The Large-scale Geography of Internet Round Trip Times

Raúl Landa, Richard G. Clegg, Joao T. Araújo, Eleni Mykoniati, David Griffin, Miguel Rio

Department of Electronic and Electrical Engineering
University College London

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement 248565.
Motivation

• We are interested in estimating RTTs
  – In a lightweight, flexible manner
  – Between “arbitrary” points in the Internet
  – Without the deployment of software probes

• Possible research benefits:
  – Clustering of network prefixes by RTT between them
  – Improved simulation of overlays and their underlying ISPs
  – Statistical prediction of RTT values
Basic Measurement Technique

• We reimplement **TurboKing**
  – Estimate RTT by inserting measurement point in the middle of recursive DNS queries

• Two phases:
  – DNS Server Collection
  – RTT Measurement Collection

TurboKing Measurement

- DNS Client
- DNS Server
- TurboKing Measurement Process
- Nameserver 1
- Nameserver 2
TurboKing Measurement: First Step

- DNS Client
- DNS Server
- Nameserver 1
- Nameserver 2

Estimating RTT to Nameserver 1

Query: 4.f9f9b8fd.measurement.net?
TurboKing Measurement: Second Step

Estimating RTT to Nameserver 2 via Nameserver 1

DNS Error

Nameserver 1

Nameserver 2

Nameserver 2 is authoritative for 4.f9f9b8fd.measurement.net.

DNS TTL = 0

4.f9f9b8fd.measurement.net?
Our TurboKing Implementation

- Written in **Java 5 SE**

- **Server Discovery** Thread Pool
  - ~75 threads
  - Throughput of ~5 new servers per second with 4 cores

- **Measurement** Thread Pool
  - ~60 threads
  - Throughput of ~10 measurements per second with 1 core
DNS Server Collection

- Random IP Addresses
- IP Address Lists
- BitTorrent Announce Messages
- Port 53 Probing
- Reverse DNS
- Forward DNS
- SOA or NS Records
DNS Server Collection

- Random IP Addresses
- Port 53 Probing
- IP Address Lists
- BlueTack Blocklists
- Nameserver Lists
- iPlane Host Lists
- Top 200 Torrents from e.g. The Pirate Bay, isoHunt, etc.
- Reverse DNS
- Forward DNS
- SOA or NS Records
DNS Server Collection

Word Lists

Search Query Creation

Google
Yahoo
Bing

Server Hostnames

Forward DNS

SOA or NS Records

Hostname Lists
DNS Server Collection

Word Lists

Search Query Creation

Google
Yahoo
Bing

Server Hostnames

Forward DNS

SOA or NS Records

Alexa Top Hosts

Queries with 2 to 4 terms

Word lists for password cracking:
41 word lists in different languages and pertaining to different areas of knowledge
DNS Server Collection

SOA or NS Records
→
Recursive traversal of DNS tree
→
SOA or NS Records
DNS Server Set

Total

- ~350k DNS Servers
- ~22k Autonomous Systems
- ~220 Countries

Measurement Set (Recursive, Non-Forwarding)

- ~54k Servers
- ~5.5k AS
- ~3.4k Cities
- ~189 Countries
- ~99.6% Internet users

- Of ~216k BGP routing prefixes, ~21k include at least one measurement server
<table>
<thead>
<tr>
<th>Subcontinental Zone</th>
<th>Servers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa</td>
<td>519</td>
</tr>
<tr>
<td>Central Asia and the Middle East</td>
<td>1,490</td>
</tr>
<tr>
<td>Asia Pacific and China</td>
<td>7,730</td>
</tr>
<tr>
<td>Indian Subcontinent</td>
<td>449</td>
</tr>
<tr>
<td>North America (North)</td>
<td>21,276</td>
</tr>
<tr>
<td>North America (South) and the Caribbean</td>
<td>526</td>
</tr>
<tr>
<td>Oceania</td>
<td>1,116</td>
</tr>
<tr>
<td>South America (West)</td>
<td>270</td>
</tr>
<tr>
<td>South America (East)</td>
<td>1,333</td>
</tr>
<tr>
<td>Eastern Europe</td>
<td>6,798</td>
</tr>
<tr>
<td>Western Europe</td>
<td>12,953</td>
</tr>
</tbody>
</table>
Geolocation Errors

• Mismatch between latitude/longitude and country
  – Created a spatial index of the \( \approx 123k \) lat./long. for cities with more than \( 1k \) inhabitants in the GeoNames database
  – Servers were resolved to countries from lat./long. using the spatial index

