Measurement Science for Complex Information Systems


and various students
Plan for Presentation

• Introduce NIST project to develop Measurement Science for Complex Information Systems

• Show an application of measurement science to compare seven alternate congestion-control algorithms for the Internet
What are complex systems?

Large collections of interconnected components whose interactions lead to macroscopic behaviors

- Biological systems (e.g., slime molds, ant colonies, embryos)
- Physical systems (e.g., earthquakes, avalanches, forest fires)
- Social systems (e.g., transportation networks, cities, economies)
- Information systems (e.g., Internet, Web services, compute grids)
What is the problem?

No one understands how to measure, predict or control macroscopic behavior in complex information systems.

“[Despite] society’s profound dependence on networks, fundamental knowledge about them is primitive. [G]lobal communication … networks have quite advanced technological implementations but their behavior under stress still cannot be predicted reliably…. There is no science today that offers the fundamental knowledge necessary to design large complex networks [so] that their behaviors can be predicted prior to building them.”

— Network Science, NRC report released in 2006
What is the new idea?

*Leverage models and mathematics from the physical sciences* to define a systematic method to measure, understand and control macroscopic behavior in the Internet and distributed software systems built on the Internet

**Technical Approach**

- Evaluate models and analysis methods
  - Computationally tractable?
  - Reveal macroscopic behavior?
  - Establish causality?
- Evaluate distributed control techniques
  - Can economic mechanisms elicit desired behaviors?
  - Can biological mechanisms organize elements?
Previous NIST Groundwork (2000-2005)

Preliminary investigation to identify hard technical issues


Yuan and Mills, *Monitoring the Macroscopic Effects of Distributed Denial of Service (DDoS) Flooding Attacks*, October 2005


Complex Dynamics in Communications Networks, December 2005

(including *Macroscopic Dynamics in Large-Scale Data Networks* by Yuan and Mills)
Why is this hard? Why can we succeed?

<table>
<thead>
<tr>
<th>Hard Issues</th>
<th>Plausible Approaches</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1. Model scale</td>
<td>A1. Scale-reduction techniques</td>
</tr>
<tr>
<td>H2. Model validation</td>
<td>A2. Sensitivity analysis &amp; key comparisons</td>
</tr>
<tr>
<td>H3. Tractable analysis</td>
<td>A3. Cluster analysis and statistical analyses</td>
</tr>
<tr>
<td>H4. Causal analysis</td>
<td>A4. Evaluate analysis techniques</td>
</tr>
<tr>
<td>H5. Controlling behavior</td>
<td>A5. Evaluate distributed control regimes</td>
</tr>
</tbody>
</table>

Project Start Date: October 2006
Multidisciplinary Project

Disciplinary Expertise

<table>
<thead>
<tr>
<th>Modeling Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analytical</td>
</tr>
<tr>
<td>Simulation</td>
</tr>
<tr>
<td>D. Genin</td>
</tr>
<tr>
<td>C. Dabrowski</td>
</tr>
<tr>
<td>F. Hunt</td>
</tr>
<tr>
<td>K. Mills</td>
</tr>
<tr>
<td>V. Marbukh</td>
</tr>
</tbody>
</table>

Experiment Design Methods

| J. Filliben               |

Data Analysis Methods

| D. Y. Cho                |
| J. Filliben              |

Visualization Methods

| J. Hagedorn              |
| C. Houard                |

Problem Domains

Computational Grids

Internet

Problem Approaches

Fluid-Flow Modeling
Markov Modeling
Mesoscopic Modeling
Mean-Field Approximation
Perturbation Analysis
Orthogonal Fractional Factorial Design
Sensitivity Analysis
Clustering Analysis
Principal Components Analysis
Correlation Analysis
Multidimensional Data Visualization

May 20, 2009
Innovations in Measurement Science
Sample Challenge Problems Under Investigation

- Predict effect on global behavior and user experience from adopting proposed replacement congestion-control algorithms for the Internet* (Filliben, Cho, Houard & Mills)
- Evaluate accuracy of proposed fluid-flow models for TCP, characterize limits of applicability of such models and propose improved analytical models (Genin & Marbukh)
- Devise efficient Markov models to accurately simulate large-scale systems and apply perturbation analysis to predict system changes that could lead to undesired behaviors (Dabrowski & Hunt)
- Investigate the use of economic methods for resource allocation in large distributed systems (e.g., computational grids and networks) (Dabrowski, Marbukh & Mills)

*Today I use this challenge problem to illustrate some of our approaches
Sample Artifacts Produced by the Project

**MesoNet** — a medium scale network simulator that includes seven congestion control algorithms: BIC, CTCP, FAST, HSTCP, H-TCP, Scalable TCP and TCP Reno

**EconoGrid** — a detailed simulation model of a standards-based Grid compute economy

**Flexi-Cluster** — a flexible simulation model of a compute cluster that includes alternate, replaceable functions for pricing, admission control, scheduling and queue management

**Markov-Model Rewriter** — software to systematically perturb a Markov model with bounds defined by a user

**DiVisa** — software for interactive exploration of multidimensional data
Challenge Problem: Study of Proposed Replacement Congestion-Control Algorithms for the Internet

- Modeling the network
  - Parameter state-space reduction techniques
  - Response state-space reduction techniques
  - Orthogonal fractional-factorial experiment design
  - Sensitivity analysis

- Modeling congestion-control algorithms
  - Unified model with phase and procedure alignment
  - Validation against empirical measurements

- Comparing congestion-control algorithms
  - Cluster analysis
  - Detailed analysis of individual responses
  - Condition-response summary analysis
  - Causality analysis – through domain expertise
The Modeling State-Space Problem

\[ y_1, \ldots, y_m = f(k \cdot x_1, \ldots, k \cdot x_n) \]

Response State-Space \quad Stimulus State-Space

<table>
<thead>
<tr>
<th>( n )</th>
<th>Number of inputs (i.e., stimulus factors)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( k )</td>
<td>Factor range (i.e., number of values each factor can assume)</td>
</tr>
<tr>
<td>( m )</td>
<td>Number of outputs (i.e., responses)</td>
</tr>
</tbody>
</table>
Parameter State-Space Reduction

\[ k^n \]

- modeling brilliance

\[ k^{(n-r_1)} \]

- factor clustering

\[ k^{(n-r_1-r_2)} \]

- 2-level experiment design

\[ 2^{(n-r_1-r_2)} \]

- OFF (orthogonal fractional factorial) experiment design

\[ 2^{(n-r_1-r_2-r_3)} \]

SCIENTIFIC DOMAIN EXPERTISE

STATISTICAL EXPERIMENT DESIGN EXPERTISE
Parameter State-Space Reduction: MesoNet Example

