iToM:
An Internet Topology Mapping Project

Kamil Sarac (ksarac@utdallas.edu)

Department of Computer Science
The University of Texas at Dallas

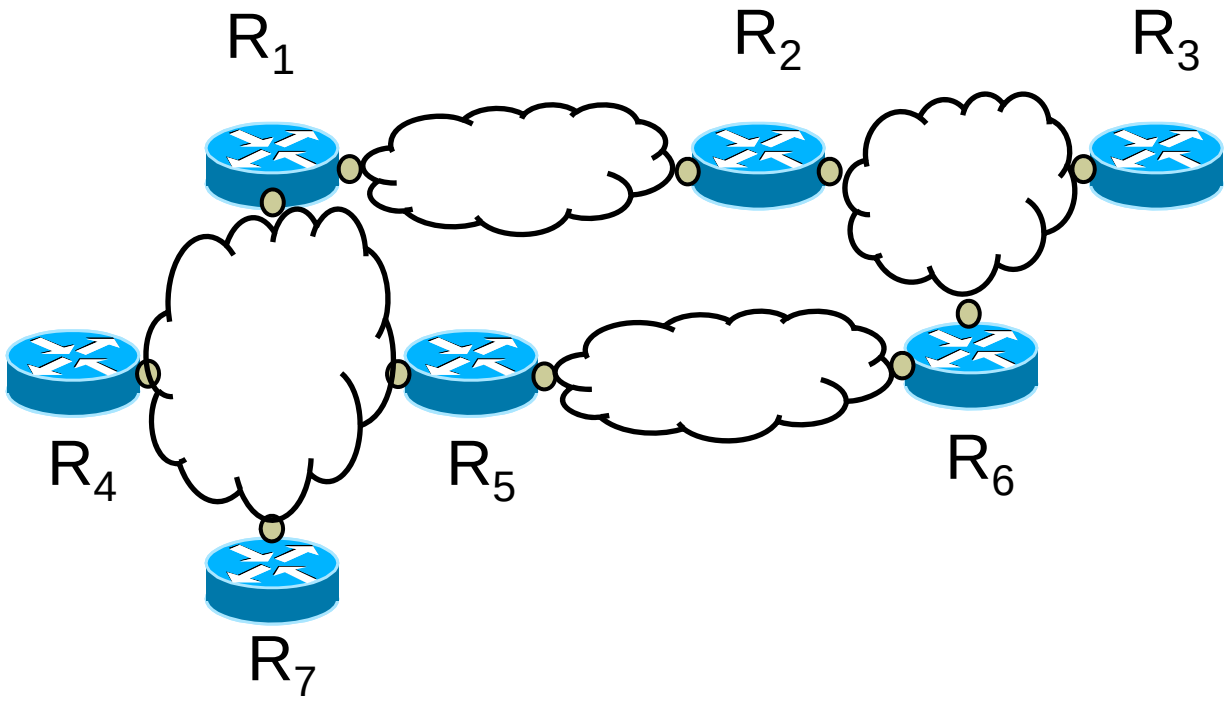
Internet topology measurement/mapping

- Need for Internet topology measurement
  - Help with network management or surveillance
  - Robustness with respect to failures/attacks
  - Comprehend spreading of worms/viruses
  - Relevant in active defense scenarios
  - Scientific discovery
    - Scale-free (power-law), Small-world, Rich-club, Disassortativity,…