• Since we are focusing in large-scale geography, small geolocation errors are unimportant
RTT Measurement Collection

• In our TurboKing implementation
  – Each RTT estimation consists of 10 individual measurements performed at 10 sec intervals
  – The median is taken as an estimation of the true RTT

• Current dataset:
  – ~200M individual RTT measurements
  – ~19M full RTT estimations
    • ~20M, with ~7% affected by TTL issues
  – ~50GB in size
Measurements as Random Variates

- We model our data as a random variable
- We approximate the density of this variable with a multidimensional histogram

<table>
<thead>
<tr>
<th>Variable</th>
<th>Interpretation</th>
<th>Bins used</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_p$</td>
<td>Pairs of /8 prefixes</td>
<td>14,189</td>
</tr>
<tr>
<td>$X_a$</td>
<td>Pairs of Top AS numbers</td>
<td>215,392</td>
</tr>
<tr>
<td>$X_z$</td>
<td>Pairs of subcontinental zones</td>
<td>66</td>
</tr>
<tr>
<td>$X_c$</td>
<td>Pairs of countries</td>
<td>2,648</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_l$</td>
<td>Common prefix length</td>
</tr>
<tr>
<td>$X_d$</td>
<td>Great circle distance</td>
</tr>
<tr>
<td>$X_t$</td>
<td>Round trip time (RTT)</td>
</tr>
</tbody>
</table>
Toolbox: Least Squares Median Line

- How well does a function $f(x_d)$ approximate the marginal distribution median for each $x_d$?

$$E_{\Phi}^2(f(x_d)) = \int_0^\infty \left( \int_0^\infty \Phi_{(d,t)}(x_d, x_t) \, dx_t - \int_{f(x_d)}^\infty \Phi_{(d,t)}(x_d, x_t) \, dx_t \right)^2 \, dx_d$$

- We are interested in the special case where $f(x_d)$ is linear. We formulate the following optimisation problem.

$$\min_{x_t = \alpha x_d + \beta} E_{\Phi}^2(\alpha x_d + \beta)$$
Toolbox: Least Squares Median Line

• We require a goodness of fit measure. We simply take the $R^2$ statistic and reformulate it. Let $\hat{m}$ be the median RTT for a given $x_d$, then we have that:

$$R^2_{\Phi} = 1 - \frac{E^2_{\Phi}(\alpha x_d + \beta)}{E^2_{\Phi}({\hat{m}})}$$

• Similarly to $R^2$, $0 < R^2_{\Phi} < 1$. Moreover,
  - $R^2_{\Phi} = 0$: The linear fit accounts for essentially no data variability
  - $R^2_{\Phi} = 1$: The linear fit perfectly explains all median RTT variability for each $x_d$. 
The conditional entropy of a random variable $X_k$ given a set of variables $S(i, j, \ldots) = \{X_k, X_j, \ldots\}$ quantifies the remaining entropy (uncertainty) in $X_k$ given that the values of the variables in $S(i, j, \ldots)$ are known.

\[ H(X_k | (i, j, \ldots)) = \sum_{x_k, x_i, x_j, \ldots \in S} \Phi_{S(k, i, j, \ldots)} \log \frac{\Phi_{S(i, j, \ldots)}}{\Phi_{S(k, i, j, \ldots)}} \]

This can be shown to be equivalent to

\[ H(X_k | S(i, j, \ldots)) = H(S(k, i, j, \ldots)) - H(S(i, j, \ldots)) \]
We can use the conditional entropy to estimate the quality of estimators. For *quantitative* variables:

$$\text{MMSE} \leq \frac{1}{2\pi e} \exp \left( 2\mathcal{H}(X_k | S(i, j, \ldots)) \right)$$

For *categorical* variables:

$$P_e \geq \frac{\mathcal{H}(X_k | S(i, j, \ldots))}{\log N} - 1$$
RTT-Distance Marginal Distribution
RTT-Distance Marginal Distribution

\[ R^2_\Phi = 0.881 \]

- Speed of Light in Vacuum
- Speed of Light in Fibre
- \( \chi_t = 0.018 \chi_d + 20.0 \)
Non-linear features

$R^2_\Phi = 0.881$

Great Circle Distance (Thousands of Km)

RTT (ms)
Subcontinental Zone Decomposition

North America (North) – South America (East)

\[ R^2_\Phi = 0.825 \]

\[ x_t = 0.018x_d + 32.6 \]
Subcontinental Zone Decomposition

\[ R^2_\Phi = 0.748 \]

\[ x_t = 0.017x_d + 6.9 \]

