Strategies for Sound Internet Measurement

Vern Paxson

In the proceedings of ACM Internet Measurement Conference, 2004

***Presented By*,**
Ramya Raghavendra
CS595

Disclaimers

- **There are no new research results**
- **A number of points apply to large scale systems in general**
- **Unfortunately, just all of the strategies involve extra work ("discipline")**
- **There is no easy answer to the question "how much extra work is merited?"**

From the author's presentation in ICIR, October 26, 2004

Outline

- Background
- Motivation
- Dealing with errors and imperfections
- Dealing with large volumes of data
- Ensuring reproducible analysis
- Making dataset publicly available
- Summary

Why simulations are not enough?

- Constructed, abstracted model of the real world
- Simulations are only as good as the simulator
- Danger of oversimplification
- Can model only foreseen properties

Why measurements are important

- **Measurements are a "reality check"**
- **Question implicit assumptions**
- **But…**
  - Conducting sound Internet measurements is **difficult**
  - Soundness in measurement process and quality of analysis
- **Emphasis on general principles as opposed to particul ar techniques and tools**

Difficulties of Measurements

- **To measure the global Internet, wide cooperation is need ed.**
  - Designing meaningful experiments
  - Securing permission
  - Testing that the tools work correctly
  - Reducing raw data
  - Explore data and present results
- **ISPs are reluctant to coordinate their efforts**
- **Tremendous growth of Internet**
  - Results get obsolete very rapidly
  - Conducting scalable measurement is difficult

Dealing With Errors and Imperfections

- Precision – errors inherent in the basic design of a tool
- Meta-data
- Accuracy – errors incurred during application of t ool
- Misconception – are we measuring what we belie ve we are?
- Calibration – strategy to detect and correct error s

Precision

- Maximum exactness that a tool's design permits
  - Consider a tcpdump timestamp:
  - 1092704424.276251 IP 192.168.0.122.22 > 192.168.0.1 16.1213:
  - How precise is it?
  - Answer: at most to 1 μsec. But perhaps much less.
- When reporting measurements, indicate the tool's precision
- Formulate error bars

Precision (Cont'd)

- **Notion applies to discrete measurements too**
  Ex: Copies of packets, web server logs etc
- **Depends:**
  - Snapshot length limits total data
  - *Filtering* does too
- **Pitfalls**
  - Reporting simplistic precisions
- **When does extra effort matter?**
  - Unfortunately no crisp rules
  - Discuss whether or not it matters

Strategy #1: Meta-data

- **Preserving the information during the course of a nalysis**
- **Many convenient data formats lack a way to anno tate measurements, e.g.**
  Today's web server log, Tcpdump
  - Simple, line-oriented structure
  - no API available for retrieving or setting the value

```
15:39:57.123143.248.140.136 - - [08/May/2005:21:18:39 -0700] "OPTIONS / HTTP/1.1" 302 286 "-
15688        " "Microsoft-WebDAV-MiniRedir/5.1.2600"
15:39:57.123143.248.140.136 - - [08/May/2005:21:18:39 -0700] "OPTIONS /apache2-default/ HTTP
8) ack 53 wi/1.1" 200 - "-" "Microsoft-WebDAV-MiniRedir/5.1.2600"
15:39:57.123143.248.140.136 - - [08/May/2005:21:18:39 -0700] "OPTIONS /apache2-default/ HTTP
8) ack 53 wi/1.1" 200 - "-" "Microsoft-WebDAV-MiniRedir/5.1.2600"
15:39:57.123143.248.140.136 - - [08/May/2005:21:22:29 -0700] "PROPFIND /sjlee HTTP/1.1" 405
8) ack 53 wi304 "-" "Microsoft-WebDAV-MiniRedir/5.1.2600"
15:39:57.132143.248.140.136 - - [08/May/2005:21:22:29 -0700] "PROPFIND /sjlee HTTP/1.1" 405
17112        304 "-" "Microsoft-WebDAV-MiniRedir/5.1.2600"
```

Key Consideration for Meta-data

- **Uniform format for the meta-data is needed**
- **Data have a lifetime beyond what the researcher initially envisions**
  - Good data is hard to gather
  - Revisiting datasets in new contexts
- **It is beneficial to retain meta-data information ev en when doing so is not of immediate benefit.**

Accuracy

- **How well does the measured abstraction indeed match the a ctual phenomenon?**
- **Much broader problem than precision**

E.g.) Clocks can:

- **Arbitrarily off from true time**
- **Jump backward or forward**
- **Run arbitrarily fast or slow**
- **Fail to move**

E.g.) Packet filters can:

- **Fail to record packets ("drops")**
- **Fail to report drops**
- **Report drops that did not occur**
- **Reorder packets**
- **Duplicate packets**
- **Record wrong packets**

Inaccuracy rises meta-data problem??

Measurement tools are in adequately recording fa ilure information, e.g.