Subnets as first class citizens in network layer Internet topology maps

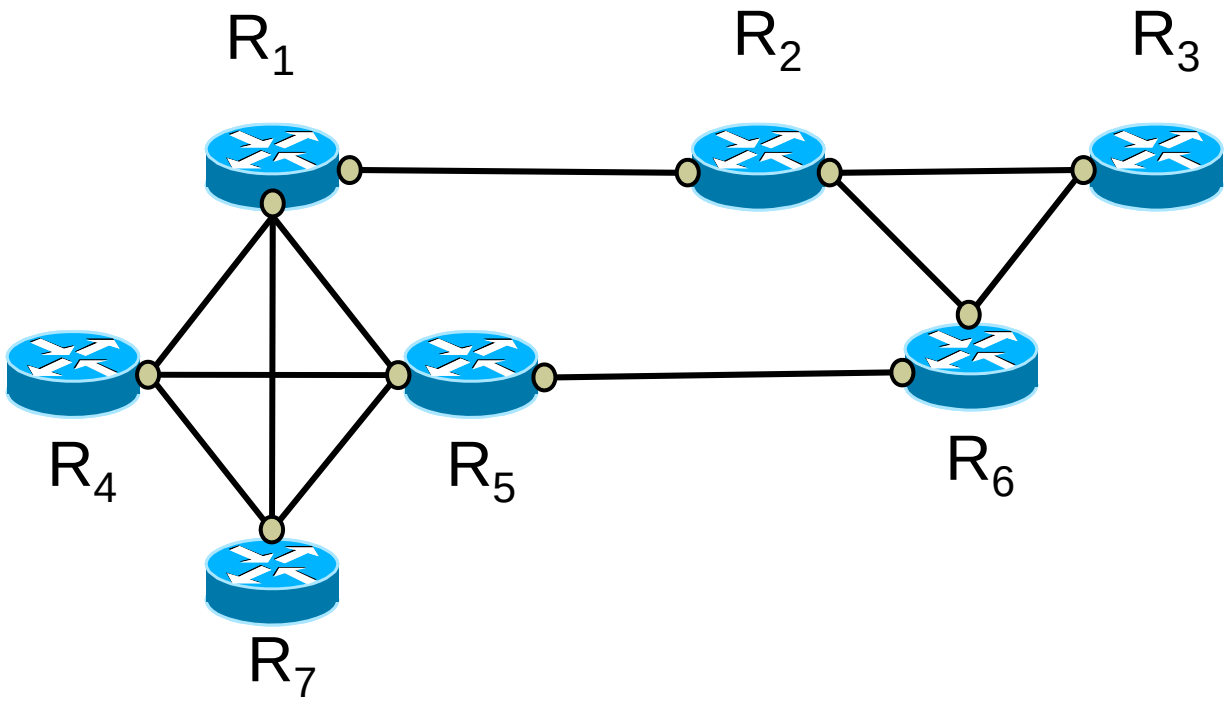
Network layer Internet topology maps

- A sample IP network segment view at Layer 3
  - A number of routers connected via subnets



A router level map at layer 3

- A corresponding router level map view

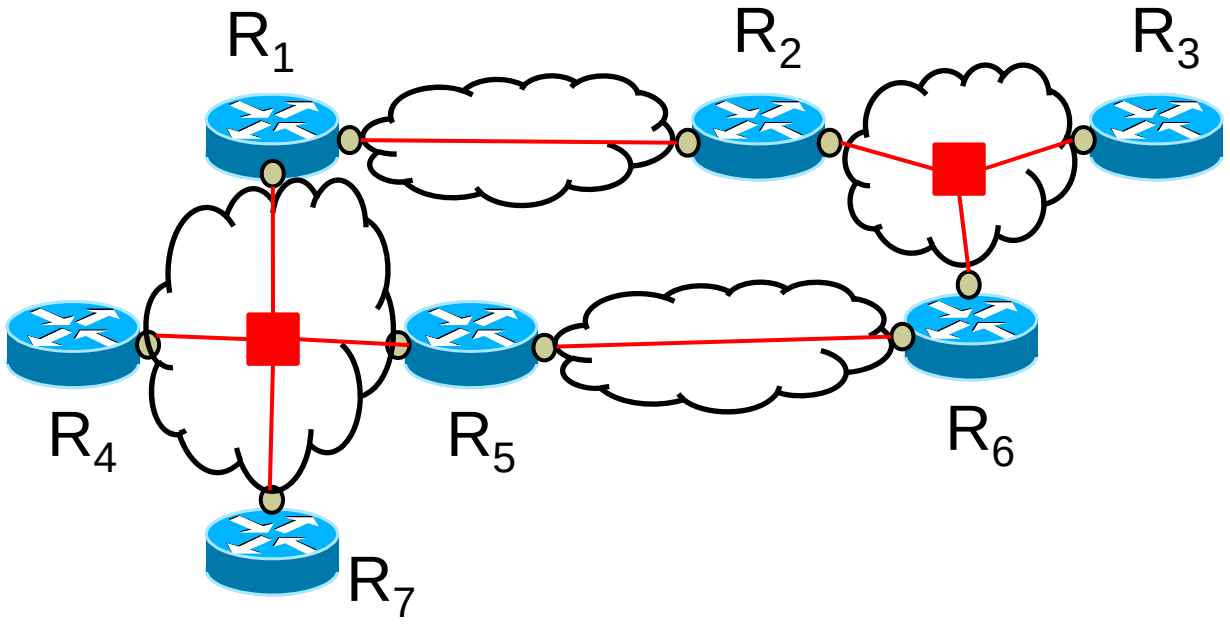


How to build router level network maps ?