$(2^{32})^{1000} \rightarrow O(10^{9633})$  

$r_1 = 944$

$(2^{32})^{56} \rightarrow O(10^{539})$

$r_2 = 36$

$(2^{32})^{20} \rightarrow O(10^{192})$

$k = 2$

$2^{20} \rightarrow O(10^6)$

$r_3 = 12$

$2^8 \rightarrow 256$

$10^{80}$ = atoms in visible universe
$2^{20-12}$ OFF Design Improves Computational Feasibility of Searching Parameter State Space

<table>
<thead>
<tr>
<th>Assumptions</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Processing Time for One Run</td>
<td>8 CPU-Hours</td>
</tr>
<tr>
<td>Number of Available CPUs</td>
<td>48</td>
</tr>
</tbody>
</table>

2-Level Experiment Design Requires 20 Years*

\[
(2^{20} \text{ runs} \times 8 \text{ CPU-Hours Per Run})/48 \text{ CPUs} = 174,762.67 \text{ Hours}
\]

OFF Experiment Design Reduces Requirement To Under 2 Days

\[
(2^8 \text{ runs} \times 8 \text{ CPU-Hours Per Run})/48 \text{ CPUs} = 42.67 \text{ Hours}
\]

*If we had 1,000 CPUs to dedicate to this problem, then we could compute the $2^{20}$ runs in just under 1 year
Response State-Space Reduction

- Principal Components Analysis
  - $m$ responses ($y_1, \ldots, y_m$)
  - $m - d1$ responses
- Correlation Analysis
  - $m - d2$ responses
- Domain Expertise
  - $m - d3$ responses

DATA ANALYSIS EXPERTISE

SCIENTIFIC DOMAIN EXPERTISE
Response State-Space Reduction: MesoNet Example

22 responses ($y_1, \ldots, y_{22}$)

Principal Components Analysis

22 – 17 responses

Correlation Analysis

22 – 15 responses

Domain Expertise

22 – 14 responses

Kevin

Jim
# 22-Dimension Response State Space from MesoNet

## Macroscopic Network Behavior

<table>
<thead>
<tr>
<th>Response</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_1$</td>
<td>Active Flows – flows attempting to transfer data</td>
</tr>
<tr>
<td>$y_2$</td>
<td>Proportion of potential flows that were active: Active Flows/All Sources</td>
</tr>
<tr>
<td>$y_3$</td>
<td>Packets entering the network per measurement interval</td>
</tr>
<tr>
<td>$y_4$</td>
<td>Packets leaving the network per measurement interval</td>
</tr>
<tr>
<td>$y_5$</td>
<td>Loss Rate: $y_4/(y_3+y_4)$</td>
</tr>
<tr>
<td>$y_6$</td>
<td>Flows Completed per measurement interval</td>
</tr>
<tr>
<td>$y_7$</td>
<td>Flow-Completion Rate: $y_6/(y_6+y_1)$</td>
</tr>
<tr>
<td>$y_8$</td>
<td>Connection Failures per measurement interval</td>
</tr>
<tr>
<td>$y_9$</td>
<td>Connection-Failure Rate: $y_8/(y_8+y_1)$</td>
</tr>
<tr>
<td>$y_{10}$</td>
<td>Retransmission Rate</td>
</tr>
<tr>
<td>$y_{11}$</td>
<td>Congestion Window per Flow</td>
</tr>
<tr>
<td>$y_{12}$</td>
<td>Window Increases per Flow per measurement interval</td>
</tr>
<tr>
<td>$y_{13}$</td>
<td>Negative Acknowledgments per Flow per measurement interval</td>
</tr>
<tr>
<td>$y_{14}$</td>
<td>Timeouts per Flow per measurement interval</td>
</tr>
<tr>
<td>$y_{15}$</td>
<td>Smoothed Round-Trip Time</td>
</tr>
<tr>
<td>$y_{16}$</td>
<td>Relative queuing delay: $y_{15}/(41*x)$</td>
</tr>
</tbody>
</table>

## User Experience

<table>
<thead>
<tr>
<th>Response</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_{17}$</td>
<td>Average Throughput for active D-D Flows</td>
</tr>
<tr>
<td>$y_{18}$</td>
<td>Average Throughput for active D-F Flows</td>
</tr>
<tr>
<td>$y_{19}$</td>
<td>Average Throughput for active D-N Flows</td>
</tr>
<tr>
<td>$y_{20}$</td>
<td>Average Throughput for active F-F Flows</td>
</tr>
<tr>
<td>$y_{21}$</td>
<td>Average Throughput for active F-N Flows</td>
</tr>
<tr>
<td>$y_{22}$</td>
<td>Average Throughput for active N-N Flows</td>
</tr>
</tbody>
</table>

May 20, 2009
Innovations in Measurement Science
Correlation Analysis Identifies 7 Dimensions

<table>
<thead>
<tr>
<th>Response</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>y4</td>
<td>Average number of packet output per measurement interval</td>
</tr>
<tr>
<td>y6</td>
<td>Average number of flows completed per measurement interval</td>
</tr>
<tr>
<td>y10</td>
<td>Average retransmission rate</td>
</tr>
<tr>
<td>y15</td>
<td>Average smoothed round-trip time</td>
</tr>
<tr>
<td>y17</td>
<td>Average instantaneous throughput for D-D flows</td>
</tr>
<tr>
<td>y20</td>
<td>Average instantaneous throughput for F-F flows</td>
</tr>
<tr>
<td>y22</td>
<td>Average instantaneous throughput for N-N flows</td>
</tr>
</tbody>
</table>
Principal Components Analysis Suggests 4 Dimensions

**PC1 – Congestion**

<table>
<thead>
<tr>
<th>Response</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1</td>
<td>Average number of active flows</td>
</tr>
<tr>
<td>y2</td>
<td>Proportion of possible flows that are active</td>
</tr>
<tr>
<td>y5</td>
<td>Loss rate</td>
</tr>
<tr>
<td>y7</td>
<td>Flow-completion rate</td>
</tr>
<tr>
<td>y10</td>
<td>Retransmission rate</td>
</tr>
<tr>
<td>y11</td>
<td>Average congestion window</td>
</tr>
<tr>
<td>y12</td>
<td>Window-increase rate</td>
</tr>
<tr>
<td>y13</td>
<td>Negative-acknowledgment rate</td>
</tr>
<tr>
<td>y14</td>
<td>Timeout rate</td>
</tr>
<tr>
<td>y19</td>
<td>Average instantaneous throughput for D-N flows</td>
</tr>
<tr>
<td>y21</td>
<td>Average instantaneous throughput for F-N flows</td>
</tr>
<tr>
<td>y22</td>
<td>Average instantaneous throughput for N-N flows</td>
</tr>
</tbody>
</table>

**PC2 – Delay**

<table>
<thead>
<tr>
<th>Response</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>y15</td>
<td>Smoothed round-trip time</td>
</tr>
<tr>
<td>y16</td>
<td>Relative queuing delay</td>
</tr>
</tbody>
</table>

