6.896 Lecture 4: Methodology II: Simulation, Implementation and Analysis

Hari Balakrishnan

September 21, 1998

1 Outline

- **Readings: CK74, Pax96**
- Finish Bolot's paper (from last class)
- Simulation overview
- Original TCP/IP paper (Cerf/Kahn, 1974)
- Internet routing behavior (Paxson, 1996)

2 Simulation overview

- Simulation: In valuable tool for performance analysis and for feasibility studies. Enables better design and (later) implementation, "time travel," system characterization. But, at the same time, must remember its limitations and pitfalls.
- Discrete-event simulations: Most common for computer systems and network simulators. Natural model to use here.
- Key components are:
  - *Scheduler:* schedule event for time t, execute for time δt, cancel a previously scheduled event. Can also hold an event for later scheduling.
  - *Clock:* global simulation variable. Can increment either **unit time** or **event-driven**.
  - *Event queue:* Describes system state. Each event has time at which it will execute. Usually a heap or calender queue.
  - *Event handler:* Invoked when event fires.
  - *User input:* In the form of topologies, delays, traffic etc. for network simulations. Often in a scripting language.
  - *Tracing:* Log interesting stuff to files.
  - *Termination:* Either when there are no more events or when a final "finish" event is called.
  - *Randomization:* Multiple runs with random variations of certain parameters (e.g., traffic patterns, protocol parameters, etc.)
- *ns* uses all of these components.

3 Original design of TCP/IP

3.1 Motivation

- Problem: many different networks, how to *internetwork* them together.
- Clearly *tremendous* impact of this work!
- Written as a design paper, but enough detail to code into implementation.

3.2 Challenges

1. Different internal addressing.
2. Different MTUs.
3. Different delays and bandwidths. (heterogeneity in characteristics and need for congestion control, but they didn't know it at the time).
4. Packet loss.
5. Different routing and management protocols within networks.

3.3 Highlights of solution

- Key idea: **gateways.** Reject translation in favor of gateways.
- IP over everything—beginnings of protocol hourglass model evident in first section.
- Addressing common to entire internetwork. Internal characteristics of networks are of course different.
- Standard packet header format. But notice mixing of transport information in routing header. (And some fields like ttl are missing).
- Gateways perform *fragmentation* if needed, *reassembly* at end hosts. Q: why don't gateways reassemble?
- **TCP:** Process level communication via the TCP. And in fact, they had the TCP do reassembly.
- Draw picture of processes and TCP. How should multiple processes communicate?
  1. Use same IP packet for multiple TCP segments?
  2. Different packets for different segments?
- 2 simpler and better than 1. (when might 1 be better?). Also, they didn't anticipate multiple connections between hosts!
- Demux key is port number (typically one process per port).
- Byte-stream TCP, so use ES and EM flags. EM, ES, SEQ and CT adjusted by gateways. I.e., gateways look through and modify transport information. End-to-end checksums for error detection.

- Sliding window, cumulative ACK-based ARQ protocol (with timeouts) for reliability and flow control. Similar to CYCLADES and ARPANET protocols. Q: advantages of cumulative acks?
- Window-based flow control. Also achieve process-level flow control.
- Implementation details/recommendations for I/O handling, TCBs.
- First-cut connection setup and release scheme. (Heavily modified subsequently to what it is today).
- Way off the mark on accounting issues! (The problem is a lot harder than they anticipated).
- One-line summary: A great paper that details an internetworking protocol of great significance. Enough details to actually implement it, but not a dry document to read in any way!

4 Announcements

- Copies of CJ89 outside my door later today. Also available on web page
- Next class: congestion control principles. Read CJ89 for sure, background from Peterson & Davie or from a textbook. CJ89 articulates the principles that motivate almost all known good schemes today.
- How many people are planning to use athena? Class locker?
- Start thinking about projects and partners. Will start putting list of suggestions this week.
- Questions?

5 End-to-end Internet routing behavior

5.1 Motivation

- Large, complex system. How does routing *behave*. Question interesting on several fronts.
- Not performance, but *behavior*. Not diagnosis, but characterization. Not with router support, but end-to-end.
- Three main goals:
  - Characterize routing pathologies.
  - Characterize route stability.
  - Characterize route symmetry.
- Scarce related work until Paxson's thesis work: Chinoy at SDSC and (more recently) Labovitz et al. at Michigan.
- Major experimental effort with lots and lots of trace data.

5.2 Background, tools, methodology

- Autonomous systems (draw pic). BGP is exterior routing protocol today. About 1000 ASs in 1995-96. (We'll study BGP later in the term).
- *traceroute*. Written by Jacobson to diagnose routes. Invaluable tool for diagnosis and route data collection. Uses ttl and ICMP error messages from routers to determine routes. Q: Why do IP networks have ttl? Traceroute from *ftp://ftp.ee.lbl.gov*.
- NPD infrastructure. (Traceroute servers exist on the Internet today — check out http://www.uniplus.ch/direct/app_trce.htm) 37 sites ran NPD in all.
- N hosts implies $O(N^2)$ paths. This is good!
- Two data sets: D_1 and D_2. D_2 had more types of measurements based on observations and analysis of D_1. E.g., symmetric experiments, multiple measurement rates. (Why?)
- *Exponential sampling:* Want to observe "true" state of the Internet, but can only sample. How to do this? Well, memoryless property of the exponential distribution leads to the PASTA principle, which applies in most general situations.

5.3 Routing pathologies

1. **Loops:** Forwarding loop, information loop (e.g., counting-to-infinity), traceroute loop. NPDs obtain traceroute loops. Two kinds of loops: less than 3h and more than 12h.
2. **Erroneous routing:** pathology where US to Europe went via Israel! But by and large, it works without errors.
3. **Midstream change:** Does happen, with bimodal recovery times.
4. **Fluttering:** Rapid oscillation in routing, could lead to out-of-ordering. Does happen; in one case it wasn't localized. Aside: other research shows that out-of-ordering has increased because of dispersity between tightly coupled routers.
5. **Infrastructure failures:** Overall network availability rate between 99.5% and 99.8%.
6. **Too many hops:** traceroute default is 30. Median hop length is 16, well-modeled by a discrete Gaussian.
7. **Temporary outages:** Do happen, long ones last over 30s. No good explanation for difference between D_1 and D_2.
8. **Outage patterns:** Outages correlated with load.
9. **Trends:** most pathologies getting worse with time!

5.4 Route stability

- *Prevalence:* how often are you likely to encounter a given route?
- *Persistence:* how long before a route changes? Most interesting for the dominant route. Independent concepts.

- Why are these important? (path predictability, learning about network, end-to-end path assessment, real-time state migration).
- In general, Internet paths are dominated by a single route, but there are significant site-to-site variations.
- Persistence at multiple time scales, seconds (load-balancing flutters) to minutes (tightly coupled routers) to 10s of minutes to hours (intra-network changes) to days.
- Two-thirds of all paths had stable routes.

5.5 Route symmetry

- Why is this important? (one-way delay, complicates measurement and debugging, asymmetric technologies—not in paper).
- Why does it happen? Technologies and/or metrics different, cold-potato/hot-potato routing between ISPs for cross-country traffic.
- Paired measurements in D_2 allow this to be studied.
- Result: asymmetry exists, is significant (and is likely to increase in future).

5.6 Conclusion

Impressive large-scale analysis of routing behavior. Strong methodology and fairly thorough documentation of pathologies, stability and symmetry characteristics.

Time for another such study, more extensive with performance characteristics too. IPPM working group (ippm-request@advanced.org).