Sensing the Internet through crowdsourcing

E. Gregori[1], L. Lenzini[2], V. Luconi[2], A. Vecchio[2]

(1) Istituto di Informatica e Telematica
Italian National Research Council (CNR)
Italy

(2) Dip. di Ingegneria dell'Informazione
University of Pisa
Italy

The graph of the Internet

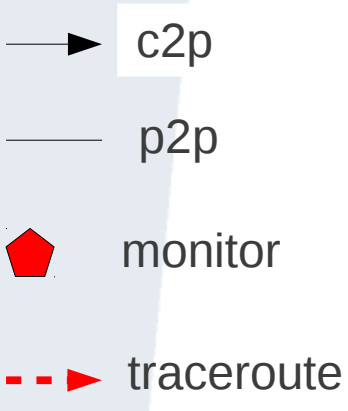
- Having a graph of the Internet is useful when designing new protocols and networks, or to understand its evolution.
- But... Internet Service Providers (ISPs) typically operate as commercial entities and are reluctant to publicly reveal their network structure and properties.
- Research focusing on methods for the discovery of the Internet topology gained momentum in the last years.
- Two approaches:
  - Passive methods: rely on Border Gateway Protocol (BGP) information and Internet registries
  - Active methods: based on traceroute and its evolutions

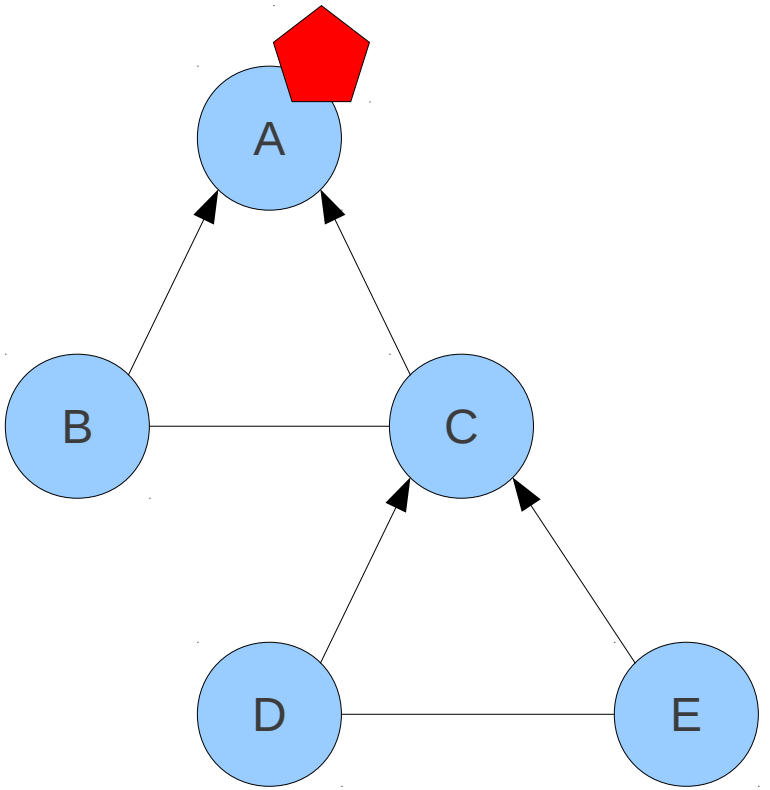
The graph of the Internet

- Passive methods
  - Pros: no traffic is generated
  - Cons: BGP paths do not cover the entire Internet (visibility constraints, route aggregation, hidden sub-optimal paths and policy filtering); information not always complete or up to date.
- Active methods:
  - Pros: selective analysis of “obscure” parts of the Internet
  - Cons: generate traffic
- We focused on active methods
- Graph at the autonomous system (AS) level

Motivation

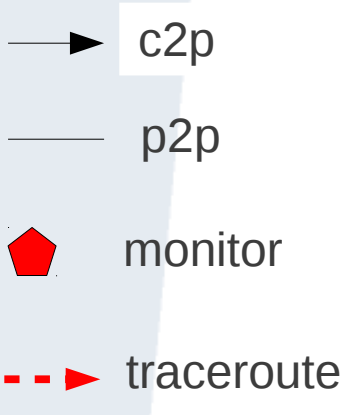
- Measuring the Internet from a set of fixed observation points (also called monitors) can still lead to unsatisfactory results: since such observation points are usually placed in proximity of the core of the Internet they are unable to provide detailed information about the fringes of the network.