- Involves topology data collection and topology construction
- How to collect topology data ?
  - Traceroute – a network debugging and diagnostic tool
  - End-to-end traces from k vantage points to n destinations where (typically) $k << n$
- How to construction topology maps?
  - Resolving alias IP addresses
  - Resolving anonymous routers

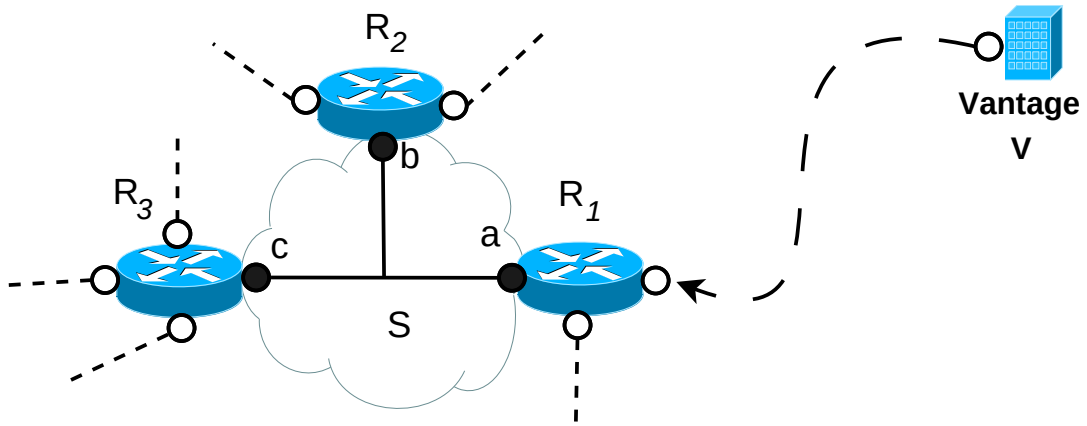
A network layer view incl. routers & subnets

- Not all subnets are created equal !
- *Can we discover layer 3 view of subnets ?*
  - *List of alive IP addresses*
  - *Subnet number as a.b.c.d/x*



How to discover a subnet?

- ExploreNET – an active probing based tool
  - Given an IP address t, discovers the subnet S hosting t
  - Labels S with its observable subnet mask
  - A black box using a set of heuristics for subnet inference



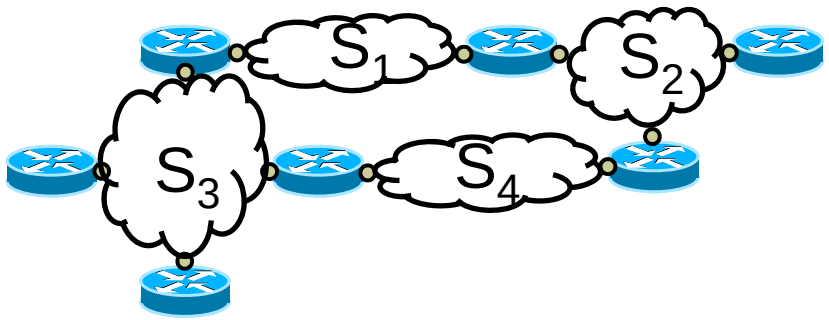
How to discover a subnet?

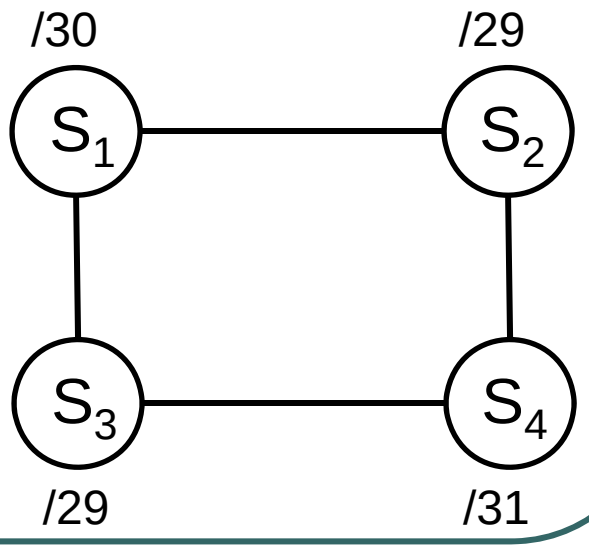
```
# sudo ./xnet -d 24.173.8.28


Network Number : [Network IP Address - Hop Distance List]
-------------- :  ------------------------------------
24.173.8.24/29 : [24.173.8.26 - 2, 24.173.8.25 - 3,

                  24.173.8.29 - 3, 24.173.8.28 - 3]
```

- ExploreNET accuracy rates (experimental)
  - 94.9% for Internet2
  - 97.3% for GEANT
  - 93.0% for global public Internet (w.r.t. mrinfo data)
- Probing cost is within $2|S|$ to $7|S| + 7$

Why know subnets?