**PC3 – Throughput for Advantaged Flows**

<table>
<thead>
<tr>
<th>Response</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>y3</td>
<td>Packets input</td>
</tr>
<tr>
<td>y4</td>
<td>Packets output</td>
</tr>
<tr>
<td>y17</td>
<td>Average instantaneous throughput for D-D flows</td>
</tr>
<tr>
<td>y18</td>
<td>Average instantaneous throughput for D-F flows</td>
</tr>
<tr>
<td>y20</td>
<td>Average instantaneous throughput for F-F flows</td>
</tr>
</tbody>
</table>

**PC4 – Aggregate Throughput**

<table>
<thead>
<tr>
<th>Response</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>y3</td>
<td>Packets input</td>
</tr>
<tr>
<td>y4</td>
<td>Packets output</td>
</tr>
<tr>
<td>y6</td>
<td>Flows completed per measurement interval</td>
</tr>
</tbody>
</table>
Domain Expert Selects 8 Dimensions

<table>
<thead>
<tr>
<th>Response</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1</td>
<td>Average number of active flows</td>
</tr>
<tr>
<td>y4</td>
<td>Average number of packet output per measurement interval</td>
</tr>
<tr>
<td>y6</td>
<td>Average number of flows completed per measurement interval</td>
</tr>
<tr>
<td>y10</td>
<td>Average retransmission rate</td>
</tr>
<tr>
<td>y15</td>
<td>Average smoothed round-trip time</td>
</tr>
<tr>
<td>y17</td>
<td>Average instantaneous throughput for D-D flows</td>
</tr>
<tr>
<td>y20</td>
<td>Average instantaneous throughput for F-F flows</td>
</tr>
<tr>
<td>y22</td>
<td>Average instantaneous throughput for N-N flows</td>
</tr>
</tbody>
</table>

Domain Expert Sides with the Correlation Analysis and Adds One Response (y1)
Sensitivity Analysis as a Model Validation Step

What parameters (or combinations of parameters) determine a model’s responses?

Does the influence of parameters on responses make sense to a domain expert?

Can any unexpected responses be explained and do the explanations make sense?
### 11 of 20 MesoNet Parameters Selected for Sensitivity Analysis

<table>
<thead>
<tr>
<th>Network Parameters</th>
<th>User Behavior</th>
<th>Protocol Parameters</th>
<th>Characteristics of Sources &amp; Receivers</th>
<th>Simulation Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topology</td>
<td>Think Time</td>
<td>Initial Congestion Window</td>
<td>Number of Sources</td>
<td>Measurement-Interval Size 200 ms</td>
</tr>
<tr>
<td>Propagation Delay</td>
<td></td>
<td>Initial Slow-Start Threshold</td>
<td>Distribution of Sources</td>
<td>Number of Measurement Intervals 6000</td>
</tr>
<tr>
<td>Speed</td>
<td></td>
<td>Type of Slow Start Regime</td>
<td>Number of Receivers</td>
<td>Random-Number Seed 200000</td>
</tr>
<tr>
<td>Buffer Sizing</td>
<td></td>
<td>Limited slow start</td>
<td>Distribution of Receivers</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Distribution of Network-Interface Speeds</td>
<td></td>
</tr>
</tbody>
</table>

*selected parameter
other parameters fixed

- **Network Parameters**: Topology, Propagation Delay, Speed, Buffer Sizing
- **User Behavior**: Think Time, File-Size Distribution, Pattern of Long-Lived Flows
- **Protocol Parameters**: Initial Congestion Window, Initial Slow-Start Threshold, Type of Slow Start Regime
- **Characteristics of Sources & Receivers**: Number of Sources, Distribution of Sources, Number of Receivers, Distribution of Receivers, Distribution of Network-Interface Speeds
- **Simulation Control**: Measurement-Interval Size 200 ms, Number of Measurement Intervals 6000, Random-Number Seed 200000

May 20, 2009
Innovations in Measurement Science
Representative Heterogeneous Four-Tier Topology Selected

Fast access routers

Directly-connected access routers

Normal access routers

May 20, 2009
Innovations in Measurement Science
2^{11-5} Orthogonal Fraction Factorial (OFF) Experiment Design

Full Factorial Design Requires $2^{11} = 2048$ runs

We can only afford $2^6 = 64$ runs; thus,

we require a $2^{11-5}$ OFF Experiment Design

No confounding of main effects with two-term interactions

Main effects may be confounded with three-term and higher interactions
Values Selected to Replace +1 and -1 in Design Template

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Factor</th>
<th>+1</th>
<th>-1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Network Factors</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiplier for Propagation Delay</td>
<td>x1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Backbone Router Speed (ppms)*</td>
<td>x2</td>
<td>400</td>
<td>800</td>
</tr>
<tr>
<td>Buffer Sizing Algorithm</td>
<td>x3</td>
<td>RTTxC</td>
<td>RTTxC/SQR(n)</td>
</tr>
<tr>
<td><strong>User Factors</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average File Size (packets)</td>
<td>x4</td>
<td>100</td>
<td>50</td>
</tr>
<tr>
<td>Average Think Time (ms)</td>
<td>x5</td>
<td>5000</td>
<td>2000</td>
</tr>
<tr>
<td>Probability of a Larger File*</td>
<td>x6</td>
<td>0.01</td>
<td>0.02</td>
</tr>
<tr>
<td><strong>Source &amp; Receiver Factors</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Probability of Fast Network Interface*</td>
<td>x7</td>
<td>0.2</td>
<td>0.4</td>
</tr>
<tr>
<td>Multiplier on Number Sources Per Access Router</td>
<td>x8</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Distribution of Sources</td>
<td>x9</td>
<td>P2P</td>
<td>WEB</td>
</tr>
<tr>
<td>Distribution of Receivers</td>
<td>x10</td>
<td>P2P</td>
<td>WEB</td>
</tr>
<tr>
<td><strong>Protocol Factors</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial Slow-Start Threshold (packets)</td>
<td>x11</td>
<td>1.07x10^9</td>
<td>43</td>
</tr>
</tbody>
</table>

*By convention +1 is coded with larger value and -1 is coded with smaller value. Here, coding was reversed in three cases. This makes no difference in the construction of the experiment, but must be managed during data analysis.
## Speed of All Routers Derived from Speed of Backbone Routers (Topology Tiers 1 to 3)

<table>
<thead>
<tr>
<th>Router Type</th>
<th>Parameter</th>
<th>Equation</th>
<th>+1</th>
<th>-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Backbone</td>
<td>x2</td>
<td>(= \frac{x2}{R2/R4}x)</td>
<td>400 ppms</td>
<td>800 ppms</td>
</tr>
<tr>
<td>POP</td>
<td>R2 (= 4)</td>
<td>(= \frac{x2}{R2}x)</td>
<td>100 ppms</td>
<td>200 ppms</td>
</tr>
<tr>
<td>Typical Access</td>
<td>R3 (= 10)</td>
<td>(= \frac{x2}{R2/R4}x)</td>
<td>10 ppms</td>
<td>20 ppms</td>
</tr>
<tr>
<td>Fast Access</td>
<td>FA (= 2)</td>
<td>(= \frac{x2}{R2/R4xFA}x)</td>
<td>20 ppms</td>
<td>40 ppms</td>
</tr>
<tr>
<td>Directly Connected Access</td>
<td>DC (= 10)</td>
<td>(= \frac{x2}{R2/R4xDC}x)</td>
<td>100 ppms</td>
<td>200 ppms</td>
</tr>
</tbody>
</table>
# Number & Distribution of Sources & Receivers (Tier 4)

