SAND2009-6712C

Quality of Name Resolution in the Domain Name System

ICNP '09

Oct 15, 2009

Casey T. Deccio
Sandia National Laboratories

Chao-Chih Chen
Prasant Mohapatra
UC Davis

Jeff Sedayao
Krishna Kant
Intel Corporatation

This research was supported in part by the National Science Foundation under the grant CNS-0716741.

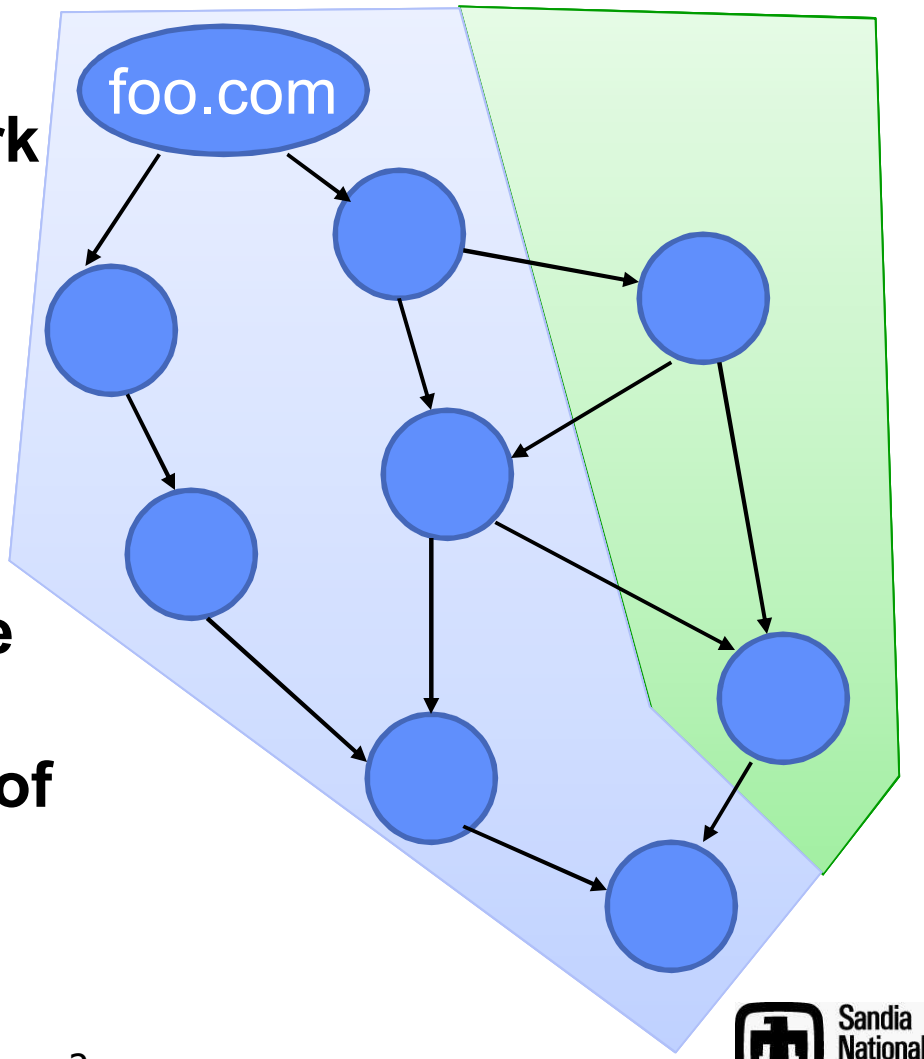
Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

NNSA National Nuclear Security Administration

Sandia National Laboratories

Objectives

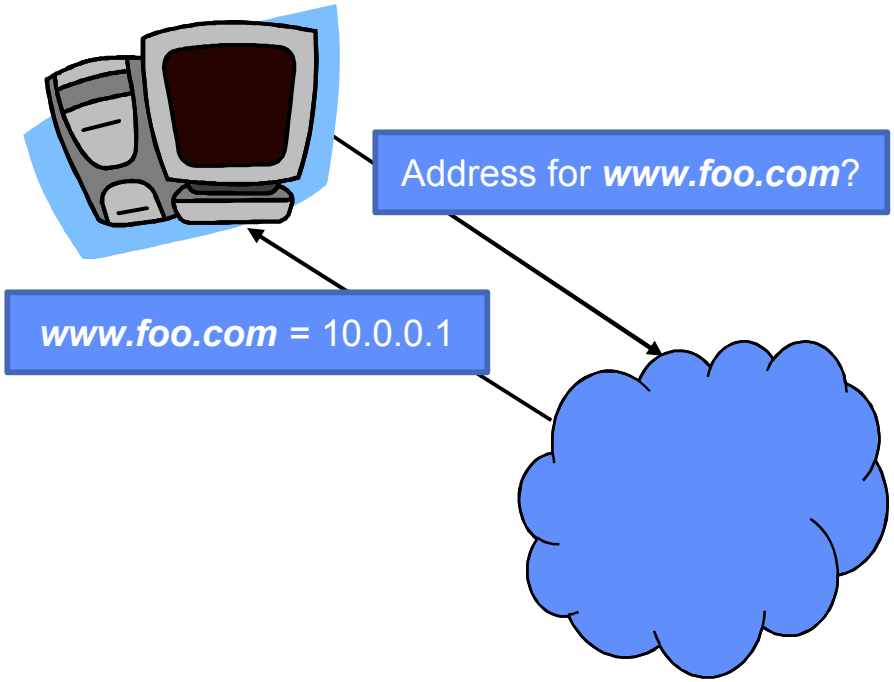
- Understand the network of dependencies for a domain name
- Quantify the impact of domain name dependencies
- Identify the namespace *within* and *without* administrative control of a domain