North America (North) – South America (East)
Subcontinental Zone Decomposition

Eastern Europe – Asia Pacific

North America (North) – Asia Pacific

\[ R^2_{\Phi} = 0.100 \]

\[ -x_t = 0.004x_d + 214.9 \]

\[ R^2_{\Phi} = 0.250 \]

\[ -x_t = -0.002x_d + 317.8 \]
Subcontinental Zone Decomposition

\[ R^2_\Phi = 0.628 \]

\[ -x_t = -0.013x_d + 429.5 \]

Eastern Europe – Asia Pacific
Subcontinental Zone Decomposition

Western Europe – Oceania

\[ \frac{R^2}{\Phi} = 0.773 \]

\[ x_t = -0.020x_d + 667.2 \]
Variation in Median Line Slope

\[ X_d[\mathbb{b}, \mathbb{b}'] \]

\[ X_d[\mathbb{a}, \mathbb{a}'] \]

\[ Z \quad \mathbb{Z} \quad Z' \]

\[ b \quad a \quad a' \quad b' \]

\[ R^2 = 0.748 \]

\[ -x_t = 0.017x_d + 6.9 \]
Variation in Median Line Slope

$X_d[a, a']$

$X_d[b, b']$

$R^2_\phi = 0.100$

$- x_t = 0.004x_d + 214.9$

Great Circle Distance (Thousands of Km)
Variation in Median Line Slope

\[ X_d[a, a'] \]

\[ X_d[b, b'] \]

\[ R^2_\phi = 0.628 \]

\[ x_t = -0.013x_d + 429.5 \]
Distribution of Median Line Slopes
Subcontinental Zone Decomposition

$R^2_\Phi = 0.628$

$x_t = -0.013x_d + 429.5$

Reflection Line
Main Line

Eastern Europe – Asia Pacific
Unfolding: Eastern Europe – Asia Pacific

\[ x_t = -0.018x_d + 458.0 \]
Unfolding: Eastern Europe – Asia Pacific

\[ x_t = -0.018x_d + 458.0 \]

\[ x_t = 0.018x_d + 20.0 \]
Unfolding: Eastern Europe – Asia Pacific

\[ x_t = 0.018x_d + 20.0 \]
Unfolded RTT/Distance Density

$$R^2 = 0.937$$

- Speed of Light in Vacuum
- Speed of Light in Fibre
- $$x_t = 0.016x_d + 22.3$$
Large-scale Circuitousness Measures

- Large scale routing distance excess $\sigma$

$$\sigma(X_z) = D_z - X_d = 2(X^*_d - X_d)$$

- Total distance ratio $\rho$

$$\rho(X_z) = \left(\frac{.65c}{2}\right) \left[\alpha_U (1 + \frac{\sigma(X_z)}{X_d}) + \frac{\beta_U}{X_d}\right]$$
# Large-scale Circuitousness Measures

<table>
<thead>
<tr>
<th>Zone 1</th>
<th>Zone 2</th>
<th>$\sigma$ (km)</th>
<th>$\rho$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asia Pacific</td>
<td>Western Europe</td>
<td>7,410</td>
<td>3.08</td>
</tr>
<tr>
<td>Asia Pacific</td>
<td>Eastern Europe</td>
<td>9,796</td>
<td>3.7</td>
</tr>
<tr>
<td>Oceania</td>
<td>Western Europe</td>
<td>2,702</td>
<td>1.98</td>
</tr>
<tr>
<td>S. A. (East)</td>
<td>Asia Pacific</td>
<td>973</td>
<td>1.79</td>
</tr>
<tr>
<td>Asia Pacific</td>
<td>Central Asia/Middle East</td>
<td>11,348</td>
<td>3.94</td>
</tr>
<tr>
<td>Oceania</td>
<td>Eastern Europe</td>
<td>5,685</td>
<td>2.33</td>
</tr>
<tr>
<td>Asia Pacific</td>
<td>Indian Subcontinent</td>
<td>4,110</td>
<td>3.14</td>
</tr>
<tr>
<td>Oceania</td>
<td>Central Asia/Middle East</td>
<td>7,623</td>
<td>2.57</td>
</tr>
<tr>
<td>Africa</td>
<td>Oceania</td>
<td>18,973</td>
<td>4.53</td>
</tr>
<tr>
<td>S. A. (East)</td>
<td>South America (West)</td>
<td>7,187</td>
<td>4.77</td>
</tr>
<tr>
<td>Oceania</td>
<td>South America (West)</td>
<td>3,608</td>
<td>2.15</td>
</tr>
<tr>
<td>Africa</td>
<td>South America (West)</td>
<td>11,208</td>
<td>3.41</td>
</tr>
</tbody>
</table>
Large-scale Circuitousness Measures

