Internet measurements at complexnetworks.fr

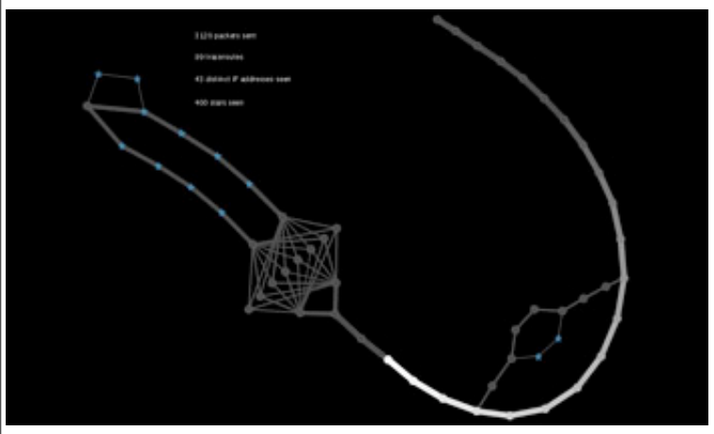
Guillaume Valadon - http://valadon.complexnetworks.fr

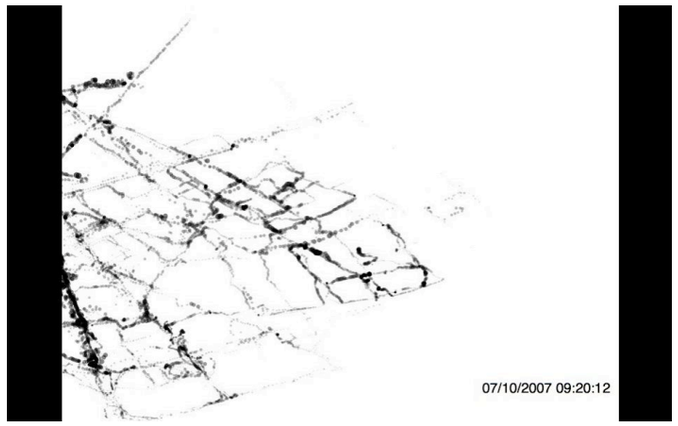
LIP6 (CNRS - UPMC)

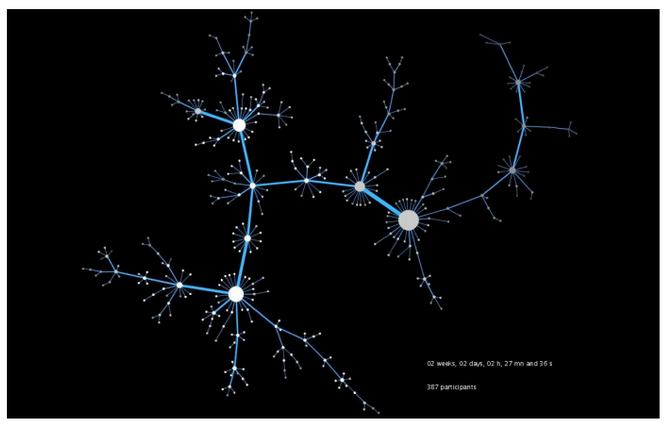
Complex Networks team
http://complexnetworks.fr

The team

- http://complexnetworks.fr : plots & videos
  - 4 permanent members : Jean-Loup Guillaume, Matthieu Latapy, Bénédicte Le Grand, Clémence Magnien
  - 2 postdocs, 9 Ph.D. students







- Focus & interests:
  - Internet topology, P2P networks, social networks
  - measurements
  - analysis

Outline

1. Internet topology measurements

   Frédéric Ouedraogo, Clémence Magnien, Matthieu Latapy

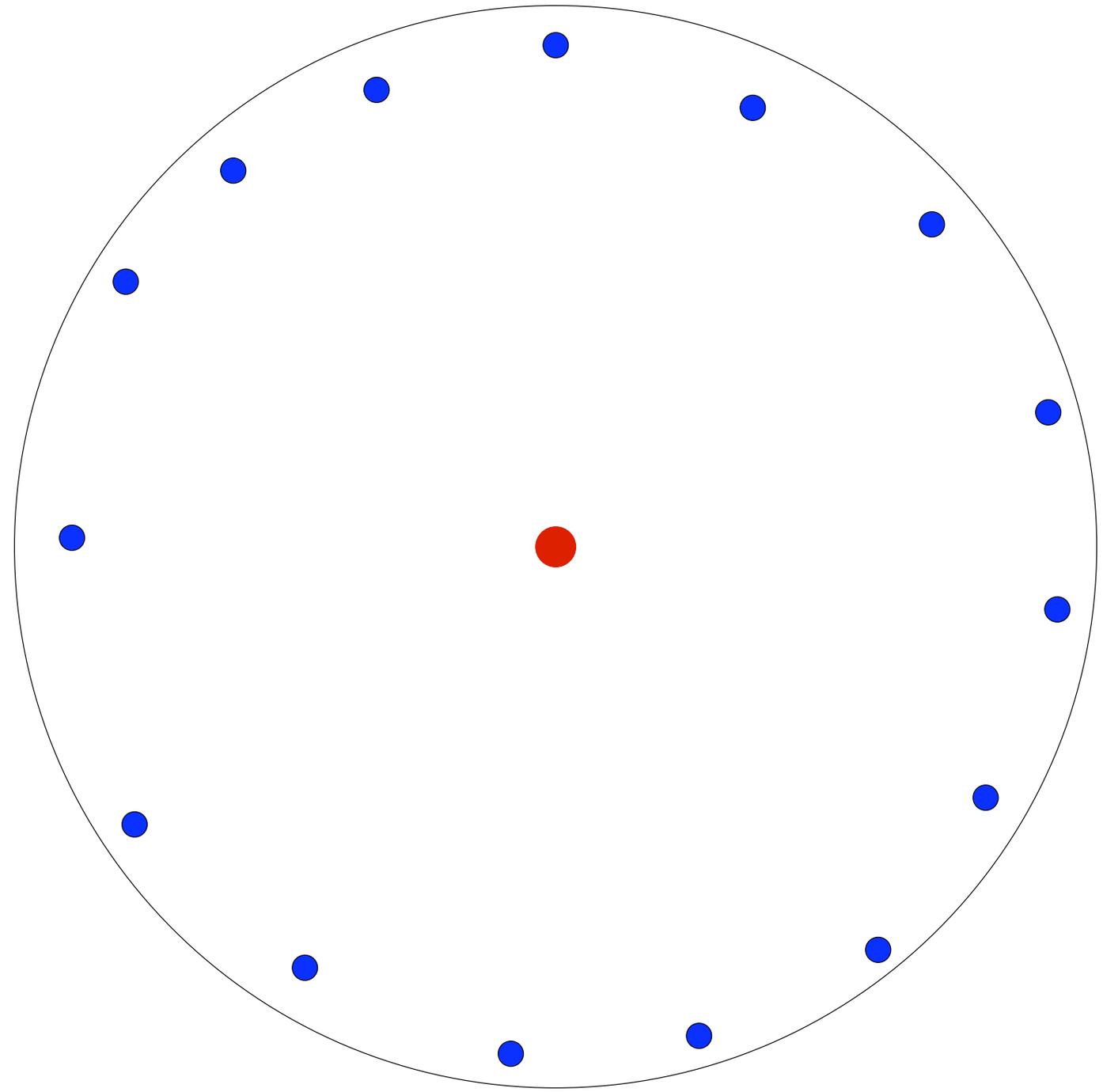
2. eDonkey measurements: server side

3. eDonkey measurements: honeypot

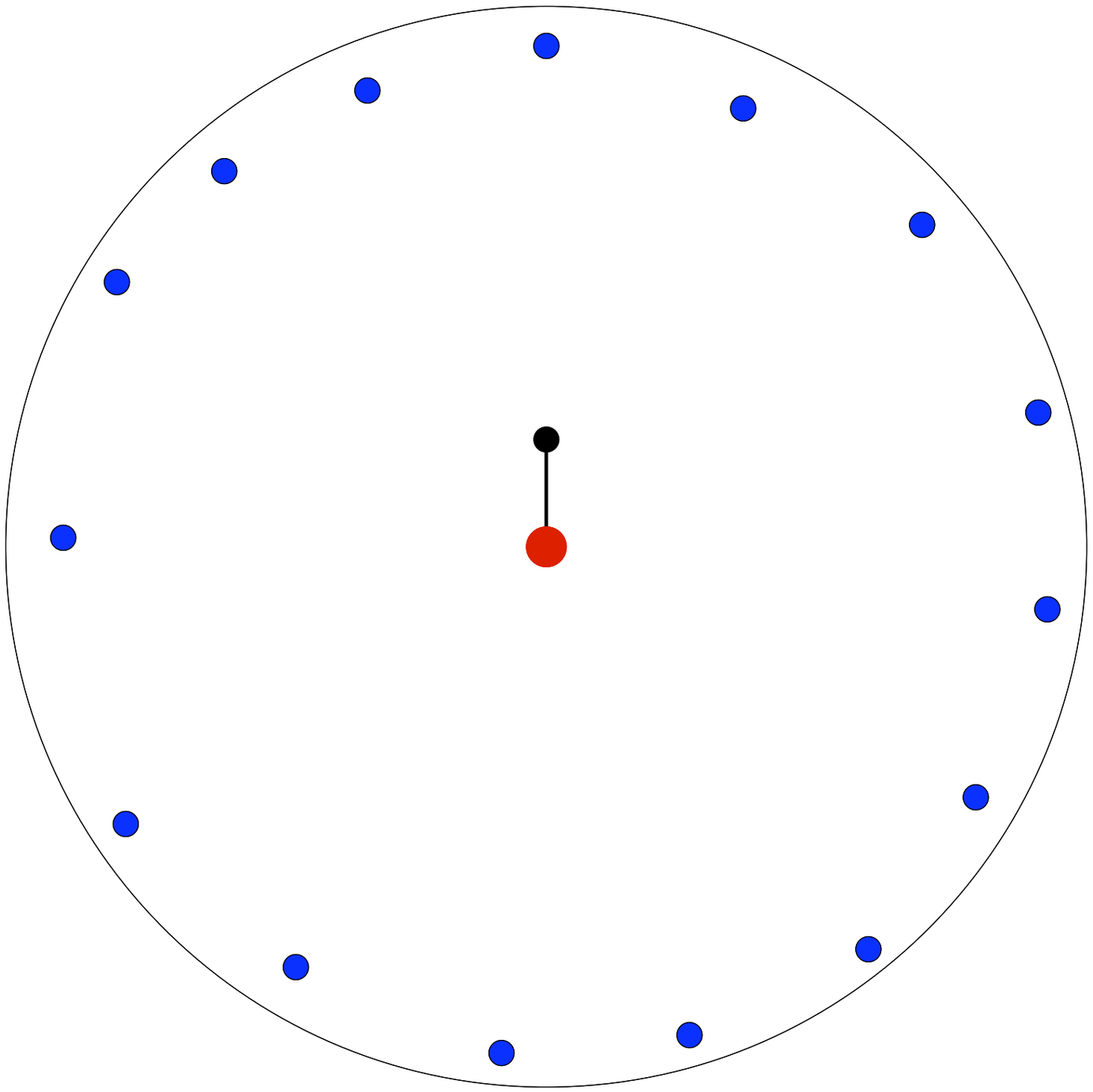
Context

- IP topology of the Internet : using traceroute-like tools
- few sources, high numbers of destinations
- measures :
  - long & high cost,
  - bias : fake links & missed links

