Sibyl
A Practical Internet Route Oracle

Ítalo Cunha
P. Marchetta, M. Calder, Y-C. Chiu
B. Schlinker, B. Machado, A. Pescapè
V. Giotsas, H. Madhyastha, E. Katz-Bassett
Traceroute is Widely Used

“The number one go-to tool is traceroute.”
NANOG Network operators troubleshooting tutorial, 2009.

Lots of use cases
- Topology mapping
- AS relationship inference
- Route performance and inflation
- Locating congestion
- Identifying outages
- Detecting prefix hijacks
Traceroute is Widely Used

“The number one go-to tool is traceroute.”
NANOG Network operators troubleshooting tutorial, 2009.

<table>
<thead>
<tr>
<th>Lots of use cases</th>
<th>Lots of vantage points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topology mapping</td>
<td>PlanetLab</td>
</tr>
<tr>
<td>AS relationship inference</td>
<td>Ark</td>
</tr>
<tr>
<td>Route performance and inflation</td>
<td>RIPE Atlas</td>
</tr>
<tr>
<td>Locating congestion</td>
<td>Traceroute servers</td>
</tr>
<tr>
<td>Identifying outages</td>
<td>MobiPerf, Dasu, BISmark</td>
</tr>
<tr>
<td>Detecting prefix hijacks</td>
<td></td>
</tr>
</tbody>
</table>
Traceroute is Widely Used

“The number one go-to tool is traceroute.”
NANOG Network operators troubleshooting tutorial, 2009.

Lots of use cases
- Topology mapping
- AS relationship inference
- Route performance and inflation
- Locating congestion
- Identifying outages
- Detecting prefix hijacks

Lots of vantage points
- PlanetLab
- Ark
- RIPE Atlas
- Traceroute servers
- MobiPerf, Dasu, BISmark

But traceroute only supports one query:
“What is the path from vantage point s to destination d?”
Next-gen measurements

What we do

What we want to do

Give me paths like X.

Here are some paths.

Unified Probing Platform

RIPE Atlas 1
RIPE Atlas 2
MobiPerf 1
MobiPerf 2

RIPE Atlas 1
RIPE Atlas 2
MobiPerf 1
MobiPerf 2

Dest  Path

Dest1  Path

Dest2  MeasrDroid

MobiPerf 1
MobiPerf 2
Goal

Provide support for rich queries on Internet paths
Paths that go through Sprint’s Chicago PoP to USC:

^.*[Sprint&Chicago].*[USC]$ 

From NANOG: “Problem between Level3 in LA and GTT in Seattle?”

^.*[Level3&LA].*[GTT&Seattle]}.${
Limited VPs $\rightarrow$ Limited Path Coverage
More VPs → Richer Path Coverage
Combining Platforms Improves Coverage
Combining Platforms Improves Coverage

Support for multiple measurement platforms
Rate Limits ➔ Cannot Issue All Measurements

![Graph showing the cumulative distribution function (CDF) of the number of Autonomous Systems (ASes) seen on paths with rate limits.]
Rate Limits → Cannot Issue All Measurements

Need to target probes intelligently
Optimize Use of Probing Budget

In each round, allocate probing budget to best serve queries

\[
\max_{T_r} f(T_r), \text{ where }
\]

\[
T_r = \bigcup_{V \in \mathcal{V}} T_{r,V}
\]

and

\[
f(T_r) = \sum_{q \in Q} f_q(T_r)
\]

subject to

\[
\forall V \in \mathcal{V}, |T_{r,V}| \leq C_V
\]

Pick traceroutes \( T_r \) that maximize the number of answered queries

Subject to the rate limits of each platform \( V \)
“I suspect problems on peering between GBLX-AT&T on way to Akamai. Give me a matching path.”
“I suspect problems on peering between GBLX-AT&T on way to Akamai. Give me a matching path.”
Train system to recognize unlikely predictions

- Features include:
  - Peering relationship at splice point
  - Path length inflation vs shortest prediction

Evaluation shows system can identify measurements more likely to match queries
Prediction is effective: Sibyl satisfies 81% as many queries as an Oracle that knows which candidates match each query.

Important to assess likelihood: Sibyl satisfies 264% more than Randomly selecting among spliced candidates.
Future Work

Improve path prediction and ranking
- Better formalism, richer training sets

Balance between serving current queries and expected benefit in serving future queries
- Fill in gaps in routing knowledge
- Refresh stale knowledge

Unify queries over historical and live data
- “Give me a path that used to look like X but now looks like Y.”

Queries over path performance
- Latency, bandwidth, loss, length