A (of nameservers)
    - 1.3M AAAA (of nameservers); got 12.6k (1%) answers with IPv6 addresses
    - 6.5M PTR
  - 8.2M response packets
    - 63% had AA bit set
    - 2.8% (233k) had AAAA glue record(s) in additional section



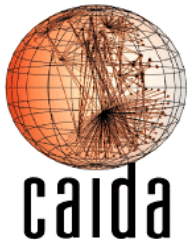
Conclusion

- Internally, we have implemented and deployed a scalable DNS lookup service
- The service enhances our security and topology data analyses
- Low effort required to always do DNS lookups as integral part of data collection process
- Quickly scale for large time-critical security events



Future Work

- Make lookup results available
- Make lookup service software available



Links & Thanks

- Archipelago: http://www.caida.org/projects/ark/
- Topology and (in the future) DNS lookup results: http://www.caida.org/data/

Much thanks to Young Hyun and David Moore for their feedback and assistance.