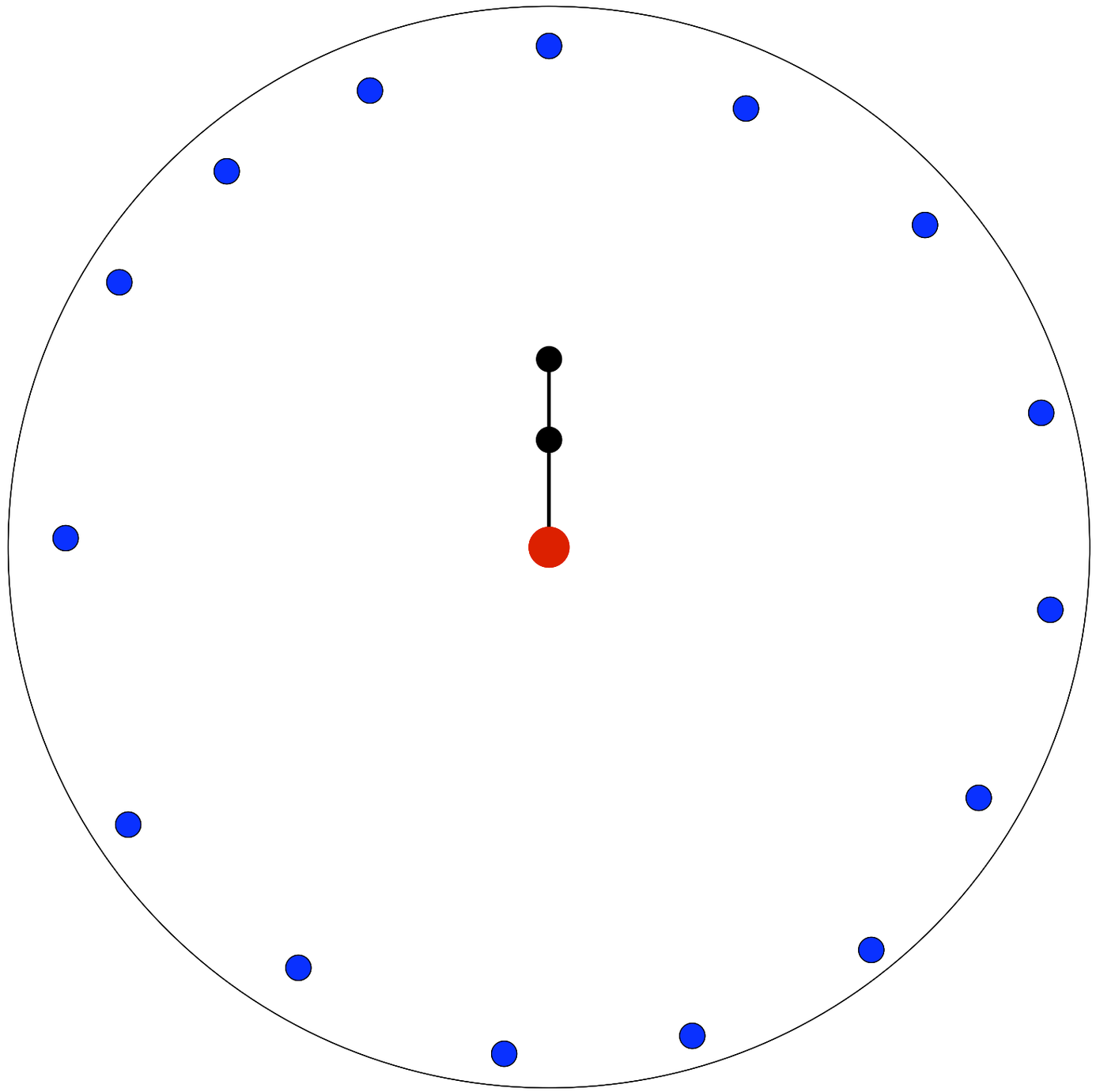
Traceroute measurements



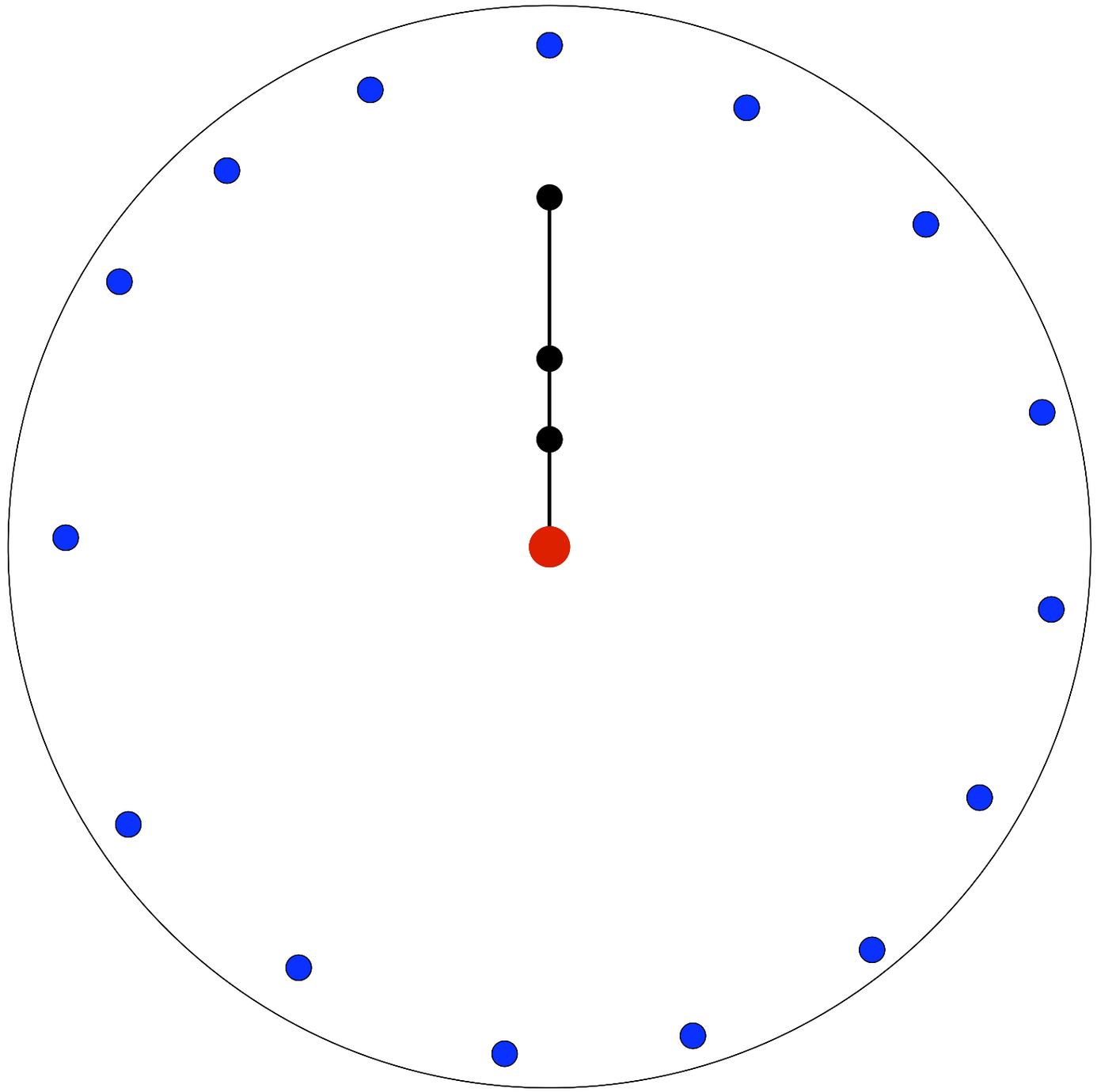
Traceroute measurements



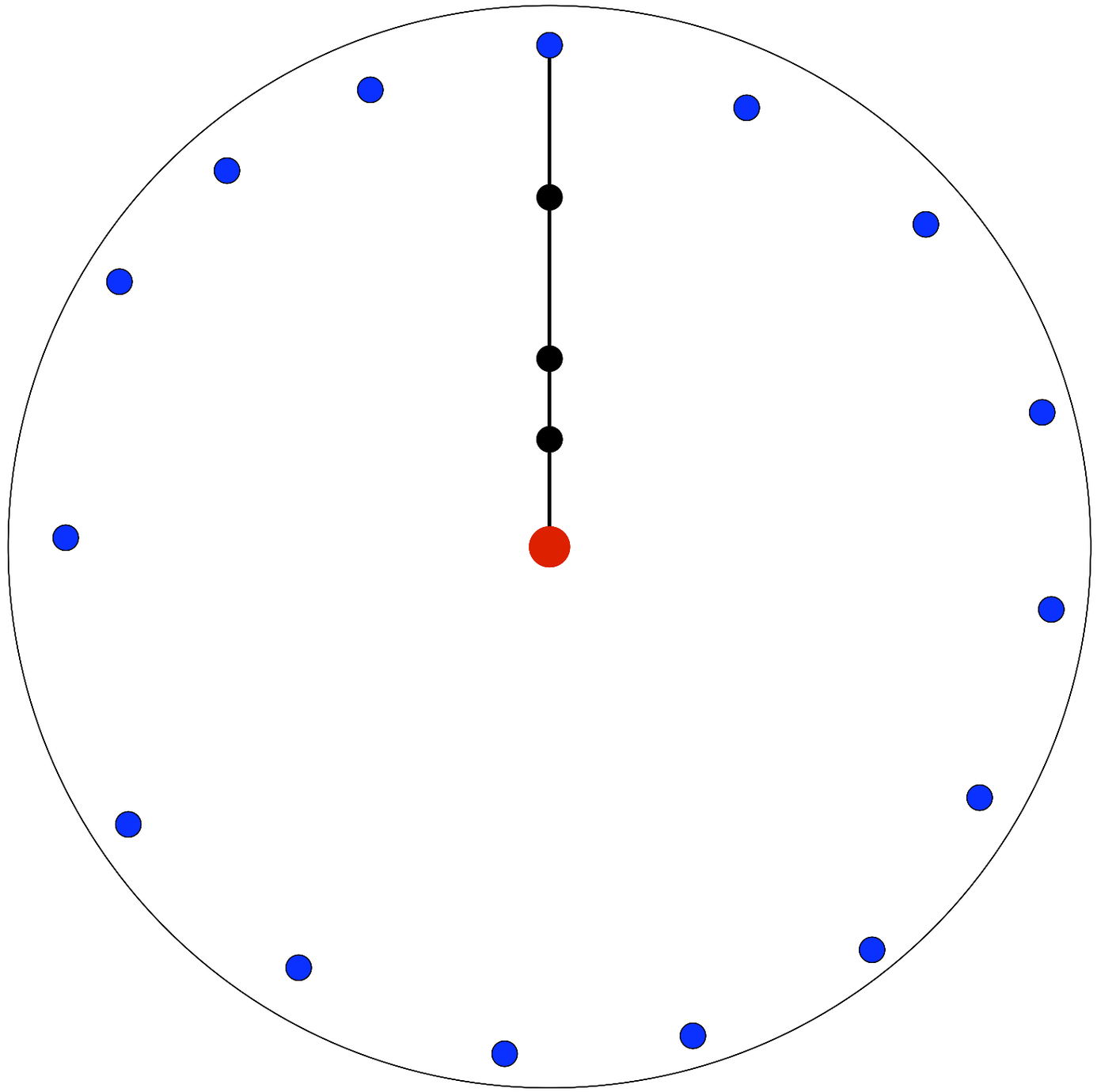
Traceroute measurements



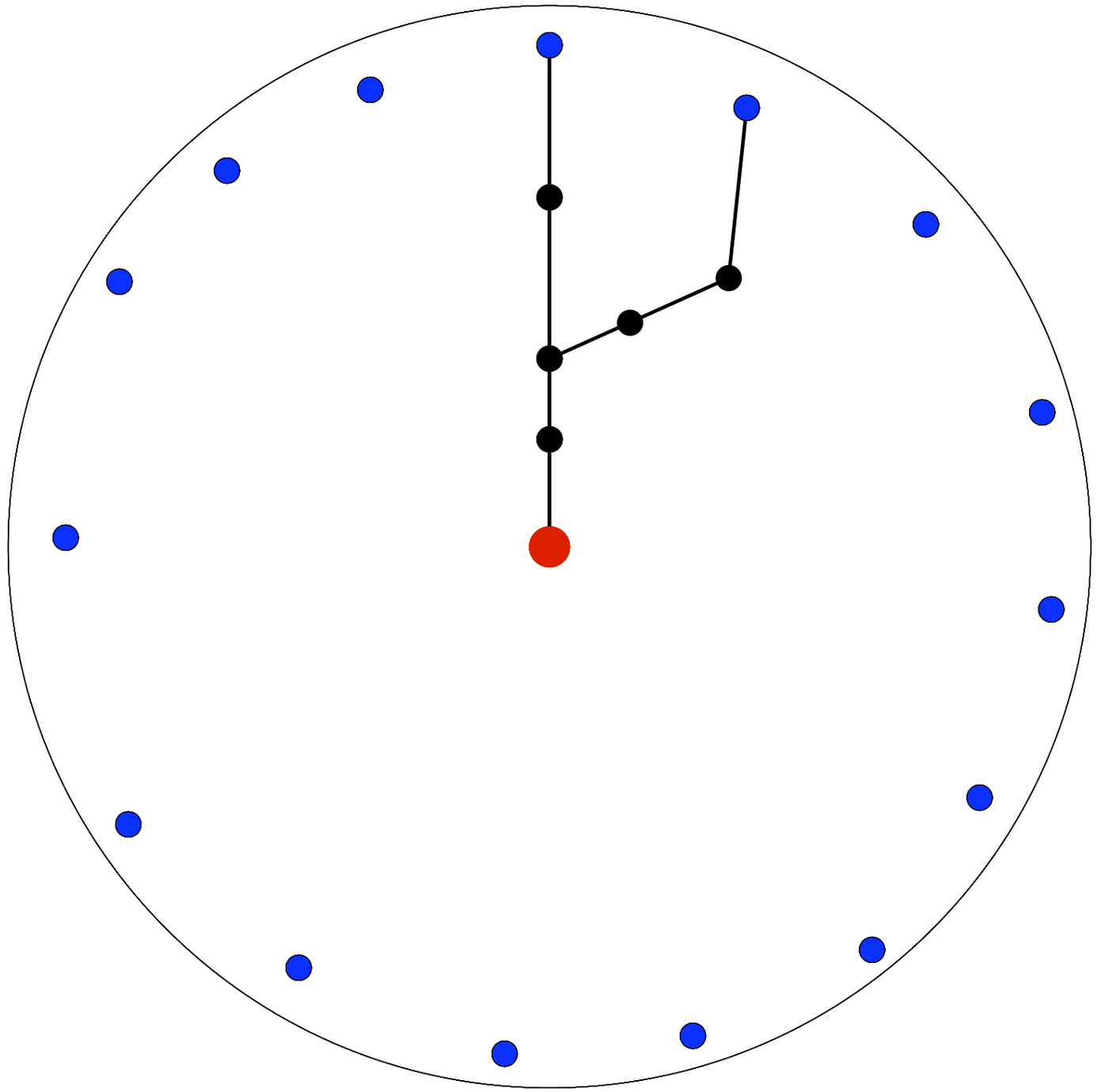
Traceroute measurements



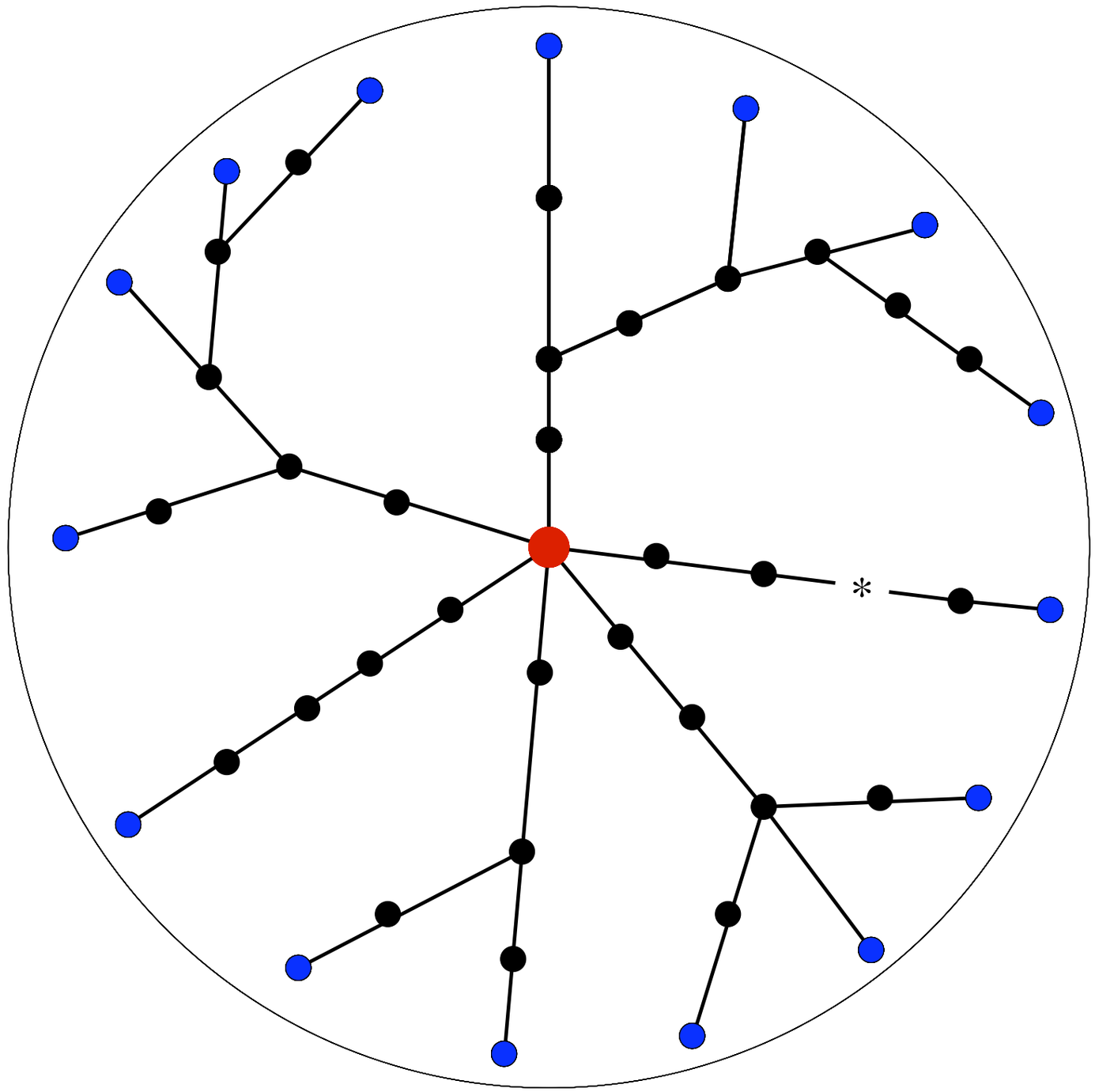
Traceroute measurements



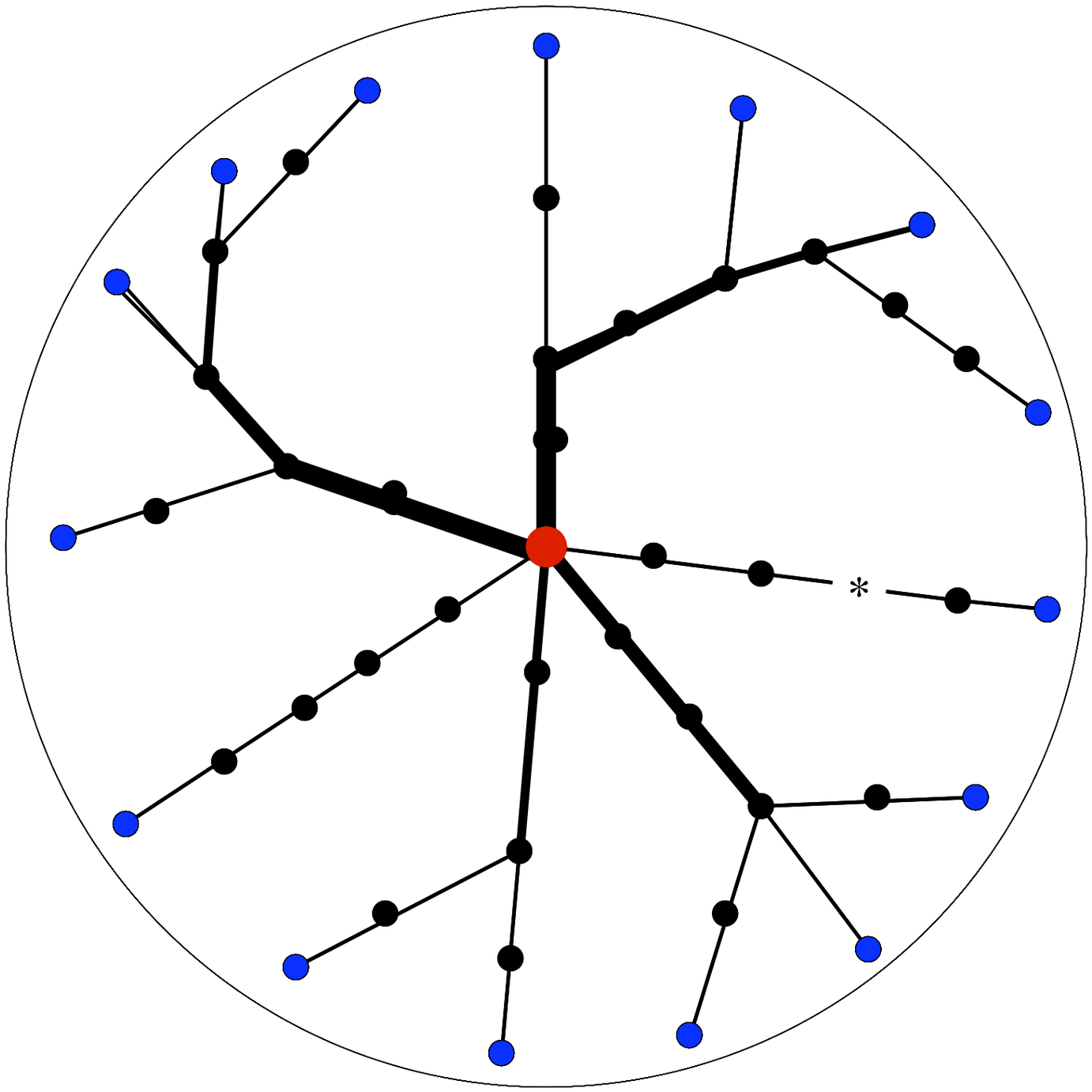