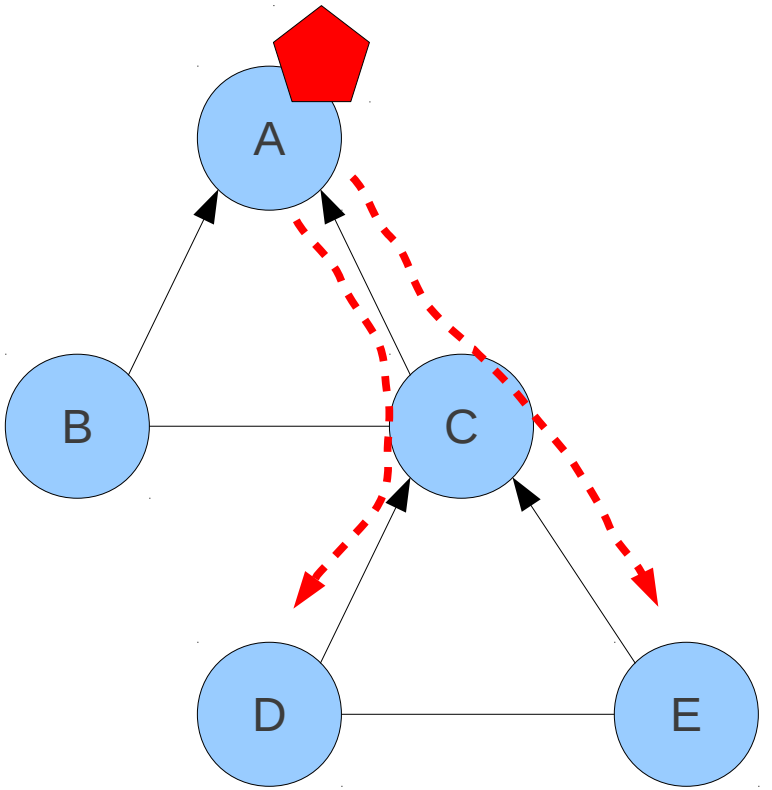




Motivation

- Measuring the Internet from a set of fixed observation points (also called monitors) can still lead to unsatisfactory results: since such observation points are usually placed in proximity of the core of the Internet they are unable to provide detailed information about the fringes of the network.

