Network Measurement Virtual Observatory: An Integrated Database Environment for Internet Measurements and Data Analysis

Tamás Sebők, Zsófia Kallus, Sándor Laki, Péter Mátray, József Stéger, Péter Hága, László Dobos, István Csabai, Gábor Vattay

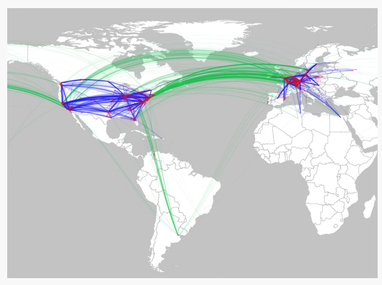
Eötvös Loránd University, Department of Physics of Complex Systems, Budapest, Hungary

Internet tomography techniques work by setting up servers with precisely synchronized clocks all around the world to measure ping delay times and to determine the structure and dynamics of the multiply connected computer network. The experiments require coordination of hundreds of computers, billions of measurements and software capable of gathering and analyzing measurement data. The high number of these micro-experiments makes Internet tomography a particularly interesting big data science.

We present the Network Measurement Virtual Observatory (nmVO), an integrated system built around a relational database system, custom web services and various open software developed primarily for other scientific fields. Its aim is to serve collaborating research groups sharing their measurement results through standardized archives. The tight integration of the data acquisition process with data analysis and the final data products makes it possible to perform historic and real-time observations of the Internet using simple database queries. A case study of the nmVO is our geographic localization service called Spotter.

X-raying the Internet

- Internet tomography
  - Build graph structure from traceroutes and ping delay data
  - Topology of the network
  - Bandwidth along edges



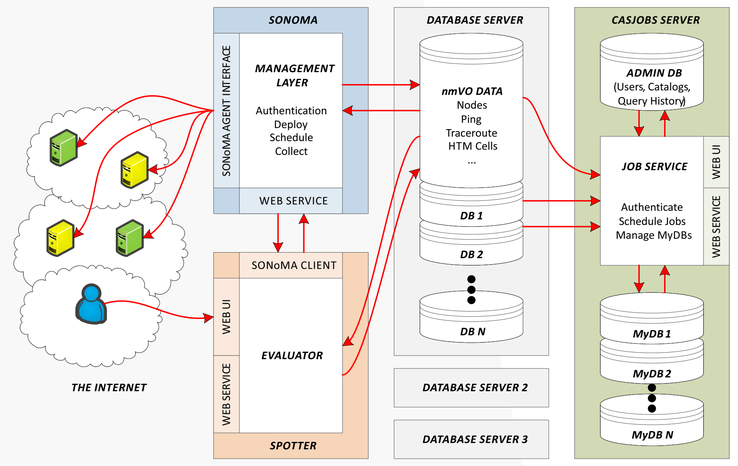
The Concept of nmVO

- Virtual Observatories: data warehouses for scientific big data
- nmVO: a dynamic data warehouse:
  - network measurements: millions of micro-measurements
  - conducting measurement requires large amount of input data
  - raw data collection directly into the database
  - data reduction is done inside the database
  - new data gets published instantly after reduction
  - re-analyze measurements later, if necessary
  - use archival data to analyze long-term dynamics of Internet
- nmVO is built on well-known scientific database software
  - CasJobs, originally built for the astronomical SkyServer
  - primary user interface: SQL-based batch job system
  - Hierarchical Triangular Mesh used for geographical data

nmVO in practice

- Multi-TB SQL database
  - collect and store raw data
- Reduction and analysis tasks are written in SQL
  - run data reduction and analysis inside DB
  - store results in the same DB
  - users can run the processing on the server where data reside
- Integrated with data collection services
  - initiate new Internet measurements right from SQL!
  - great potential, e.g. collect missing data on-the-fly
- Access to remote datasets helps incorporate data from other Internet measurement projects

