Measurement Based Geolocation in the ETOMIC Infrastructure

Sándor Laki*, Péter Mátray, Péter Hága, István Csabai and Gábor Vattay

Communication Networks Laboratory
Eötvös Loránd University
Budapest, Hungary

*E-mail: laki@complex.elte.hu
## Agenda

- **Introduction**
- Path-latency Model
- Velocity of Signal Propagation in Network
- Geographic Constraints
- Data Collection
- Performance Analysis
- A „Case Study”
- Summary
Motivation

• Location information can be useful to both private and corporate users
  – Targeted advertising on the web
  – Restricted content delivery
  – Location-based security check
  – Web statistics

• Scientific applications
  – Measurement visualization
  – Network diagnostics
Geolocation in General

• Passive geolocation
  – Extracting location information from domain names
  – DNS and WhoIS databases
  – Commercial databases
    • MaxMind, IPLigence, Hexasoft
  – Large and geographically dispersed IP blocks can be allocated to a single entity

• Active geolocation
  – Active probing
  – Measurement nodes with known locations (landmarks)
  – GPS-like multilateration (CBG)
Measurement Based Geolocation

• Active measurements
  – **Network Delays**
    • *Delays can be transformed to geographic distance*
      – Round Trip Time (ping)
      – One-way delay
    • Effects of over and underestimation
  – **Topology**
    • Network-path discovery
      – Traceroute with fixed port pairs
    • Interface clustering
      – Mercator, etc.
Presentation Outline

• Introduction

• **Path-latency Model**

• Velocity of Signal Propagation in Network

• Geographic Constraints

• Data Collection

• Performance Analysis

• Summary
Why do we need a latency model?

• The basis of active methods is to transform delays to geographic distance

• The model aims to decompose the overall packet delay to linkwise components

• Approximating propagation delays along a network path leads to more precise distance estimations...
Modelling Packet Delays

• A packet delay \(d\) can be divided into...
  - Queuing delay \(D_q\)
  - Processing delay \(D_{pc}\)
  - Transmission delay \(D_{tr}\)
  - Propagation delay \(D_{pg}\)

• A given path: \(n_0 \rightarrow n_1 \rightarrow n_2 \rightarrow \ldots \rightarrow n_H\)

• The overall packet delay for a network path \((s = n_0 \text{ and } d = n_H)\):

\[
d(s, d) = \sum_{i=1}^{H} (D_{pc}^i + D_q^i + D_{tr}^i + D_{pg}(n_{i-1}, n_i))
\]
How to Estimate Propagation Delays

• Assumptions in the model
  – **No queuing** \((D_q = 0)\)
  – The per-hop processing and transmission delays can be approximated by a global constant \((d_h)\)
    • \(d_h = D_{pc} + D_{tr}\)
  – Based on the literature and our observations:
    » \(d_h = 100\,\mu s\)

---

**The one-way propagation delay** along a given path:

\[
D_{pg}(s, d) = d(s, d) - H \cdot d_h
\]
An Extra Cost - ICMP Generation Time

- In case of ICMP based RTT measurements **an extra delay appears** at the target node
  - ICMP Echo Reply Generation Time \((D_g)\)
- The overall Round Trip Delay:
  
  \[ d(s, d, s) = d_{fw} + D_g + d_{bw} \]

Is it possible to measure this \(D_g\) delay component?

Yes, there’s a way…
ICMP Generation Times

Less than 1%

$D_g = 300 \mu s$ to avoid distance underestimation
Presentation outline

• Introduction
• Path-latency Model
• **Velocity of Signal Propagation in Network**
• Geographic Constraints
• Data Collection
• Performance Analysis
• Summary
Distance Approximation

• An upper approximation of geographical distance from source $s$ to destination $d$:

$$s(s, d) \leq s^* = c \cdot r \cdot \sum_{i=1}^{H} D_{pg}(n_{i-1}, n_i) = c \cdot r \cdot D_{pg}(s, d)$$

• where $r$ is the velocity of signal propagation in network [in $c$ units]
The maximum velocity we measured in network was 0.47!

The velocity of signal propagation in a copper cable is ~0.66-0.7!

