Dependability Benchmarking of Network-Based Applications Viewed as Testing of End-User Quality of Service

Mladen A. Vouk

http://renoir.csc.ncsu.edu/Faculty/Vouk
Mladen A. Vouk
Professor
Department of Computer Science,
Box 8206
North Carolina State University
Raleigh, NC 27695, U.S.A.
Tel: 919-515-7886, FAX: 919-515-6497
vouk@csc.ncsu.edu

Software Engineering Laboratory
Computer-Based Education Laboratory
Multimedia Laboratory
Center for Advanced Computing and Communications
Support

- NSF
- NCSU CACC, and other NCSU sources
- IBM
- Alcatel
- Bell South
End-User: Can I “Depend” on this application, system, net... to support in an appliance-like fashion what I do?
Key Trick

• Eliminate or standardize “non-essential” elements of the environment in which the software operates

• Build a “standard” benchmarking environment.

• Specific example - network-based software applications.
Questions

• **What is an End-User view?**
• **What is a workflow-based view (scenarios)?**
• **What is an appliance-based view?**
• **What is QoS?**
• **What is End-User QoS?**
• **What is an SLA and SLS?**

• Some answers and examples...
End-User Quality Hyper-Surface

- Throughput
- "Ease of Use" Appliance-Like
- Security
- Support
- Availability of Content
- User needs and Fault-Tolerance?
- Availability of System
- Response
- Services, Functions, ...

North Carolina State University
Department of Computer Science
Framework: Benchmark + User Elements

- **End-User Workflows**: E.g., Professor, University, ISP ...
- **SLA/SLS**: E.g., problem solving environment, DB
- **Application**: E.g., CORBA-based
- **Middleware**: E.g., IP
- **OS**: E.g., SONET, ATM, “Light” ..
- **Com Protocols**: E.g., IP
- **“Wire” Infrastructure**: E.g., IP
“Standard” Elements + User’s Elements

- User benchmarks (profile, what constitutes a failure for this user, etc.)
- SLA/SLS benchmarks
- Application benchmarks
- Middleware benchmarks
- OS benchmarks
- Network and Infrastructure benchmarks
Questions

• What is an End-User view?
• What is a workflow-based view (scenarios)?
• What is an appliance-based view?
• What is QoS?
• What is End-User QoS?
• What is an SLA and SLS?

• Some answers and examples.
Network-Based Education

• Modern learning environments are integrated collections of cooperating workflows, programs, tools, clients, and intelligent agents.
• They are distributed.
• User satisfaction is critical.
(Education) Workflows

• A series of (user-level) structured activities and computations that arise in education, training and learning, in general.
Cooperating Workflows

• Education workflows are expected to coexist, cooperate and even meld with other user workflows (e.g., business workflows, scientific workflows, legislative workflows).

• As such they must support compatible interfaces and expectations about quality of the delivered services.
“Horizontal” Workflows

- Quality of Service Needs
- Many Other Issues
- Assessment
- Research
- Legal
- Workflow Integration
- Education

Users

+ Business

+ Legal

Legal

Workflow Integration

Education

Many Other Issues
Operational Profile

System Software Developer
Skill: S/W Expert

Application Package Developer (Author)
Skill: Pedagogical and Content Expert

Package Selector (Instructor)
Needed Skill: Knowledge of User Needs

Naive User
Needed Skill: Can Find Terminal

Advanced User
Needed Skill: Uses SuperComputers

Support and Generation of Information

Feedback and Error Reports

10
10^3
10^5
10^7
# Flow Modes

<table>
<thead>
<tr>
<th>Synchronous Capture</th>
<th>Synchronous Delivery</th>
<th>Asynchronous Delivery</th>
</tr>
</thead>
<tbody>
<tr>
<td>Live 1 or 2-way interactions</td>
<td>Live recording, On-demand playback</td>
<td></td>
</tr>
<tr>
<td>“Tapes,” “Live-like” broadcast</td>
<td>“Tapes,” On-demand playback</td>
<td></td>
</tr>
</tbody>
</table>
Major Support Modes

• System Development
• Courseware Development and Management - Author Mode
• Curriculum Development and Management - Instructor Mode
• Auto-Adaptive Teaching
• Centralized Administration and Control
• Local and Remote Storage and Processing
• End-User Quality of Service Engines
• Networking (the “Glue”)

Interaction/Delivery Models

- **TV model**
  - Full motion video (e.g., MPEG2)
  - Possible two-way interaction
  - Usually high overhead

- **Data model**
  - Text, graphics, voice, possibly compressed video
  - Interaction and collaboration
  - Low to medium overhead
When Audio Only?

• In many situations about 80-95% of relevant information is captured through text, voice, and low-bandwidth interactions and animations.

• Full motion video approach may be needed only infrequently, but when it is needed it is essential.
When Video?