Sandia National Laboratories

Overview

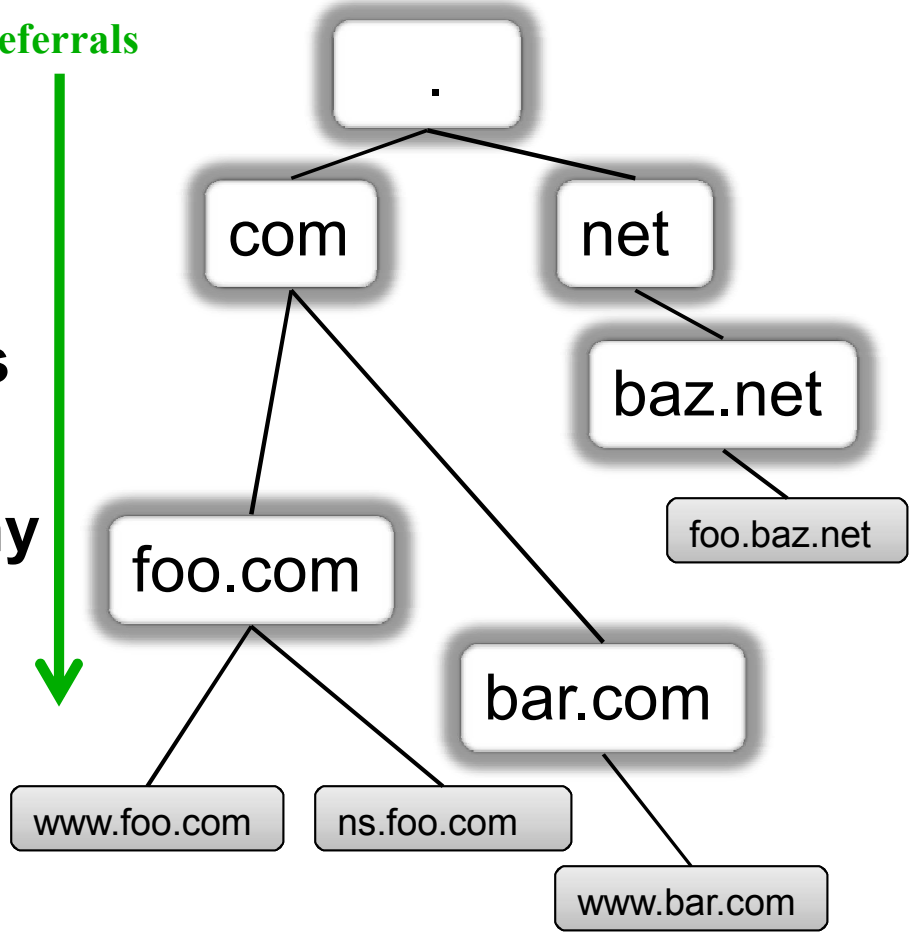
- Background
  - DNS fundamentals
  - Name dependencies
- DNS dependency model
  - Domain name influence
  - Metrics for analysis
- Survey of DNS namespace
  - Data collection
  - Analysis and results



Sandia National Laboratories

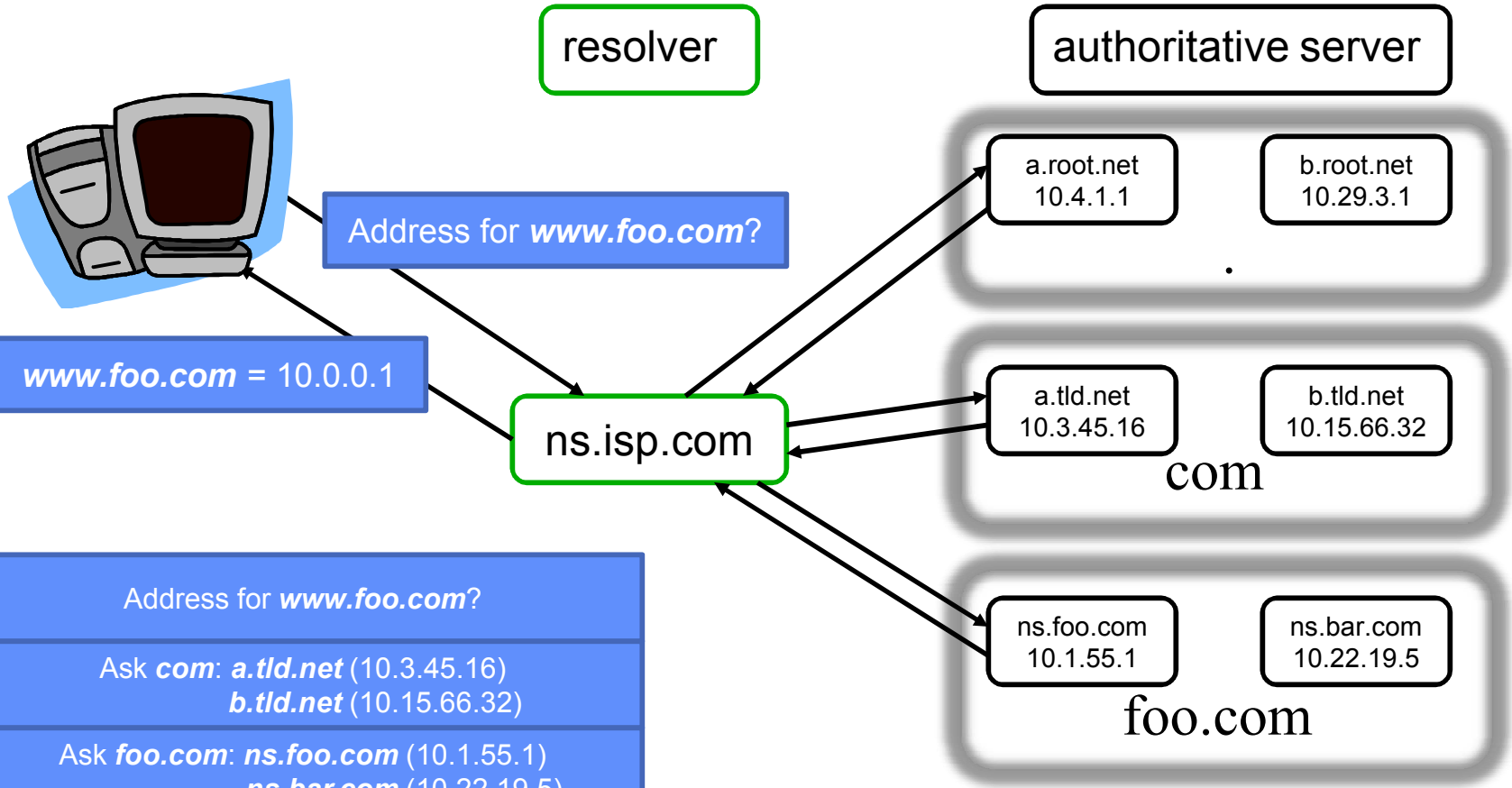
Name resolution in DNS

- *Resolver*:
  - has questions
- *Authoritative server*:
  - has answers/referrals
- Resolvers begin queries at the top of the hierarchy
- Authoritative servers refer to delegated subdomain namespace



Sandia National Laboratories

Resolving www.foo.com



Address for ***www.foo.com***?
Ask ***com***: ***a.tld.net*** (10.3.45.16) ***b.tld.net*** (10.15.66.32)
Ask ***foo.com***: ***ns.foo.com*** (10.1.55.1) ***ns.bar.com*** (10.22.19.5)
www.foo.com = 10.0.0.1

