Methodologies for Networking Research
Measurement

J. Padhye, V. Firoiu, D. Towesley, and J. Kurose
“Reality Check”

Are our assumptions reasonable? Is our mathematical model a good estimation of the real world?
Figure 7: manic to baskerville

Figure 8: pif to imagine

Figure 9: pif to manic

Figure 10: void to alps
Experimentation

e.g., V. Jacobson. “Congestion Control and Avoidance"
Deal with implementation issues

Sometimes unforeseen complexities (e.g. own research experience in Unreliable TCP)
Understand the Behavior of Systems

Some systems are too complex to understand with “thought experiments” alone.
Trace data from four simultaneous TCP conversations using congestion avoidance over the paths shown in figure 7.
Analysis

D. Chiu and R. Jain, "Analysis of the increase and decrease algorithms for congestion avoidance in computer networks,"

Explore with Complete Control

We can understand the basic forces that affect the system. e.g. TCP throughput is inversely proportional to $\sqrt{p}$
Simplify complex systems

But, if too simplified, important behavior could be missed (TCP throughput without timeout)
Simulation

K. Fall and S. Floyd, "Simulation-based comparison of Tahoe, Reno, and SACK TCP,"

S. Floyd, V. Jacobson, "Random Early Detection Gateways for Congestion Avoidance,"
Check Correctness of Analysis

If a simulation uses the same assumptions/model as the analysis, this simply verifies the correctness of the mathematical derivations.
Check Correctness of Analysis

Simulation can relax some assumptions, use more complex models, etc. to test the limits of analysis.

(Real measurement/experiments still needed to check the usefulness of analysis results)
Explore Complex Systems

Some systems are too difficult/impossible to analyzed (e.g. Internet)
Helps Develop Intuition
“Difficulties in Simulating the Internet”

Sally Floyd, Van Paxson
ACM/IEEE TON, 9(4) August 2001
Why is Internet hard to simulate?
1. Internet is diverse
End-hosts: phones, desktops, servers, iPod, Wii
Links: Ethernet, WiFi, Satellite, Dial-up, 3G
Transport: TCP variants, UDP, DCCP
Applications: games, videos, web, ftp, bittorrent
2. Internet is huge
570,937,778

Number of Hosts as of July 2008
681,064,561

Number of Hosts as of July 2009
https://www.isc.org/solutions/survey/history
3.

Internet is changing
Why is Internet hard to simulate?

1. Heterogeneous
2. Huge
3. Changing
What Internet topology should you use in your simulation?

How are end hosts connected? What are the properties of the links?
Topology changes constantly
Companies keep info secrets
Routes may change
Routes may be asymmetric
You will need to simulate over a wide range of connectivity and link properties
Which TCP version to use?
Using “fingerprinting”, 831 different TCP implementations and versions are identified.
Which to use?
Which to ignore?
What applications to run?

What type of traffic to generate?

How congested should the network be?

Asia Packet Loss (%) Past 24 Hours

North America Packet Loss (%) Past 24 Hours
How congested should the network be?
Example from Sally Floyd: RED vs DropTail
Example from Sally Floyd: TFRC for VoIP
We can focus our simulation on dominant technology/application today.
TCP: NewReno SACKS
OS: Windows Linux
Applications: Web, FTP
What about tomorrow?
WiMax?
Sensors?
Virtual World?
DCCP?
How to verify the simulation is correct?
Looking for Invariants
1. Diurnal Patterns
<table>
<thead>
<tr>
<th>hour</th>
<th>#constrained</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>139 2.5%</td>
</tr>
<tr>
<td>01</td>
<td>144 2.6%</td>
</tr>
<tr>
<td>02</td>
<td>146 2.6%</td>
</tr>
<tr>
<td>03</td>
<td>140 2.5%</td>
</tr>
<tr>
<td>04</td>
<td>119 2.1%</td>
</tr>
<tr>
<td>05</td>
<td>89 1.6%</td>
</tr>
<tr>
<td>06</td>
<td>69 1.2%</td>
</tr>
<tr>
<td>07</td>
<td>55 1.0%</td>
</tr>
<tr>
<td>08</td>
<td>45 0.8%</td>
</tr>
<tr>
<td>09</td>
<td>40 0.7%</td>
</tr>
<tr>
<td>10</td>
<td>40 0.7%</td>
</tr>
<tr>
<td>11</td>
<td>42 0.8%</td>
</tr>
<tr>
<td>12</td>
<td>51 0.9%</td>
</tr>
<tr>
<td>13</td>
<td>57 1.0%</td>
</tr>
<tr>
<td>14</td>
<td>68 1.2%</td>
</tr>
<tr>
<td>15</td>
<td>75 1.3%</td>
</tr>
<tr>
<td>16</td>
<td>77 1.4%</td>
</tr>
<tr>
<td>17</td>
<td>92 1.6%</td>
</tr>
<tr>
<td>18</td>
<td>98 1.8%</td>
</tr>
<tr>
<td>19</td>
<td>105 1.9%</td>
</tr>
<tr>
<td>20</td>
<td>108 1.9%</td>
</tr>
<tr>
<td>21</td>
<td>113 2.0%</td>
</tr>
<tr>
<td>22</td>
<td>124 2.2%</td>
</tr>
<tr>
<td>23</td>
<td>134 2.4%</td>
</tr>
</tbody>
</table>

U Waterloo Data 24 Oct 2007
2. Self-Similar Traffic
The traffic is bursty regardless of time scale
Wikipedia
3. Poisson Session

Arrival

\[ f(k; \lambda) = \frac{\lambda^k e^{-\lambda}}{k!}, \]
Remote logins, starting FTP, beginning of web surfing etc.
(so are dead light bulbs, spelling mistakes, etc.)
4. Log-normal Duration

\[ f(x; \mu, \sigma) = \frac{e^{-(\ln x - \mu)^2/(2\sigma^2)}}{x\sigma \sqrt{2\pi}} \]
5. Heavy Tail Distributions

\[ P[X > x] \sim x^{-\alpha}, \]
Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes, by Mark E. Crovella and Azer Bestavros
1. Looking for Invariants
2. Explore Parameter Space
Change one parameter, fix the rest
Explore a wide range of values
3. Use Traces
e.g. collects traces of web sessions, video files, VoIP traffic
Use it to simulate the traffic source
But must be careful about traffic shaping and user/application adaptation.
e.g. traces collected during non-congested time should not be used to simulate congested networks.
4. publish simulator script for others to verify
Conclusion
Simulation is useful but needs to do it properly
Be careful about your simulation model: you want it to be as simple as possible, but not simpler.
Be careful about your conclusion: “A is 13.5% better than B” is probably useless.
“A is 13.5% better than B under these environment” is better but not general
Not really for quantitative results, but more for
understanding the dynamics, illustrate a point, explore unexpected behavior.