1. A more complete network layer picture of the underlying network
2. An alternative layer 3 view of the Internet map where
   - subnets are nodes
   - routers are links





Statistical sampling for studying characteristics of networks

Why statistical sampling?

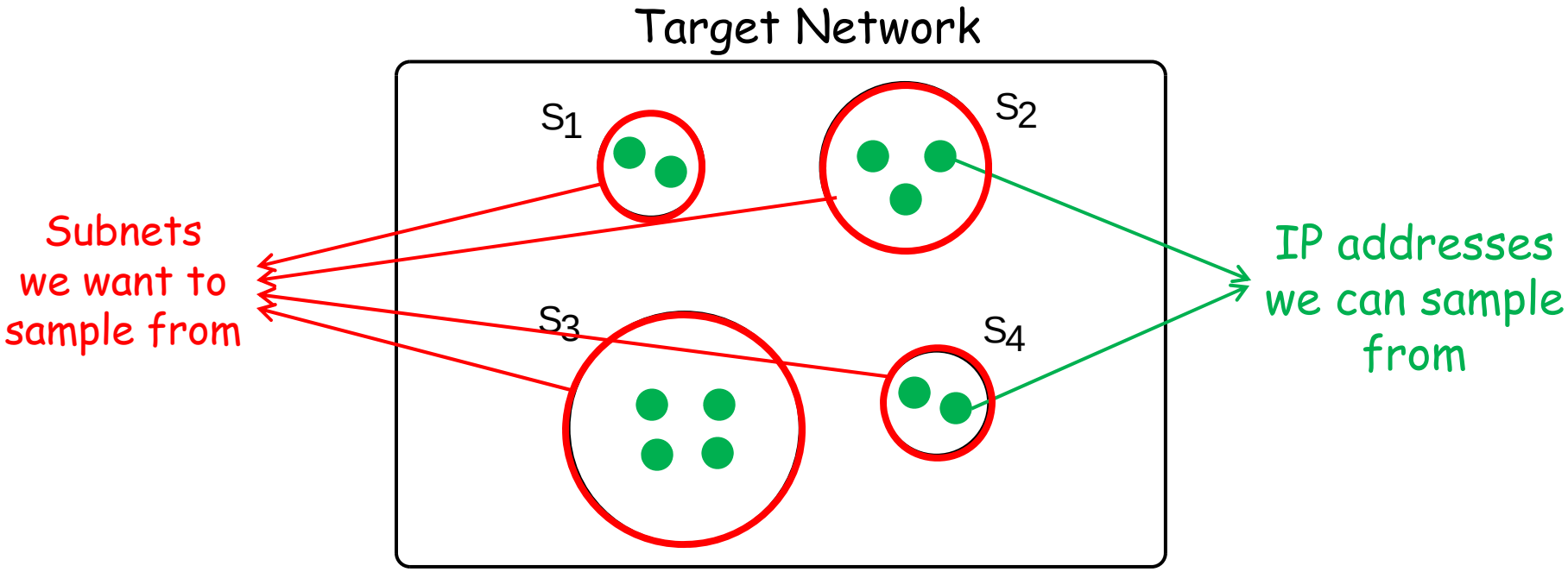
- Difficult to collect complete topology map
  - Internet/ISP topologies (eg. subnet level maps)
  - Social network graphs (eg. Facebook)
- Statistical sampling as a viable solution
- Challenges in statistical sampling
  - Sampling error vs. non-sampling error
  - Unresponsive units
  - Discrepancy between sampling & observation units
- *Goal: develop good (unbiased) estimators*

Statistical sampling of subnets in a network

- Subnet characteristics of interest
  - *Number of subnets*
  - *Subnet prefix length distribution*
  - *Mean subnet prefix length*
  - *IP address utilization ratio*

Subnet sampling

- Non-uniform sampling of subnets due to degree discrepancy



S_3 is twice as likely to be sampled as compared to S_1

Challenges in network sampling

- How to design an effective sampling scheme?
  - What is the impact of the characteristics under study?
  - What sampling/entity selection method to use ?
    - Random selection, crawling, forest fire, etc
  - What objects to sample ?
    - Nodes, links, cliques, end-to-end paths, etc
  - How to overcome application domain specific limitations to sampling ?
    - Mismatch between selection units and observation units in sampling

Thank you