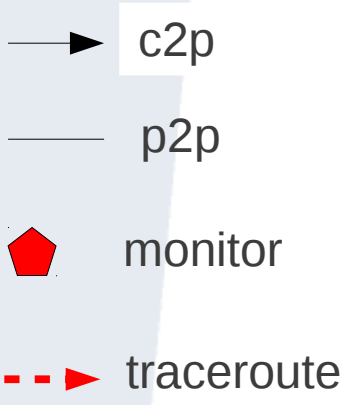


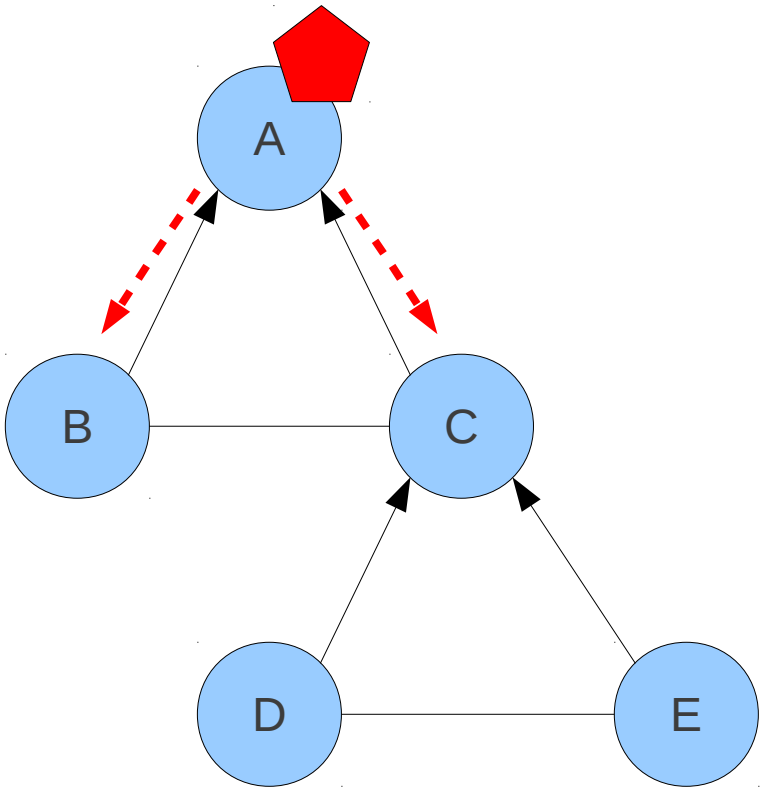


D-E cannot be detected

Motivation

- Measuring the Internet from a set of fixed observation points (also called monitors) can still lead to unsatisfactory results: since such observation points are usually placed in proximity of the core of the Internet they are unable to provide detailed information about the fringes of the network.



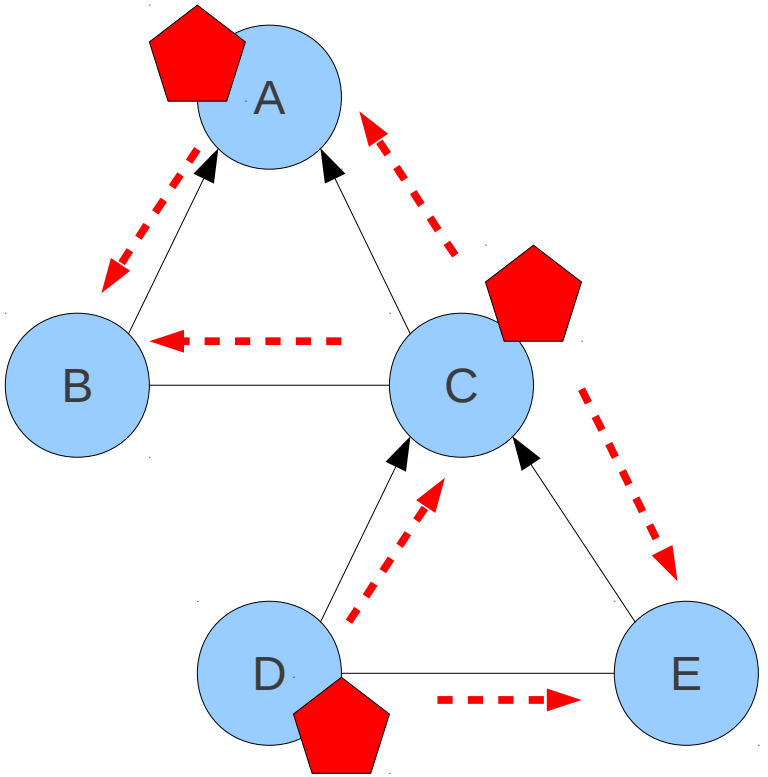


B-C cannot be detected

Motivation

- Links can be detected
  - with a large number of monitors
  - omnidirectional probes, limited radius

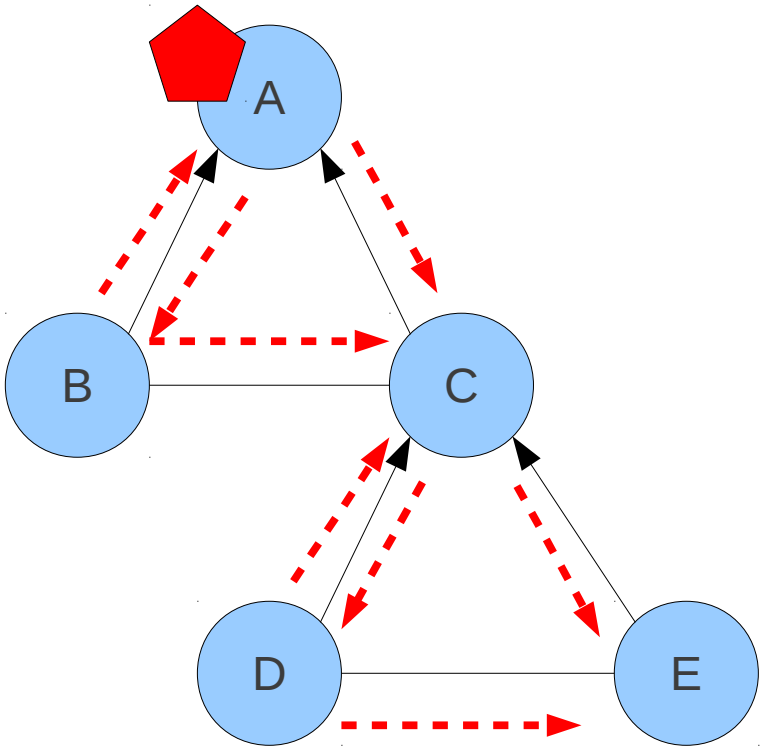
Crowdsourcing!