Traceroute measurements



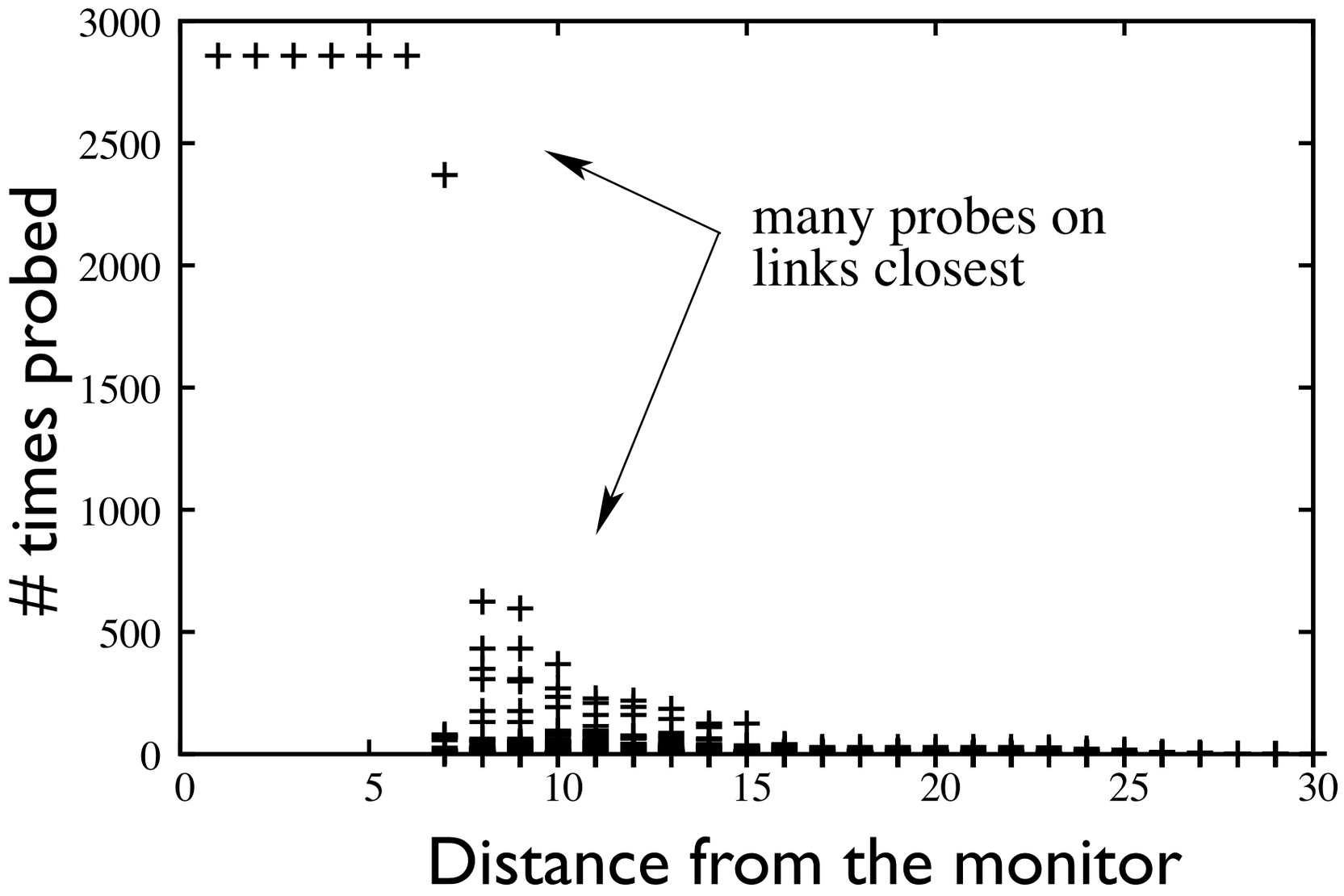
Traceroute measurements



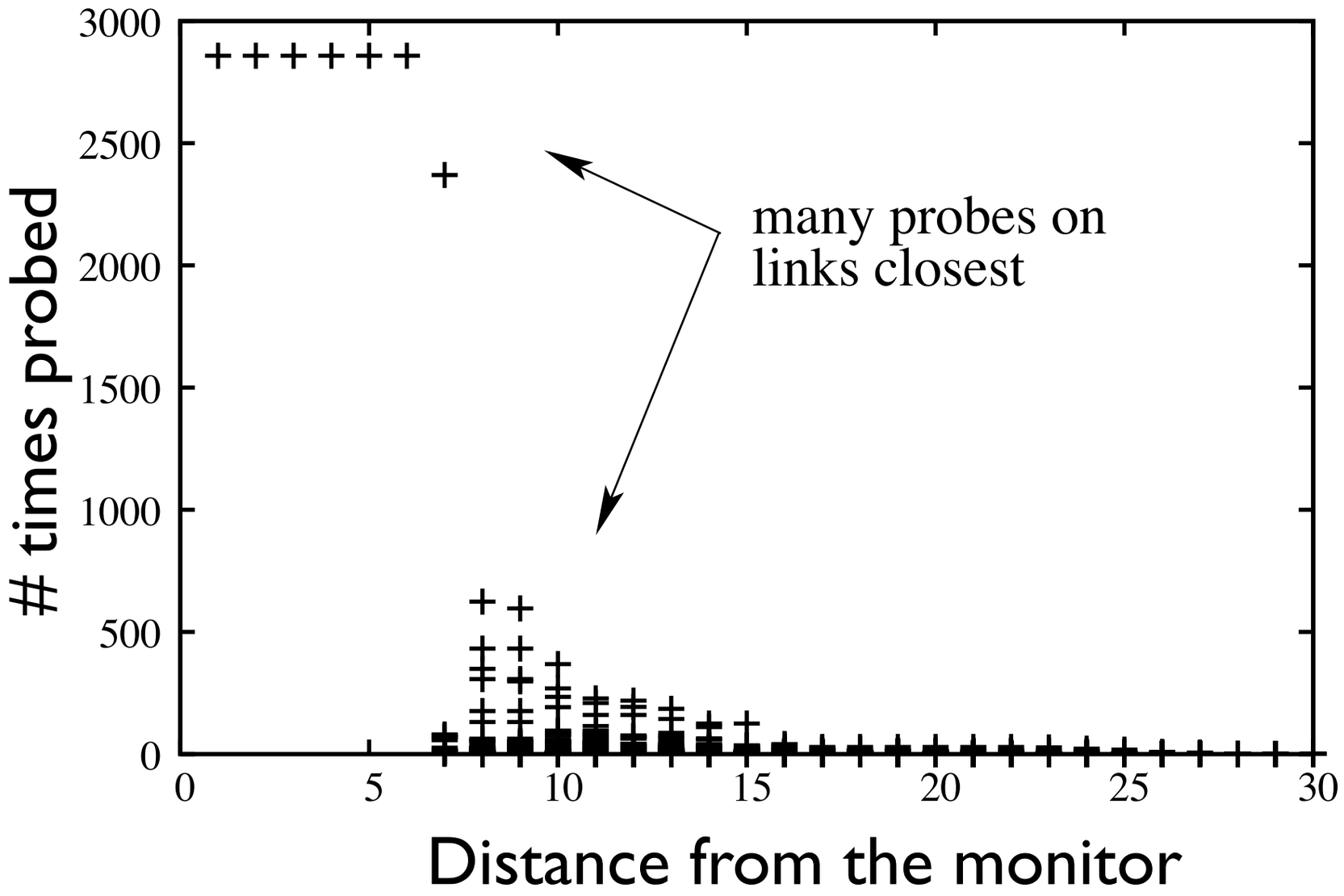
Traceroute measurements



Traceroute: unbalanced load



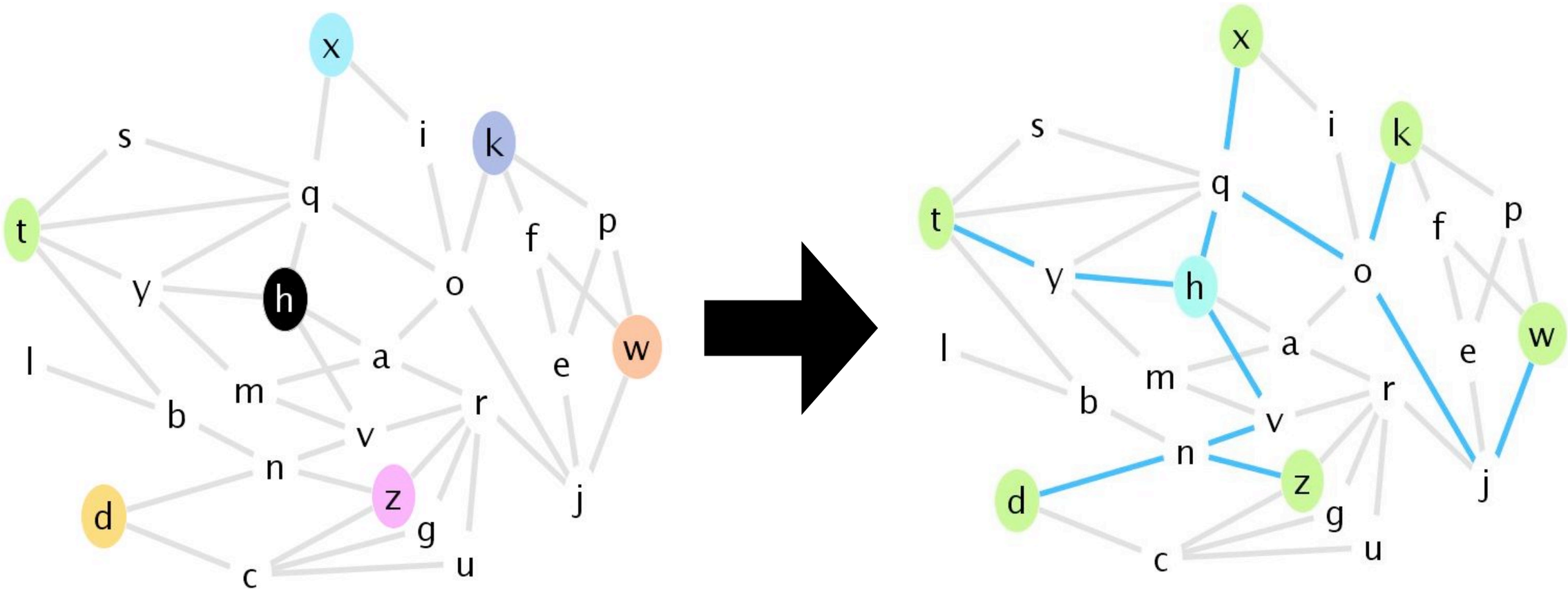
Traceroute: unbalanced load



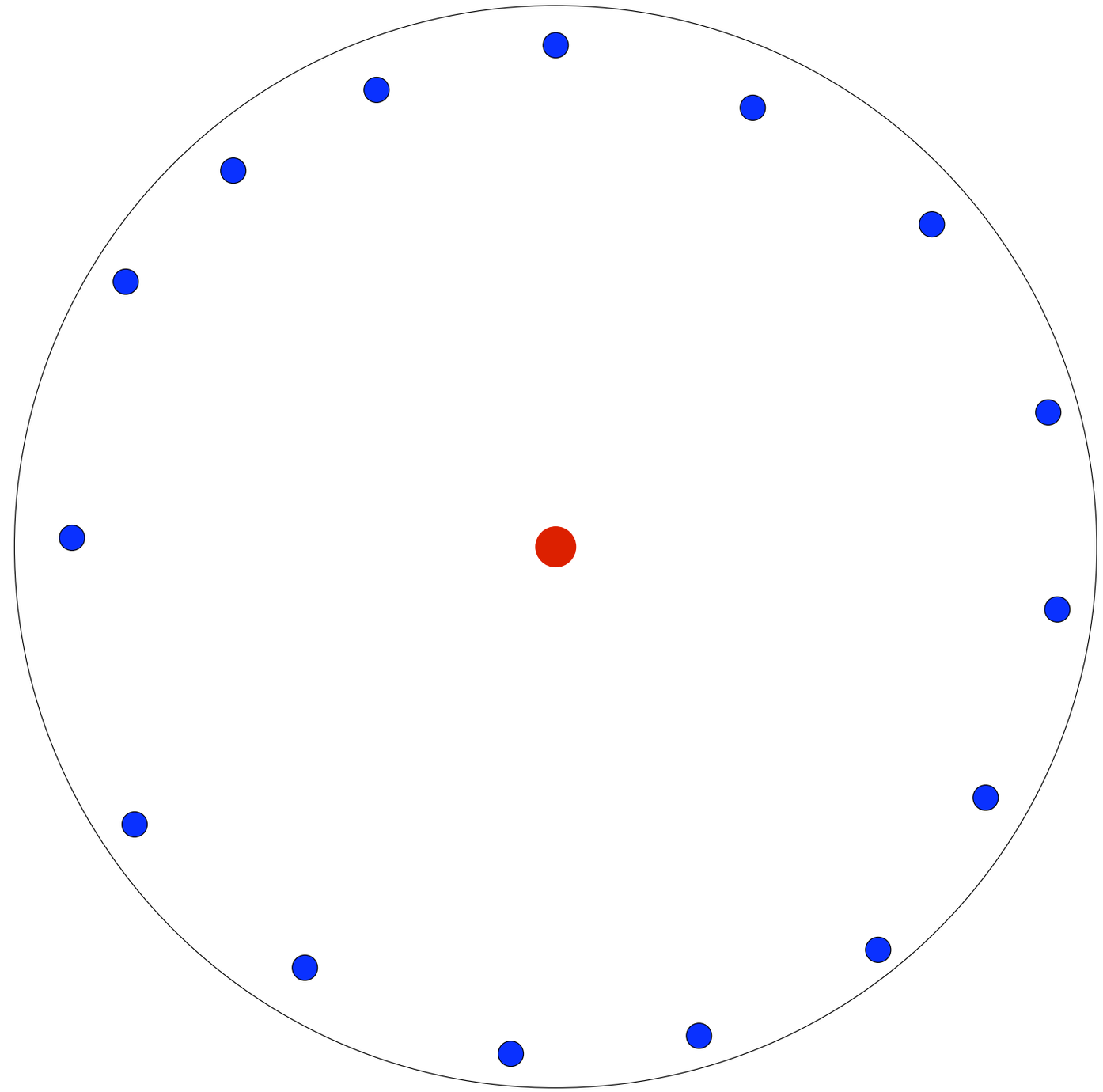
Traceroute limitations: unbalanced load, information redundancy, obtained view is not a tree

