Measurement Challenges for VoIP Infrastructures

Prasad Calyam,

OARnet and Internet2

pcalyam@internet2.edu

Internet2 VoIP Workshop,Indianapolis, October 2003

Topics of Discussion

- Introduction
- Motivations for VoIP Measurements
- E2E Application Measurements
- E2E Network Measurements
- Best practices for VoIP Deployment
- Our research related to VoIP
- H.323 Beacon Tool for troubleshooting
- Conclusion

New-Generation Internet Architecture…

- Video/Voice convergence with the IP-based Data Networks (H.323, SIP, …)
  - Cost Saving
  - Ability to integrate into applications and services IP networks provide
- But, Video/Voice traffic and Data traffic have opposite requirements from the network!
  - Data traffic is asynchronous (delays are acceptable) and extremely sensitive to error
  - Video/Voice traffic is synchronous (significant delays are unacceptable) and more tolerant to errors

What are we up against?

How do we replace a system that traditionally guaranteed 99.999% reliability with a system that is built on a "best effort service" philosophy!!!

- 99.999% availability = 315 seconds per year, calculated as: Availability=Mean Time Between Failures (MTBF)/Total Time where; Total Time=MTBF + Mean Time To Repair (MTTR)

OR

In a single year, you could expect no more than 5 minutes total cumulative outage!

- Data networks average around 98.5% uptime for an average of 131 hours/year; Network downtime caused by power outages, server crashes, software failures, network congestion, user error, …

The essential Mantras…

- Thinking End-to-End is critical in deploying VoIP based networks
  - Installing good end-points, application-service equipment, training end-users and admins, …
    - E2E Application Measurements
  - Good network design, lots of bandwidth, understanding network dynamics and adapting to the changes
    - E2E Network Measurements
- Keeping the IP network “Carrier Grade” from the end-user’s perspective
  - High reliability, short call setup time, high speech quality, i.e., no perceptible echo, noticeable delay and annoying voices on the line

Understanding the performance factors of a VoIP System

- Human Factors
  - Individual perception of audio/video quality, Lack of training to use the system effectively, …
- Device Factors
  - VoIP endpoints, gateways,MCUs, Routers, Firewalls, NATs, Modems, Operating System, Processor, memory, …
- Network Factors
  - Delay, Jitter, Packet loss, Throughput, BER, …

E2E Application Measurements

- Two approaches to evaluating the performance of voice coder in terms of its ability to preserve the signal quality
    - Objective Measurements
    - Subjective Measurements

Objective Measurements

- Calculate signal-to-noise (SNR) ratio and provide a quantitative value of how well the reconstructed voice approximates the original voice
  - Examples: Mean Square Error (MSE) distortion, frequency weighted MSE, segmented SNR, articulation index, Perceptual Evaluation of Speech Quality (PESQ; ITU-T P.862), Perceptual Analysis Measurement System (PAMS), …
- Most popular objective measure used today is based on the “E-Model” (ITU-T G.107)
  - Measured data circuit impairments (i.e. packet loss, jitter, and latency) contribute to an “R factor” from 0 to 100 that is mapped over time to obtain an equivalent MOS score

Objective Measurements (Contd.)

+)They provide means for automated measurements and useful for initial design and simulations of coding techniques

-)They do not necessarily give an indication of speech quality as perceived by the human ear

Subjective Measurements

- They are conducted by playing a sample of voice to a number of listeners and their judgment of the quality of the voice is used as a metric *Can you hear me now?*
- They are carried out in different environments to simulate real life conditions such as noisy, multiple speakers, etc..
  - Results in terms of overall quality, listening effort, intelligibility and naturalness
- Some techniques
  - Diagnostic Rhyme Test (DRT) – “those dose” tests
  - Mean Opinion Score Rankings (MOS) *Popular!*

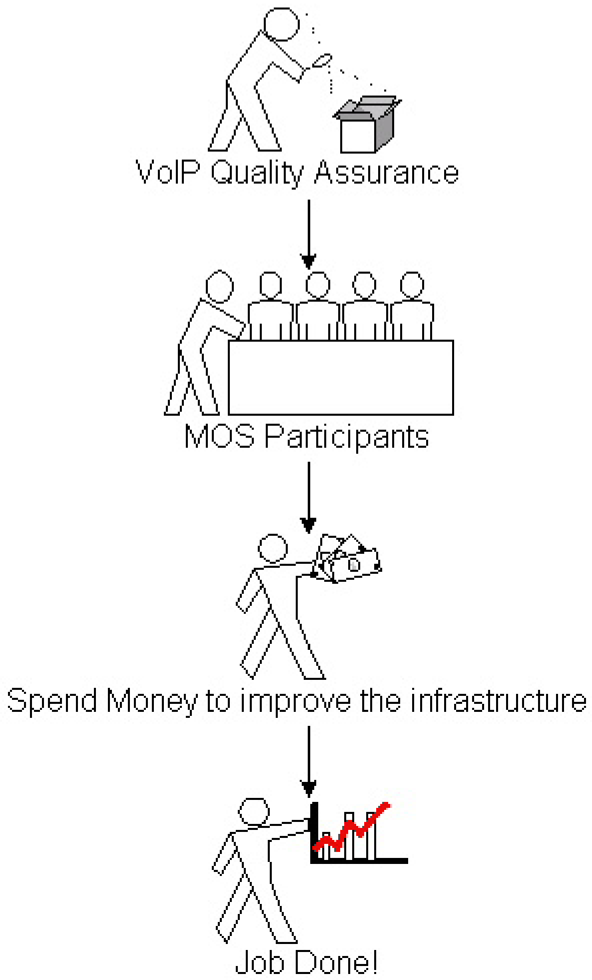
MOS Rankings (ITU-T P.800)

Quality Scale	Score	Listening Effort Scale
Excellent	5	No effort required
Good	4	No appreciable effort required
Fair	3	Moderate effort required
Poor	2	Considerable effort required
Bad	1	No meaning understood with reasonable effort

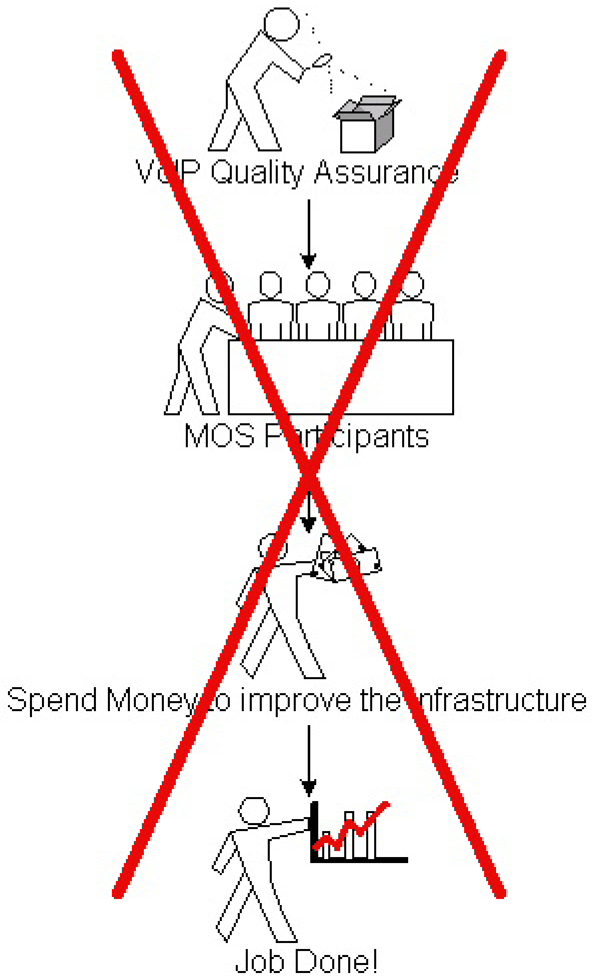
- Performance of Popular Voice Coders on MOS Scale
  - Coding techniques are such that speech quality degrades as data rate reduces. However, the relationship is not linear!

CODEC	Data Rate	MOS Score
G.711	64Kbps	4.3
G.726	32Kbps	4.0
G.723	6.3Kbps	3.8
G.728	16Kbps	3.9
G.729	8Kbps	4.0
GSM	13Kbps	3.7

E2E Network Measurements



E2E Network Measurements (Contd.)



E2E Network Measurements (Contd.)

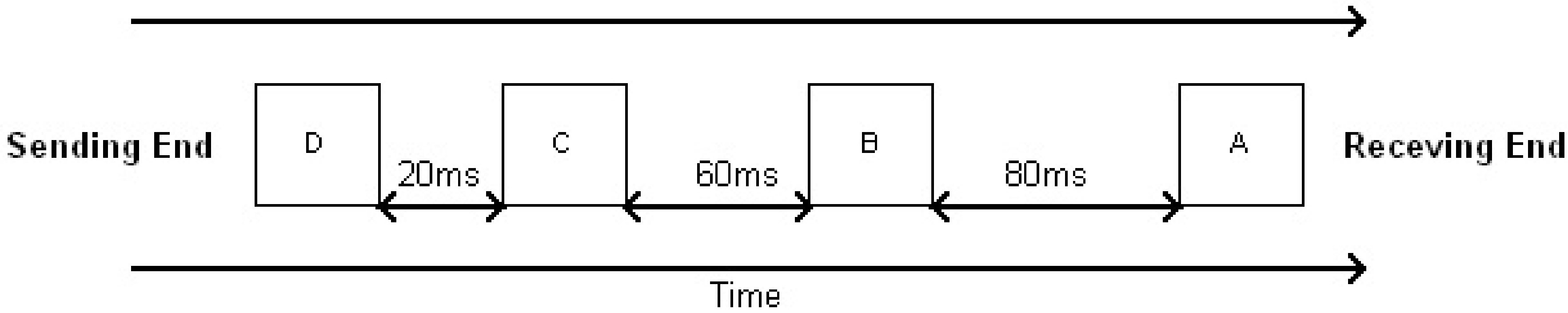
- The network contributes to performance variations which your customers will complain!
- The network dynamics is caused by the route changes, competing traffic, congestion, …
- Three network metrics to worry about
  - Delay, Packet loss and Jitter
- RFC 1889 defines how to obtain these values using information in real-time traffic packets