Sandia National Laboratories

Resolver needs server addresses

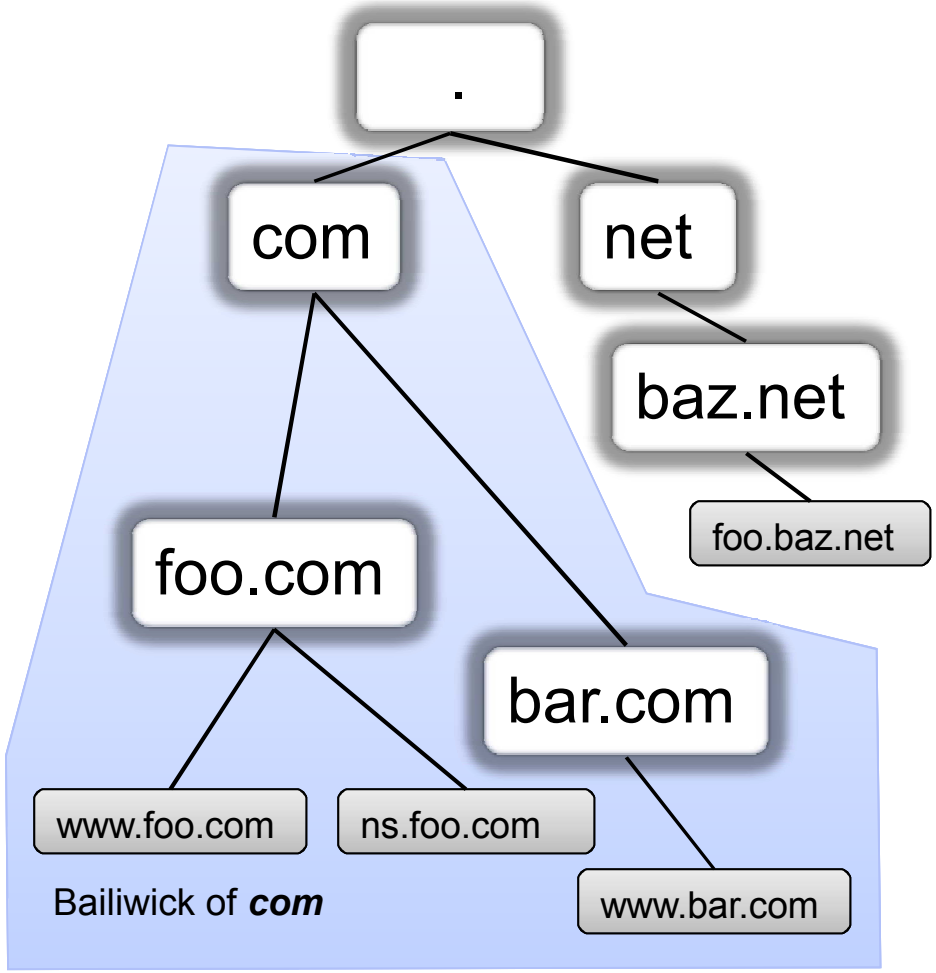
- ***Names* used to designate servers authoritative for zone:**

  foo.com. NS ns.foo.com.
- **Resolver needs *address* to query server:**

  ns.foo.com → 10.1.55.1
- **Authoritative servers may provide addresses (i.e., *glue records*) for names in-bailiwick (subdomains):**

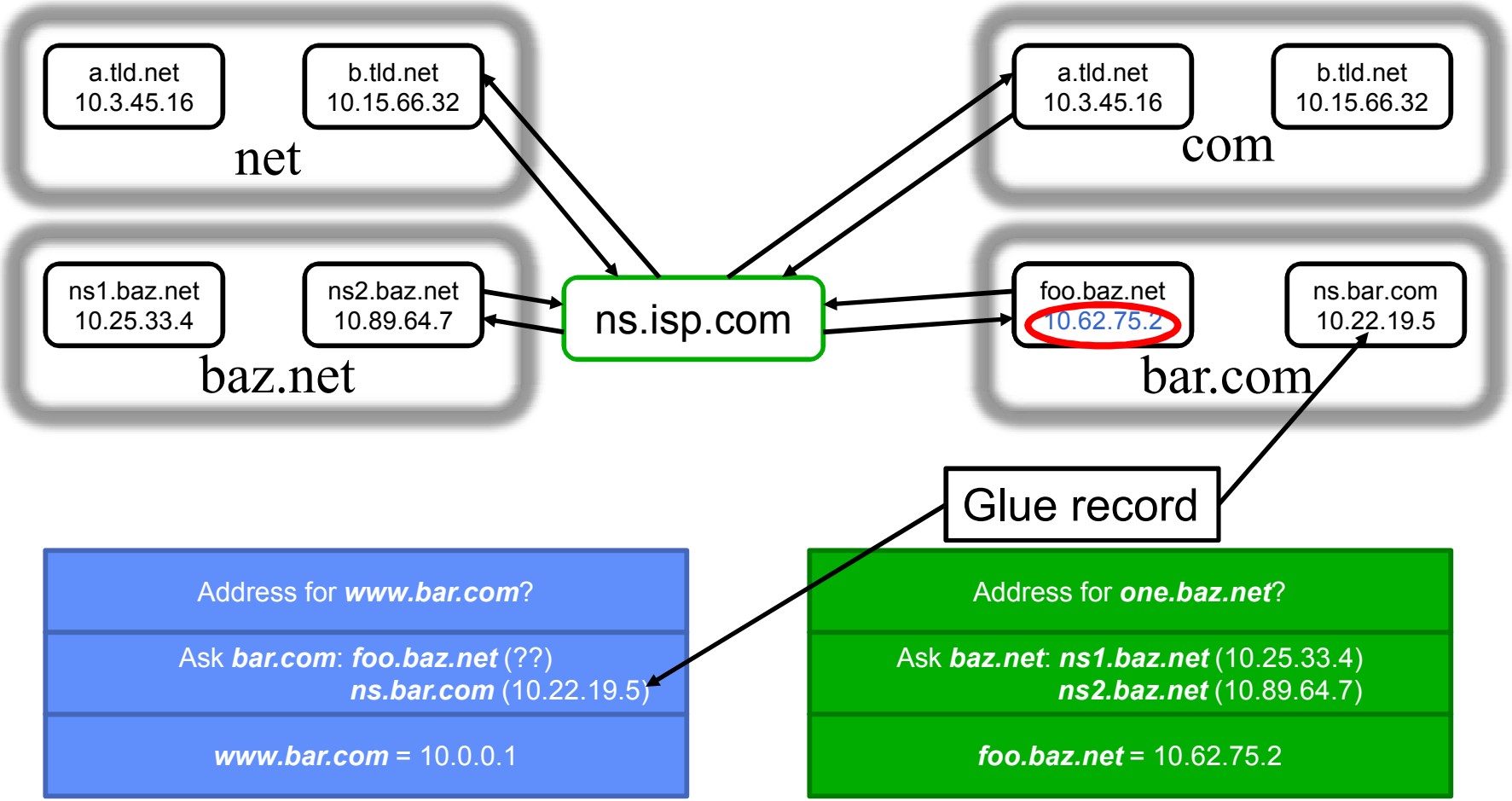
  com. provides address for ***ns.bar.com.***
- **Other names must be resolved by resolver:**