Tcpdump

- **produces an end-of-run summary of the total number of drops.**
- **No report on the drops by the tap itself**
- **Cannot associate drops with the time of occurrence**
- **Records them separately, requires association**

Misconception

- **Errors in equating what we are <u>actually measuring</u> with what we <u>wish to measure</u>**

Examples of Misconception

- Measuring TCP packet loss
  - by counting retransmitted packets
  - Risks overlooking the problem
    - Unnecessary retransmission
    - Packets replicated by the network
- Computing the distribution of TCP connection sizes
  - By capturing SYN and FIN packets
    - Using the difference between the sequence numbers
    - Miss/fail to include packets SYNs or FINs
      - Very largest connections started before we start tracing
      - Not terminated when we finish

Significant Misconceptions

- Vantage point
  - The location of exactly where a measurement is performed c an significantly skew the interpretation of the measurement
- Representativeness
  - Internet properties vary a great deal across different points and different times
- General strategy against misconceptions
  - To gather more than one type of dataset
    - From a different location
    - From a different time
  - Seek out early peer review

Calibration

- A set of techniques that help with detecting prob lems of inaccuracy, misconception, errors in anal ysis.
  - Examining outliers and spikes
  - Employing self-consistency checks
  - Measuring facets of the same phenomenon different wa ys and comparing
  - Evaluating synthetic data

Not achieve perfection but achieve confidence

Examining outliers and spikes

- Outliers
  - unusually low or high values
- Spikes
  - values that repeat a great deal
- Represent "corner cases" at the extremes of me asurement where problems often manifest
- Reflect measurement errors, analysis errors, mis conception

Employing self-consistency checks

Testing whether properties that must hold do in f act hold, e.g.

How to find evidence of packet filter drops:

Use cumulative TCP ACK properties

- To see whether each ACK present in a trace at the point i n time it was sent
- All of the data up to the sequence number of ACK has be en seen previously in the trace

Comparing multiple measurements

- Measuring the same phenomenon different ways and comparing the results
- Conducting <u>additional</u> measurement
- <u>Multiple</u> *versions* of analysis

Evaluating Synthetic Data

- Hand-edit some measurements to introduce changes
- Test whether software processes it correctly
- Monte Carlo simulations
  - Simulate multiple random instances

Dealing with large volume of data

- Large scale leads to a number of potential difficulties
- System limitations such as disk space, maximum file sizes, number of files on a volume, ….
- Many systems used for statistical analysis have upper bounds on the amount of data they can process
- Strategy : Indepth analysis of small subsets

Ensuring reproducible analysis

- **The analysis derived from the data is reproducibl e**
- **Strategy: Structured analysis**
  - **Single master script**
  - **Cataloging notebook**
- **Benefits of structuring analysis**
  - **Reproduce the results, Minimizing the headaches**
  - **Explore the analysis of the data in a consistent fashion**

Making datasets publicly available

- **Dearth of publicly available datasets**
- Strategy: periodically analyze ongoing measurements
  - **let's you discover when data acquisition *broken***
  - **ensures you're collecting (some) *meta-data***
- Problems
  - **Reluctance, legal impediments**
  - **Privacy, security and business sensitivities - anonymization**
- Strategy: package analysis for "data reduction requests"
  - **send data analysis software to dataset holder**
  - **they run it, inspect results, & return them**

Summary

- **The basic problem of dealing with imperfection measurement**
  - **Precision / Accuracy**
  - **Meta-data**
  - **Misconception**
- **Strategies for soundness**
  - **Summary of Strategies:**
  - Strategy #1: *maintain meta-data*
  - Strategy #2: *run your intended methodology by colleagues*
  - Strategy #3a: *examine outliers and spikes*
  - Strategy #3b: *employ self-consistency checks*
  - Strategy #3c: *compare multiple measurements/computations*
  - Strategy #4: *structure for reproducible analysis*
  - Strategy #5: *periodically analyze ongoing measurements*
  - Strategy #6: *package analysis for "data reduction requests"*
  - Strategy #7: *subsample large datasets, assess variability*

Worth to pursue …

- **Data management using databases and <u>version control</u>**
- **Scriptable analysis environments for exploration and <u>reproducibility</u>**
- **Tools investigate <u>differences</u>**
- **<u>Portable</u> measurement management tools, environments**
- **<u>Publication</u> of measurement data**

Thank you!

Comparing Multiple Measurements

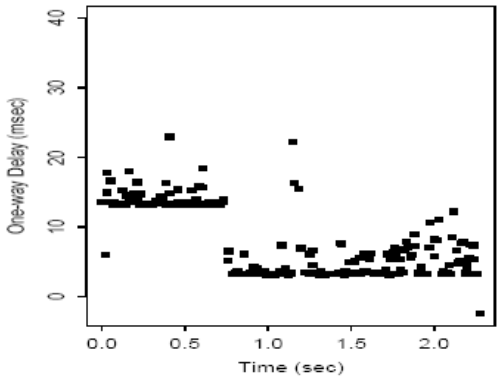

Figure 1: One-way transit time step that could be due to either a routing change or a clock adjustment.

Comparing Multiple Measurements

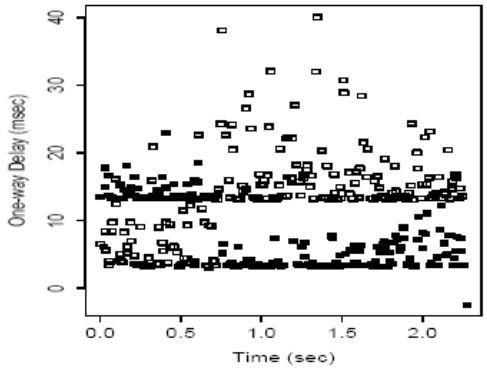

Figure 2: Incorporating additional measurements resolves the change as due to a clock adjustment.