End-to-end Delay Components

SENDER SIDE	NETWORK	RECEIVER SIDE
Compression Delay	Propagation Delay	Resynchronization Delay
Transmission Delay	Processing Delay	Decompression Delay
Electronic Delay	Queuing Delay	Presentation Delay

- Delay is the amount of time that a packet takes to travel from the sender's application to reach the receiver's destination application
  - Caused by codecs, router queuing delays, …
- One-way delay requirement is stringent for VoIP to maintain good interaction between ends
- Good (0ms-150ms), Acceptable (150ms-300ms), Unacceptable (> 300ms) [ITU-T G.114]

Understanding jitter…



- Jitter is the variation in delay of the packets arriving at the receiving end
  - Caused by congestion, insufficient bandwidth, varying packet sizes in the network, out of order packets, …
- Excessive jitter may cause packet loss in the receiver jitter buffers thus affecting the playback of the voice stream
- Good (0-20ms), Acceptable (20ms-50ms), Unacceptable (>50ms) [http://www.wainhouse.com/files/papers/wr-qos-in-ip-networks.pdf]

Understanding Packet Loss…

- Packet Loss is the packets discarded deliberately (RED, TTL=0) or non-deliberately by intermediate links, nodes and end-systems along a given transmission path
    - Caused by line properties (Layer 1), full buffers (Layer 3) or late arrivals (at the application)
- Good(0%-0.5%), Acceptable (0.5%-1.5%), Unacceptable (>1.5%) [http://www.adec.edu/nsf/Summary%20Test%20H.323.v7.pdf]

Some Network Measurement Techniques to measure VoIP QoS

- Continuous measurements for delay, jitter and packet loss
  - Active measurements using beacons/PMPs/ Verifiers, tools in soft switches, one-way delays with GPS, traceroutes, …
  - Simulating VoIP traffic and producing in-depth reports (historical data, real-time, predictions, …)
  - Measuring at gateways, end-points and at other strategic network points
- Call generators
  - Provide repeatable call sequences and analyzing the logs to stress test the network

Best practices for VoIP Deployment

- Make Voice Quality a Top Priority in the network
- Test Gateways under Full Load
- Perform Live LAN/WAN Audit
- Quality Affected by Network or VoIP Gear
- Verify Echo Cancellation Issues with Manufacturer
- Employ Real-time Indication of VoIP Quality
- Test Quality with all CODECs
- Get Adequate Training from Vendor

Reference: http://www.ct-labs.com

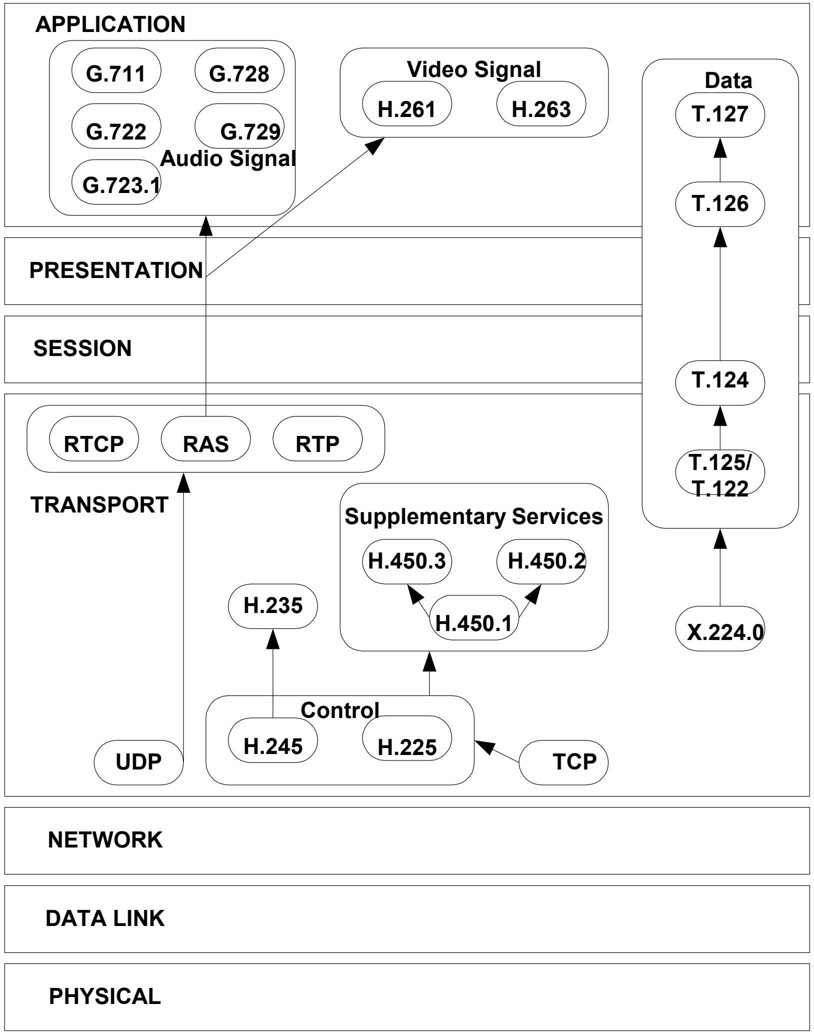
Our Research Objectives

- "To measure the behavior of an H.323 Videoconferencing system with the variations in the network behavior, study end-to-end delay contributions and suggest a model for large-scale multipoint H.323-based Videoconferencing"
  - Performance Bounds Testing, End-to-End Delay Testing
- "To develop a tool that can be used to monitor and measure the performance of H.323 Videoconference sessions to identify and troubleshoot performance problems in a H.323 Videoconferencing system"
  - H.323 Beacon

H.323 Overview

- H.323 is an umbrella standard that defines how real-time multimedia communications such as Videoconferencing can be supported on packet switched networks (Internet)
- Devices: Terminals, Gateways, Gatekeepers and MCUs
- Codecs: H.261, H.263, G.711, G.723.1
- Signaling: H.225, H.245
- Transport Mechanisms: TCP, UDP, RTP and RTCP
- Data collaboration: T.120
- Many others…

H.323 Protocol Stack



Performance Evaluation of H.323 Videoconference Traffic

- Performance Bounds Testing
  - Testing Environment
  - Metrics
  - Experiment Methodology
  - Results
- End-to-end Delay Testing
  - Testing Environment
  - Experiment Methodology
  - Results
- Recommendations for large multipoint H.323 Videoconferencing systems

Why Bounds Testing?

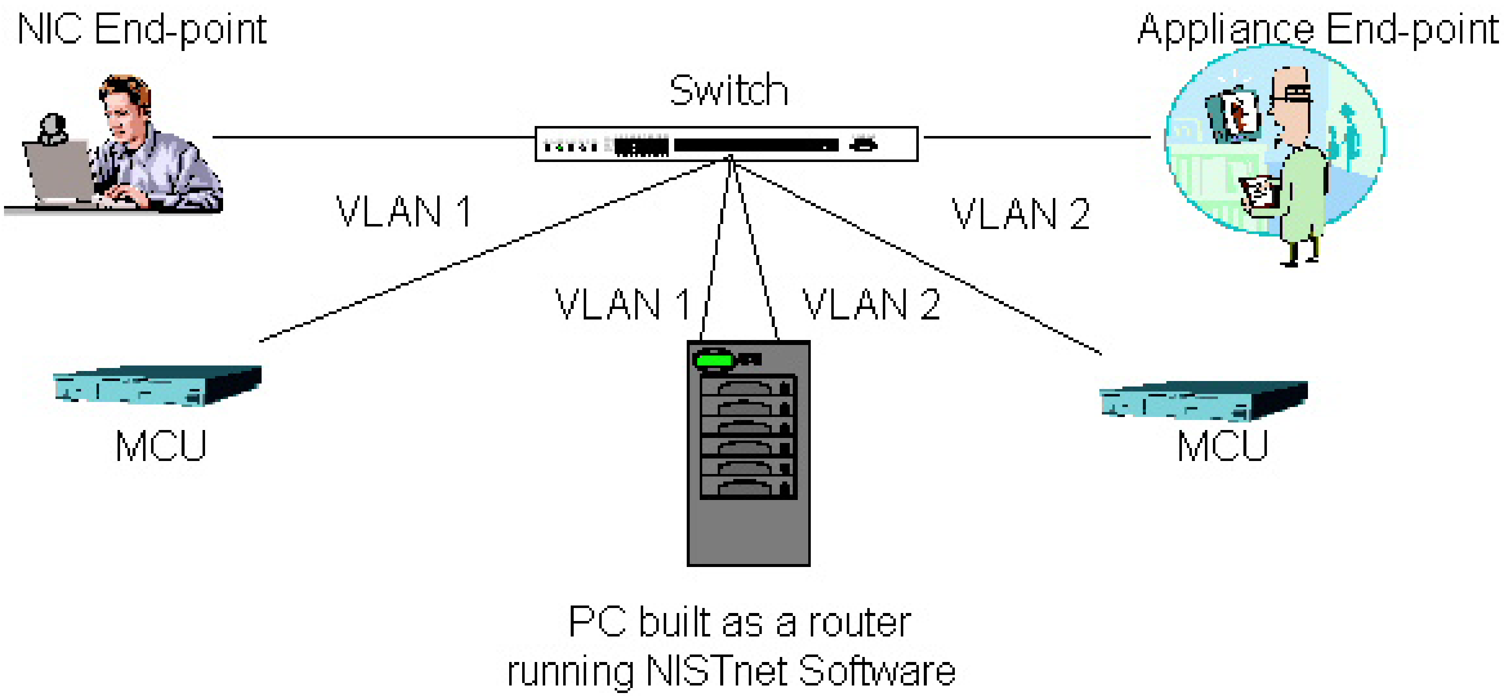
- Just network metrics such as bandwidth, frame rate, etc… cannot quantify H.323 audio/video performance; H.323 adapts!
- Every application has many idiosyncrasies and requires network parameters to be within certain bounds to achieve acceptable performance
- Regulation of one network parameter influences other network parameters
  - Sharp variations in jitter values leads to a significant increase in packet loss
- It is necessary to understand application behavior in an isolated environment with the variations in the network parameters to make provisions in the network, without affecting other best-effort traffic