  com. provides only name for ***foo.baz.net***; resolver must look up address



Sandia National Laboratories

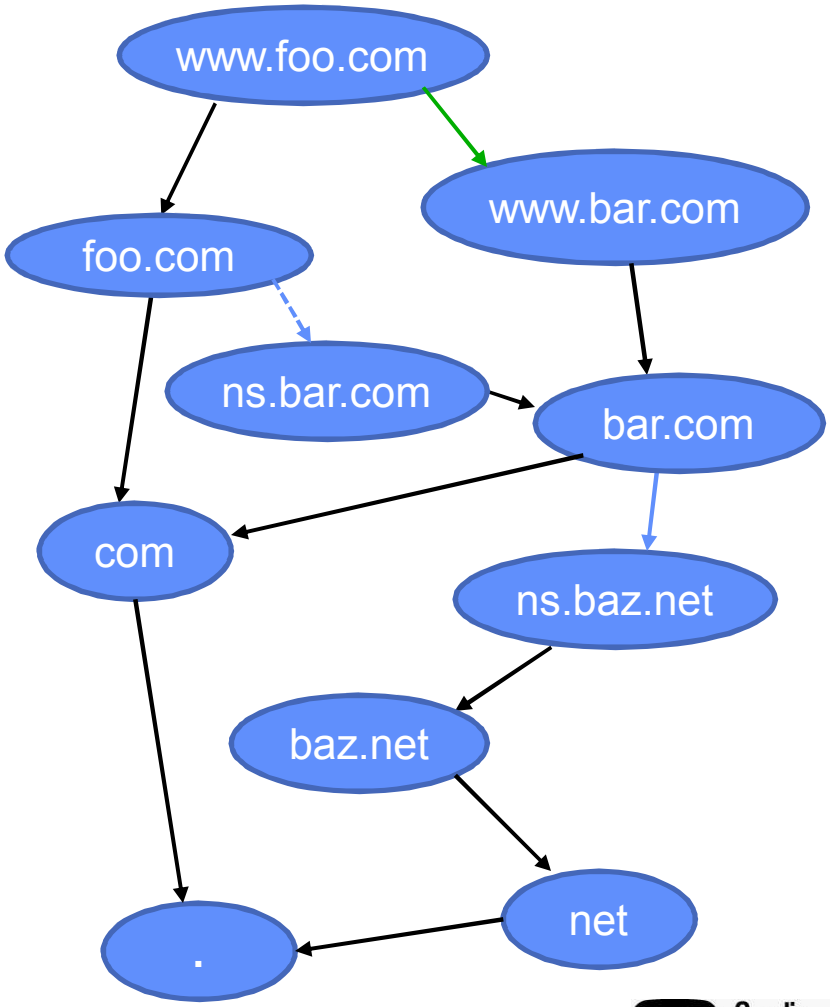
Name resolution example



Sandia National Laboratories

Name dependency graph

- Nodes = domain names
- Edges = dependencies
  - Child to parent
  - Alias to target
  - Zone to NS targets
- D. J. Bernstein, “Notes on the Domain Name System”
  - Dependencies in DNS
- Ramasubramanian, et al., “Perils of Transitive Trust …”
  - Size of dependency graph
- Pappas, et al., “Impact of Configuration Errors on DNS Robustness”

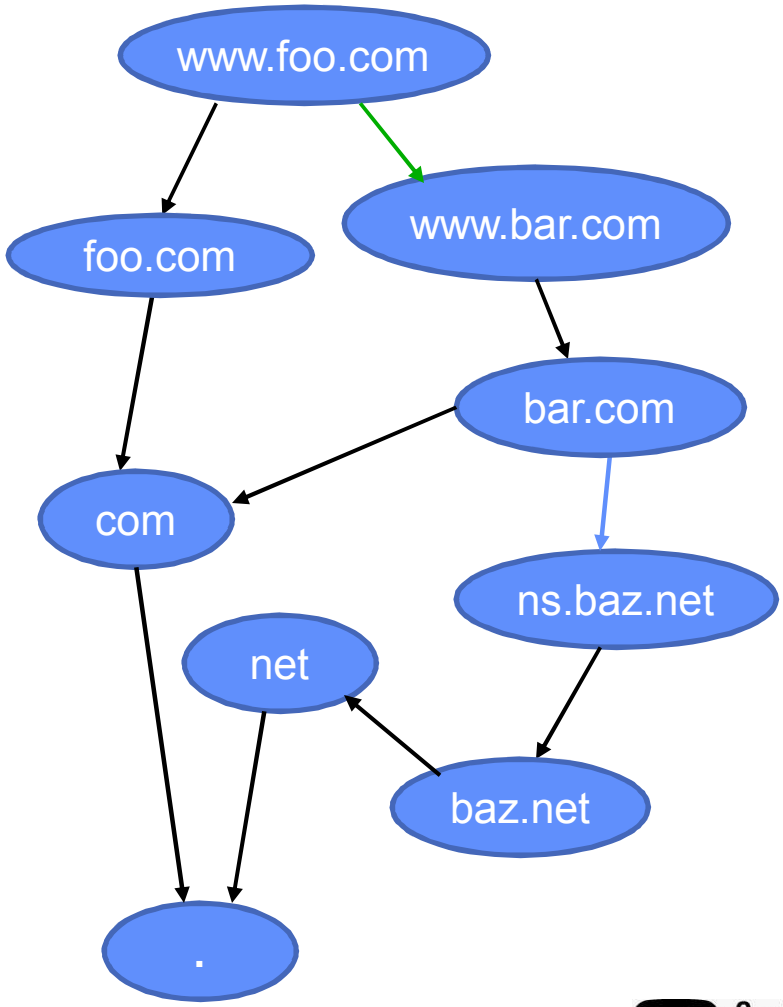


Sandia National Laboratories

Robustness is determined by graph

- Graph properties affect:
  - ~~Potential attack target: authenticity~~
  - Availability
  - Performance