## Sources

<table>
<thead>
<tr>
<th>x8</th>
<th>x9</th>
<th>x10</th>
<th>Total Sources</th>
<th>% under D Routers</th>
<th>% under F Routers</th>
<th>% under N Routers</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>P2P</td>
<td>P2P</td>
<td>27,800</td>
<td>4.32</td>
<td>20.14</td>
<td>75.54</td>
</tr>
<tr>
<td>3</td>
<td>P2P</td>
<td>P2P</td>
<td>41,700</td>
<td>4.32</td>
<td>20.14</td>
<td>75.54</td>
</tr>
<tr>
<td>2</td>
<td>WEB</td>
<td>WEB</td>
<td>18,560</td>
<td>6.46</td>
<td>48.27</td>
<td>45.25</td>
</tr>
<tr>
<td>3</td>
<td>WEB</td>
<td>WEB</td>
<td>27,840</td>
<td>6.46</td>
<td>48.27</td>
<td>45.25</td>
</tr>
<tr>
<td>2</td>
<td>P2P</td>
<td>WEB</td>
<td>27,800</td>
<td>4.32</td>
<td>20.14</td>
<td>75.54</td>
</tr>
<tr>
<td>3</td>
<td>P2P</td>
<td>WEB</td>
<td>41,700</td>
<td>4.32</td>
<td>20.14</td>
<td>75.54</td>
</tr>
<tr>
<td>2</td>
<td>WEB</td>
<td>P2P</td>
<td>18,560</td>
<td>6.46</td>
<td>48.27</td>
<td>45.25</td>
</tr>
<tr>
<td>3</td>
<td>WEB</td>
<td>P2P</td>
<td>27,840</td>
<td>6.46</td>
<td>48.27</td>
<td>45.25</td>
</tr>
</tbody>
</table>

## Receivers

<table>
<thead>
<tr>
<th>x8</th>
<th>x9</th>
<th>x10</th>
<th>Total Receivers</th>
<th>% under D Routers</th>
<th>% under F Routers</th>
<th>% under N Routers</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>P2P</td>
<td>P2P</td>
<td>111,200</td>
<td>4.32</td>
<td>20.14</td>
<td>75.54</td>
</tr>
<tr>
<td>3</td>
<td>P2P</td>
<td>P2P</td>
<td>166,800</td>
<td>4.32</td>
<td>20.14</td>
<td>75.54</td>
</tr>
<tr>
<td>2</td>
<td>WEB</td>
<td>WEB</td>
<td>146,400</td>
<td>2.45</td>
<td>11.47</td>
<td>86.06</td>
</tr>
<tr>
<td>3</td>
<td>WEB</td>
<td>WEB</td>
<td>219,600</td>
<td>2.45</td>
<td>11.47</td>
<td>86.06</td>
</tr>
<tr>
<td>2</td>
<td>P2P</td>
<td>WEB</td>
<td>146,400</td>
<td>2.45</td>
<td>11.47</td>
<td>86.06</td>
</tr>
<tr>
<td>3</td>
<td>P2P</td>
<td>WEB</td>
<td>219,600</td>
<td>2.45</td>
<td>11.47</td>
<td>86.06</td>
</tr>
<tr>
<td>2</td>
<td>WEB</td>
<td>P2P</td>
<td>111,200</td>
<td>4.32</td>
<td>20.14</td>
<td>75.54</td>
</tr>
<tr>
<td>3</td>
<td>WEB</td>
<td>P2P</td>
<td>166,800</td>
<td>4.32</td>
<td>20.14</td>
<td>75.54</td>
</tr>
</tbody>
</table>
Resulting Distribution of Flow Classes

<table>
<thead>
<tr>
<th>x8</th>
<th>x9</th>
<th>x10</th>
<th>% DD Flows</th>
<th>% DF Flows</th>
<th>% DN Flows</th>
<th>% FF Flows</th>
<th>% FN Flows</th>
<th>% NN Flows</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>P2P</td>
<td>P2P</td>
<td>0.186</td>
<td>1.74</td>
<td>6.52</td>
<td>4.05</td>
<td>30.43</td>
<td>57.06</td>
</tr>
<tr>
<td>3</td>
<td>P2P</td>
<td>P2P</td>
<td>0.186</td>
<td>1.74</td>
<td>6.52</td>
<td>4.05</td>
<td>30.43</td>
<td>57.06</td>
</tr>
<tr>
<td>2</td>
<td>WEB</td>
<td>WEB</td>
<td>0.159</td>
<td>1.92</td>
<td>6.67</td>
<td>5.53</td>
<td>46.74</td>
<td>38.95</td>
</tr>
<tr>
<td>3</td>
<td>WEB</td>
<td>WEB</td>
<td>0.159</td>
<td>1.92</td>
<td>6.67</td>
<td>5.53</td>
<td>46.74</td>
<td>38.95</td>
</tr>
<tr>
<td>2</td>
<td>P2P</td>
<td>WEB</td>
<td>0.106</td>
<td>0.99</td>
<td>5.57</td>
<td>2.31</td>
<td>26.00</td>
<td>65.01</td>
</tr>
<tr>
<td>3</td>
<td>P2P</td>
<td>WEB</td>
<td>0.106</td>
<td>0.99</td>
<td>5.57</td>
<td>2.31</td>
<td>26.00</td>
<td>65.01</td>
</tr>
<tr>
<td>2</td>
<td>WEB</td>
<td>P2P</td>
<td>0.279</td>
<td>3.38</td>
<td>6.83</td>
<td>9.72</td>
<td>45.58</td>
<td>34.18</td>
</tr>
<tr>
<td>3</td>
<td>WEB</td>
<td>P2P</td>
<td>0.279</td>
<td>3.38</td>
<td>6.83</td>
<td>9.72</td>
<td>45.58</td>
<td>34.18</td>
</tr>
</tbody>
</table>

Flow classes defined by relative locations of source & receiver
Influence of other Factors on Average Buffer Sizes

<table>
<thead>
<tr>
<th>x1</th>
<th>x2</th>
<th>Backbone Router Buffers (avg.)</th>
<th>POP Router Buffers (avg.)</th>
<th>Access Router Buffers (avg.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>400</td>
<td>16277</td>
<td>4070</td>
<td>647</td>
</tr>
<tr>
<td>2</td>
<td>400</td>
<td>32553</td>
<td>8139</td>
<td>1294</td>
</tr>
<tr>
<td>1</td>
<td>800</td>
<td>32553</td>
<td>8139</td>
<td>1294</td>
</tr>
<tr>
<td>2</td>
<td>800</td>
<td>65106</td>
<td>16277</td>
<td>2588</td>
</tr>
</tbody>
</table>