Bounds Testing: Testing Environment

- An experimental LAN was created which was isolated from noise and external traffic
- Network Simulator
  - Linux Mandrake 7.2 Kernel recompiled and optimized for the device to be a router
    - Other specs: Intel Pentium III 733 MHz processor, 256MB RAM, Two EtherPro 10/100 NICS
  - Nistnet 2.1.0 Emulator Software
    - Verified Nistnet system reliability prior to test, using Spirent SmartBits™
    - Latency and Jitter parameters were met with in a +/- 1 msec deviation
- Codecs: H.261 Video and G.711 Audio

Bounds Testing: Test Setup

- Point-to-point client test
- Test with multiple clients connected via a single MCU
- Test with multiple clients cascaded via cascaded MCUs



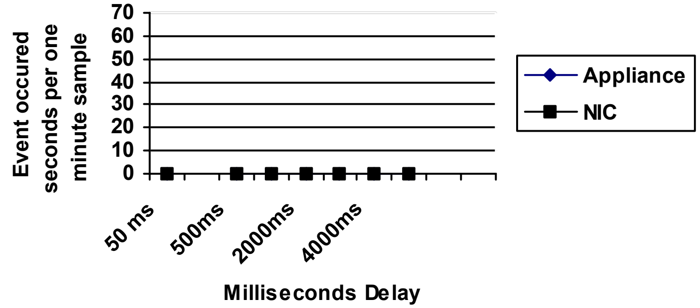
Bounds Testing: Metrics

- Events
  - Spatial Augmentation: Video artifacts such as tiles added to the picture
  - Spatial Depreciation: Parts of the picture or objects in the picture seen to be missing
  - Temporal Distortion: Over time, the "flow" of an event is distorted by missing data
  - Audio Augmentation: Audio artifacts added to audio stream such as pops, clicks and hisses
  - Audio Depreciation: Parts of the audio noticed to be missing
- Terms used to record performance
  - Tile Pulse: Tiling of pixels occurring at fairly regular intervals
  - Frames Freeze: Picture remains still for a period of time
  - Clipped Speech: Audio with missing phonemes commonly at the beginning or end of a word

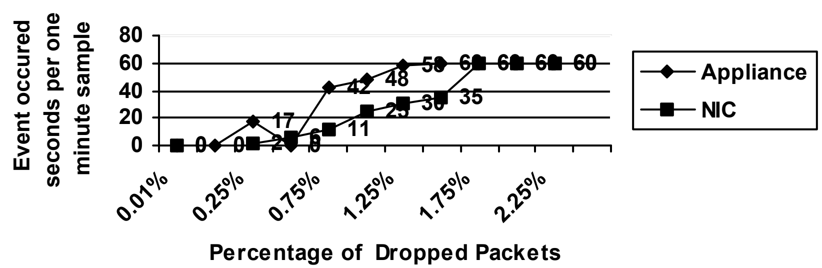
Bounds Testing: Experiment Methodology

- Initiate a call and observe the audio/video events that appear during a 60 second sampling period
- Based on the frequency and duration of the events, come up with a MOS ranking
- Repeat for different values of latency, jitter and packet loss configured on NISTnet
- Test for repeatability

Bounds Testing: Sample plots



Events on latency variations for point-to-point client test



Events on packet drop variations for point-to-point client test

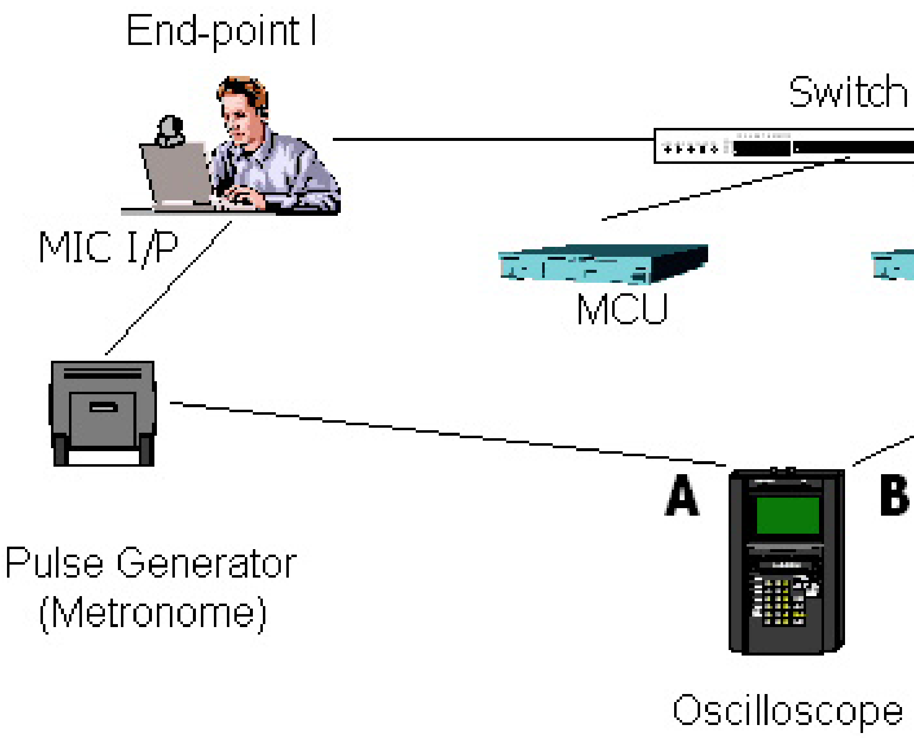
Bounds Testing: Conclusions

- Latency though annoying to users, does not affect the H.323 protocol itself
  - Latency may be translated into packet loss and jitter in the buffers and intermediate routers that handle the H.323 traffic and may result in the deterioration of the call quality
- Packet Loss is tolerated by the H.323 protocol to a certain extent
  - Packet loss must be below 1% in point-to-point and below 0.75% when using cascaded MCUs for the H.323 audio/video to be acceptable to an end-user
  - For the packet loss values above the aforementioned values, the call was terminated sometimes, showing that the H.323 protocol failed to maintain the session
- Jitter causes the *most* distress to the H.323 protocol
  - When a single MCU is used to place a call, it was found to smoothen the jitter
  - However, in a cascaded MCU scenario, the H.323 audio/video was found to be more intolerant, as shown by the increase in the events

Why End-to-end Delay Testing?

- To study the effect of the various H.323 Videoconferencing system components on the overall end-to-end delay and identify the bottlenecks in the system
- To characterize the end-to-end delay of point-to-point and multi-point H.323 Videoconferences based on the end-to-end delay at different bandwidth settings
  - Audio is the reference for end-to-end delay
    - Audio is constantly sampled (64kbps) – PCM!
    - End-to-end delay for video differs with the scene being captured by the camera
    - Video stream is subjected to lip synchronization with the audio stream

End-to-End Delay Testing: Test Setup



SCOPE I/P A:

METRONOME I/P

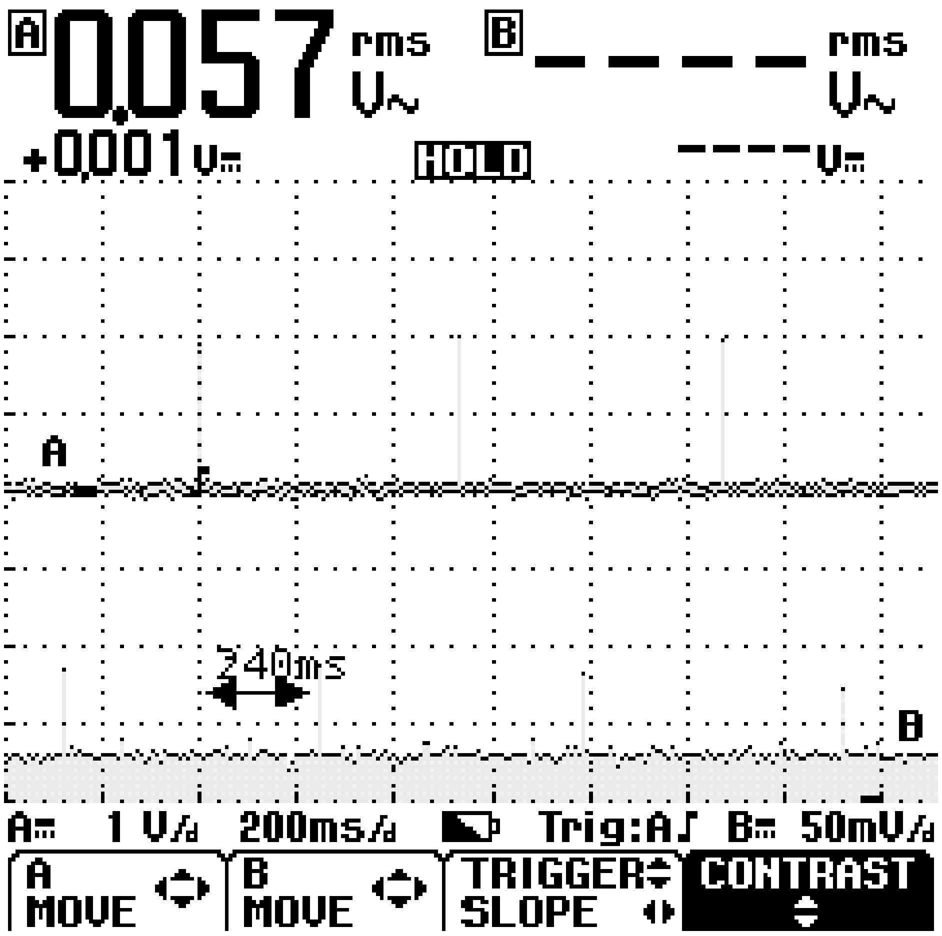
SCOPE I/P B:

ENDPOINT 2 AUDIO O/P

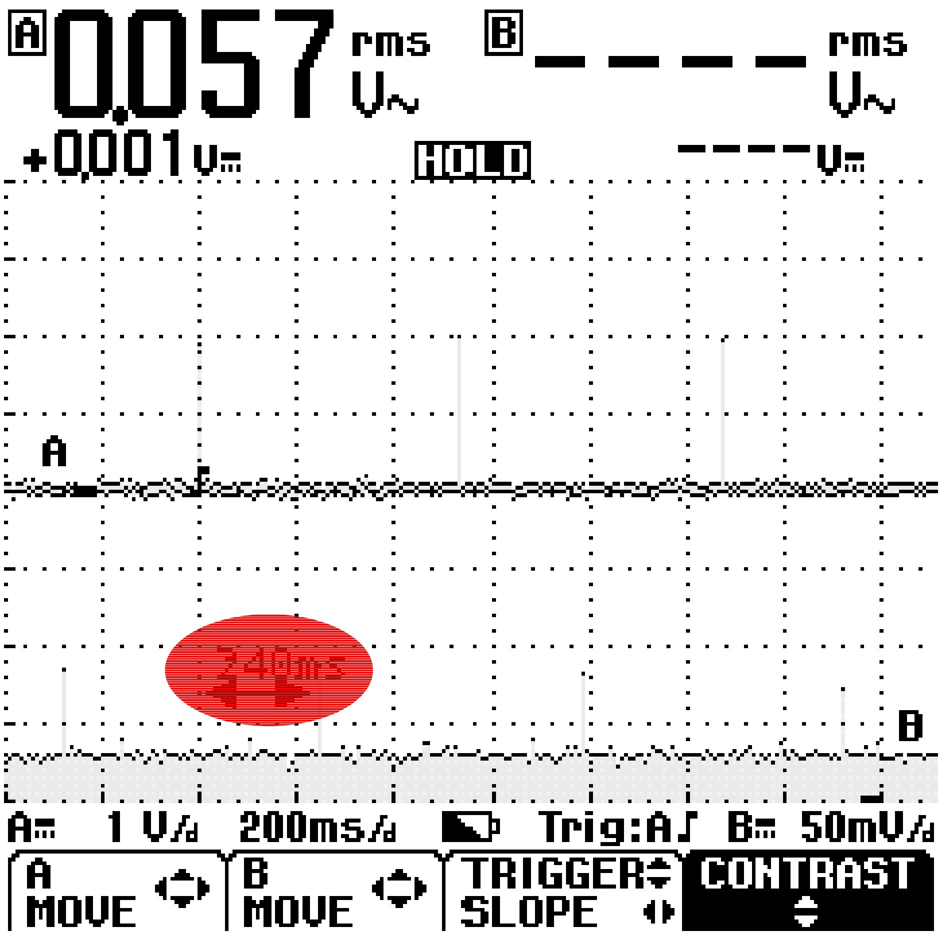
End-to-end Delay Testing: Experiment Methodology

- Measurement of end-to-end delay for 3 modes in an isolated LAN:
  - Point-to-point mode
  - Via an MCU mode
  - Via multiple MCUs mode
- Assumption:
  - Delay contribution due to encoding/decoding in codecs and MCU processing (if present in the path)
  - Switch propagation delay negligible (~1-3ms)
- Dial setting of Metronome (pulse generator) for trace to be discernable: 113
- Mute microphone of endpoint-2
- Place calls at different dialing speeds: 256Kbps, 384Kbps, 512Kbps, 768Kbps
- Record the end-to-end delay from the oscilloscope waveforms

Oscilloscope Waveform



Oscilloscope Waveform

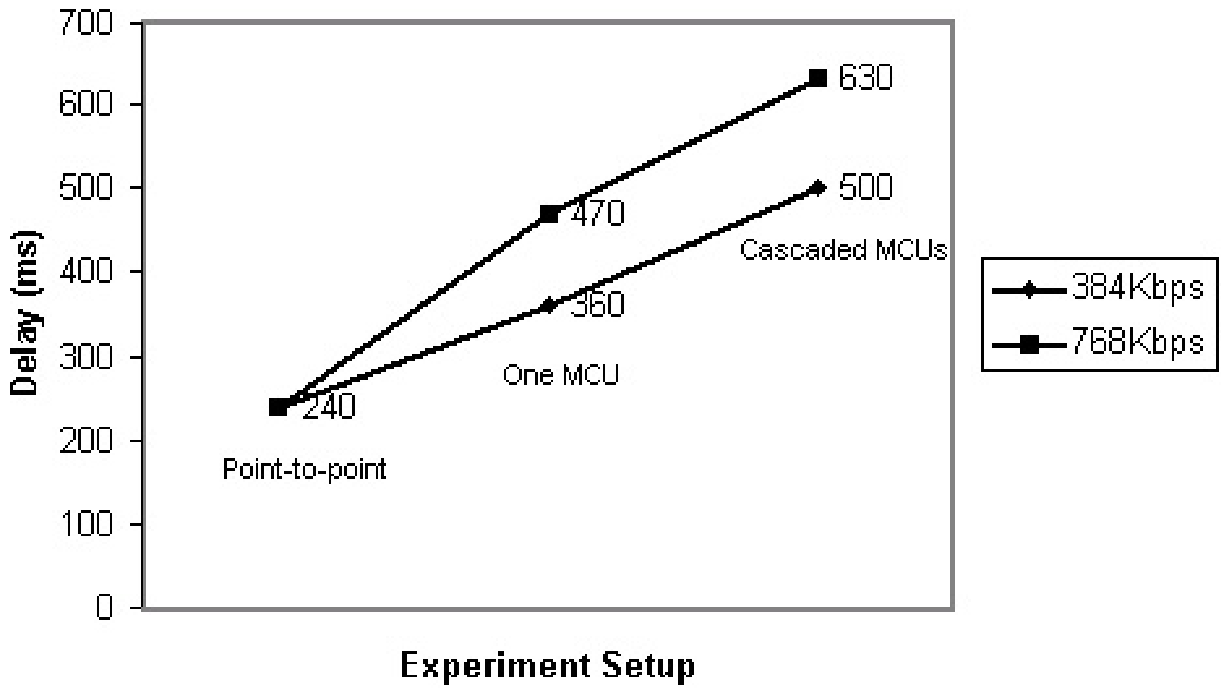


End-to-end Delay Testing: Results Summary

Dialing Speed	Setup	Delay in ms
384K	Point-to-point	240
	Single MCU	360
	Two Cascaded MCUs	500
768K	Point-to-point	240
	Single MCU	470
	Two Cascaded MCUs	630

Delay values recorded for popular dialing speeds

End-to-end Delay Testing: Results Summary (Contd.)



Graph showing the change in the delay values for popular dialing speeds

End-to-end Delay Testing: Conclusions

- End-system delays are much larger than the network delays in a H.323 Videoconferencing system
- The encode-decode delay in a point-to-point settings is ~240ms and independent of the dialing speed
- The minimum delay contribution of an MCU is on the order of ~120ms and the value increases with the increase in the dialing speed
- The delay introduced by cascaded MCUs is significant to the overall end-to-end delay of a session

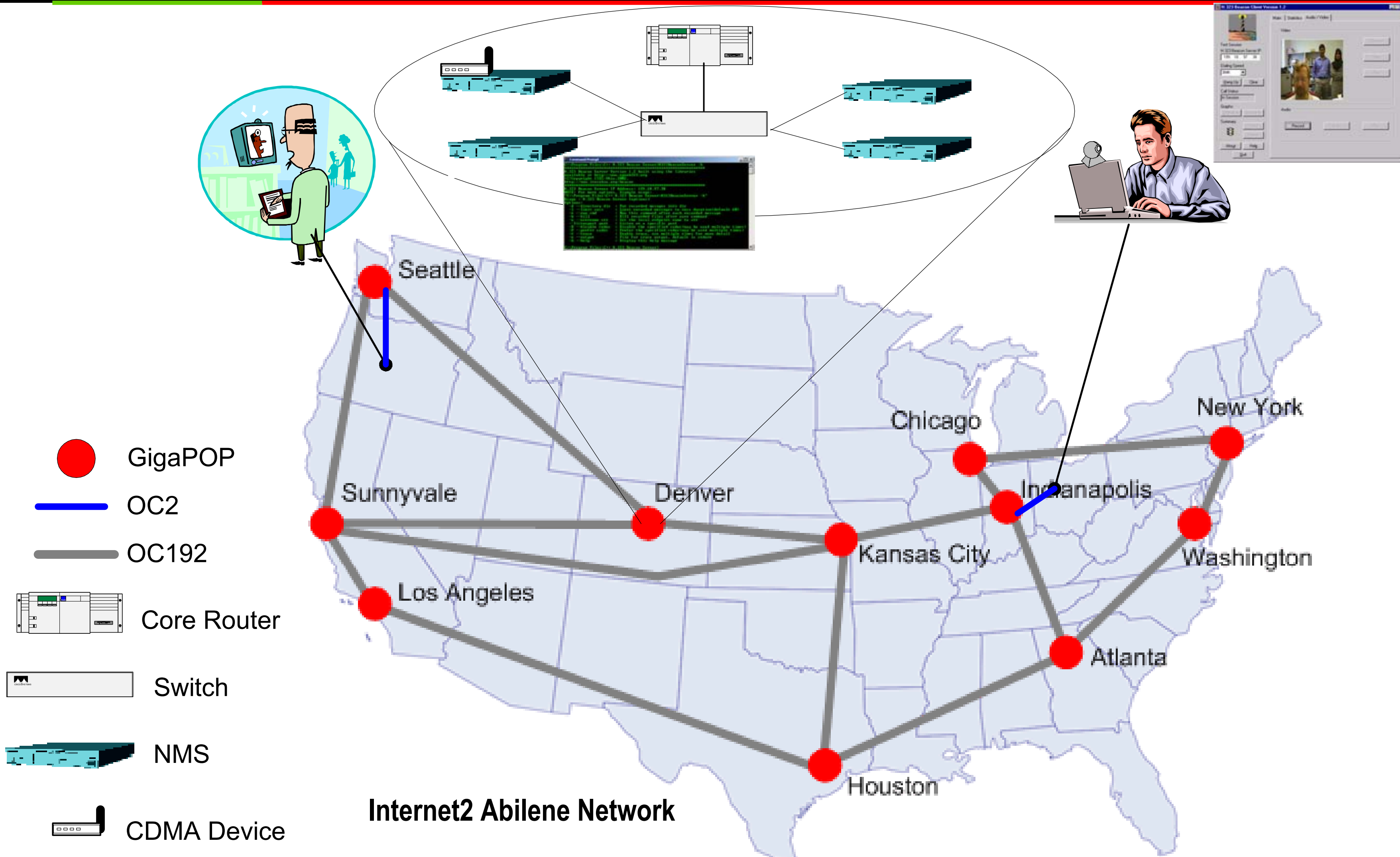
Recommendations for large multipoint H.323 videoconferences