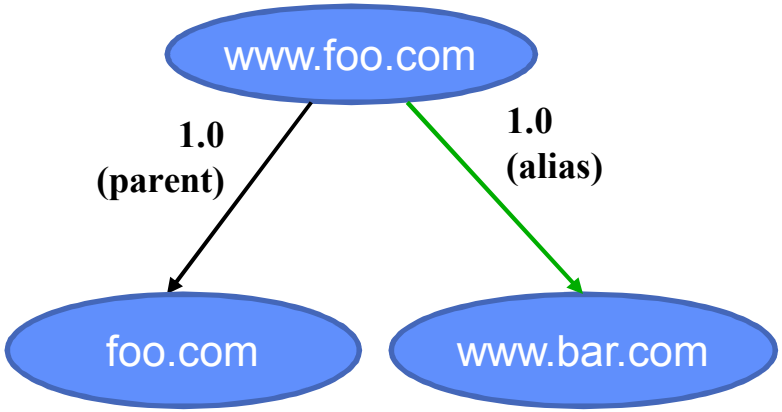
What about DNSSEC?



Sandia National Laboratories

How much influence does a name have?

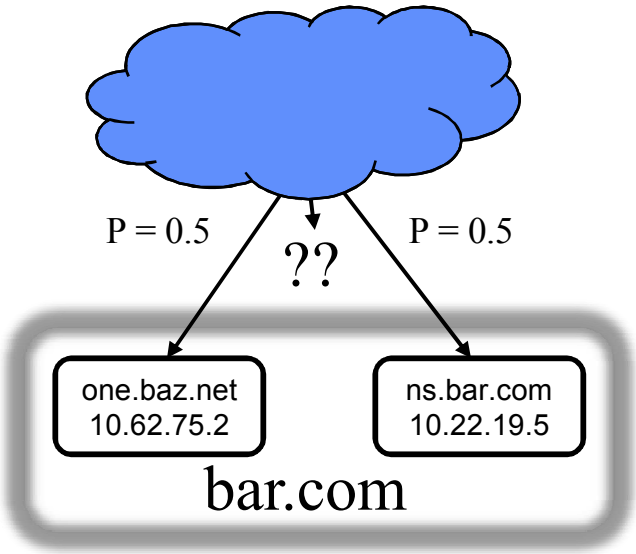
- Influence = probability that name is used in resolving another name
- Direct influence measured using edge weights
- Edge weight values:
  - Child always dependent on parent
  - Alias always dependent on target



Sandia National Laboratories

Which server is selected for query?

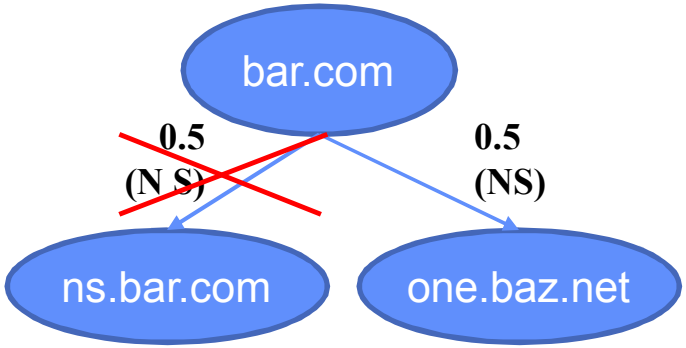
- Zones generally advertise more than one authoritative server
- Resolvers “learn” to use server with best history
- From diverse locations, each server is selected with equal probability



Sandia National Laboratories

NS dependency edge weights

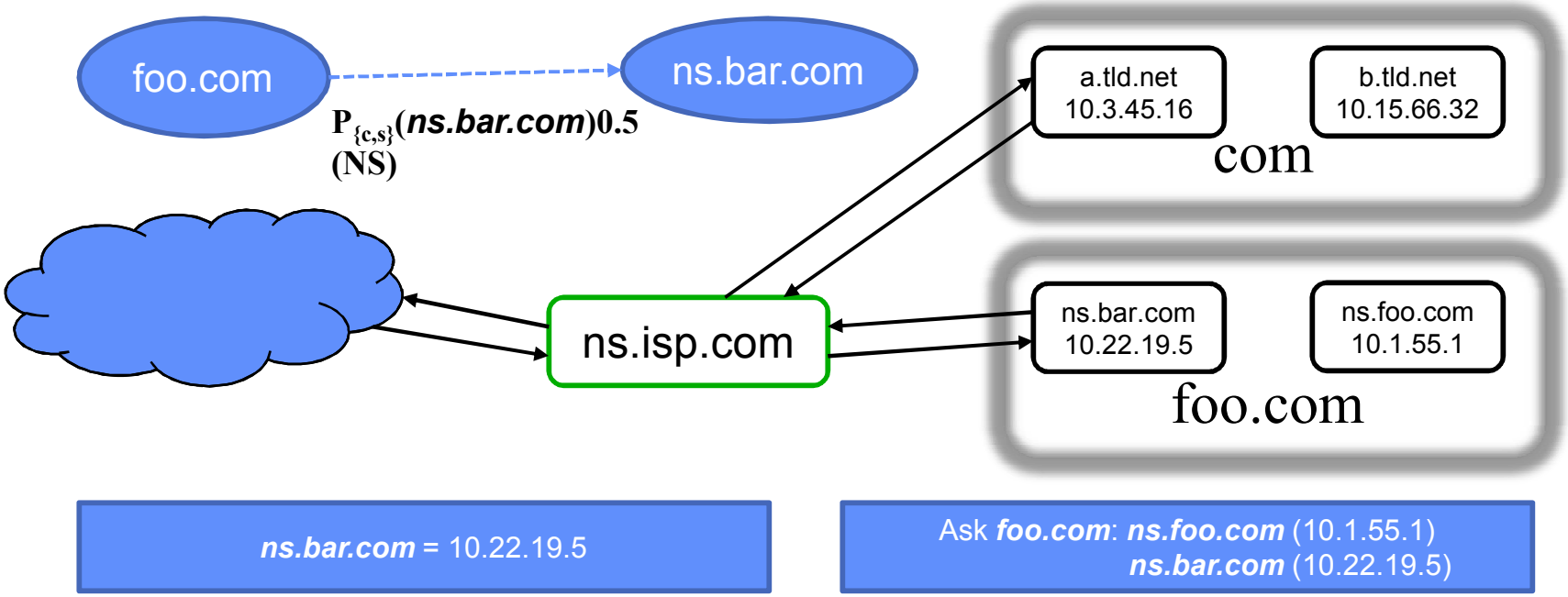
- Addresses provided by ancestor zones (*glue records*) eliminate name dependencies
- Such edges excluded from graph