$RTT \times C$

<table>
<thead>
<tr>
<th>x1</th>
<th>x2</th>
<th>Backbone Router Buffers (avg.)</th>
<th>POP Router Buffers (avg.)</th>
<th>Access Router Buffers (avg.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>400</td>
<td>182</td>
<td>68</td>
<td>27</td>
</tr>
<tr>
<td>2</td>
<td>400</td>
<td>364</td>
<td>135</td>
<td>53</td>
</tr>
<tr>
<td>1</td>
<td>800</td>
<td>364</td>
<td>135</td>
<td>53</td>
</tr>
<tr>
<td>2</td>
<td>800</td>
<td>728</td>
<td>270</td>
<td>105</td>
</tr>
</tbody>
</table>

$RTT \times C / SQR(n)$
Sensitivity Analysis Relies on Main Effects Plots

- Mean Response Y (Congestion Window Per Connection (average))

Plot with data points representing the mean response Y for different conditions.

- Factors:
  - PDM, BRS, QSA, AvFSWO, AvTh, PrLF, PrFH, SFSR, SDist, RDist, SST

- Data:
  - k = 11
  - n = 64
  - Various values for each factor indicating the response.
Sensitivity Analysis Driven by Correlation Analysis

Active Flows
Retransmission Rate
Avg. TP NN Flows
SRTT
Aggregate TP
Flow Completion Rate
Avg. TP DD Flows
Avg. TP FF Flows

Seven Responses Chosen by Correlation Analysis + 1
Major Factors Influencing Model Behavior

(1) Network Speed
(2) File Size
(3) Think Time
(4) Number of Sources
(5) Propagation Delay
(6) Distribution of Sources

(7) Buffer Size – small buffer size reduces delay variability & larger buffer size has greater effect under high network speed

Correlation-based Analysis

<table>
<thead>
<tr>
<th></th>
<th>x1</th>
<th>x2</th>
<th>x3</th>
<th>x4</th>
<th>x5</th>
<th>x6</th>
<th>x7</th>
<th>x8</th>
<th>x9</th>
<th>x10</th>
<th>x11</th>
</tr>
</thead>
<tbody>
<tr>
<td>y4</td>
<td>9.5</td>
<td>1</td>
<td>8</td>
<td>3</td>
<td>2</td>
<td>9.5</td>
<td>11</td>
<td>4</td>
<td>5</td>
<td>6.5</td>
<td>6.5</td>
</tr>
<tr>
<td>y6</td>
<td>11</td>
<td>1</td>
<td>6</td>
<td>2</td>
<td>3</td>
<td>9</td>
<td>10</td>
<td>4</td>
<td>5</td>
<td>7.5</td>
<td>7.5</td>
</tr>
<tr>
<td>y10</td>
<td>7</td>
<td>1.5</td>
<td>1.5</td>
<td>5</td>
<td>5</td>
<td>10</td>
<td>10</td>
<td>5</td>
<td>3</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>y15</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>8</td>
<td>10</td>
<td>6</td>
<td>7</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>y17</td>
<td>1</td>
<td>8</td>
<td>10.5</td>
<td>2</td>
<td>8</td>
<td>5</td>
<td>10</td>
<td>3</td>
<td>10.5</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>y20</td>
<td>1</td>
<td>3</td>
<td>8</td>
<td>5</td>
<td>4</td>
<td>10</td>
<td>11</td>
<td>6</td>
<td>2</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>Average Rank</td>
<td>5.25</td>
<td>2.92</td>
<td>5.83</td>
<td>3.5</td>
<td>4.5</td>
<td>8.58</td>
<td>9.50</td>
<td>4.67</td>
<td>5.42</td>
<td>8.17</td>
<td>7.67</td>
</tr>
<tr>
<td>Ordinal Rank</td>
<td>5</td>
<td>1</td>
<td>7</td>
<td>2</td>
<td>3</td>
<td>10</td>
<td>11</td>
<td>4</td>
<td>6</td>
<td>9</td>
<td>8</td>
</tr>
</tbody>
</table>

Principal Components-based Analysis

<table>
<thead>
<tr>
<th></th>
<th>x1</th>
<th>x2</th>
<th>x3</th>
<th>x4</th>
<th>x5</th>
<th>x6</th>
<th>x7</th>
<th>x8</th>
<th>x9</th>
<th>x10</th>
<th>x11</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC1</td>
<td>10.5</td>
<td>1</td>
<td>6</td>
<td>5</td>
<td>2</td>
<td>10.5</td>
<td>9</td>
<td>4</td>
<td>3</td>
<td>7.5</td>
<td>7.5</td>
</tr>
<tr>
<td>PC2</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>7</td>
<td>3</td>
<td>8.5</td>
<td>8.5</td>
<td>4</td>
<td>11</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>PC3</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>6</td>
<td>8</td>
<td>10.5</td>
<td>7</td>
<td>4</td>
<td>9</td>
<td>10.5</td>
</tr>
<tr>
<td>PC4</td>
<td>5</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>10</td>
<td>11</td>
<td>7</td>
<td>9</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>Average Rank</td>
<td>5.13</td>
<td>2.50</td>
<td>4.00</td>
<td>3.25</td>
<td>9.25</td>
<td>9.75</td>
<td>5.50</td>
<td>6.75</td>
<td>8.63</td>
<td>7.25</td>
<td></td>
</tr>
<tr>
<td>Ordinal Rank</td>
<td>5</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>10</td>
<td>11</td>
<td>6</td>
<td>7</td>
<td>9</td>
<td>8</td>
</tr>
</tbody>
</table>
**Unified Model for Congestion-Control Algorithms**

**LIFE OF A TCP FLOW**

---

**CONNECTION PHASE**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>synint</td>
<td>Timeout interval for initial SYN</td>
</tr>
<tr>
<td>synmax</td>
<td>Maximum number of SYNs to send</td>
</tr>
<tr>
<td>synsent</td>
<td>Number of SYNs that have been sent</td>
</tr>
<tr>
<td>synmo</td>
<td>Timeout for current SYN</td>
</tr>
<tr>
<td>time</td>
<td>Current time</td>
</tr>
</tbody>
</table>

**Initiate Connection**

\[
\begin{align*}
\text{SYNMAX} & \leftarrow 3 \\
\text{SYNINT} & \leftarrow 3 \ s \\
\text{SYNTO} & \leftarrow \text{time} + \text{SYNINT} \\
\text{SYNSENT} & \leftarrow 1 \\
\text{send}(\text{SYN}) & \\
\end{align*}
\]