Ego-centered view

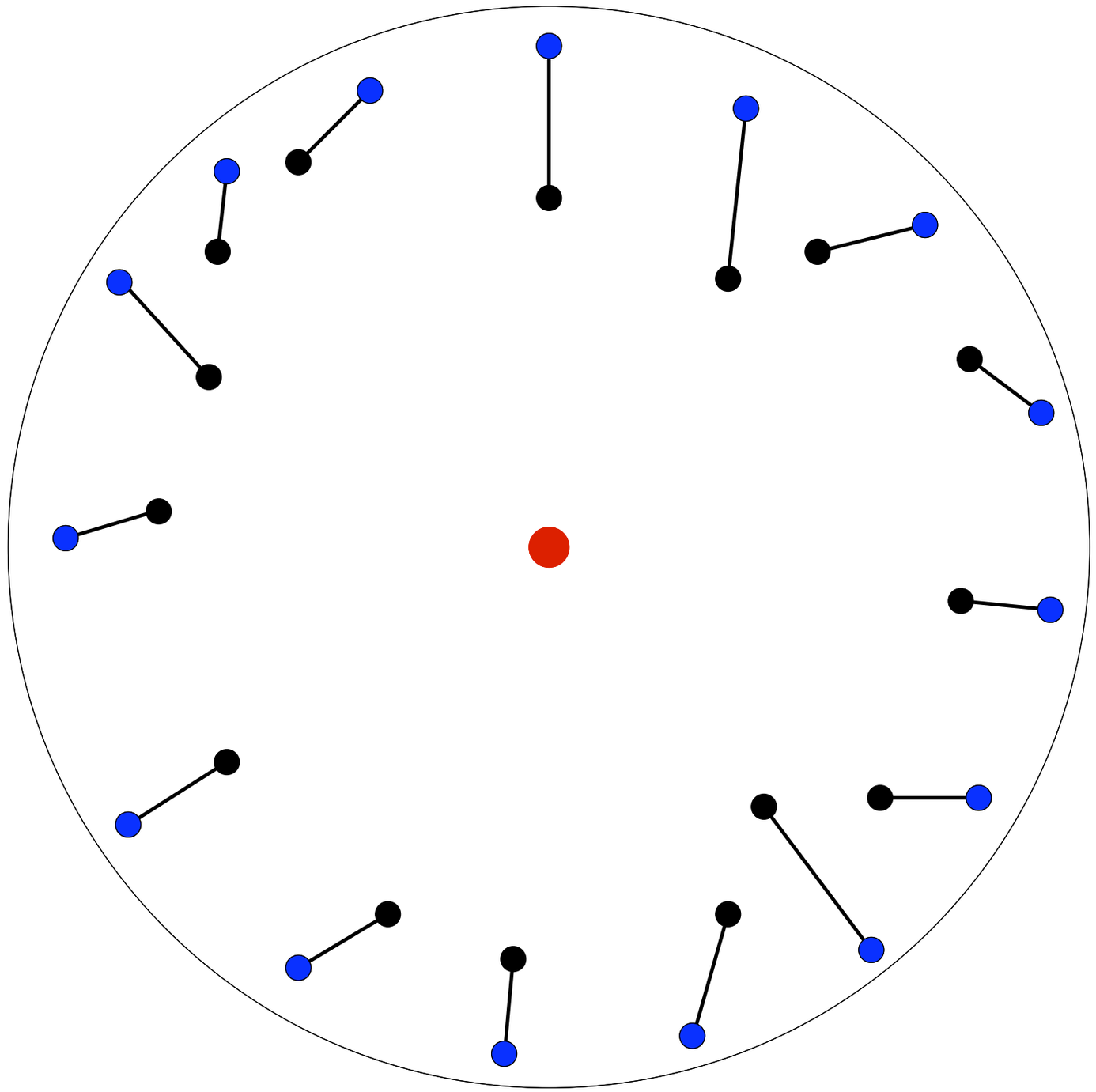
- ***tracetree*** http://data.complexnetworks.fr/Radar/
  - one source
  - fixed set of destinations
  - the result is a tree
  - fast measurement (~100 round per day)