Sandia National Laboratories

Caching of NS targets

- Resolvers cache *answer* and its *source*
- Resolver will "trust" an address from an authoritative source, over one from glue
- Dependency based on probability that name is cached



Sandia National Laboratories

Calculating level of influence

- **Level of influence:**

$$I_d(v) = P(d, \ldots, v)$$

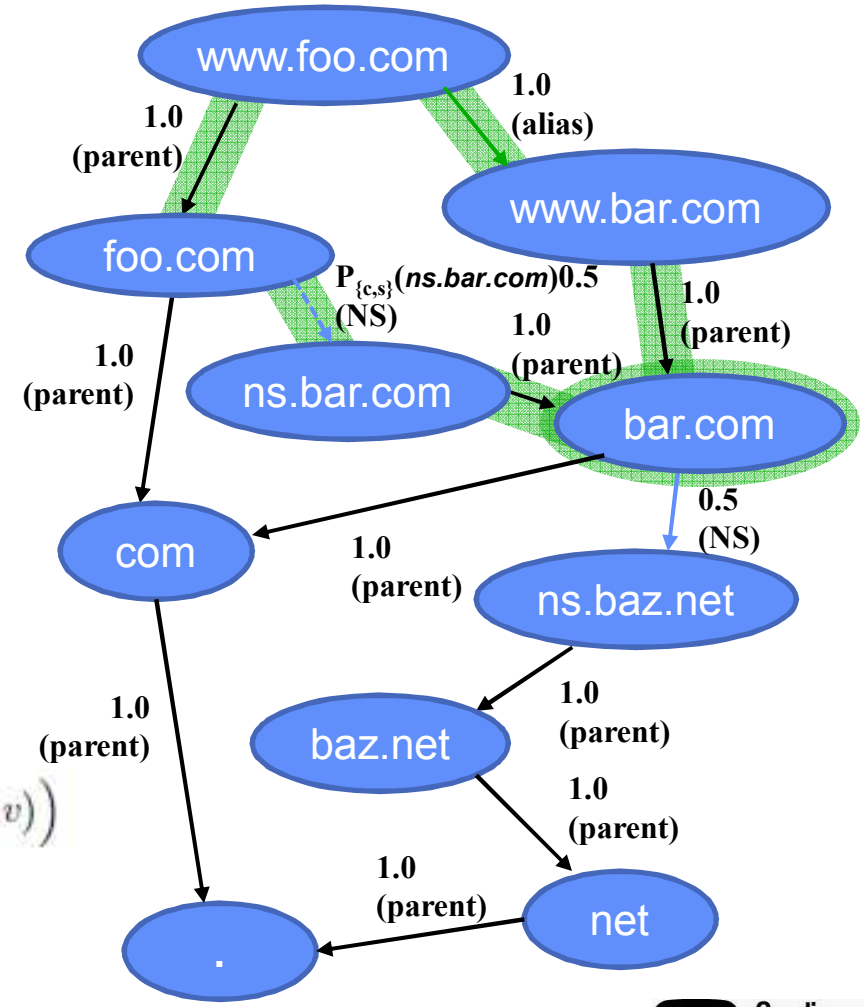
- **Path probability:**

$$P(u, j, \ldots, v) = \begin{cases} w(u,j) & \text{if } j = v \text{ (direct dep)} \\ 0 & \text{if } j = r \text{ (root)} \\ w(u,j)P(j,\ldots,v) & \text{otherwise} \end{cases}$$

- **Aggregating influence:**

$$P(u, [NS\ dep], \ldots, v) = \sum_{j \in NS_u} w(u,j)P(j,\ldots,v)$$

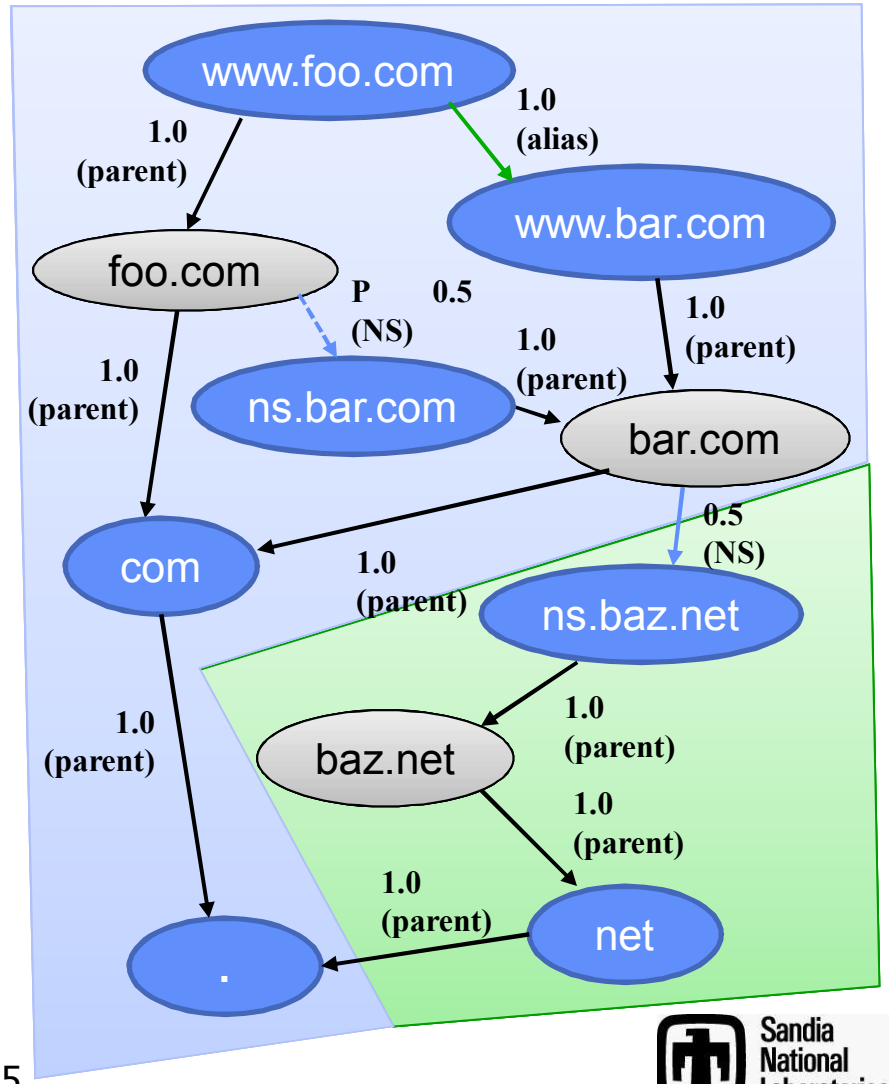
$$P(u,\ldots,v) = 1 - \Big(1 - P(u, Parent(u), \ldots, v)\Big) \Big(1 - P(u, Cname(u), \ldots, v)\Big) \Big(1 - P(u, [NS\ dep], \ldots, v)\Big)$$



Sandia National Laboratories

Which names really matter?