**Timeout**

\[
\begin{align*}
\text{Timeout} & = \begin{cases} \\
\text{SYNSENT} < \text{SYNMAX} & \\
\text{SYNINT} & \leftarrow 2 \times \text{SYNINT} \\
\text{SYNTO} & \leftarrow \text{time} + \text{SYNINT} \\
\text{SYNSENT} & \leftarrow \text{SYNSENT} + 1 \\
\text{send}(\text{SYN}) \\
\text{signal}(	ext{ConnectionFailure}) & \text{otherwise} \\
\end{cases}
\end{align*}
\]

---

**TRANSFER PHASE**

**SLOW START**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>cwnd</td>
<td>Current congestion window</td>
</tr>
<tr>
<td>cwndmax</td>
<td>Initial congestion window (we use (\text{cwnd}_{\text{max}} = 3))</td>
</tr>
<tr>
<td>act</td>
<td>Current slow-start threshold</td>
</tr>
<tr>
<td>instart</td>
<td>Threshold to switch from exponential to logarithmic increase (varies with experiment)</td>
</tr>
<tr>
<td>instmax</td>
<td>Threshold to terminate initial slow start (varies with experiment)</td>
</tr>
</tbody>
</table>

**Initiate Transfer Phase**

\[
\text{cwnd} \leftarrow \text{cwnd}_{\text{INT}} \\
\text{ssl} \leftarrow \text{ssl}_{\text{INT}} \\
\]

\[
\text{ACK} \cdot \left(\text{seq} < \text{cwnd} \right) \cdot \\
\begin{cases} \\
\text{send} = \text{cwnd} + 1 & \text{if } \text{cwnd} < \text{cwnd}_{\text{max}} \\
\text{send} = \text{cwnd} + \left\lfloor \frac{1}{\text{instmax}} \right\rfloor & \text{otherwise} \\
\end{cases}
\]

**CONGESTION AVOIDANCE**

Window increase procedures (ACK)

Window decrease procedures (Loss)

Timeout procedures

Optional periodic procedures:

CTCP, FAST, H-TCP

Optional mode switch between TCP & alternate procedures:

BIC, CTCP, HSTCP, H-TCP, Scalable TCP

---

May 20, 2009
Innovations in Measurement Science
Sample: CTCP Congestion-Avoidance Model

### Mode Switch Procedures

Select Procedures =

\[ \begin{align*}
\text{TCPcongestionAvoidance} & \quad \text{if } \text{cwnd} < \text{LW}_C \\
\text{TCPcongestionAvoidance} & \quad \text{otherwise}
\end{align*} \]

### Increase Procedures

**ACK** = \[ \frac{\text{cwnd}}{\text{cwnd + dwnd}} + \frac{1}{\text{cwnd + dwnd}} \]
\[ \text{cwnd} \leftarrow \text{cwnd + dwnd} \]

### Decrease Procedures

**Loss** = \[ \frac{\text{cwnd} + \text{dwnd}}{2} \]
\[ \text{CD}_C \leftarrow \text{true} \]
\[ \text{sst} \leftarrow \text{cwnd} \]

### Timeout Procedures

**Timeout** = \[ \text{sst} \leftarrow \max \left( \frac{\text{cwnd + dwnd}}{2}, \text{cwnd}_\text{INT} \right) \]
\[ \text{cwnd} \leftarrow \text{cwnd}_\text{INT} \]
\[ \text{dwnd} \leftarrow 0 \]
\[ \text{CD}_C \leftarrow \text{true} \]
Empirical Studies of Six Alternate Congestion-Control Algorithms

   - Covers BIC, FAST, HSTCP, H-TCP, Scalable TCP
   - Uses implementations within Linux kernel

   - Covers CTCP and TCP Illinois
   - Uses CTCP implementation in Windows® VISTA and TCP Illinois implementation within Linux kernel

PARAMETERS
- Number of flows sharing path
- Start time of each flow
- Bottleneck bandwidth
- Round-trip propagation delay
- Number of buffers
Simulated Behavior of MesoNet CTCP Model

MesoNet CTCP simulation behavior agrees with empirical results

Other MesoNet congestion-control algorithms also agree with empirical results

May 20, 2009
Innovations in Measurement Science
Comparing Alternate Congestion-Control Algorithms in a Large (up to 278,000 sources), Fast (up 192 Gbps) Homogeneous Network

7 Algorithms

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Label</th>
<th>Name of Congestion-Avoidance Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>BIC</td>
<td>Binary Increase Congestion Control</td>
</tr>
<tr>
<td>2</td>
<td>CTCP</td>
<td>Compound Transmission Control Protocol</td>
</tr>
<tr>
<td>3</td>
<td>FAST</td>
<td>Fast Active-Queue Management Scalable Transmission Control Protocol</td>
</tr>
<tr>
<td>4</td>
<td>HSTCP</td>
<td>High-Speed Transmission Control Protocol</td>
</tr>
<tr>
<td>5</td>
<td>HTCP</td>
<td>Hamilton Transmission Control Protocol</td>
</tr>
<tr>
<td>6</td>
<td>Scalable</td>
<td>Scalable Transmission Control Protocol</td>
</tr>
<tr>
<td>7</td>
<td>TCP</td>
<td>Transmission Control Protocol (Reno)</td>
</tr>
</tbody>
</table>

3 Path Classes

<table>
<thead>
<tr>
<th>Path Class</th>
<th>Flow Type</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Fast</td>
<td>DD</td>
<td>Source &amp; receiver under directly connected access router</td>
</tr>
<tr>
<td>Fast</td>
<td>DF</td>
<td>Source or receiver under directly connected access router and correspondent under fast access router</td>
</tr>
<tr>
<td></td>
<td>FF</td>
<td>Source &amp; receiver under fast access router</td>
</tr>
<tr>
<td>Typical</td>
<td>DN</td>
<td>Source or receiver under directly connected access router and correspondent under normal access router</td>
</tr>
<tr>
<td></td>
<td>FN</td>
<td>Source or receiver under fast access router and correspondent under normal access router</td>
</tr>
<tr>
<td></td>
<td>NN</td>
<td>Source &amp; receiver under normal access router</td>
</tr>
</tbody>
</table>
### Input Factors & Network Parameters

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Definition</th>
<th>PLUS (+1) Value</th>
<th>Minus (-1) Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>x1</td>
<td>Network Speed</td>
<td>8000</td>
<td>4000</td>
</tr>
<tr>
<td>x2</td>
<td>Think Time</td>
<td>5000</td>
<td>2500</td>
</tr>
<tr>
<td>x3</td>
<td>Source Distribution</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x4</td>
<td>Propagation Delay</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>x5</td>
<td>File Size</td>
<td>100</td>
<td>50</td>
</tr>
<tr>
<td>x6</td>
<td>Buffer Sizing Algorithm</td>
<td>RTT x Capacity</td>
<td>RTT x Capacity/SQR(N)</td>
</tr>
</tbody>
</table>