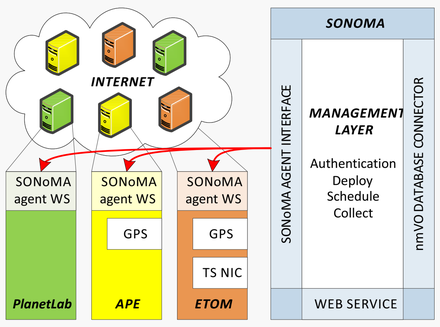
nmVO infrastructure



▲ Figure: The nmVO architecture. Services are centered around the database servers (grey) storing raw and processed measurement data. The SONoMA management layer (blue) is responsible for conducting measurements by controlling measurement software on remote server nodes around the world. The Spotter web service (orange) is built on top of SONoMA and nmVO. When a users initiates an IP localization, Spotter instructs SONoMA to collect necessary ping delay data and analizes the raw measurements. Both raw data and analyzed measurements are made instantly public in nmVO. Users can interact with the system via CasJobs (green), a SQL query interface and batch execution layer. Data processing tasks are formulated in SQL and executed by the database servers. SONoMA functions are directly available as SQL stored procedures which allow on-the-fly experiments.

SONoMA

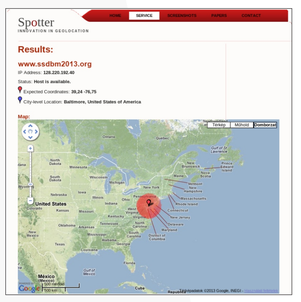
- Use Etomic and PlanetLab servers around the world
- SOAP-based web services to collect Internet measurement data
  - hides complexities of underlying infrastructure
  - easy to implement interfaces for network experiment designers
- Wide set of atomic measurements
  - ping, tracerount, chirp, train, DNS lookup etc.



◀ Figure: The SONoMA architecture. SONoMA agents run on network measurement servers world wide. SONoMA currently supports the PlanetLab and Etomic infrastructures. Etomic machines are also equipped with high precision GPS clocks for accurate timing of ping delays. The SONoMA management layer (grey) is responsible for conducting measurements and store raw and processed data directly in nmVO. SONoMA is accessible via its SOAP web service interface.

Spotter

- Geolocation service built on nmVO and SONoMA
- Determines location of host from its IP address
  - landmarks: servers with known coordinates
  - ping delay times measured from landmarks
  - location determined from ping delay times
- Raw data and measuments published in nmVO



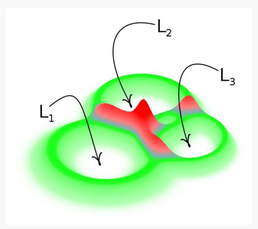


Figure: Results page of Spotter localizing ssdbm2013.org. ▲ The accuracy of geolocation depends significantly on the available landmarks (servers with known coordinates).

◀ Figure: Spotter uses a complex statistical model based on geographic coordinates of landmarks and the measured ping delay times is used to calculate the PDF of the approximate coordinates of the localized IP address.

Please visit http://nm.vo.elte.hu

The authors thank the partial support of the EU FP7 OpenLab project (Grant No.287581), the EU FIRE NOVI project (Grant No.257867), the European Union and the European Social Fund through project FuturICT.hu (grant no.: TAMOP-4.2.2.C-11/1/KONV-2012-0013), the OTKA 7779 and 103244, and the NAP 2005/KCKHA005 grants. EITKIC_12-1-2012-0001 project was partially supported by the Hungarian Government, managed by the National Development Agency, and financed by the Research and Technology Innovation Fund and the MAKOG Foundation.

References:

- S. Laki, P. Mátray, P. Hága, T. Sebők, I. Csabai, and G. Vattay. Spotter: A model based active geolocation service. INFOCOM, pages 3173-3181. IEEE, 2011.
- P. Mátray, I. Csabai, P. Hága, J. Stéger, L. Dobos, and G. Vattay. Building a prototype for network measurement virtual observatory. MineNet, pages 23-28. ACM, 2007.
- B. Hullár, S, Laki, J. Stéger, I. Csabai and G. Vattay. SONoMA: A Service Oriented Network Measurement Architecture. http://arxiv.org/abs/1006.1222, 2010.