• Interactions dependent on interpretation of facial expressions, body language, color, ...
• Very complex moves, demos, ...
• Exploratory data analysis and model building
• Lack of knowledge of exact learning model
• Art
• “Social” interactions
• …
Constraints

• User Needs
• Human capabilities
• Technology
• Financial limitations
• Timely delivery
• Social
• Other ...
Matching

• **Student and system capabilities must be explicitly matched.**

• **In the future education workflows will need to at least match that of other workflows with which they interact.**

• **We use existing information about other workflows to estimate user-acceptable bounds for ET-Workflow support systems (e.g., end-user QoS models).**
Human Capacity

• User interaction and ability to absorb knowledge is a strong function of computer-human interface (CHI), e.g. user cannot extract new information (learn and remember) faster than about 20 bits/second (i.e., differentiate among ~1,000,000 "symbols" each second).

• Effective information transfer rates may require different sets of "symbols" and presentation rates (from several thousand bits per second to many megabits per second).
Questions

• What is an End-User view?
• What is a workflow-based view (scenarios)?
• What is an appliance-based view?
• What is QoS?
• What is End-User QoS?
• What is an SLA and SLS?

• Some answers and examples ...
Appliance-Like

• Should not impact existing workflows in terms of attention and additional operational workload - aid not distract (e.g., as easy to use a whiteboard or a blackboard)

• Seamless integration of user-workflows and technology - business model, effort-model, end-user fault-tolerance
Questions

• What is an End-User view?
• What is a workflow-based view (scenarios)?
• What is an appliance-based view?
• What is QoS?
• What is End-User QoS?
• What is an SLA and SLS?

• Some answers and examples ...
Quality of Service (QoS)

• “Classical” (network-based)
  - Throughput, delay, loss, jitter

• Extended (user-based)
  - Reliability & Availability
  - End-User Response Time
  - Scalability
  - Usability
  - Functionality, Interoperability
  - SLA support, Security...
  - Other ...
End-User QoS Hyper-Surface

- Throughput
- Jitter
- Reliability + Downtime
- Response
Reliability and Availability

• Once a user starts a ONE HOUR session, the probability getting through that session without any problems must be above 0.95

• Criterion is based on the NovaNET experiences.
UnAvailability


System Usage (Months)

System UnAvailability

Average

Instantaneous

Model
Response Times

- End-to-end response delay can be a big problem in an education environment.

- Synchronous end-to-end interaction (round-trip) delays that consistently exceed about 250 ms are often unacceptable when the interaction is conducted in the key-stroke mode.
# Network Delays (Summer ‘95)

<table>
<thead>
<tr>
<th>Network</th>
<th>Traffic Load</th>
<th>Probability that Response Time is</th>
<th>Good</th>
<th>Acceptable</th>
<th>Poor*</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCSU</td>
<td>Low</td>
<td>0.9963</td>
<td>0.0020</td>
<td>0.0017</td>
<td></td>
</tr>
<tr>
<td>Campus</td>
<td>Medium</td>
<td>0.9889</td>
<td>0.0054</td>
<td>0.0057</td>
<td></td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>0.9566</td>
<td>0.0356</td>
<td>0.0078</td>
<td></td>
</tr>
<tr>
<td>Internet</td>
<td>Low</td>
<td>0.9682</td>
<td>0.0176</td>
<td>0.0142</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>0.9502</td>
<td>0.0130</td>
<td>0.0368</td>
<td></td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>0.7187</td>
<td>0.0458</td>
<td>0.2355</td>
<td></td>
</tr>
</tbody>
</table>

(*) Includes lost packets.
Throughput

• Principal driver is the problem solving workflow.

• Usually takes one of the two forms
  • TV Model
  • Data Model
Questions

• What is an End-User view?
• What is a workflow-based view (scenarios)?
• What is an appliance-based view?
• What is QoS?
• What is End-User QoS?
• What is an SLA and SLS?

• Some answers and examples...
**Service Level Agreement/Specs**
(How do we fulfill the following scenario?)

<table>
<thead>
<tr>
<th>Actual customer input</th>
<th>Network reaction</th>
<th>Contraced SLA</th>
<th>Contracted Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>600 kbps</td>
<td>600 kbps</td>
<td>2 Mbps</td>
<td>Super (EF)</td>
</tr>
<tr>
<td></td>
<td>800 kbps from silver</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>600 kbps from economy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.8 Mbps</td>
<td>3 Mbps</td>
<td>3 Mbps</td>
<td>Silver (AF)</td>
</tr>
<tr>
<td>8 Mbps</td>
<td>1 Mbps</td>
<td>1 Mbps</td>
<td>Economy (BE)</td>
</tr>
</tbody>
</table>
## Service Level Agreement (2)

<table>
<thead>
<tr>
<th>Actual customer input</th>
<th>Network reaction</th>
<th>Contracted SLA</th>
<th>Contracted Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>600 kbps</td>
<td>600 kbps</td>
<td>2 Mbps</td>
<td>Super (EF)</td>
</tr>
<tr>
<td></td>
<td>1.4 Mbps from silver</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.8 Mbps</td>
<td>2.4 Mbps from silver</td>
<td>3 Mpbs</td>
<td>Silver (AF)</td>
</tr>
<tr>
<td></td>
<td>600 kbps from economy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 Mbps</td>
<td>1 Mbps</td>
<td>1 Mbps</td>
<td>Economy (BE)</td>
</tr>
</tbody>
</table>
Service Level Specs - How to implement?
Do I need to/Can I Bid/Auction for Resources?