Motivation

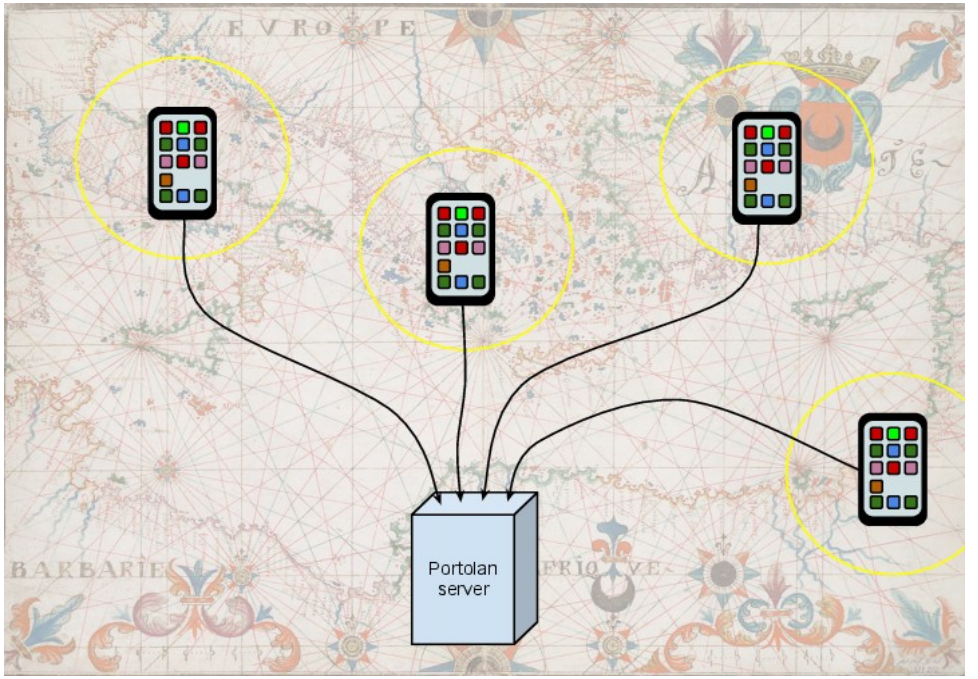
- Links can be detected
  - with a large number of monitors
  - omnidirectional probes, limited radius

Mobility!