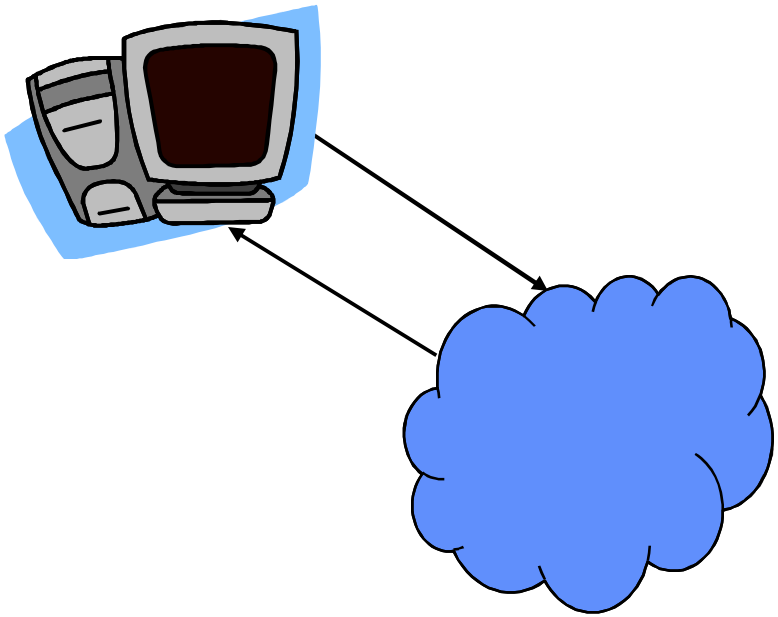
- Non-trivial zones:
  - Zones with non-zone direct dependents
  - Implies explicit configuration
- First-order zones:
  - Non-trivial zones explicitly configured by zone in question
- Third-party influence:
  - Probability of being influenced by third-party names



Sandia National Laboratories

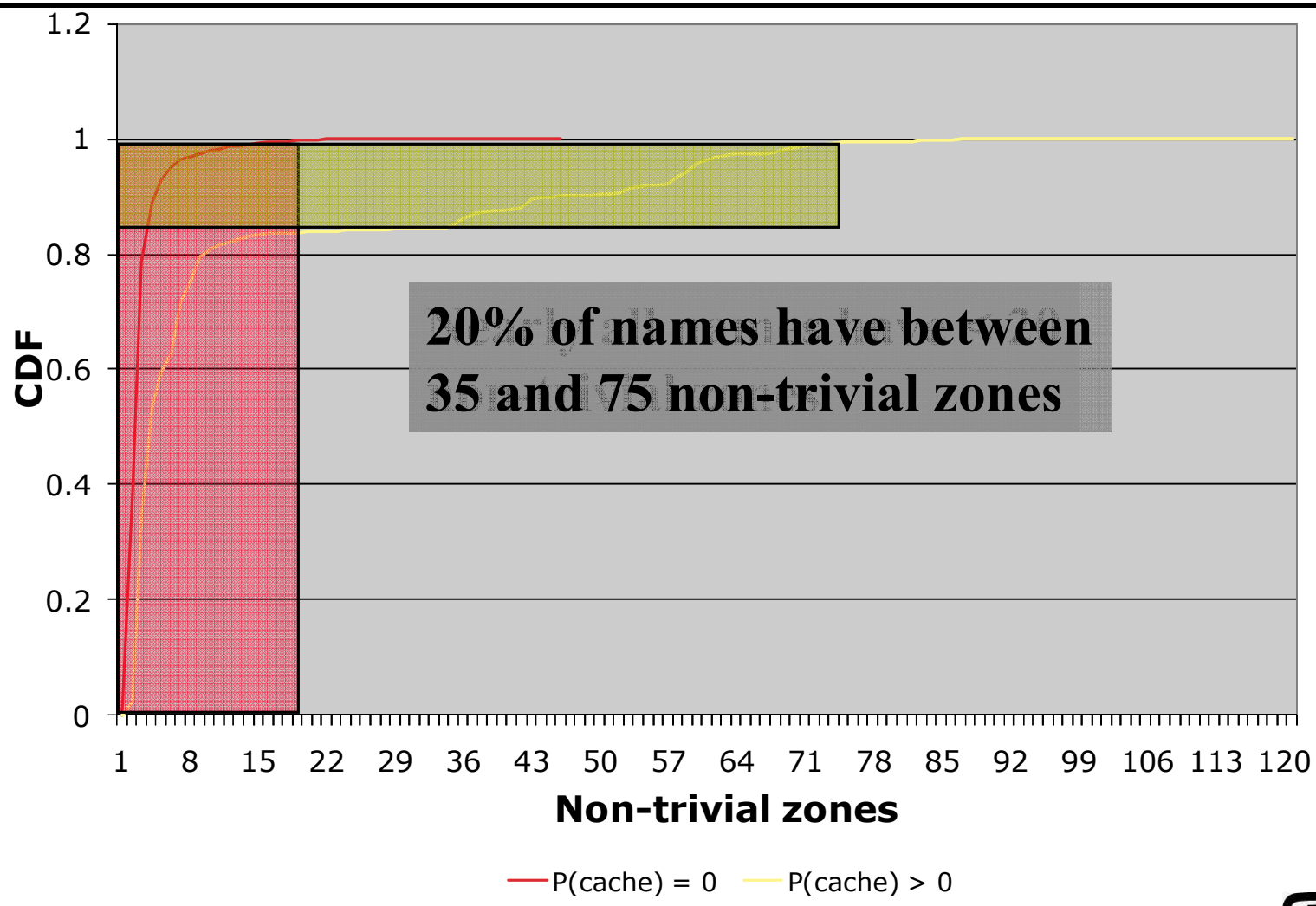
Data collection

- Extracted ~3 million names from Open Directory Project (dmoz.org)
- Collected additional 100,000 names from SC08
- Crawled dependencies of each name
- Resulting graph:
  - 8.4 million nodes
  - 22.3 million edges