- It is favorable to use MCUs with larger ports to host a conference, rather than cascading MCUs that have lesser number of ports to support participants in a H.323 Videoconference
- A co-location of all the cascaded MCUs might help in limiting the effects of latency, packet loss and jitter on the performance of H.323 audio/video traffic
  - Advantages
    - This architecture eases the network monitoring and measurement activity; helps troubleshooting problems easily and quickly
  - Shortcomings
    - Heavy load on the switch that routes traffic into MCU concentration
    - Single point of failure in case of network distress

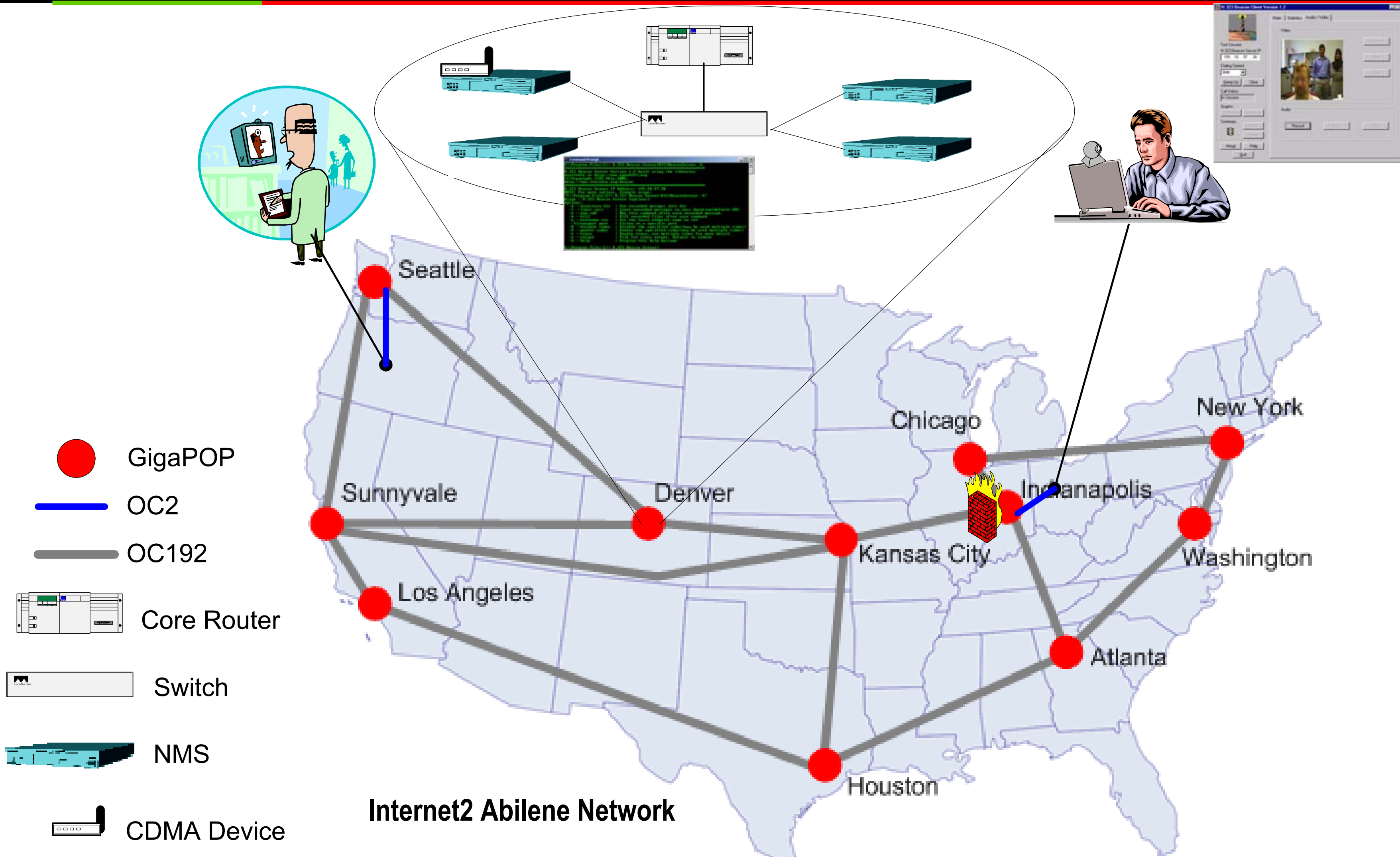
H.323 Beacon Overview

- A tool that can be used to measure, monitor and qualify the performance of an H.323 Videoconference session
- It can be used by an end-user/conference operator/network engineer as a debugging tool to troubleshoot H.323 application performance problems in the network and at the host (end-to-end)
- Provides H.323-protocol-specific evidence and other information necessary to troubleshoot a Videoconference performance problem
- It has a distributed client/server architecture
- No manual intervention is necessary for qualifying an H.323 Videoconference at the remote end
- Easy to install and use!

A scenario where the H.323 Beacon is useful!