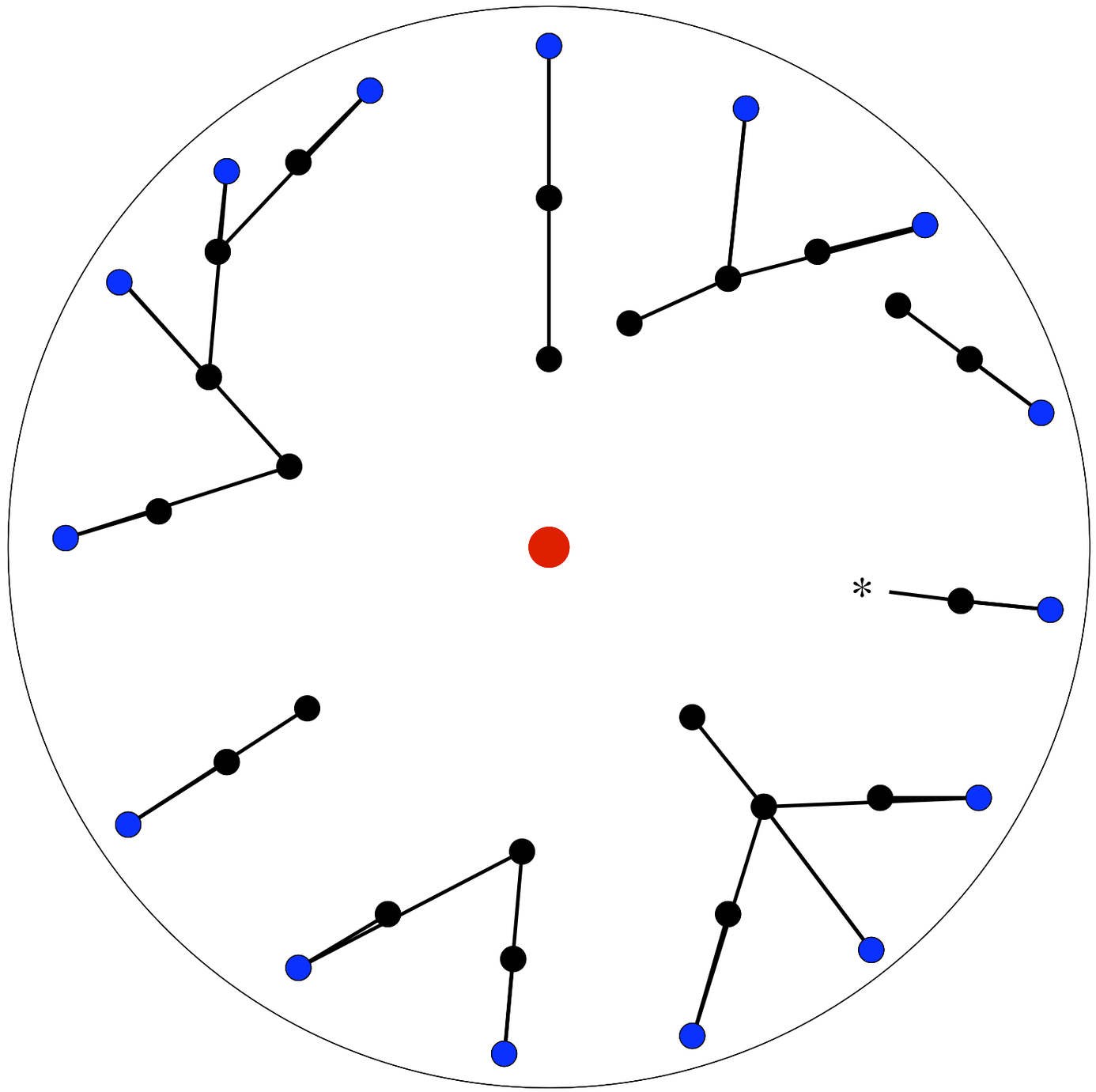
Tracetree measurements



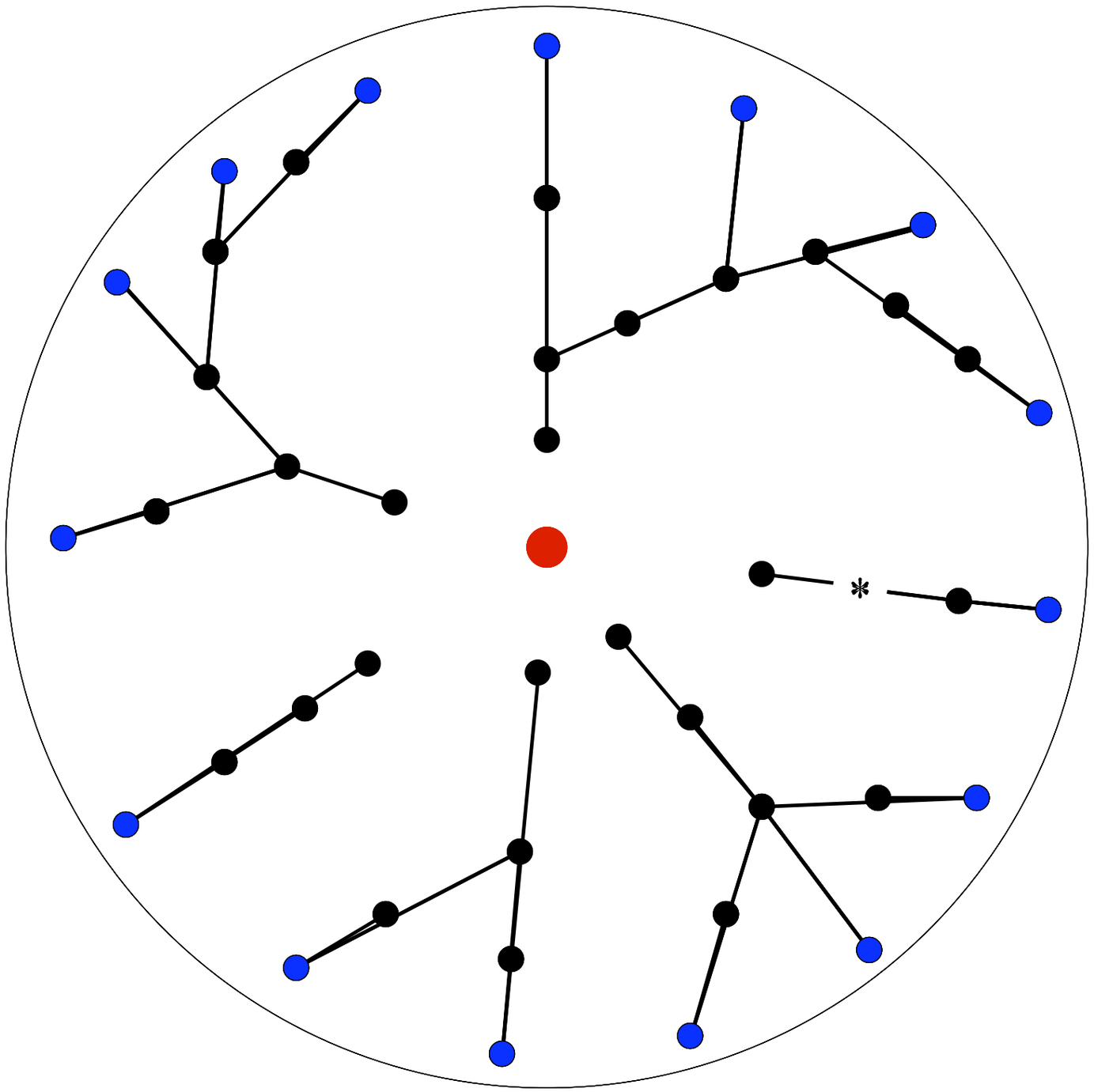
Tracetree measurements



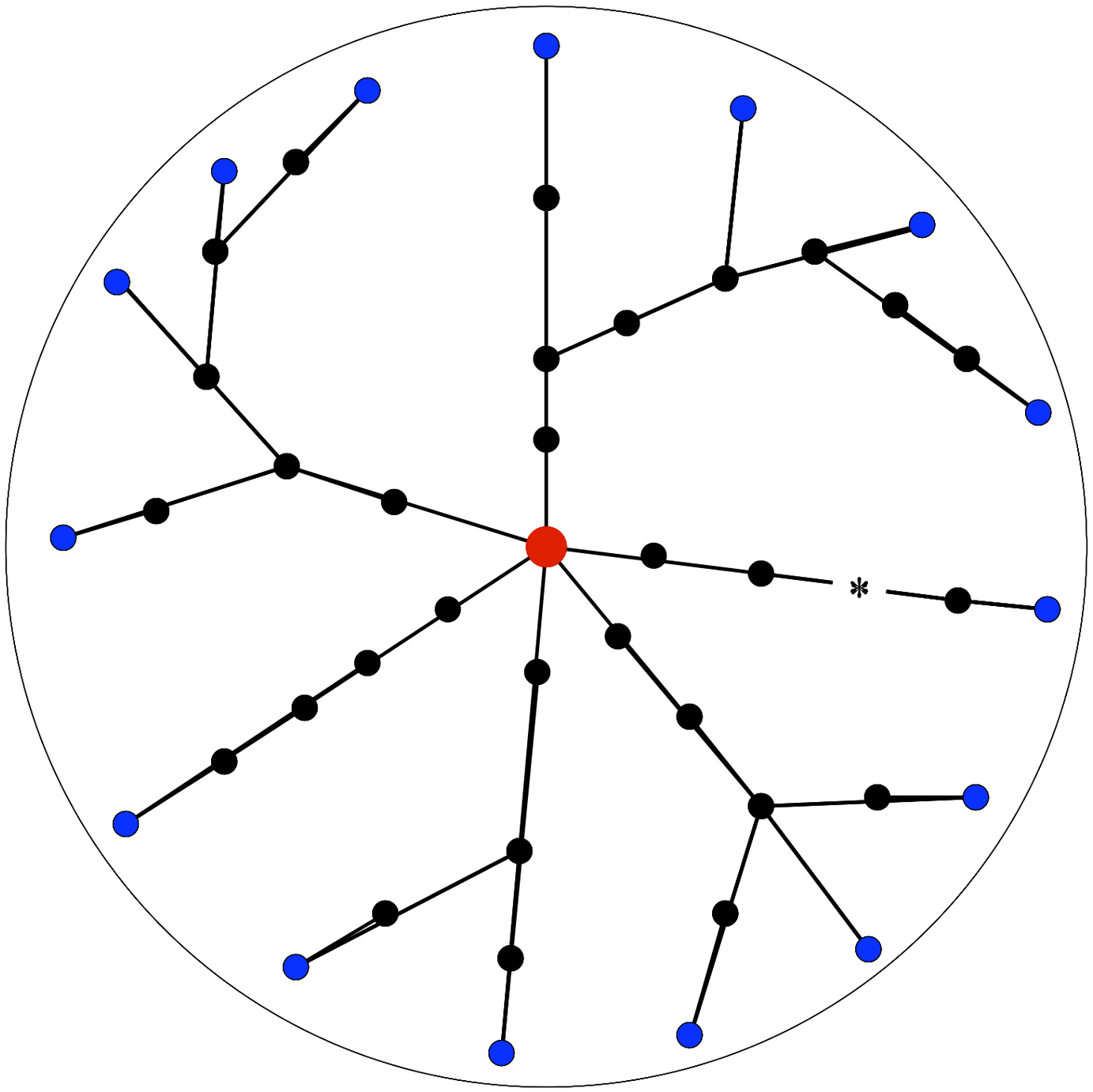
Tracetree measurements



Tracetree measurements



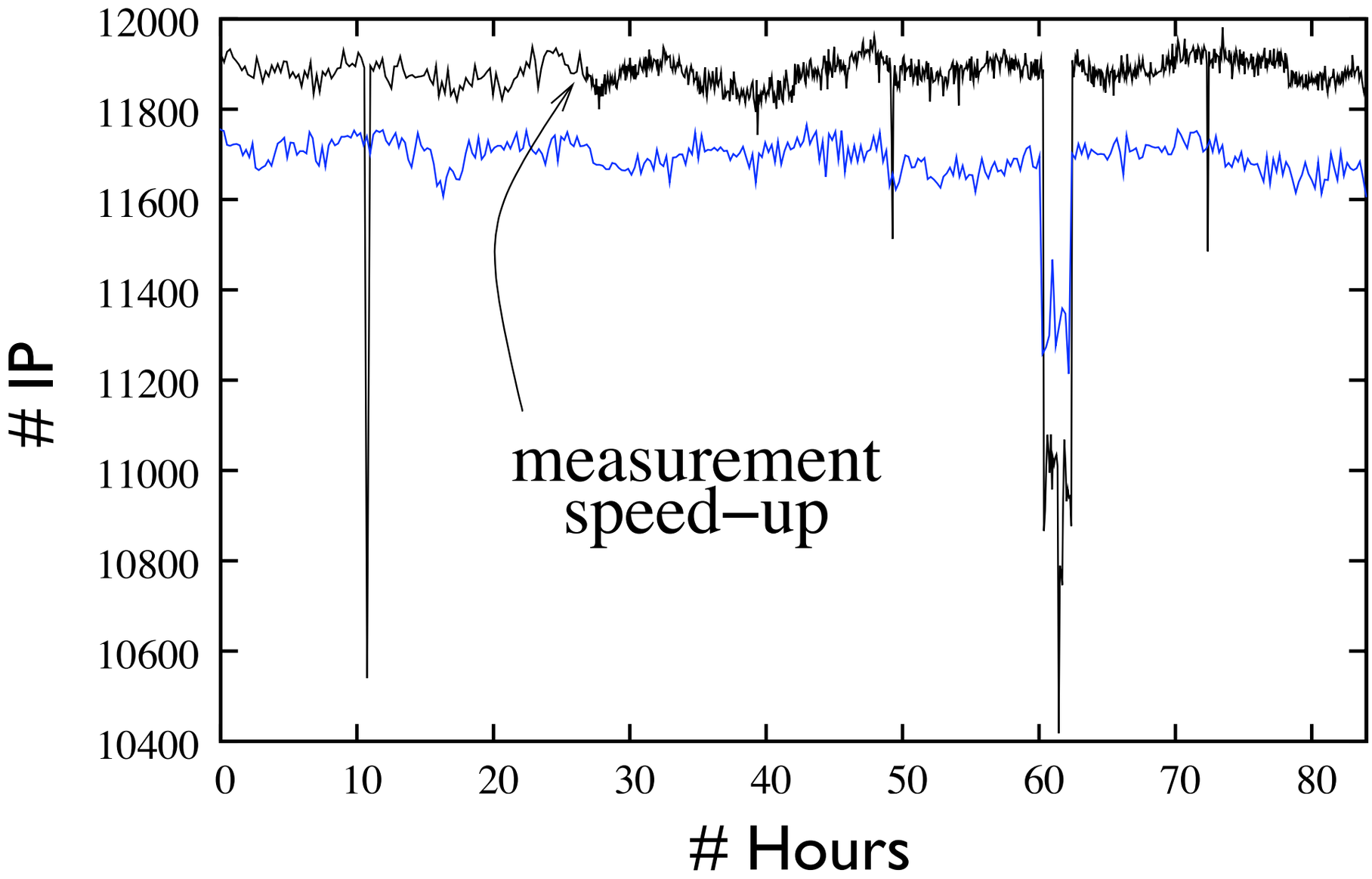
Tracetree measurements



Parameters

- Many parameters:
  - number of destinations
  - delay between rounds
  - maximum TTL ?
  - ...
- We want:
  1. high frequency
  2. large ego-centered view
  3. low network load

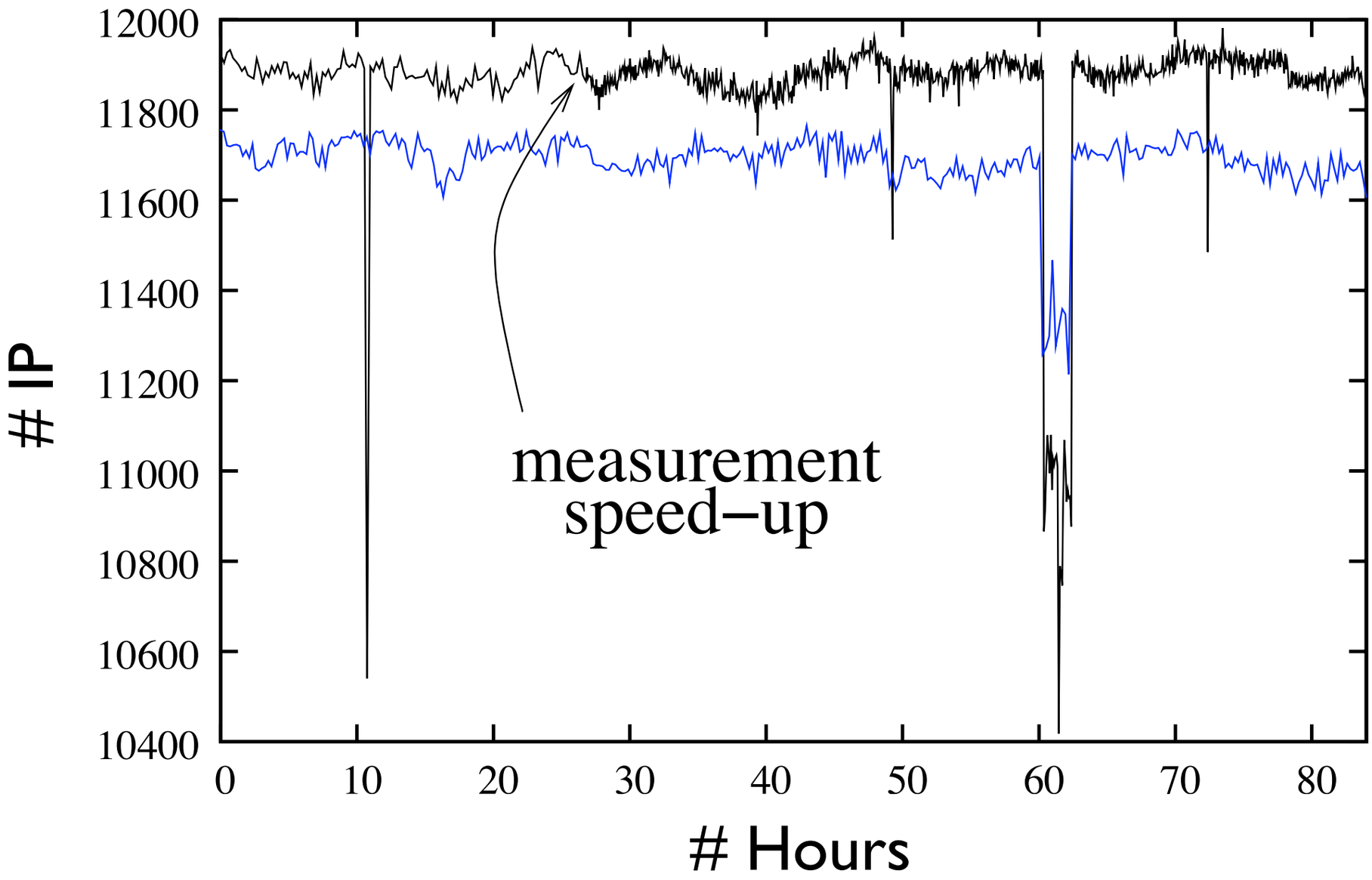
Parameters : frequency



Test monitor

Control monitor

Parameters : frequency

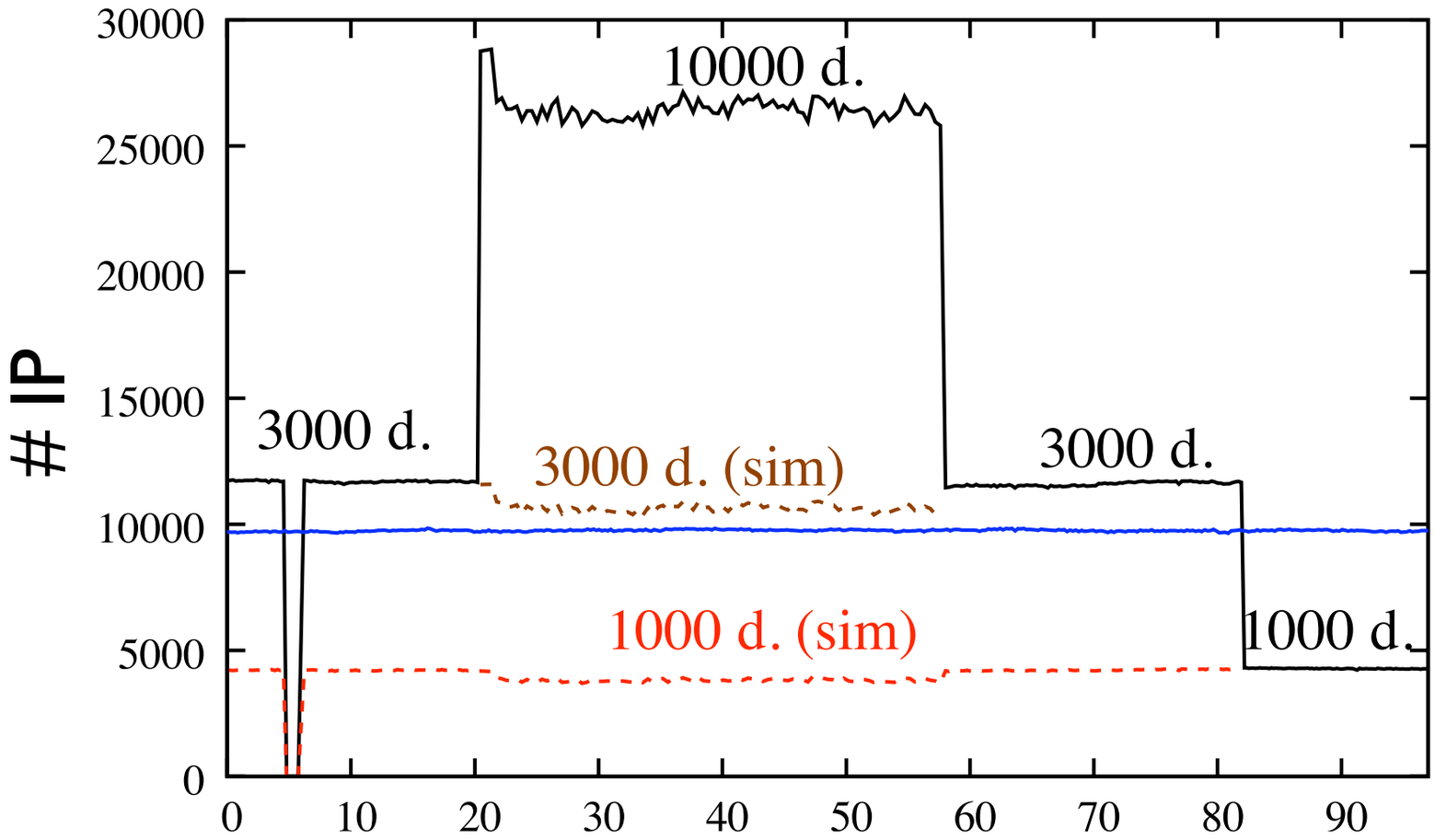


Test monitor

Control monitor

frequency has no impact on discovered addresses

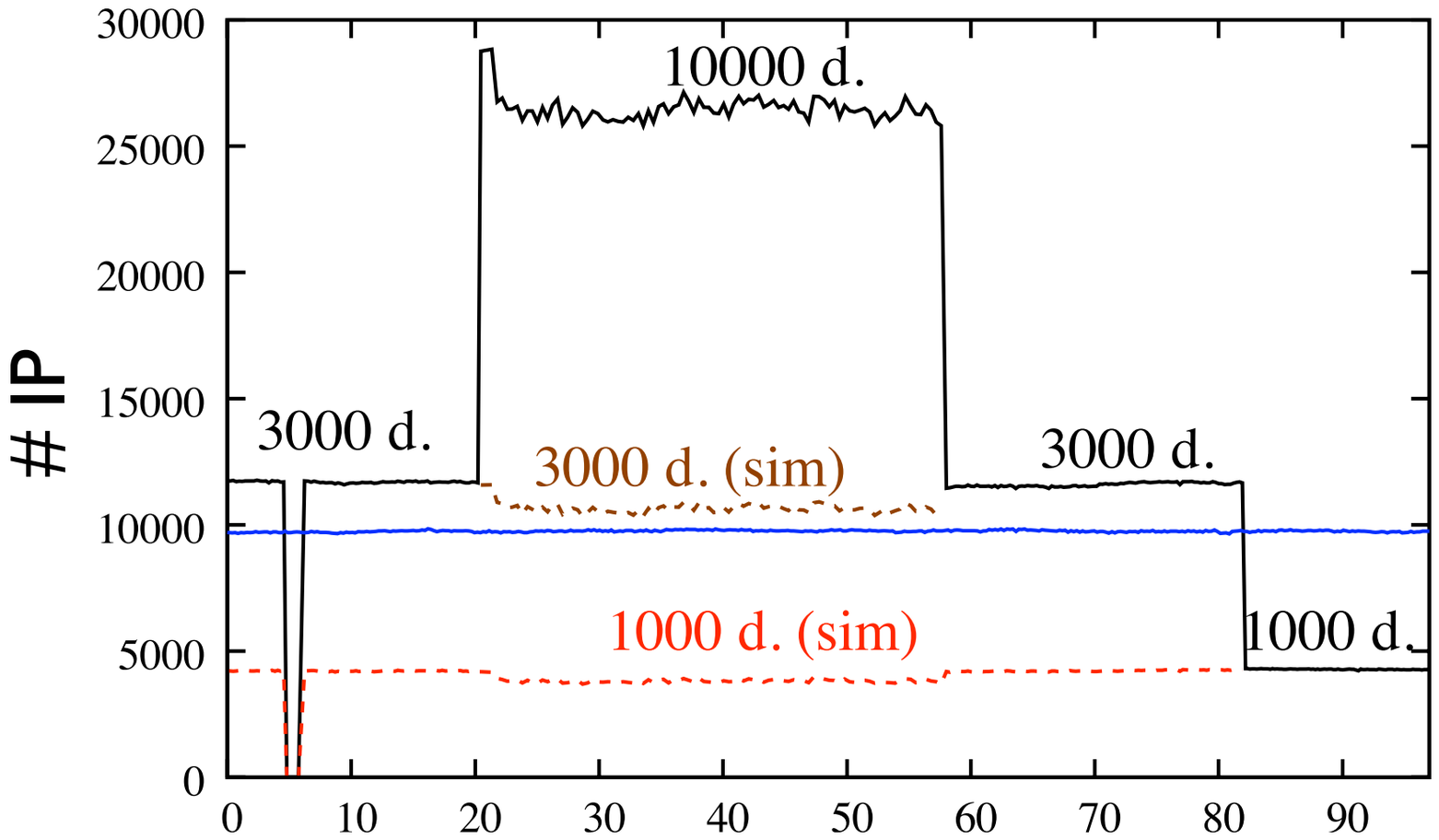
Parameters : destination number



Test monitor

Control monitor

Parameters : destination number



Test monitor

Control monitor

too many destinations == loss of efficiency

Available data

[ADN'08, ICIMP'09]