- Some routing samples between Western Europe and Asia Pacific flow through the USA; others flow through Russia/China

- \( \sigma \approx 10,000\ km \)
- \( \sigma \approx 4,000\ km \)
Large-scale Circuitousness Measures

- Routing samples between Russia and Eastern Australia flow through Western Europe and the USA

\[ \sigma \approx 5,700 \text{ km} \]
Large-scale Circuitousness Measures

- Routing samples between South Africa and Eastern Australia flow through Western Europe and the USA

- $6,000 < \sigma < 24,000 \text{ km}$
Conditional Entropy

- We choose subsets of variables from this table, and estimate the amount of information that they give about another variable.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Interpretation</th>
<th>Bins used</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_p$</td>
<td>Pairs of /8 prefixes</td>
<td>14,189</td>
</tr>
<tr>
<td>$X_\alpha$</td>
<td>Pairs of Top AS numbers</td>
<td>215,392</td>
</tr>
<tr>
<td>$X_z$</td>
<td>Pairs of subcontinental zones</td>
<td>66</td>
</tr>
<tr>
<td>$X_c$</td>
<td>Pairs of countries</td>
<td>2,648</td>
</tr>
</tbody>
</table>

Quantitative Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Interpretation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_l$</td>
<td>Common prefix length</td>
<td>32</td>
</tr>
<tr>
<td>$X_d$</td>
<td>Great circle distance</td>
<td>300</td>
</tr>
<tr>
<td>$X_t$</td>
<td>Round trip time (RTT)</td>
<td>300</td>
</tr>
</tbody>
</table>
Conditional Entropy (RTT)

- The single variable that gives most information about RTT is the *country pair* of the endpoints.
  - This gives an RMSE < ~162 ms.

- If we consider the /8 prefix pair, the country pair and the geodesic distance, the remaining uncertainty is close to the minimum achievable with 4 variables or less (2.7 bits).
  - This gives an RMSE < ~12 ms.
The single variable that gives most information about the subcontinental zone pair is the /8 prefix pair of the endpoints.

- This gives a $P_e > \sim 0.09$.

If we consider the /8 prefix pair, the geodesic distance and the RTT, the remaining uncertainty is close to the minimum achievable with 4 variables or less (.35 bits).
Conditional Entropy (Country Pair)

- Both the \textit{AS pair} and the \textit{subcontinental zone pair} give close to the maximum information about country pair in a single variable.
  - This gives a $P_e > \sim.35$.

- If we consider the AS pair, the subcontinental zone pair and the geodesic distance, the remaining uncertainty is close to the minimum achievable with 4 variables or less (.98 bits).
Conditional Entropy (Geodesic Distance)

- The single variable that gives most information about the geodesic distance is the *country pair* of the endpoints.
  - This gives a RMSE < ~1600 km.

- The single network variable that gives most information about the geodesic distance is the *AS pair* of the endpoints.
  - This gives a RMSE < ~4000 km.
Conditional Entropy (Geodesic Distance)

- If we consider the AS pair, the country pair, the common prefix length and the RTT, the remaining uncertainty is close to the minimum achievable with 4 variables or less.
  - This gives an RMSE < ~94 km.

- If we consider only network variables (AS pair, common prefix length and RTT), the remaining uncertainty is 3.25 bits.
  - This gives an RMSE < ~416 km.
Conditional Entropy (AS Pair)

- The single variable that gives most information about the AS pair is *country pair* of the endpoints.
  - This gives a $P_e > \sim .5$.

- If we consider the country pair, the common prefix length, the geodesic distance and the RTT, the remaining uncertainty is close to the minimum achievable with 4 variables or less (3.5 bits).
  - This gives a $P_e > \sim .2$. 
Conclusions

• Large-scale analysis of RTT and its related geographic and network properties
  – Novel RTT dataset comprising 19 million measurements between 54 thousand measurement points.
  – RTT measurements as realisations of a multidimensional random variable (multidimensional histogram)
  – Novel median-based linear fitting algorithm
  – Subcontinent-scale routing distance analysis
  – Conditional entropy of sets of geographic and network variables
  – Bounds on estimation and classification errors