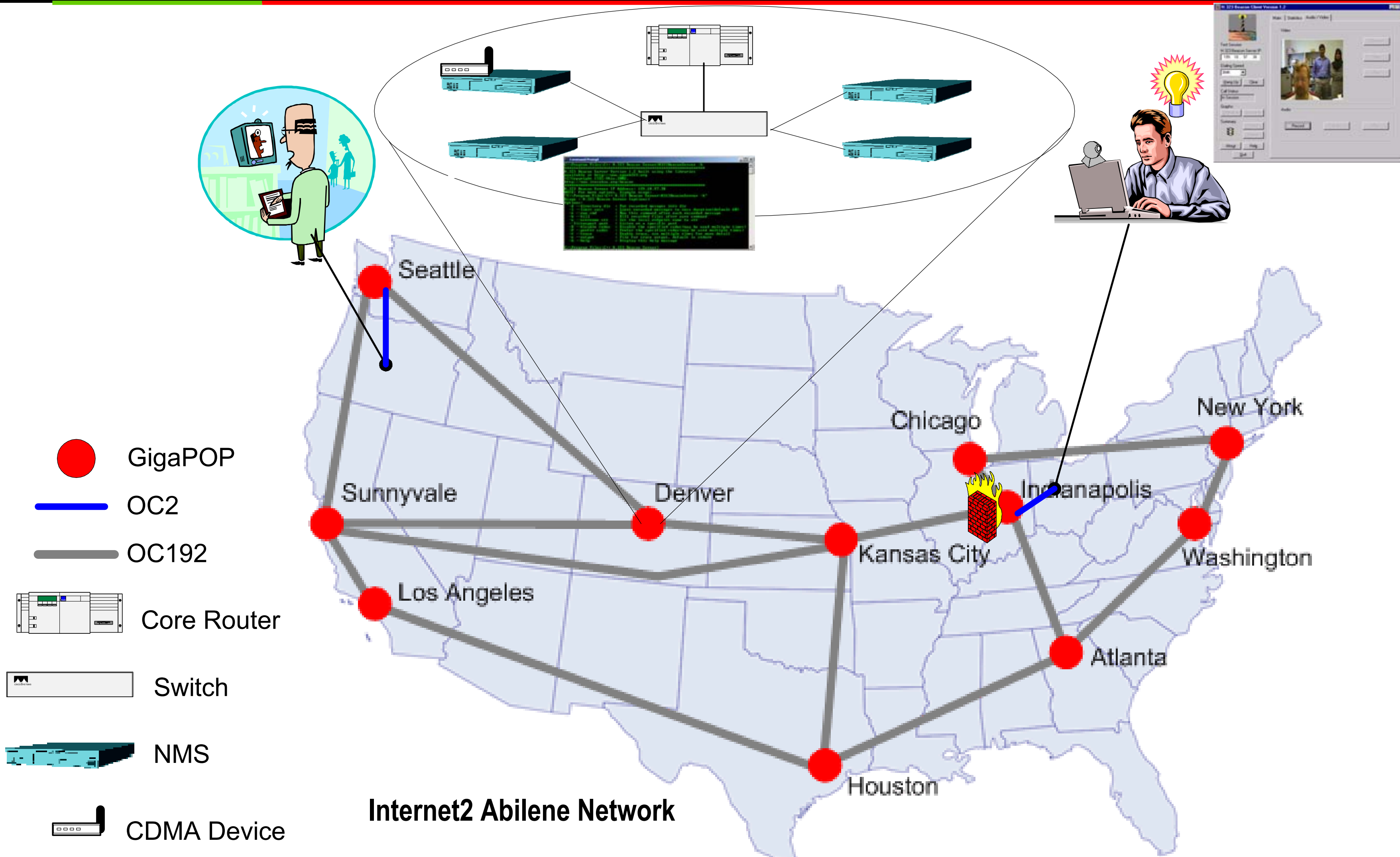


If the industry professional's network segment had a firewall…

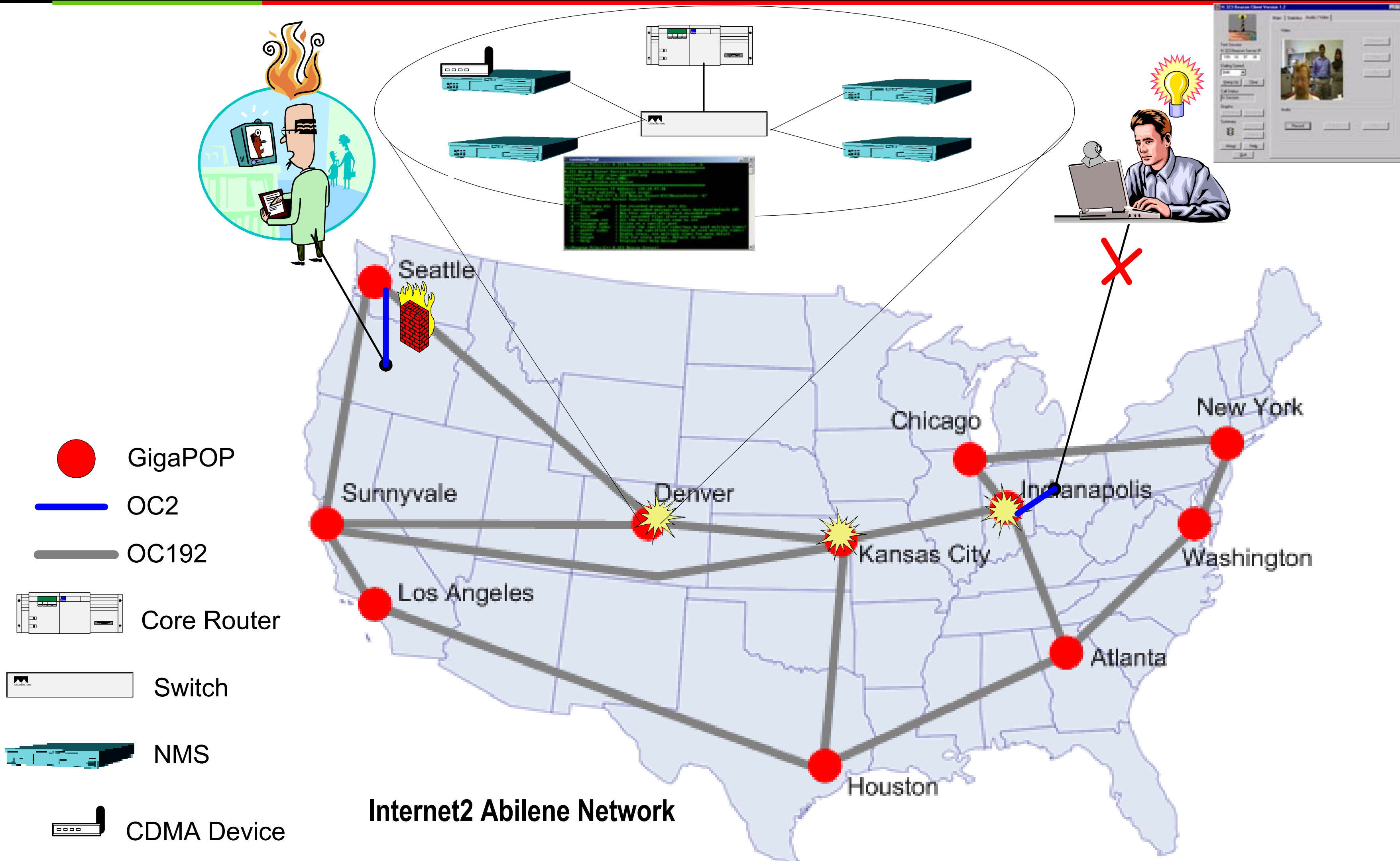


Internet2 Abilene Network

Ah.. A firewall in the path!



By suitably distributing the servers, any E2E problem can be identified!



Version 1.0 Features…

- Multi-threaded Server
- Client and Server interoperability with commercial clients
- Call Status: “In Session”, “Normal Close”, “Exception Close”
  - Call exception handling Alarms: local client has no Internet connectivity, network congestion, firewall presence, remote client/server not online, transport error, insufficient bandwidth, invalid IP address of remote client or server, …

Version 1.0 Features…(Contd.)

- Call bandwidth selection capability in client
- H.323 session statistics: Round Trip Time, Audio jitter, packet loss, packets and octets sent/received, codec information, …
- Excel sheet generation for offline graphical-viewing of statistics!
- Real time audio/video feedback: Test audio (in wav format) and video quality (in MPEG, QuickTime and AVI formats) of the end-user as seen on the remote side!
- Easy to install setup programs and help utilities…

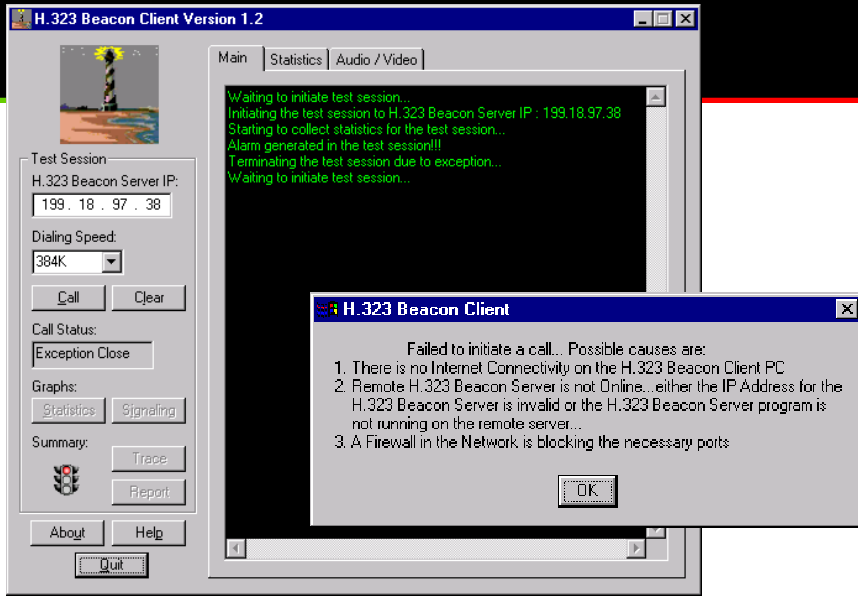
Version 1.3 Features… (Latest Version)

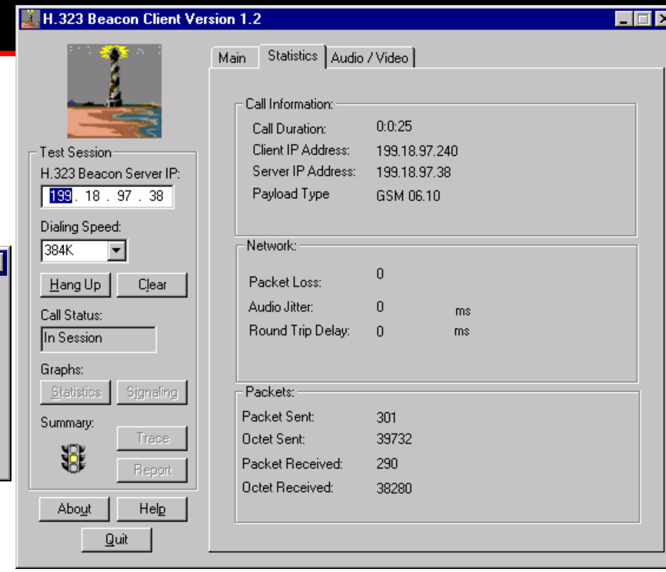
- New GUI layout that reflects the V1.3 features
  - Activities-log: real-time display of all the activities occurring during the test session
  - Traffic light functionality to indicate test session result
  - Settings tab: customize test session data folder, TCP/UDP/RTP port settings, H.225 and H.245 configs, audio/video codec selection, saving/deleting test data, graph formats (bmp/ png), …
  - Report of test session summary: periodic traceroutes in both directions, ping, 50th and 90th percentile summary of statistics, exception information (if any), …

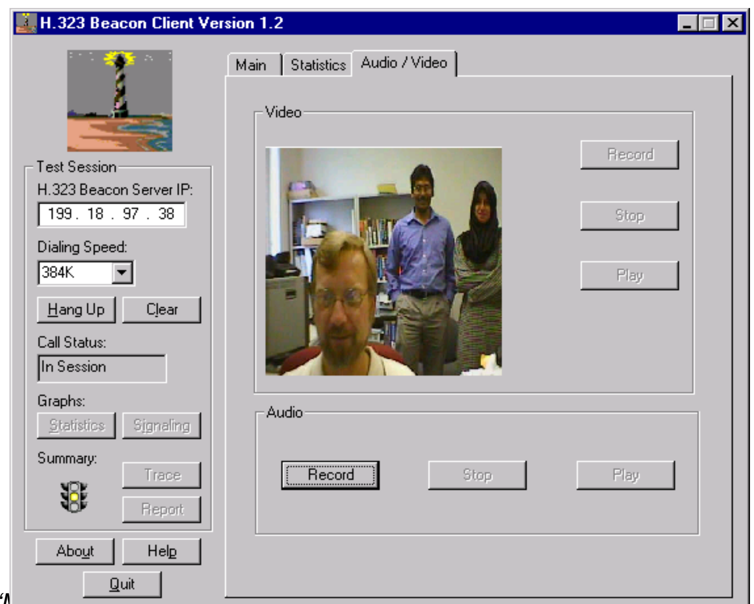
Version 1.3 Features…(Contd.)

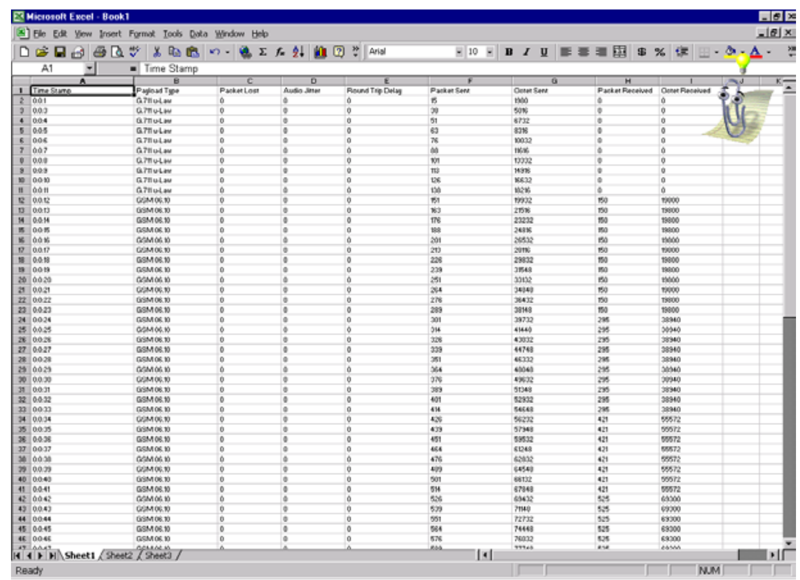
- Call Detail Record at Server to track details of test sessions conducted
- Graph plotting of statistics and ladder diagram for test session signaling message sequences
- Miscellaneous bug fixes and modifications (dynamic tab handling, deleting file locks on server for wav files, etc…)
- Other activities include-
  - LAN testing to test viability of statistics, alarms such as firewall detection, excessive packet drop indicated as congestion, …
  - Video handling and loop back feature research with C++