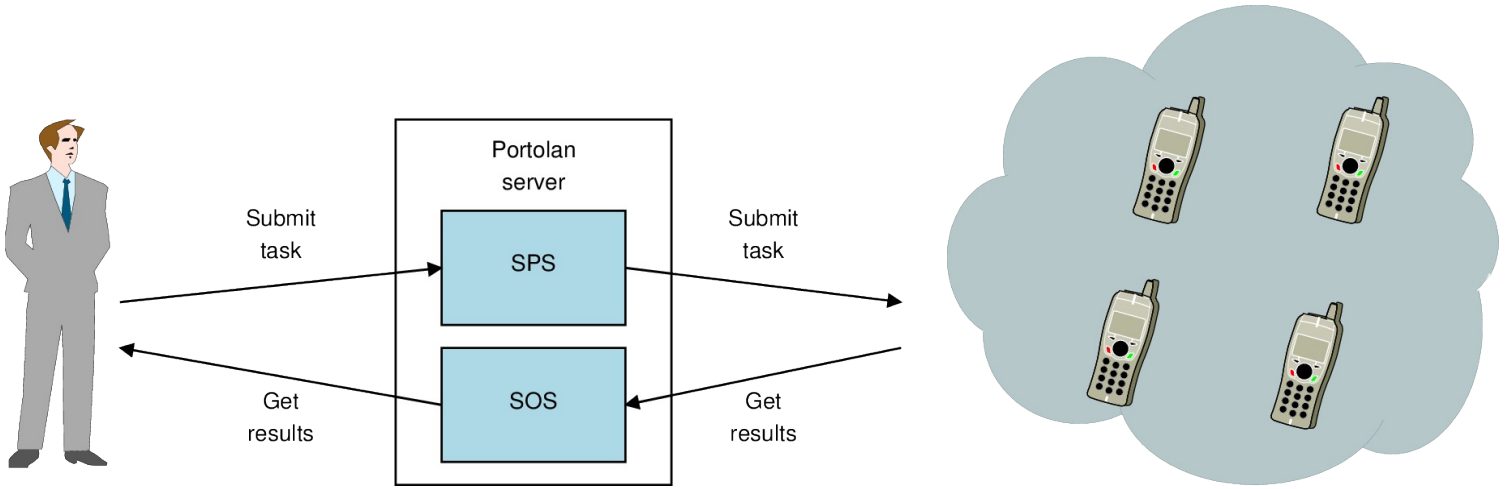


Portolan

- Portolan is a crowdsourcing-based system where smartphones play the role of mobile measuring elements.
- Each smartphone collects a number of local measures that are subsequently forwarded to a server, where they are assembled to generate a global map of the Internet.
- Mobility of nodes enables each single monitor to have different perspectives of the network, thus obtaining more detailed information.
- GPS enables geo-location of measures.

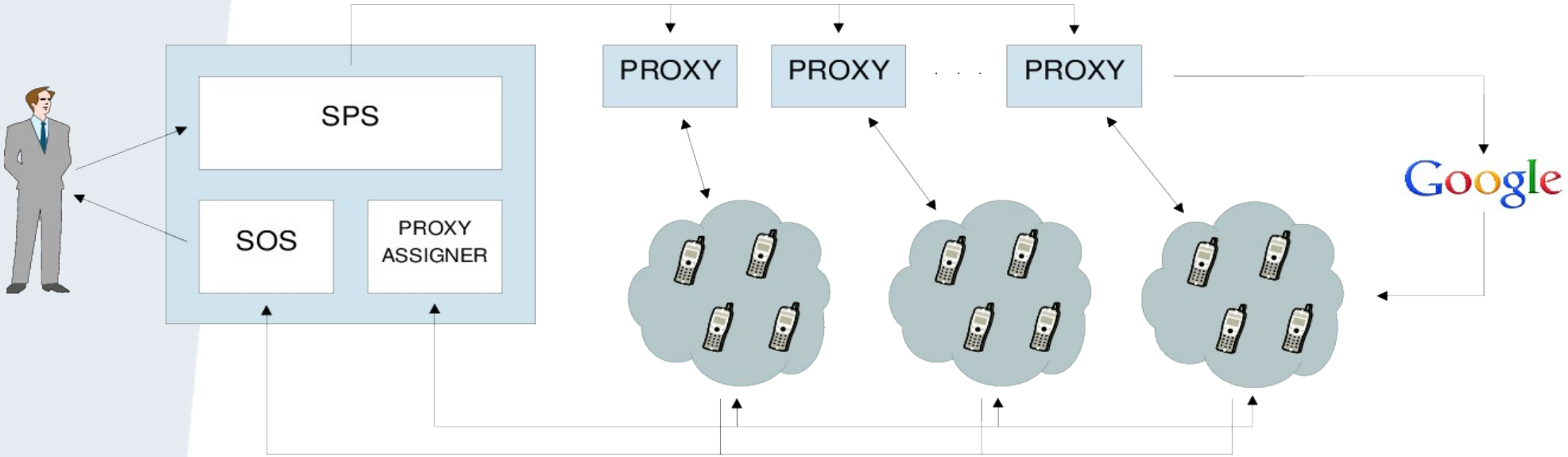


Portolan



- We used some standard specifications (Sensor Web Enablement, Open Geospatial Consortium).
  - Sensor Observation Service (SOS): repository of observations;
  - Sensor Planning Service (SPS): used for assigning tasks to sensors;
  - Observation&Measurements: for standard representation;
  - SensorML: to describe sensors' capabilities.
- Large tasks are divided into loosely coupled micro-tasks, to achieve parallelization if a large number of clients is available.