### Router Speeds

- **Backbone**: 192 Gbps (+1) / 96 Gbps (-1)
- **POP**: 24 Gbps (+1) / 12 Gbps (-1)
- **Normal Access**: 2.4 Gbps (+1) / 1.2 Gbps (-1)
- **Fast Access**: 4.8 Gbps (+1) / 2.4 Gbps (-1)
- **Directly Connected Access**: 24 Gbps (+1) / 12 Gbps (-1)

### Propagation Delays (ms)

- **PLUS (+1)**: Max 100, Avg 41, Min 6
- **Minus (-1)**: Max 200, Avg 100

### Buffer Sizes (packets)

<table>
<thead>
<tr>
<th>Router</th>
<th>Min</th>
<th>Avg</th>
<th>Max</th>
<th>Min</th>
<th>Avg</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Backbone</strong></td>
<td>325,528</td>
<td>732,437</td>
<td>1,302,110</td>
<td>1,153</td>
<td>2,606</td>
<td>4,654</td>
</tr>
<tr>
<td><strong>POP</strong></td>
<td>40,691</td>
<td>91,555</td>
<td>162,764</td>
<td>221</td>
<td>505</td>
<td>908</td>
</tr>
<tr>
<td><strong>Access</strong></td>
<td>6,470</td>
<td>14,557</td>
<td>25,879</td>
<td>91</td>
<td>207</td>
<td>369</td>
</tr>
</tbody>
</table>

### Number of Sources

- **PLUS (+1)**: 278,000
- **Minus (-1)**: 174,600

---

6 Robustness Factors

Router Speeds

Propagation Delays (ms)

Buffer Sizes (packets)
Simulation Scenario & Long-Lived Flows

Scenario

Start

10 mins.

Warm up Period

15 mins.

Time Period 1

Record Selected Totals

20 mins.

Time Period 2

25 mins.

Time Period 3

Long-Lived Flows

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Definition</th>
<th>Source Router</th>
<th>Receiver Router</th>
<th>Start Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>Long-distance flow</td>
<td>B0a</td>
<td>K0a</td>
<td>0.4 x 25 mins.</td>
</tr>
<tr>
<td>L2</td>
<td>Medium-distance flow</td>
<td>C0a</td>
<td>I0a</td>
<td>0.4 x 25 mins.</td>
</tr>
<tr>
<td>L3</td>
<td>Short-distance flow</td>
<td>E0a</td>
<td>F0a</td>
<td>0.4 x 25 mins.</td>
</tr>
</tbody>
</table>
### 32 Conditions Simulated (2^6-1 OFF Design)

<table>
<thead>
<tr>
<th>Condition</th>
<th>X1</th>
<th>X2</th>
<th>X3</th>
<th>X4</th>
<th>X5</th>
<th>X6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4000</td>
<td>2500</td>
<td>.1/.6/3</td>
<td>1</td>
<td>50</td>
<td>RTTxCapacity/SQR(N)</td>
</tr>
<tr>
<td>2</td>
<td>8000</td>
<td>2500</td>
<td>.1/.6/3</td>
<td>1</td>
<td>50</td>
<td>RTTxCapacity</td>
</tr>
<tr>
<td>3</td>
<td>4000</td>
<td>5000</td>
<td>.1/.6/3</td>
<td>1</td>
<td>50</td>
<td>RTTxCapacity</td>
</tr>
<tr>
<td>4</td>
<td>8000</td>
<td>5000</td>
<td>.1/.6/3</td>
<td>1</td>
<td>50</td>
<td>RTTxCapacity/SQR(N)</td>
</tr>
<tr>
<td>5</td>
<td>4000</td>
<td>2500</td>
<td>.3/.3/.3</td>
<td>1</td>
<td>50</td>
<td>RTTxCapacity</td>
</tr>
<tr>
<td>6</td>
<td>8000</td>
<td>2500</td>
<td>.3/.3/.3</td>
<td>1</td>
<td>50</td>
<td>RTTxCapacity</td>
</tr>
<tr>
<td>7</td>
<td>4000</td>
<td>5000</td>
<td>.3/.3/.3</td>
<td>1</td>
<td>50</td>
<td>RTTxCapacity/SQR(N)</td>
</tr>
<tr>
<td>8</td>
<td>8000</td>
<td>5000</td>
<td>.3/.3/.3</td>
<td>1</td>
<td>50</td>
<td>RTTxCapacity</td>
</tr>
<tr>
<td>9</td>
<td>4000</td>
<td>2500</td>
<td>.1/.6/3</td>
<td>2</td>
<td>50</td>
<td>RTTxCapacity</td>
</tr>
<tr>
<td>10</td>
<td>8000</td>
<td>2500</td>
<td>.1/.6/3</td>
<td>2</td>
<td>50</td>
<td>RTTxCapacity/SQR(N)</td>
</tr>
<tr>
<td>11</td>
<td>4000</td>
<td>5000</td>
<td>.1/.6/3</td>
<td>2</td>
<td>50</td>
<td>RTTxCapacity</td>
</tr>
<tr>
<td>12</td>
<td>8000</td>
<td>5000</td>
<td>.1/.6/3</td>
<td>2</td>
<td>50</td>
<td>RTTxCapacity</td>
</tr>
<tr>
<td>13</td>
<td>4000</td>
<td>2500</td>
<td>.3/.3/.3</td>
<td>2</td>
<td>50</td>
<td>RTTxCapacity/SQR(N)</td>
</tr>
<tr>
<td>14</td>
<td>8000</td>
<td>2500</td>
<td>.3/.3/.3</td>
<td>2</td>
<td>50</td>
<td>RTTxCapacity</td>
</tr>
<tr>
<td>15</td>
<td>4000</td>
<td>5000</td>
<td>.3/.3/.3</td>
<td>2</td>
<td>50</td>
<td>RTTxCapacity</td>
</tr>
<tr>
<td>16</td>
<td>8000</td>
<td>5000</td>
<td>.3/.3/.3</td>
<td>2</td>
<td>50</td>
<td>RTTxCapacity/SQR(N)</td>
</tr>
<tr>
<td>17</td>
<td>4000</td>
<td>2500</td>
<td>.1/.6/3</td>
<td>1</td>
<td>100</td>
<td>RTTxCapacity</td>
</tr>
<tr>
<td>18</td>
<td>8000</td>
<td>2500</td>
<td>.1/.6/3</td>
<td>1</td>
<td>100</td>
<td>RTTxCapacity/SQR(N)</td>
</tr>
<tr>
<td>19</td>
<td>4000</td>
<td>5000</td>
<td>.1/.6/3</td>
<td>1</td>
<td>100</td>
<td>RTTxCapacity</td>
</tr>
<tr>
<td>20</td>
<td>8000</td>
<td>5000</td>
<td>.1/.6/3</td>
<td>1</td>
<td>100</td>
<td>RTTxCapacity</td>
</tr>
<tr>
<td>21</td>
<td>4000</td>
<td>2500</td>
<td>.3/.3/.3</td>
<td>1</td>
<td>100</td>
<td>RTTxCapacity/SQR(N)</td>
</tr>
<tr>
<td>22</td>
<td>8000</td>
<td>2500</td>
<td>.3/.3/.3</td>
<td>1</td>
<td>100</td>
<td>RTTxCapacity</td>
</tr>
<tr>
<td>23</td>
<td>4000</td>
<td>5000</td>
<td>.3/.3/.3</td>
<td>1</td>
<td>100</td>
<td>RTTxCapacity</td>
</tr>
<tr>
<td>24</td>
<td>8000</td>
<td>5000</td>
<td>.3/.3/.3</td>
<td>1</td>
<td>100</td>
<td>RTTxCapacity/SQR(N)</td>
</tr>
<tr>
<td>25</td>
<td>4000</td>
<td>2500</td>
<td>.1/.6/3</td>
<td>2</td>
<td>100</td>
<td>RTTxCapacity</td>
</tr>
<tr>
<td>26</td>
<td>8000</td>
<td>2500</td>
<td>.1/.6/3</td>
<td>2</td>
<td>100</td>
<td>RTTxCapacity</td>
</tr>
<tr>
<td>27</td>
<td>4000</td>
<td>5000</td>
<td>.1/.6/3</td>
<td>2</td>
<td>100</td>
<td>RTTxCapacity</td>
</tr>
<tr>
<td>28</td>
<td>8000</td>
<td>5000</td>
<td>.1/.6/3</td>
<td>2</td>
<td>100</td>
<td>RTTxCapacity/SQR(N)</td>
</tr>
<tr>
<td>29</td>
<td>4000</td>
<td>2500</td>
<td>.3/.3/.3</td>
<td>2</td>
<td>100</td>
<td>RTTxCapacity</td>
</tr>
<tr>
<td>30</td>
<td>8000</td>
<td>2500</td>
<td>.3/.3/.3</td>
<td>2</td>
<td>100</td>
<td>RTTxCapacity/SQR(N)</td>
</tr>
<tr>
<td>31</td>
<td>4000</td>
<td>5000</td>
<td>.3/.3/.3</td>
<td>2</td>
<td>100</td>
<td>RTTxCapacity</td>
</tr>
<tr>
<td>32</td>
<td>8000</td>
<td>5000</td>
<td>.3/.3/.3</td>
<td>2</td>
<td>100</td>
<td>RTTxCapacity</td>
</tr>
</tbody>
</table>
Processor Time Requirements
(Units are CPU days)