Some Screenshots…

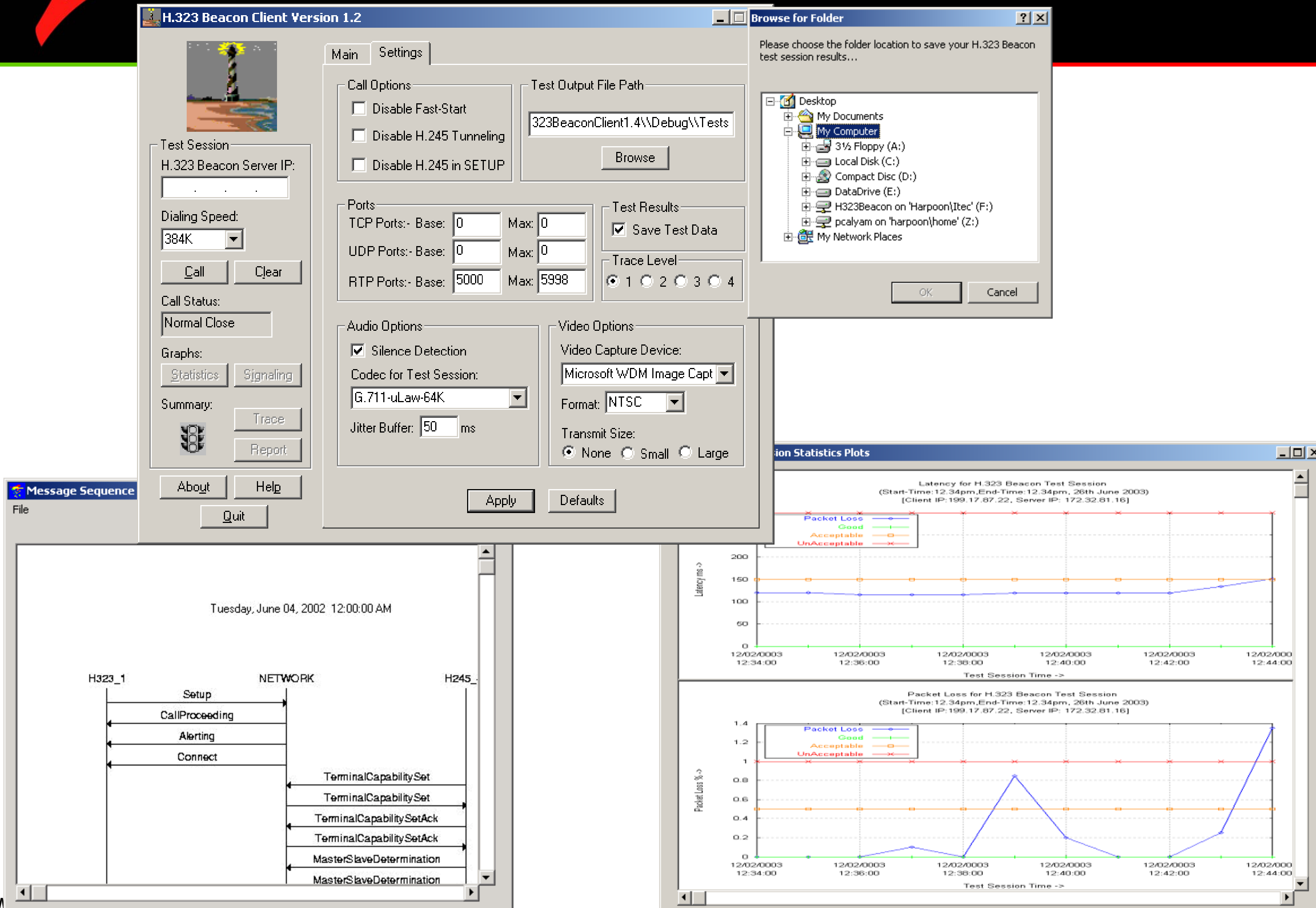




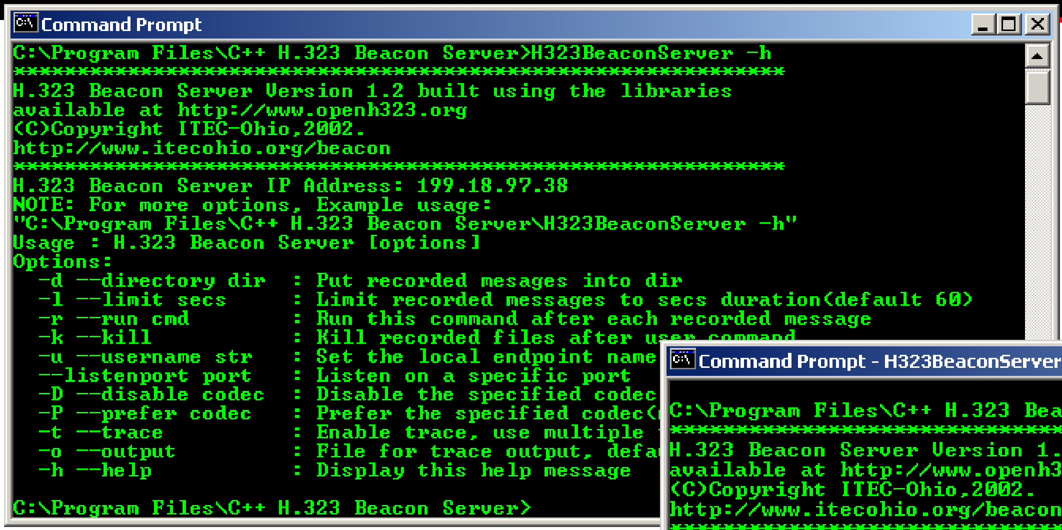


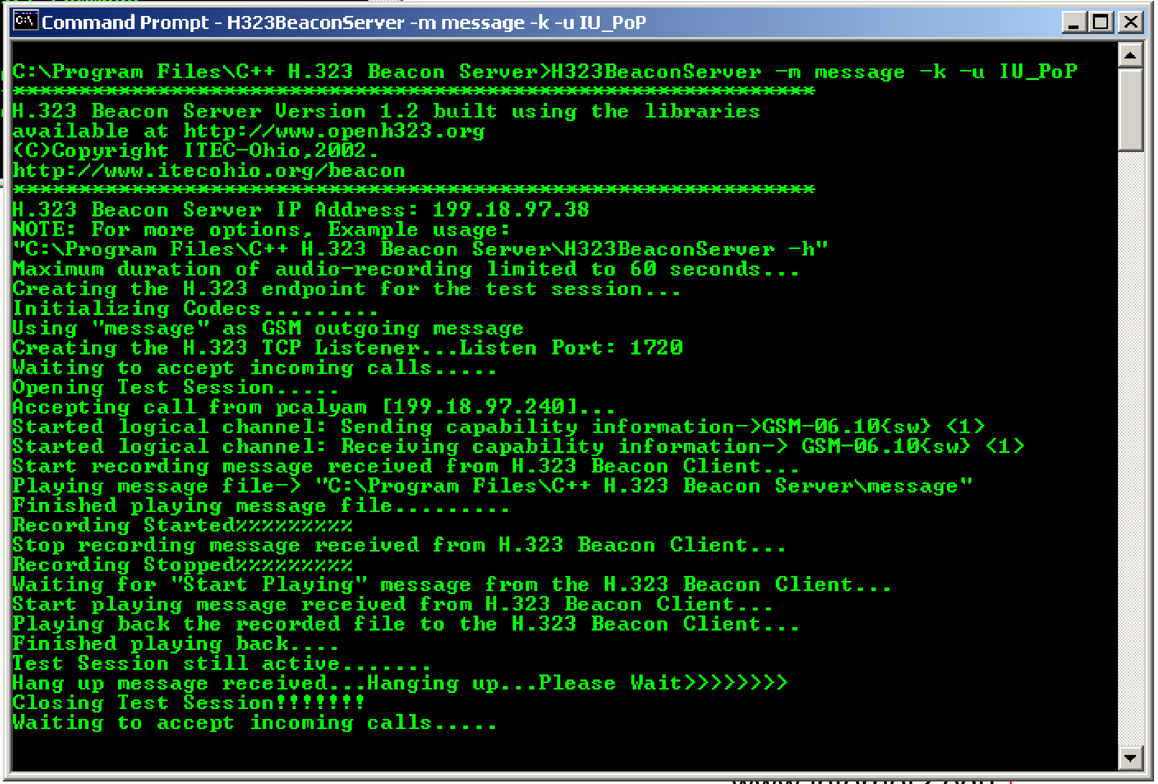


Some Screenshots… (Contd.)



Some Screenshots… (Contd.)