And the average value was 0.27

The maximum value was used to avoid distance underestimation
Presentation outline

- Introduction
- Path-latency Model
- Velocity of Signal Propagation in Network
- **Geographic Constraints**
- Data Collection
- Performance Analysis
- Summary
Solving Geolocation

• Defining geographic constraints
• We are looking for a location set where all the constrains come true
• Defining the overall tension in the system
  – A cost function
• By minimizing this function the problem can be solved
  » Non-convex optimization problem
  » Well known solutions in sensor networks
Round-Trip Time Constraint

• Using path-latency model
  – Round-trip propagation delay from a landmark
  • Upper approximation of one-way propagation delay
One-way Delay Constraint

• A novel constraint for a network path between two landmarks
  – Limiting the geographic length of a given network path

• High-precision OWD measurements

\[ \sum_{i=1}^{H-1} s(n_i, n_{i+1}) \leq s^* = c \cdot r \cdot (d(s, d) - H \cdot d_h) \]
Presentation outline

- Introduction
- Path-latency Model
- Velocity of Signal Propagation in Network
- Geographic Constraints
- Data Collection
- Performance Analysis
- Summary
An Award-winning Testbed

- **European Traffic Observatory Measurement InfrastruCture (etomic)** was created in 2004-05 within the **Evergrow** Integrated Project.
  - **Open and public** testbed for researchers experimenting the Internet
  - **18 GPS synchronized active probing nodes**
  - Equipment with Endace DAG cards
    - **High-precision end-to-end measurements**
  - **Scheduled experiments**
  - **NO SLICES**
    - » You own the resources during the experimentation

www.etomic.org
Data Collection and Evaluation

STEP 1 – Traceroute topology

- **Goal:** Localizing inter-ETOMIC routers
Data Collection and Evaluation
STEP 2 – Interface Clustering

• Goal: **Localizing inter-ETOMIC routers**
Data Collection and Evaluation
STEP 2 – Interface Clustering

- **Goal:** Localizing inter-ETOMIC routers
Data Collection and Evaluation
STEP 3 – RTT Measurements

- **Goal:** Localizing inter-ETOMIC routers
Data Collection and Evaluation
STEP 4 – OWD Measurement

• **Goal:** Localizing inter-ETOMIC routers
OneLab2 - The Next Generation of ETOMIC

OneLab2 sites:
- ≥2 PlanetLab nodes
  - New features
    - Dummybox
    - WiFi access
    - UMTS, WiMax…
- 1 ETOMIC2-COMO integrated node
  - High precision e2e measurements
  - ARGOS meas. card
  - available via ETOMIC’s CMS
- 1 APE box
  - A lightweight measurement tool
Presentation outline

• Introduction
• Path-latency Model
• Velocity of Signal Propagation in Network
• Geographic Constraints
• Data Collection
• Performance Analysis
• Summary
Performance Analysis

- **Geo-Rh**
  - Mean error: 251 km
  - Max. error: 699 km
  - StdDev: 205 km

- **Geo-RhOL**
  - Mean error: 149 km
  - Max. error: 312 km
  - StdDev: 104 km

- **Geo-R**
  - Mean error: 305 km
  - Max. error: 878 km
  - StdDev: 236 km
A case study –
Where are your YouTube videos?
A case study – Where are your YouTube videos?

• Where are YouTube’s content delivery servers?
  – The IP range: **74.125.0.0/16**
  – 8127 globally accessible IP addresses
  – 8127 nodes to be localized

• Using PlanetLab nodes as landmarks
A case study –
Where are „YouTube” servers?

- Tokyo
- Taipei
- Frankfurt
- London
- Baltimore
- Seattle
- San Francisco
- New York
- Atlanta
- Toronto
- Chicago
- Hong Kong
- Singapore
- Rotterdam, Amsterdam
- Tokyo
- New York
- Atlanta
- Baltimore
- Hong Kong
- Singapore
- Rotterdam, Amsterdam

- N=1
- 2≤N<10
- 10≤N
A case study –
Where are „YouTube” servers?

- N=1
- 2≤N<10
- 10≤N
Presentation outline

- Introduction
- Path-latency Model
- Velocity of Signal Propagation in Network
- Geographic Constraints
- Data Collection
- Performance Analysis
- **Summary**
Summary

• Estimating propagation delays more precisely
  – Separation of propagation and per-hop delays in the overall packet latency
• Velocity of signal propagation in network is much smaller than we assumed before due to curvatures
• The novel one-way delay constraints improve the accuracy of router geolocation significantly
• A „real life” study

• Plans for future extensions
  • The method can be combined with passive techniques
  • Improving latency model
  • Real-time Geolocation Service using Web Services (details in the poster session)
Thank you for your attention!

Contact: laki@complex.elte.hu

More details:
S. Laki, P. Mátray, P. Hága, P. Csabai and G. Vattay:
A Detailed Path-latency Model for Router Geolocation

http://cnl.elte.hu