<table>
<thead>
<tr>
<th>Compute Servers ws11-ws14</th>
<th>Compute Servers ws9-ws10</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIC</td>
<td>CTCP</td>
</tr>
<tr>
<td>CPU time (32 runs)</td>
<td>91.5</td>
</tr>
<tr>
<td>Avg. CPU time (per run)</td>
<td>2.86</td>
</tr>
<tr>
<td>Min. CPU time (one run)</td>
<td>1.16</td>
</tr>
<tr>
<td>Max. CPU time (one run)</td>
<td>5.94</td>
</tr>
</tbody>
</table>

Experiment required about 15 days of wall-clock time spread over 48 processors

Simulated Workload

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Flows Completed</th>
<th>Data Packets Sent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. Per Condition</td>
<td>74,033,116</td>
<td>6,912,373,746</td>
</tr>
<tr>
<td>Min. Per Condition</td>
<td>40,966,013</td>
<td>3,146,870,571</td>
</tr>
<tr>
<td>Max. Per Condition</td>
<td>154,914,953</td>
<td>11,917,420,154</td>
</tr>
<tr>
<td>Total All Runs</td>
<td>16,583,418,069</td>
<td>1,548,371,719,084</td>
</tr>
</tbody>
</table>
Characterizing the 32 Conditions Simulated

16 Uncongested and 16 Congested Conditions

Max. = 0.0018

Evolution of Flow States in Uncongested Condition 4

Evolution of Flow States in Congested Condition 5

Cluster Analysis Annotated with Congestion Level
Cluster Analyses – Algorithm 3 (FAST) Stands Out

Time Period 1

Time Period 2

Time Period 3

Aggregate Responses over 25 minutes

May 20, 2009
Innovations in Measurement Science
Approach to Detailed Analysis of Individual Responses

Plot Character = Algorithm
(Min,Max) Raw Response for Y6 = (0,0.520089)

Time Period 1
Retransmission Rate
# Approach to Summarizing Detailed Analyses of Responses

<table>
<thead>
<tr>
<th>Condition</th>
<th>Response Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Condition-Response Summary Time Period 1**

- Statistically significant and >10% Effect Filter
Condition-Response Summaries – Algorithm 3 (FAST) Stands Out

Time Period 1 – 10% Filter

Time Period 2 – 30% Filter

Time Period 3 – 30% Filter

Aggregate Responses – No Filter
Why does Algorithm 3 (FAST) Stand Out?

FAST does not respond well to congestion, where too many flows compete for insufficient buffers – leading to rapid oscillation in flow congestion windows.

Evolution of congestion window under FAST for long-lived flow L2 during 500 measurement intervals within Time Period 2 under (the most congested) condition 21.
Other Congestion-Control Algorithms Adjust Less Rapidly

Evolution of congestion window under other congestion-control algorithms for long-lived flow L2 during the same 500 measurement intervals within Time Period 2 under (the most congested) condition 21
Summary of FAST Behavior under Congestion

• Rapid oscillatory adjustment of congestion window leads to:
  – Larger congestion-window increase rate
  – Higher retransmission rate (including for SYNs)
  – Larger number of flows pending in the connecting state

• Practical implications include:
  – Flows take longer to connect
  – Flows take longer to complete
  – Goodput is lower for flows transiting congested areas
  – Fewer \((10^5 \text{ to } 10^7, \text{ depending on condition})\) flows complete in a 25 minute period
Future Work on Congestion-Control Algorithms

• Does MesoNet reveal the same behavior when modeling a smaller, slower network with a much lower initial slow-start threshold?

• How do the congestion-control algorithms compare in a relatively uncongested, heterogeneous network with a wider range of traffic classes (e.g., web objects, documents, service packs and movie downloads)?

• Does the sensitivity analysis change when considering all 20 MesoNet parameters?
Summary of Presentation

• Introduced NIST project to develop Measurement Science for Complex Information Systems

• Showed an application of measurement science to compare seven alternate congestion-control algorithms for the Internet

• Identified a potential for FAST algorithm to behave undesirably under congested conditions
  – Explained the root cause for the potential undesirable behavior
  – Explained why other congestion-control algorithms are not likely to exhibit the same potential undesirable behavior
Backup Information
Interactive Exploration of Multidimensional Data

Download from – http://math.nist.gov/mcsd/savg/software/divisa/
Sample Papers Produced by the Project (1 of 2)


Sample Papers Produced by the Project (2 of 2)


Sensitivity Analysis Driven by Principal Components Analysis

PC1 - Congestion

PC2 - Delay

PC3 - Avg. Throughput on Advantaged Flows

PC4 - Aggregate Throughput