Portolan architecture

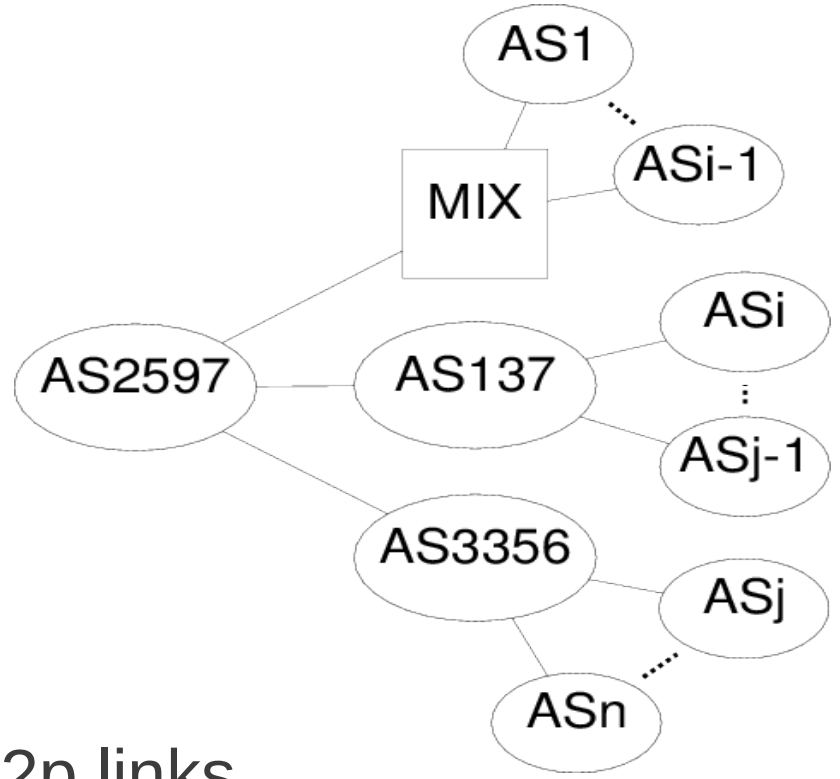


- Proxies are introduced for scalability.
- Smartphones poll proxies for micro-tasks
  - Tasks take into account dynamical properties of the client such as geo-position and network-position;
  - Clients are usually behind a NAT.
- Google's Cloud Messaging for urgent tasks and other out-of-band communication.

Sensing the Internet

- We carried out a measurement campaign to validate the system functionality, and to assess the soundness of an approach based on local measures.
- The experiment was aimed at discovering the public peering links between a specific AS (source) and a set of candidate peers (targets).
- An Internet Exchange Point (IXP) is a physical location that allows multiple ASes to exchange their traffic through its switching infrastructure.
- In many cases peering links are established at an IXP.
- Experiment:
  - 1 source AS (AS2597)
  - 1 IXP (Milan Internet eXchange, MIX)
  - 79 target ASes (all the ASes connected to MIX)

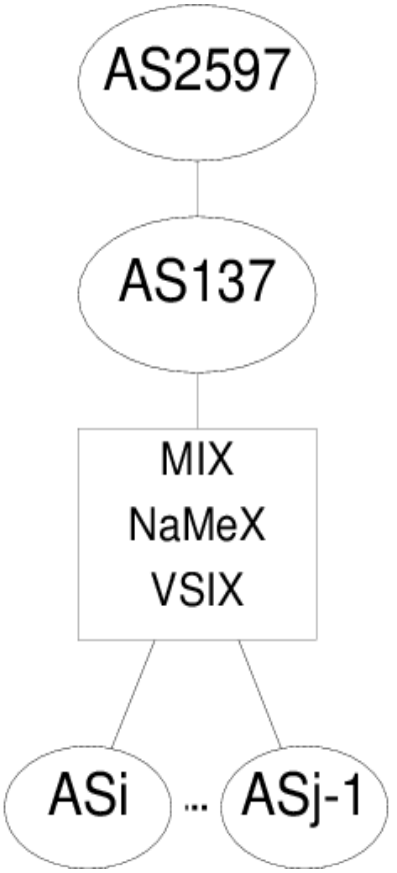
Sensing the Internet



- Portolan discovered 22 p2p links.
- Results have been compared with the set of links contained in three popular databases: CAIDA, Isolario, DIMES.
- 16 unknown links discovered (only 6 were known).