- Two parameter sets:
  - **normal:** 3000 destinations, max TTL 30, 10 minutes delay (~100 rounds / day)
  - **fast:** 1000 destinations, max TTL 15, 1 minute delay (~ 800 rounds / day)

- Available data at http://data.complexnetworks.fr/Radar/
  - several sets of random destinations
  - 150 monitors
  - several months of uninterrupted measures

Outline

1. Internet topology measurements

2. eDonkey measurements: **server side**

   Frédéric Aidouni, Matthieu Latapy, Clémence Magnien

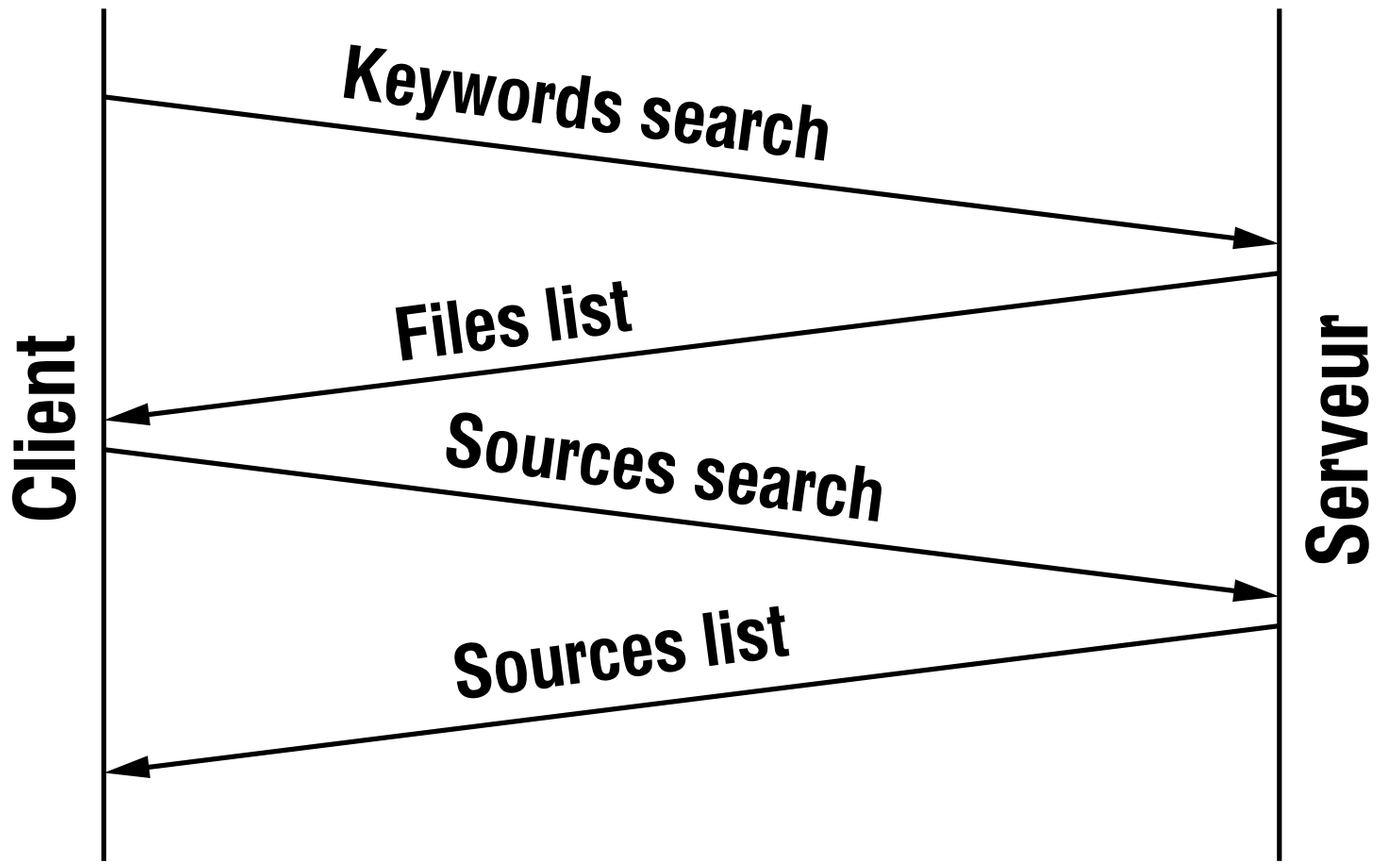
3. eDonkey measurements: honeypot

Context

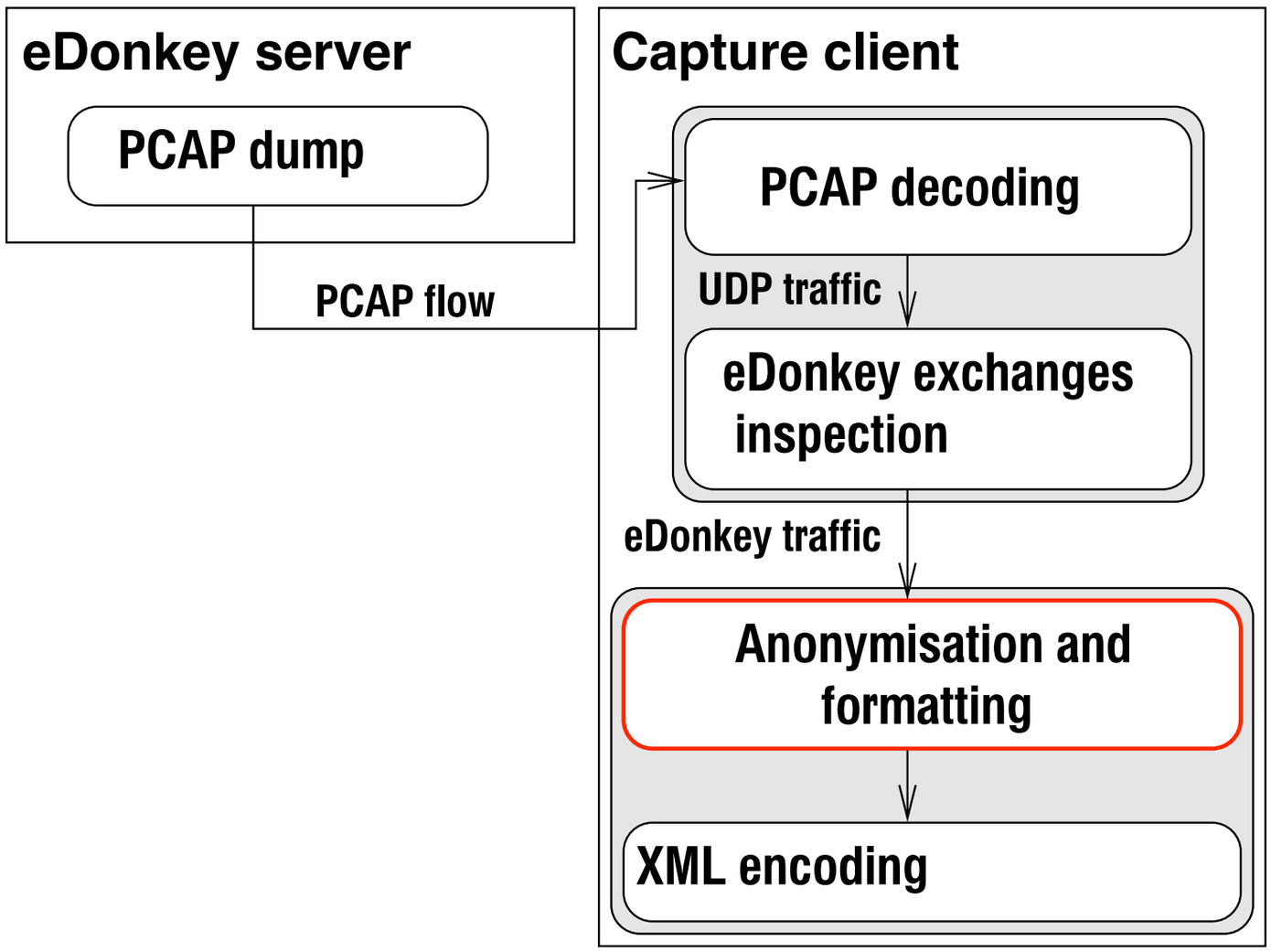
- study *exchanges* in P2P networks
  - files diffusion
  - communities of interests
  - popularity
- some motivations
  - understand users behaviour
  - develop new P2P protocols
  - blind content detection
  - detect pedophile activities
  - protocol and exchange simulations

eDonkey exchanges

1. inter-clients: file downloads
2. inter-servers: statistical data
3. clients-servers: files & sources search

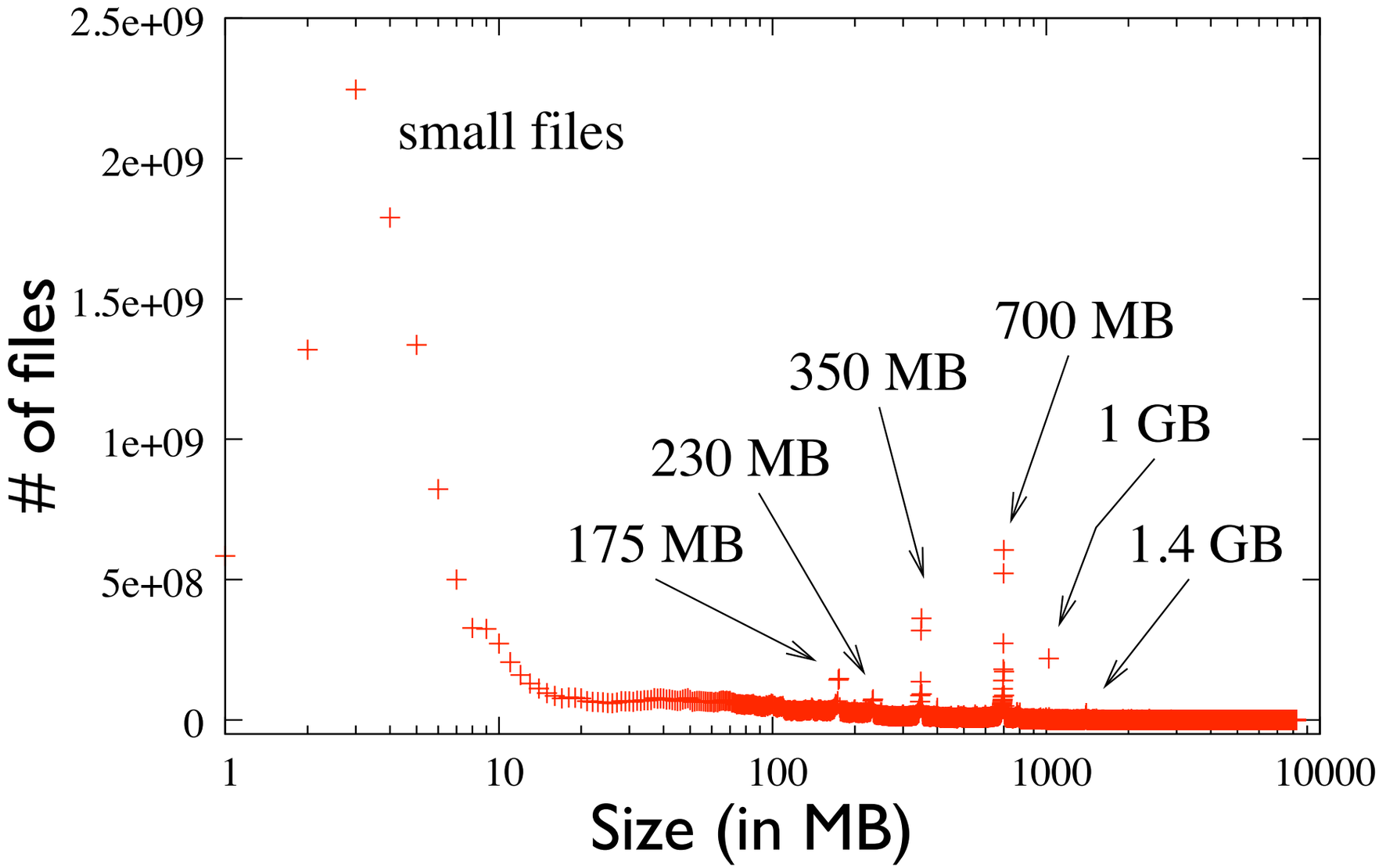


Capturing traffic on a real server




```
<opcode dir="received" TS="2786402.373146" IP="0045125351"
type="high" port="02029"><OP_GLOBSEARCHREQ>
<tags count="1"><anon-string>3108886</anon-string></tags>
</OP_GLOBSEARCHREQ></opcode>
```

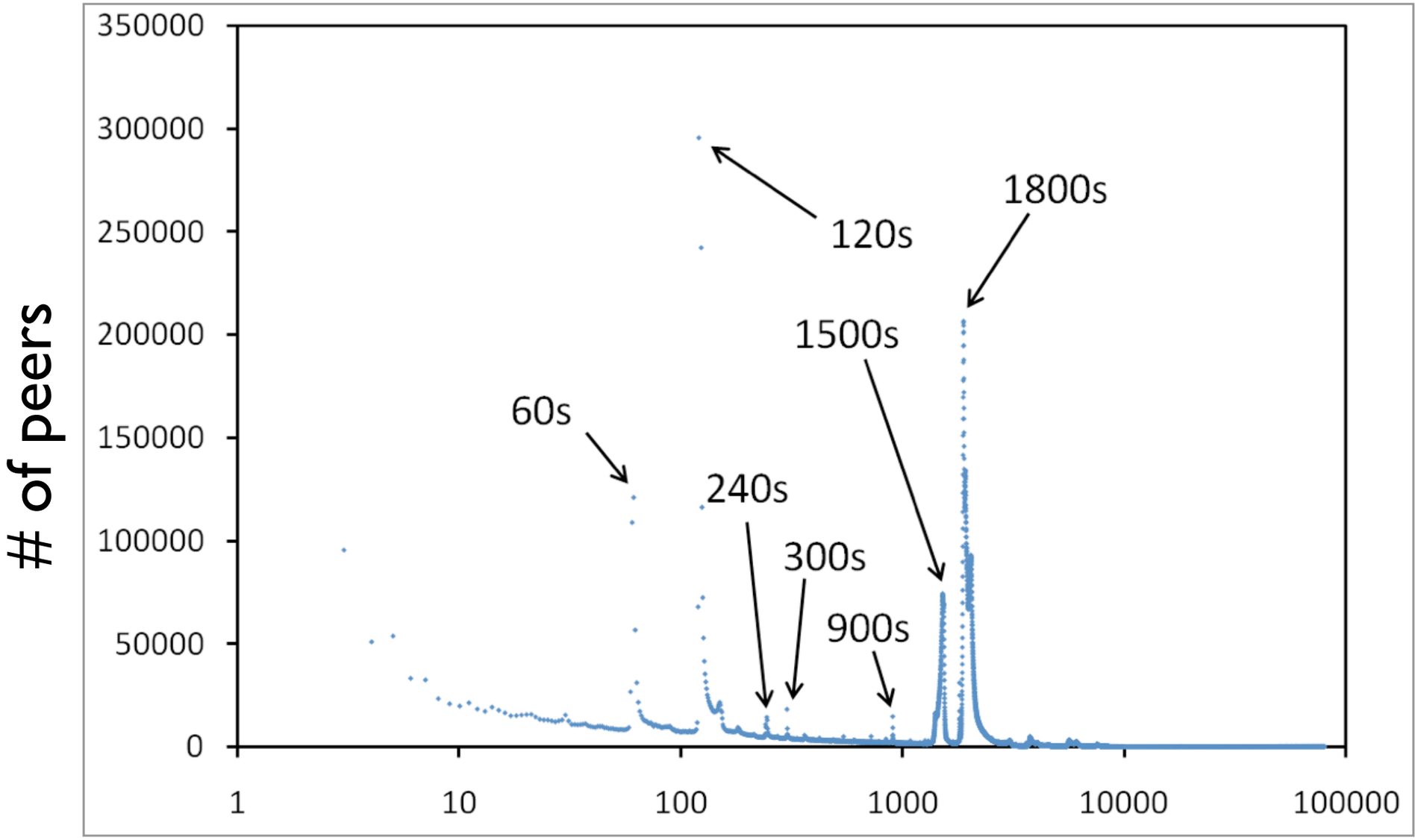
Basic analysis : files sizes



- obtained from the server answers
- CD-ROM size and fractions (1/2, 1/3, and 1/4)