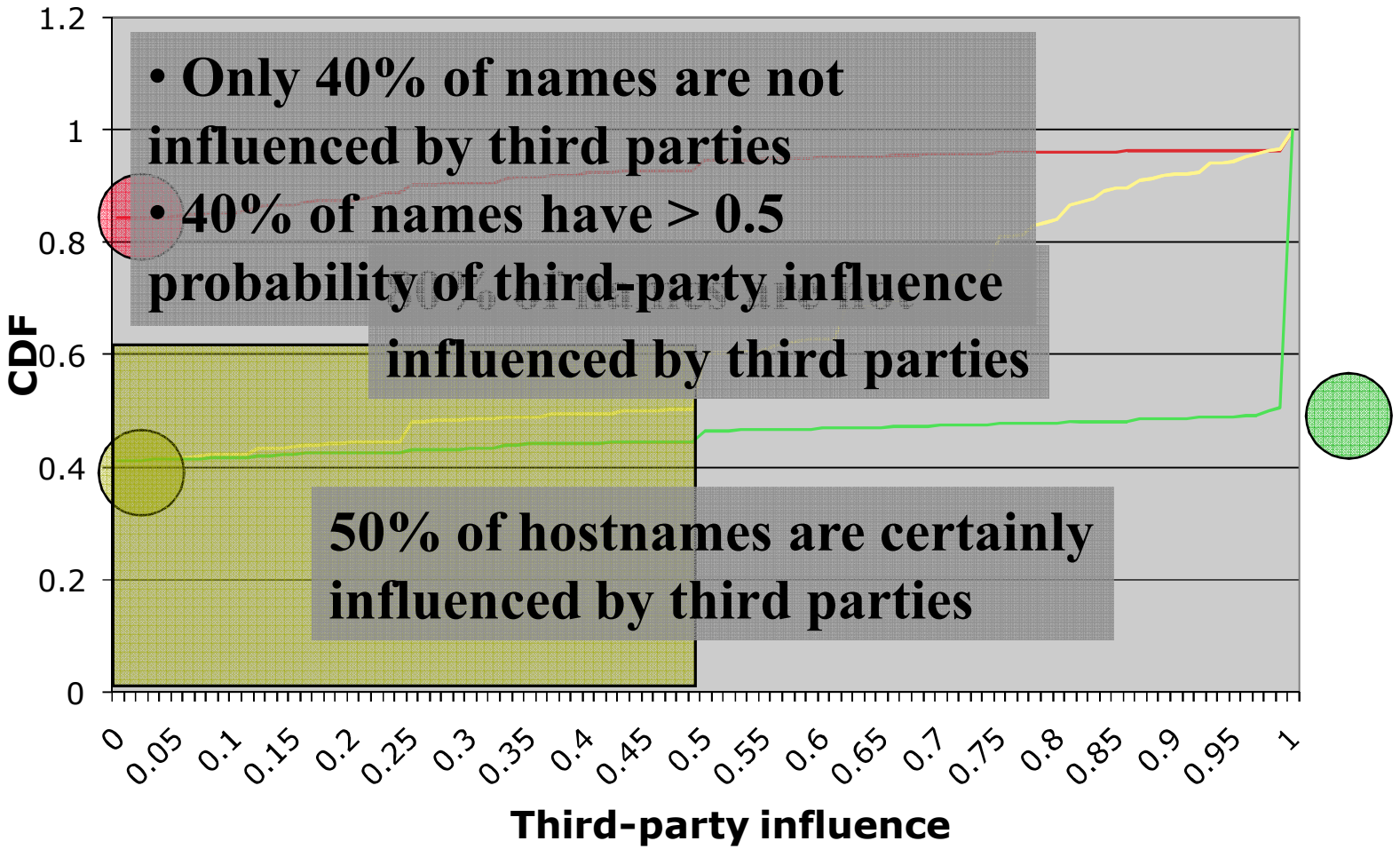
Sandia National Laboratories

Trusted computing base (zones)



Sandia National Laboratories

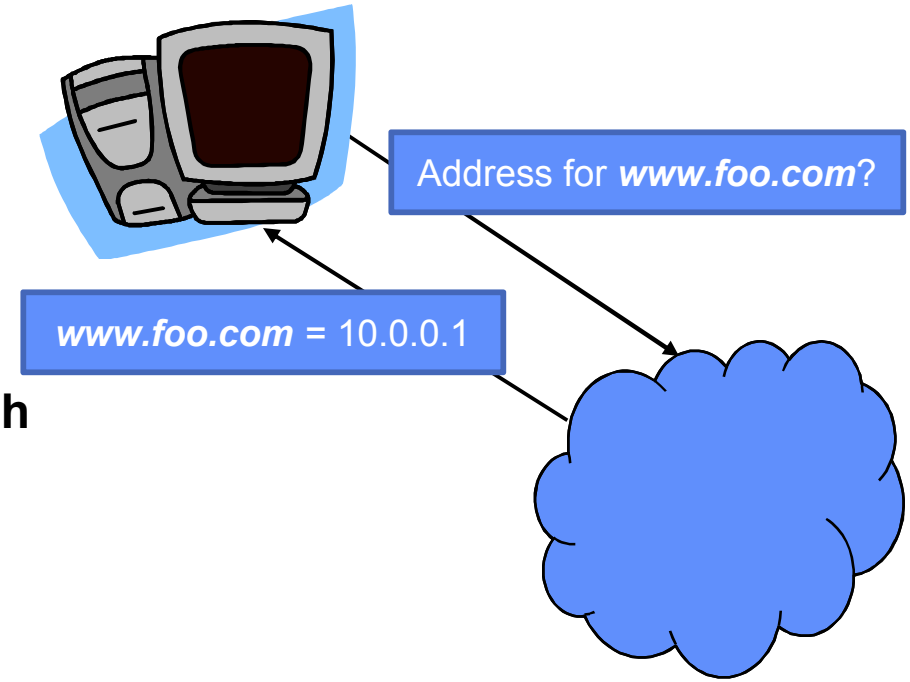
Third-party influence



Sandia National Laboratories

Summary

- DNS dependency model
  - Quantifies influence of domain names
  - Defines metrics for analysis
  - Caching of NS target names increases:
    - Number of zones in graph
    - Third-party influence
- Future work
  - Theoretical analysis of DNS misconfigurations
  - DNS availability study



Sandia National Laboratories