Example: **WLS+MPEG2+HDTV** (needs many “thin guaranteed streams with dynamically allocated 3 to 10 Mbps streams”) + RSVP?
+ COPS-based policy services +
+ DiffServ edges (Replace the WRR scheduler at node 2 with HOL scheduler & study its effect?).
+ MPLS “cloud” (what about change traffic?)
Issues

MPLS-capable

1. PDP
2. A
3. PDP
4. B

GPDP

PEP
Another Scenario

• An ISP has Points of Presence (POPs) at 40 cities in the country and a backbone presence at 5 cities. Due to the high growth rate, it has been decided to have a 4 node LAN backbone at each POP through which there is also access to the ISP backbone. Each POP hosts servers, provides dial-in access to users, and is connected to routes/access nodes from several towns. The ISP would like to have an RSVP/COPS/DiffServ/MPLS solution at the access nodes and the core nodes. Some of the requirements are that no edge node or link should be loaded to more than 75% of its capacity, failure of a single node in the network should not disrupt any user. Also, to provide a highly reliable backbone, two switches are to be installed at each backbone site. The ISP plans to lease two OC3 links from a POP to the associated backbone site and one OC12 link between the backbone sites.

• What is the recommended topology to maximize network capacity and revenue?
More on the Scenario

• The ISP estimates that 5% of the user traffic traverses outside the local POP. To provide competitive Quality of Service, the ISP projects that each ISP and POP backbone node should support 20% of change traffic.

• How does it influence the projected capacity? Do I need to bid for just-in-time resources? Can I auction some of mine?
Hierarchical Model

Edge nodes

Level-N nodes

Core
NC State Test-Bed for End-User Oriented Benchmarking of QoS for Network-Based I2 Applications
NCState.Net

- **World-class campus production network**
  - Fault-Tolerant (redundant, managed)
  - 4 Gbps backbone (plans to go to 10 Gbps)
  - 2 Gbps routing drops (fiber)
  - 1 Gbps building drops (fiber)
  - 100 Mbps floor drops (fiber)
  - 10/100 desktop (copper)
  - Nomadic computing + wireless
  - Being hardened to provide 6 to 7 9’s availability in order to support critical end-users application (e.g., VoIP)
Current NC State Production Backbone

Internet Internet2 NCNI NCREN

Remote Access Off Campus Facilities

OC48/OC3

Fault-Tolerant Nomadic Net

4 X Gigabit Enet

HLB North Campus 24 Interfaces

POE East Campus 32 Interfaces

SMDF South Campus 38 Interfaces

CMDF Cent. Campus 14 Interfaces

1 Gig

1 Gig

T1/ISDN

NCNI

NCREN

Remote Access Off Campus Facilities
Includes DiffServ and Policy-Based Routing
NC State - Alcatel Link

Video
Voice
Data

Alcatel

I2

NC State

Video
Voice
Data
• Current: multi-location distributed facility
• Future: Large centralized + distributed labs
Televator
Real-Time Stereo Projection over ATM
High Resolution Stereo and Stereo Projection
NC State WLS-Assisted Video-Access-Node Classroom

Instructor

ATM (Internet II)
MPEG-2
+ Web
+QoS

Internet

RealAudio (Video), Web

Web Lecture System

Phone Modem

Wireless

MPEG-1 or RealVideo
RealAudio

Web Lecture System

Web Lecture System

Web Lecture System

Web Lecture System

Web Lecture System

Web Lecture System

Web Lecture System

Web Lecture System

Web Lecture System

Web Lecture System

Web Lecture System
End-User QoS Probes

- Standardized EU-QoS probes in all key elements of NCState network.
- Constant/dynamic QoS reports (wrt. Benchmark metrics) transmitted to Network-Operations Center
- Matching of EU-QoS probe results with end-user SLAs and SLSs
- Corrective actions
Current NC State Production Backbone

- Internet Internet2 NCNI NCREN
- Remote Access Off Campus Facilities
- OC48/OC3
- T1/ISDN
- 4 X Gigabit Enet
- Fault-Tolerant Nomadic Net

- HLB North Campus 24 Interfaces
- POE East Campus 32 Interfaces
- SMDF South Campus 38 Interfaces
- CMDF Cent. Campus 14 Interfaces

NCSU Department of Computer Science - 59
Benchmark Probes

- **End-User Workflows**
  - E.g., Professor, University, ISP ...
  - Being constructed

- **SLA/SLS**
  - Currently Open

- **Application**

- **Middleware**

- **OS**
  - Linux, W2000, CiscoR OS

- **Com Protocols**
  - IP, TCP, UDP

- **“Wire” Infrastructure**
  - Ethernet, ATM, “light”, DiffServ