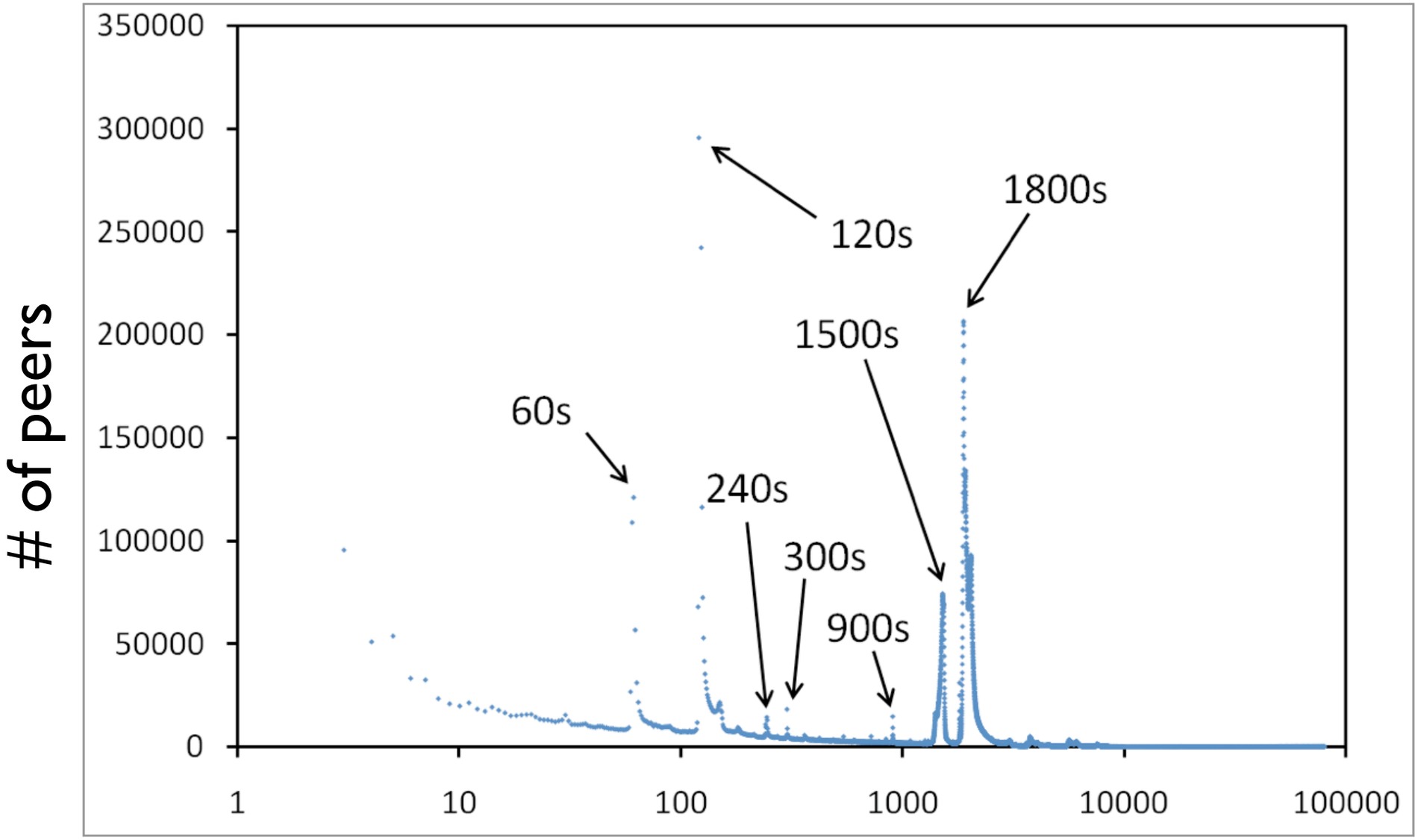
➡ related to classical sizes of storage support

Basic analysis : time between queries



Time (in seconds)

Basic analysis : time between queries



Time (in seconds)

regularities of queries

Resulting data set in numbers

[HotP2P'09]

- 10 weeks measurements
- ~500 GB of compressed XML
- ~ 10 billions messages
- ~ 90 millions clients
- ~ 280 millions of distinct files

➡ anonymized data available online at http://antipaedo.lip6.fr

Outline

1. Internet topology measurements

2. eDonkey measurements: server side

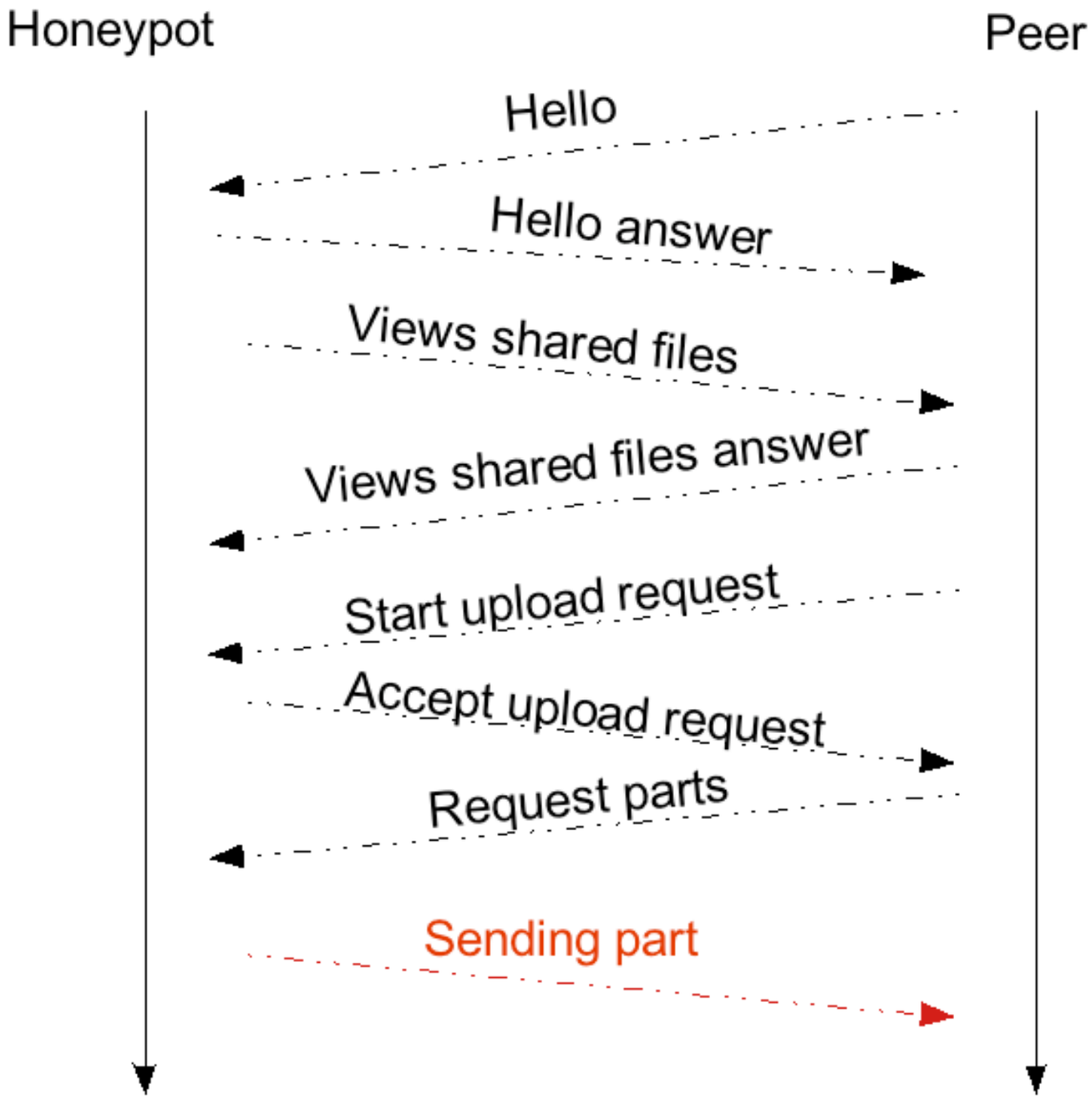
3. eDonkey measurements: **honeypot**

Oussama Allali, Matthieu Latapy, Clémence Magnien

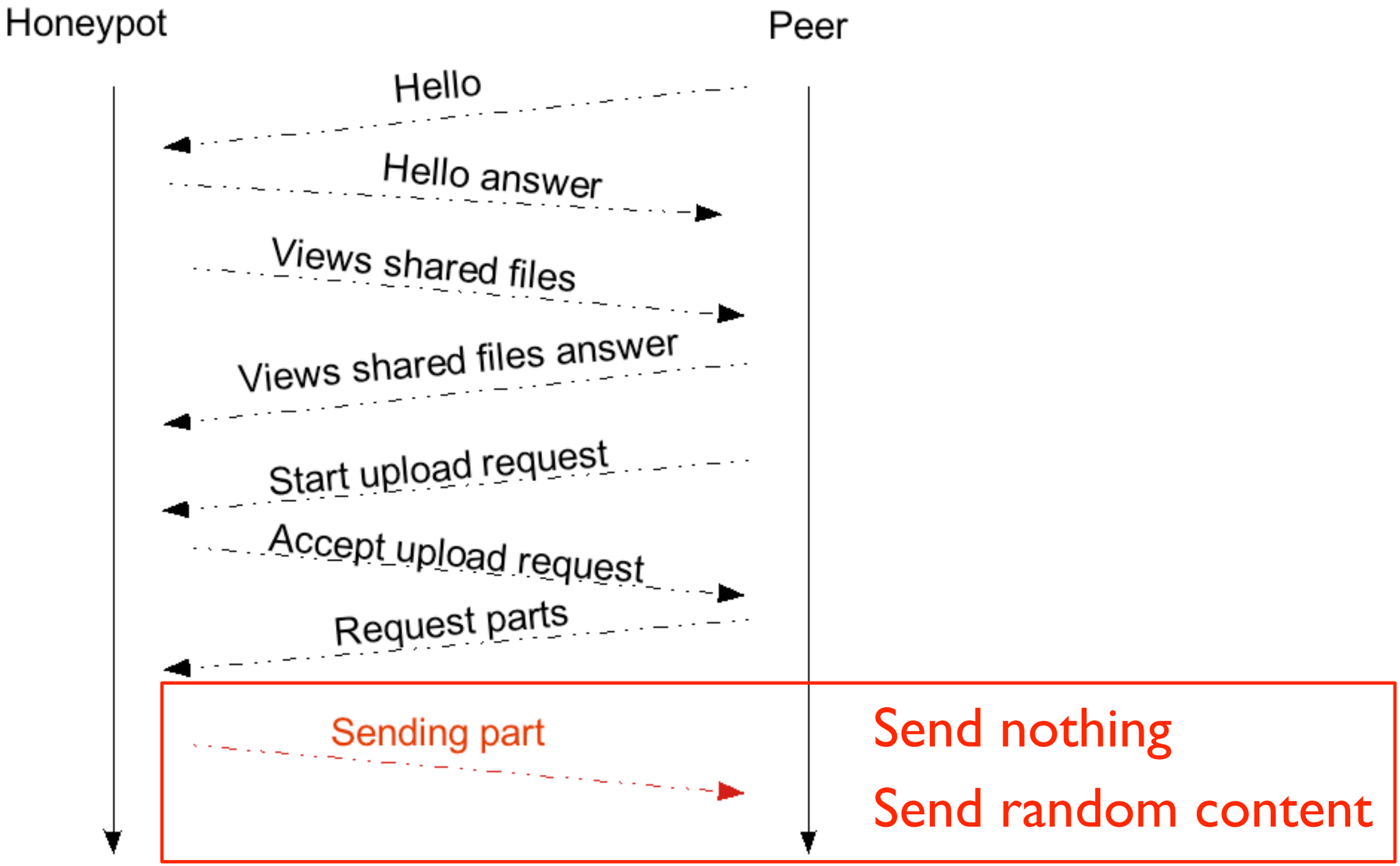
Honeypot based measurements

- eDonkey honeypot:
  - customized eDonkey client
  - announce files to a server (filename, hash, size)
  - log queries made by regular clients
- Manager:
  - control distributed honeypots
  - send commands to honeypots: server to connect, files to exchange, ...

