Internet Measurements and Public Policy: Mind the Gap

Hadi Asghari *
Michel JG van Eeten *
Milton L Mueller +

*** Delft University of Technology, Faculty of Technology, Policy and Management**
+ Syracuse University, School of Information Studies

Delft University of Technology

In Summary

- Measurement projects (often) fail to have the desired (deserved) impact on cyber policy debates
- … because of a gap between how computer scientists think of measurement data and how other researchers use it

Why This Matters

- There is a real need for large scale empirical evidence in policy research (e.g., versus ideas based on hunches, anecdotes or untested models)
- Policy makers are very receptive towards empirically backed policy research
- So: measurement science and policy research should be a fruitful marriage; however, it's often frustrating

Understanding the Gap

Measurement Experts

- Create a measurement tool
- Have it run for a while on some infrastructure

Policy Researchers

- Obtain & convert logs to dataset
- Experiment with models, add variables, extract patterns and insights for policy

Caused by different needs & values with regards to data, e.g.: granular vs aggregated, real-time vs historical, etc

TU Delft

The Ideal Measurement Set (1/5)

Measurement sets ought to keep archives and benefit from being up to date

The Ideal Measurement Set (2/5)

Providing spatially and temporally aggregated versions of the data is helpful

- Due to different units of analysis
- Also less to download; solves some privacy issues

The Ideal Measurement Set (3/5)

Measurements ought to have clear verdicts and interpretations

In practice:

i. documentation
ii. "complete audit trail" (Paxson, 2004)
iii. parsers
iv. access to authors / community

The Ideal Measurement Set (4/5)

Consistency of the measurement instrument and logs over time is important.

- E.g., have parallel versions running; monitor test infrastructure, etc.

The Ideal Measurement Set (5/5)

Data collection should be organized in a manner that promotes sample validity.

Discussion (1) – N=1

	Time Period	Aggregation & format	Logs to verdicts	Consistency	Sample size
Glasnost	2009-now	(-) Individual test logs	(-) Parsing involves many steps	(-) Multiple discontinuities	(+/-) Mixed
Spam-trap	2005-now	(+/-) Logs, daily. IP based.	(+) Relatively clear, excellent support	(+) Yes	(+) Good
DShield	2006-now	(+/-) Logs, daily. IP & port.	(+/-) False positives still unclear	(+) Yes	(+) Good
SSL-Observatory	Only 2010	(+) SQL dump; Certificates	(+) Good documentation and support	N/A	(+/-) Full, but once
Conficker sinkhole	2009-now	(-) Logs, hourly; per connect	(+) Relatively Clear	(-) Log format change	(+) Full

Discussion (2) - Costs

- Q: Great ideas! But who will pay for them?

- See it as cost of doing business, if impact important
- And not necessarily expensive, e.g. like "hallway usability testing"

Discussion (3) – Alternative View

- Q: Maybe it makes more sense to improve the technical competencies of social scientists

- A: We've tried. Definitely useful for those with some engineering background; but has limitations; e.g. using SQL is intuitive and productive, whereas text parsing is a no-no
- "Rich context" and other points still important

Acknowledgements

- We would like to thank the developers and maintainers of the measurement projects that have provided us great data for our research

- I would like to thank USENIX for graciously funding my trip to present this paper and have this discussion here

- Papers: http://homepage.tudelft.nl/r0d9v/research.html

TU Delft