Sensing the Internet

- In many cases the source AS is connected to the target ASes via another AS (AS137).
- Thus we indirectely found the p2p links of AS137.
- Portolan discovered 33 links, only 9 were already known to CAIDA, Isolario, and DIMES.
- 24 unknown links discovered.

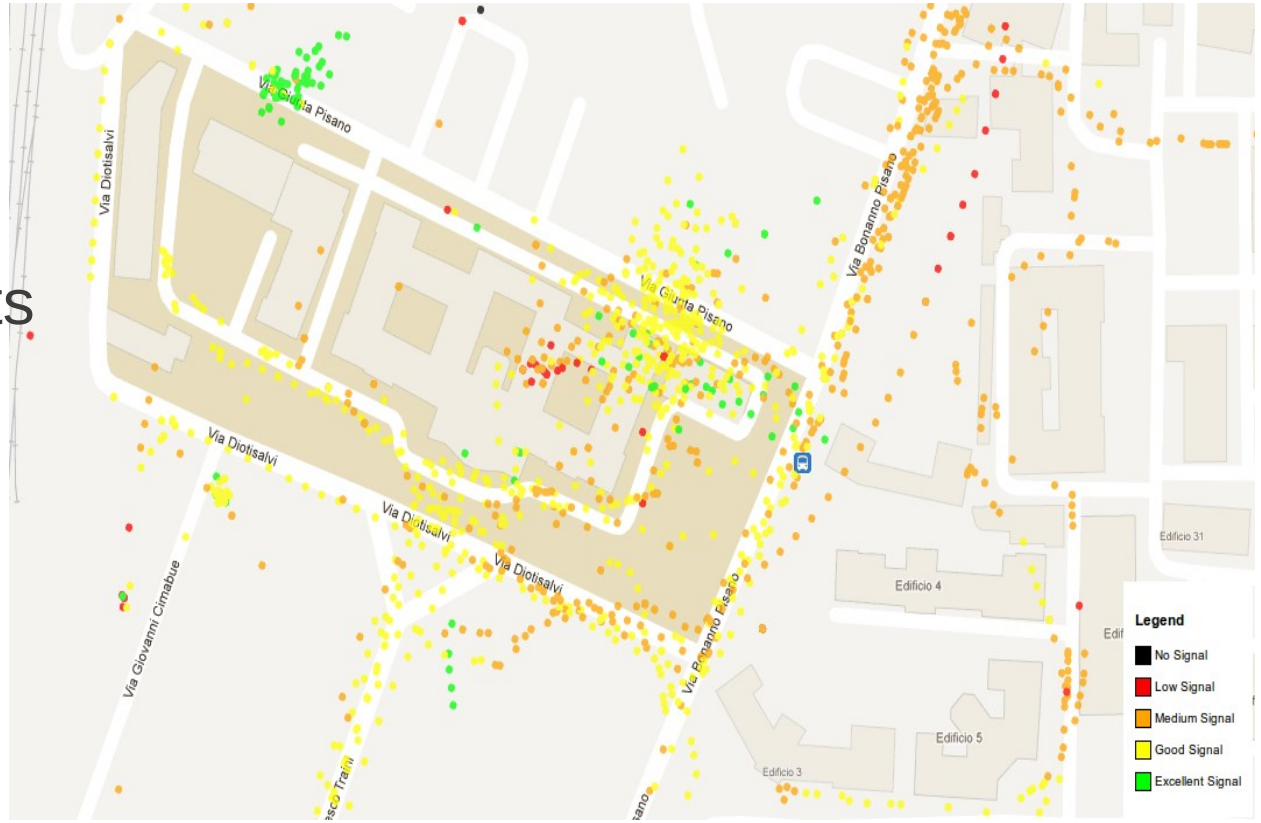


Execution cost

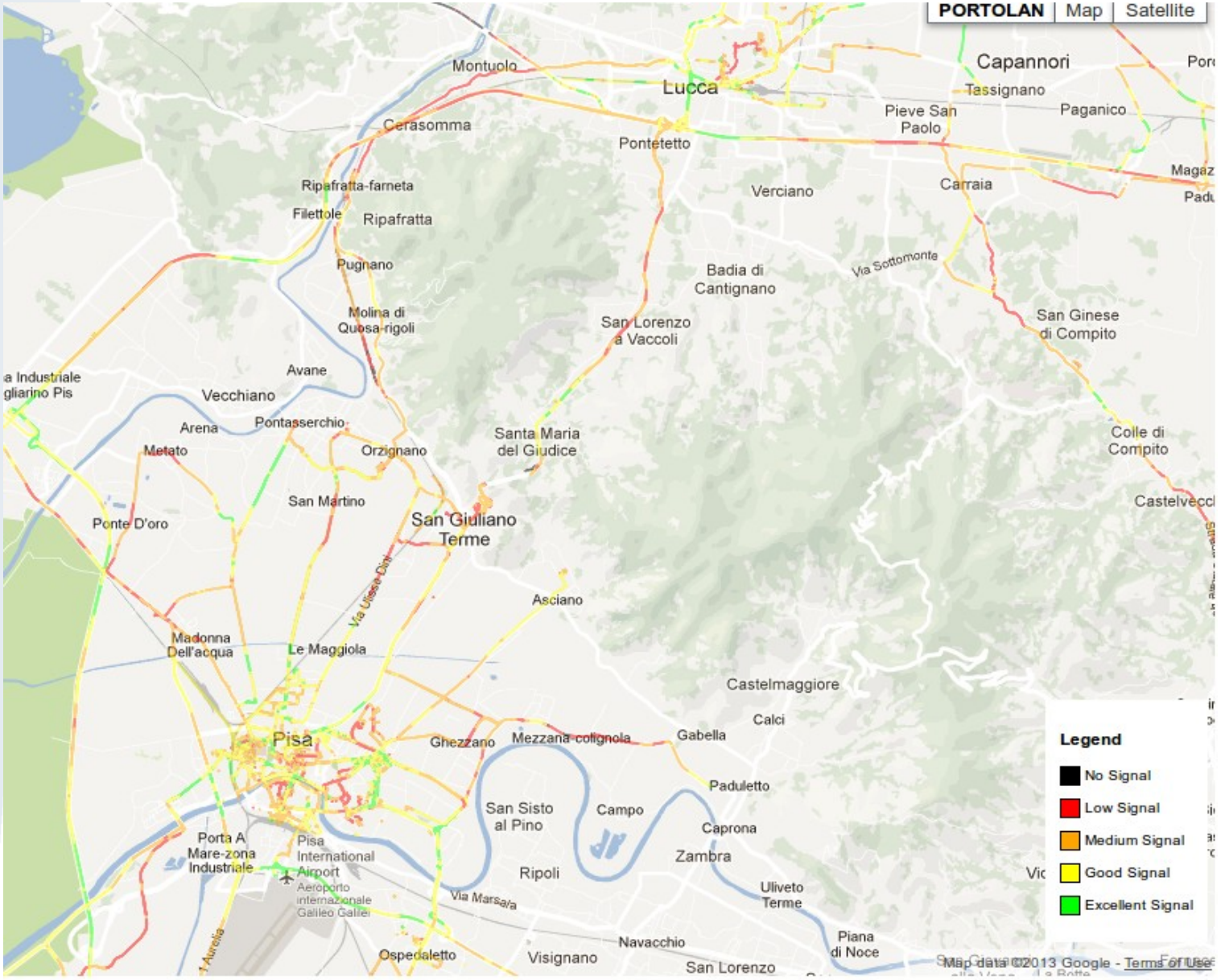
- List of target IP addresses
  - 1556 microtasks
  - Each microtask 100 IP addresses
- Duration
  - 1 traceroute: 5-10 seconds
  - 1 microtask: 8-16 minutes
  - Entire campaign: 16 days on a single device
- We optimized the execution by removing microtasks from the execution queue when they could not give additional information → 135 microtasks.
- 36 hours on a single device.
- Energy and network:
  - 1 microtask: approximately 1% of battery
  - 1 microtask: 340KB out, 440KB in

Sensing received signal strength

- Portolan has been designed as a general network measuring tool, based on crowdsourcing to solve large-scale problems.
- It is able to measure the received signal strength of mobile operators.
- Now:
  - 90 users
  - 450K points



Sensing received signal strength



Conclusion

- The amount of new links discovered by using even a single smartphone demonstrates the effectiveness of the “local” sensing approach.
- Crowdsourcing can be an effective strategy for tackling large-scale network monitoring.
- Problem:
  - Motivating users