eDonkey exchanges



eDonkey exchanges

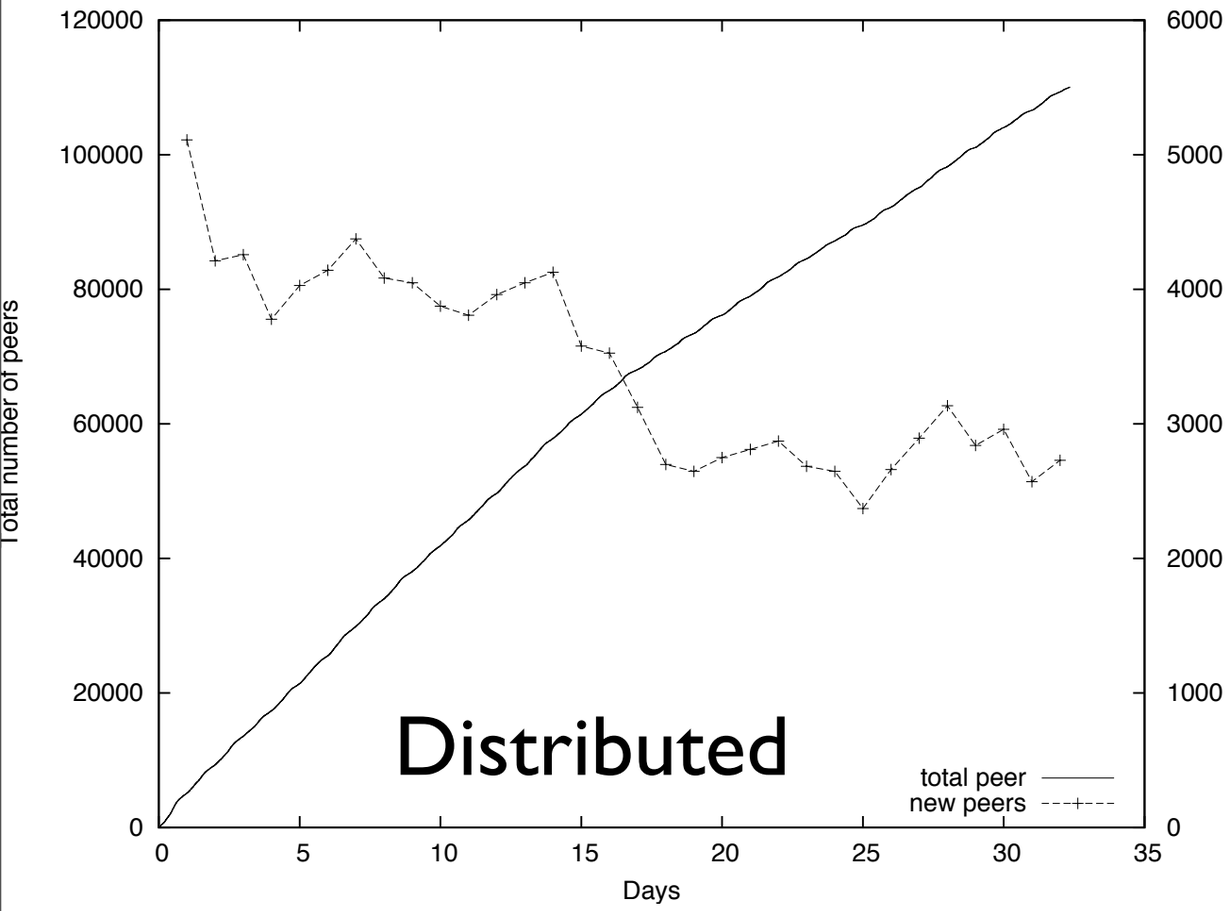


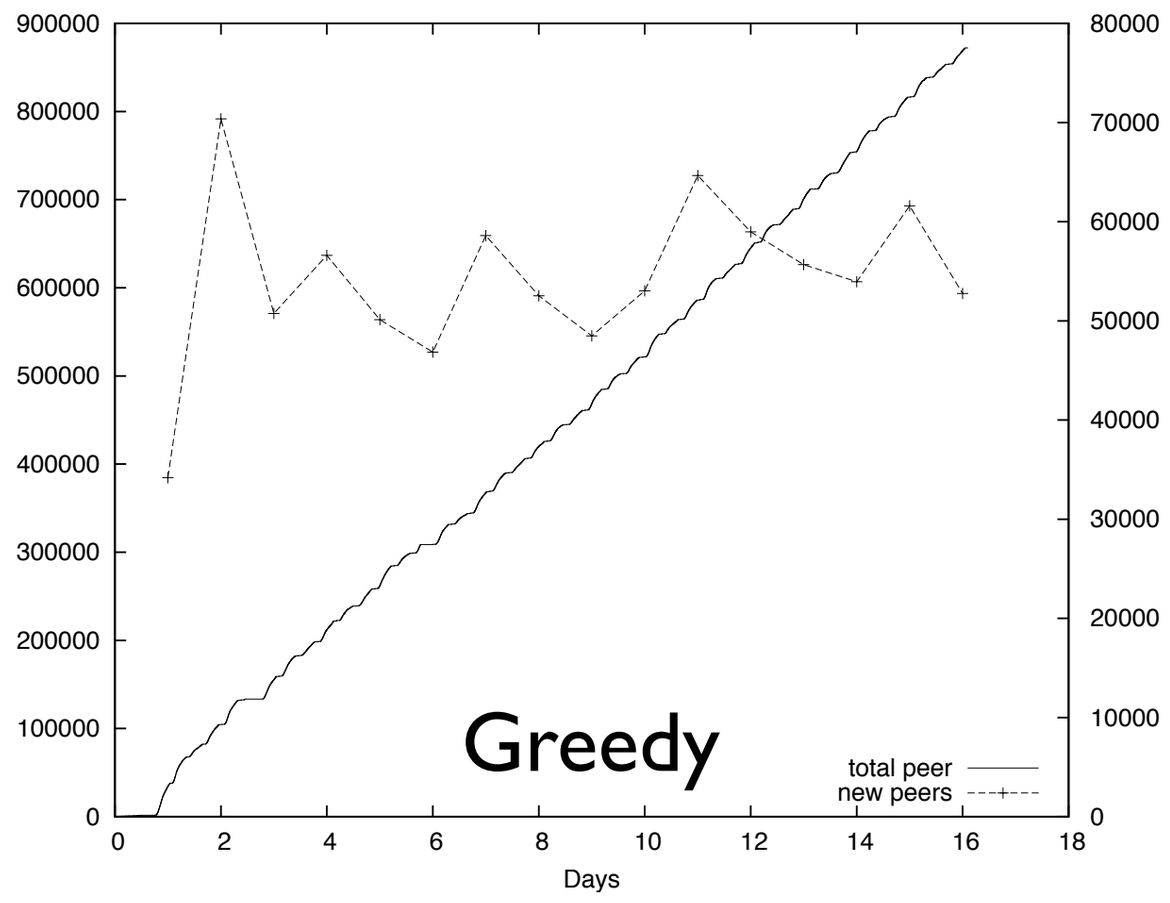
Methodology

- 24 PlanetLab nodes, running distributed honeypots:
  - 12 sending *no content*
  - 12 sending *random content*
- 1 greedy honeypot:
  - learn files during the first day
  - afterwards, announce these files

	distributed	greedy
Honeypots	24	1
Duration in days	32	15
Shared files	4	3 175
Distinct peers	110 049	871 445
Distinct files	28 007	267 047

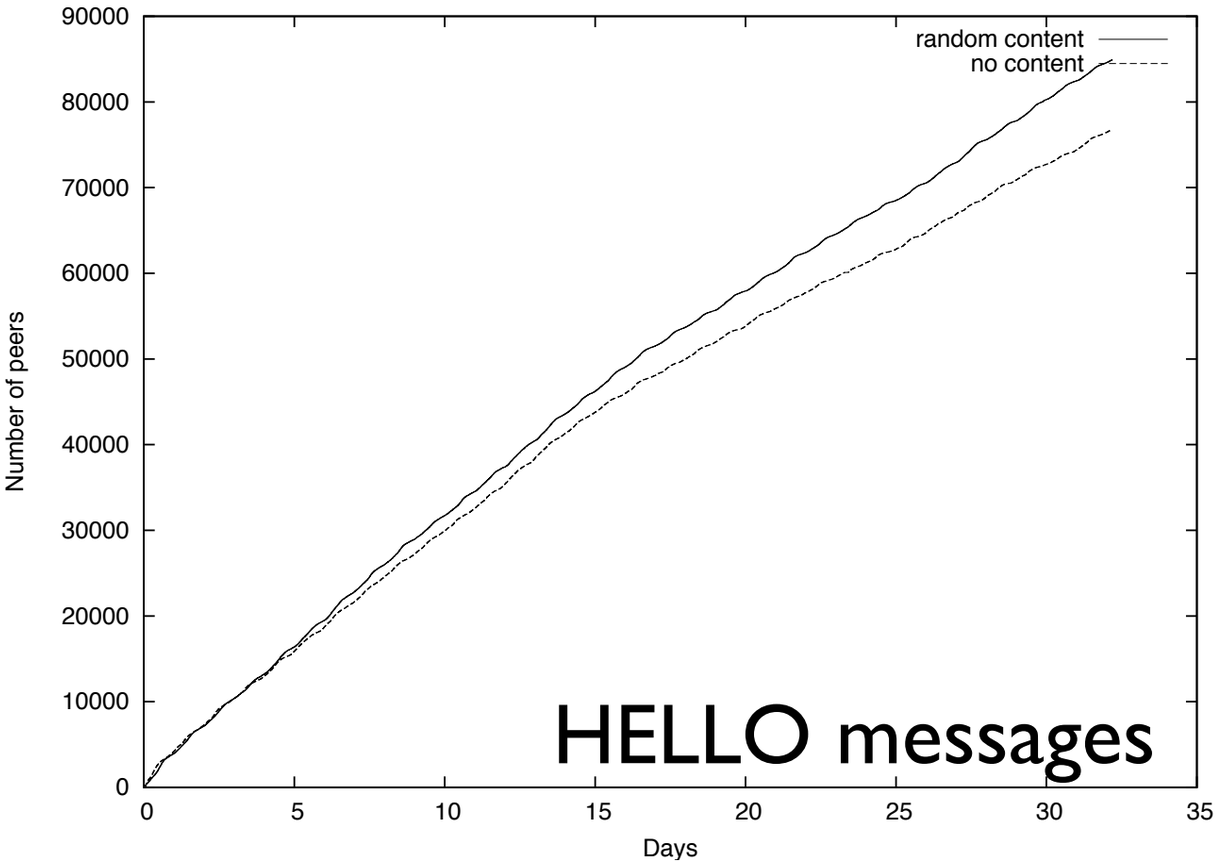
Parameters : distributed or greedy

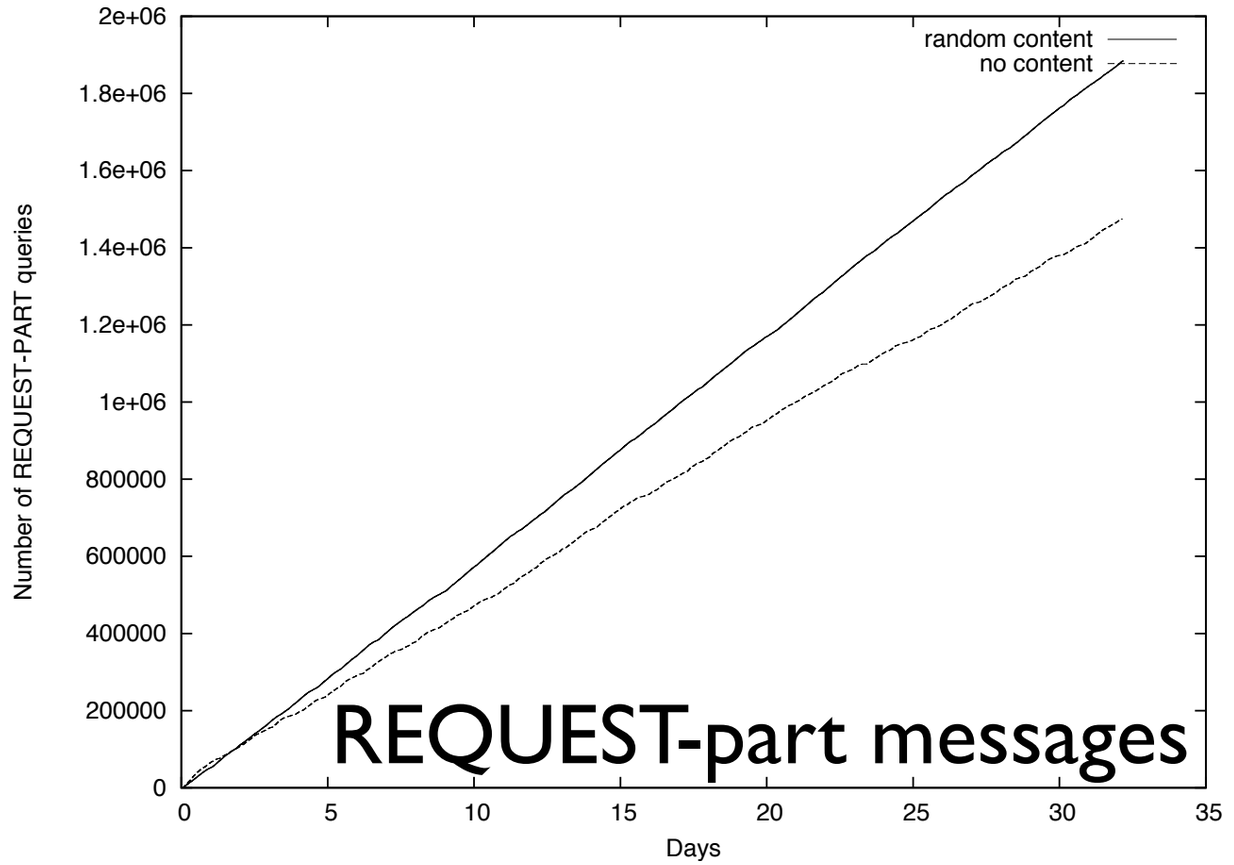




- long measurements are relevant
- effects of blacklisting and file popularity

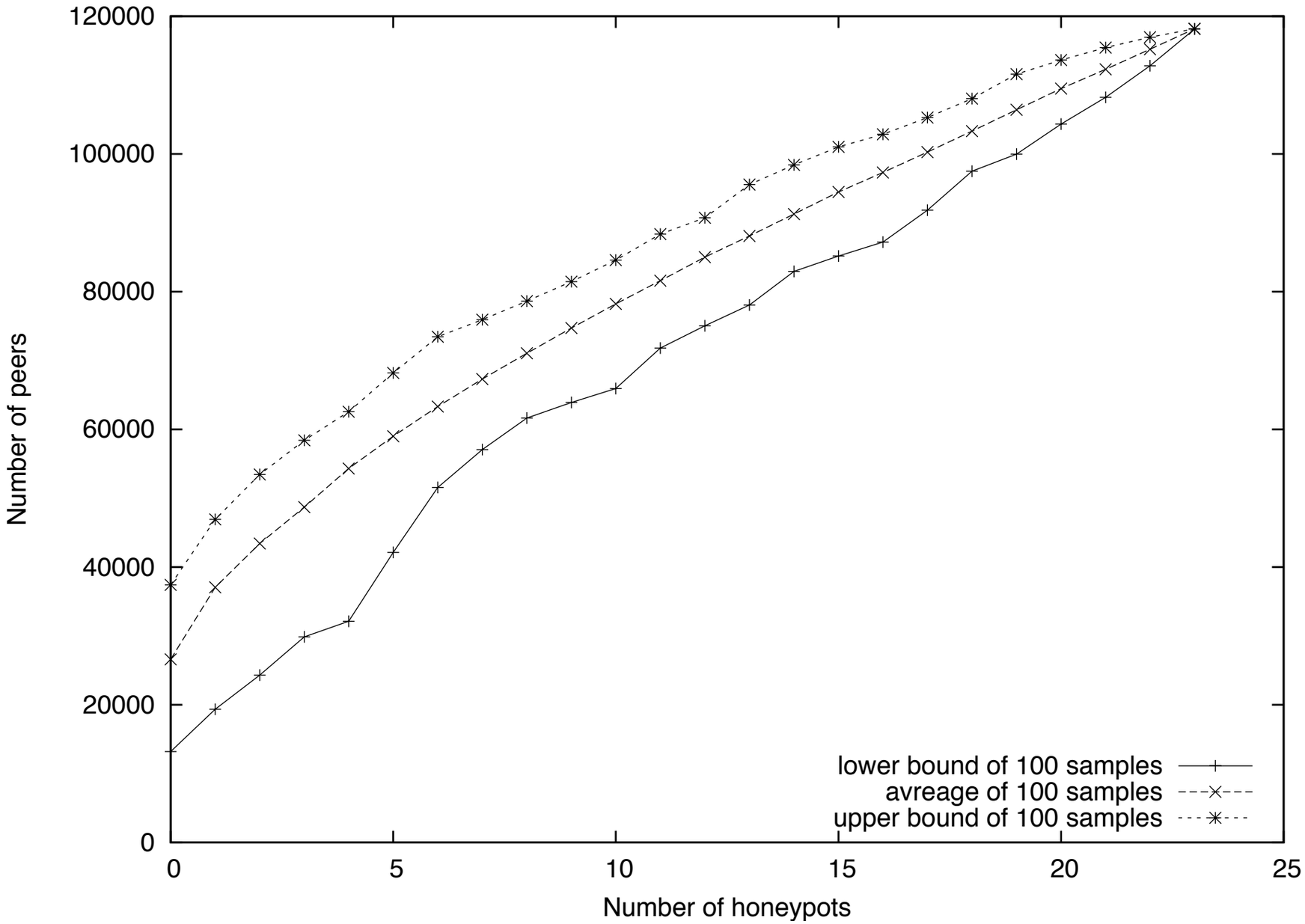
Parameters : no-content & random-content





- advantage of sending random content
- global and local blacklisting

Parameters : number of honeypots



- important benefit in using several honeypots

Conclusion

- several data sets available
  - IP topology
  - eDonkey measurement:
    - server side
    - client side
    - honeypot
- Ongoing works
  - understand topology dynamics
  - community of interests in eDonkey
  - anomaly detection in the IP topology
  - ...

Questions ?