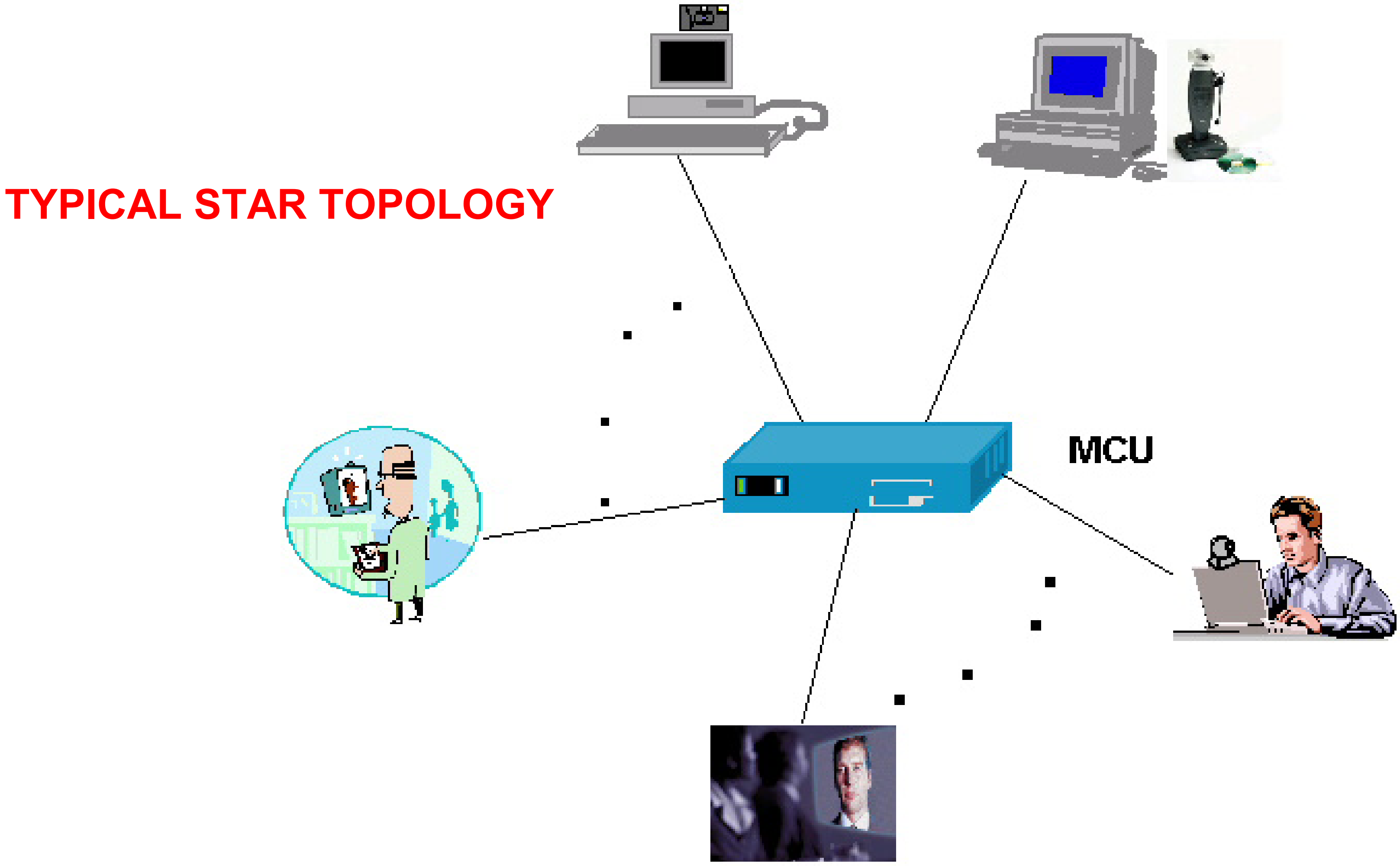




Looking Beyond…

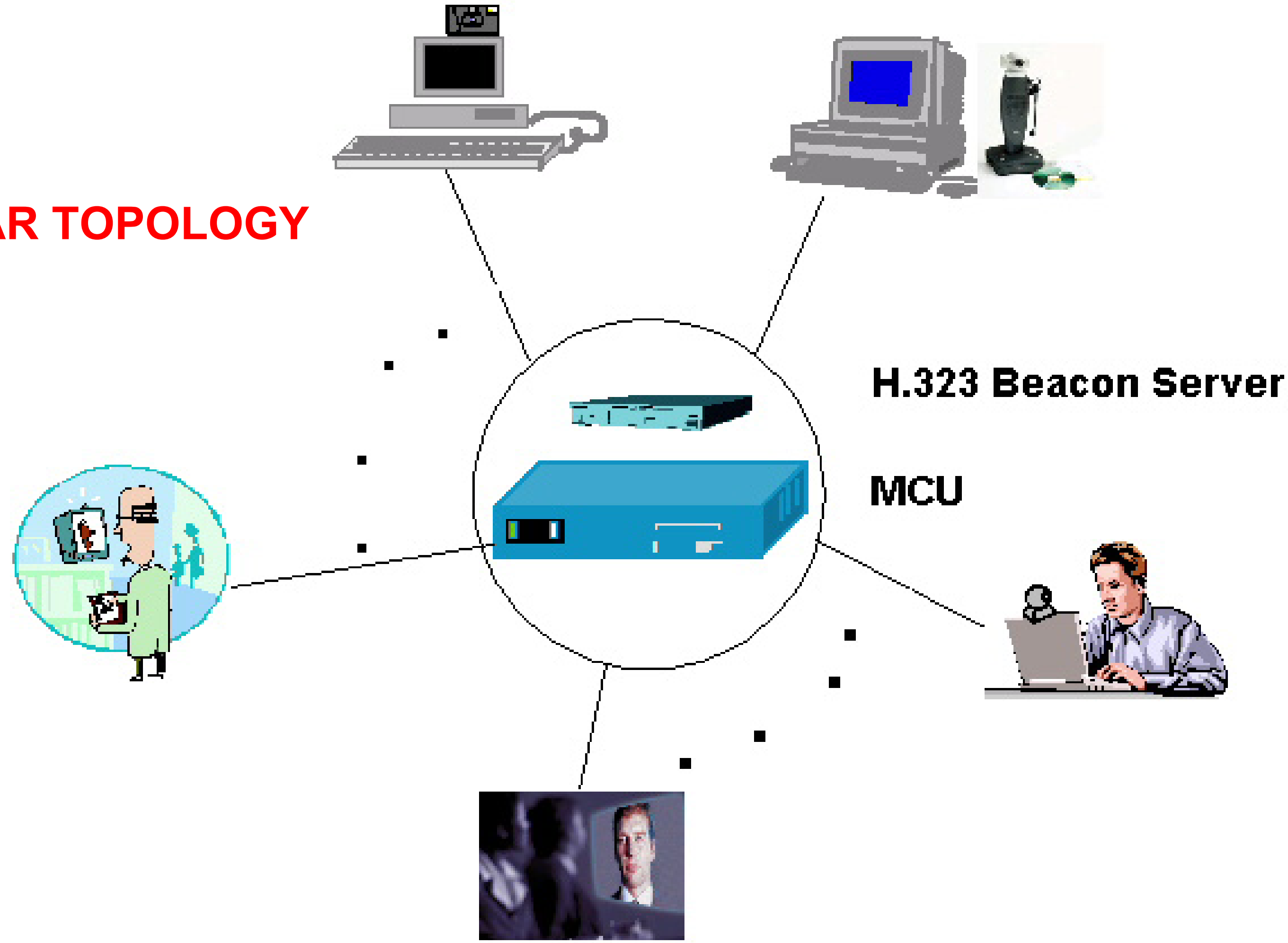
- More statistics variables such as duplicate packets, packets out of order, packets too late, frame rate,…
- Add video capability and loopback feature to C++ client and the C++ server
- *(ps:Java version has video loopback implemented!)*
- Port H.323 Beacon Server source code to run on *nix platforms
- Client-to-client testing capability
- Develop the server-to-server H.323 Beacon
  - Make the H.323 Beacon be able to generate Mbps of traffic with multiple call generation and perform more extensive graph plotting for the large data sets in a web based format
- Local Audio/Video Tuning Wizard

How to deploy this tool?



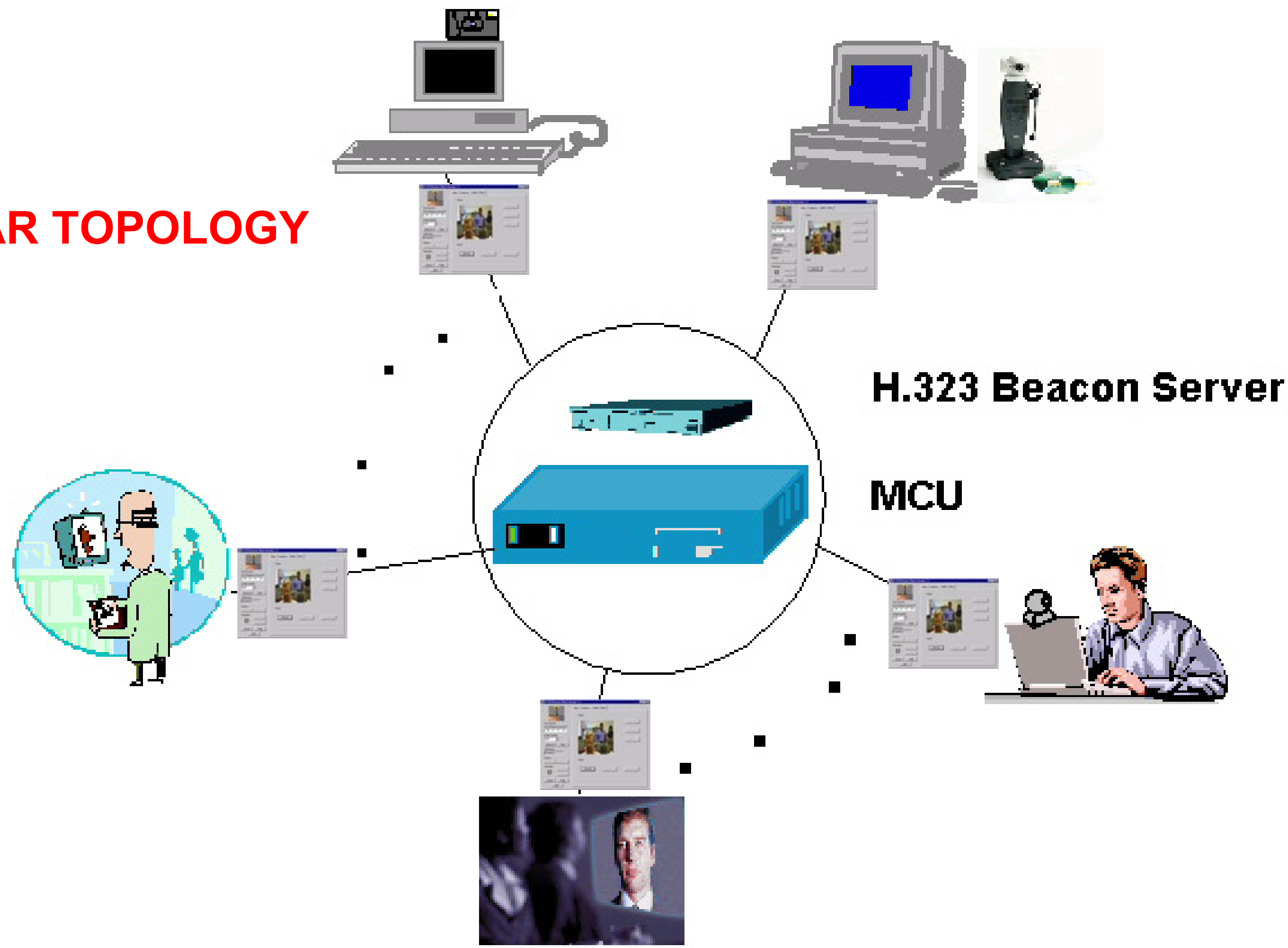
How to deploy this tool? (Contd.)

TYPICAL STAR TOPOLOGY



How to deploy this tool? (Contd.)

TYPICAL STAR TOPOLOGY



More Details…

- Project Website

http://www.itecohio.org/beacon

- Sourceforge.net

http://sourceforge.net/projects/h323beacon/

  - Latest CVS code and Releases …
  - Mailing lists (developers, users, announcements)
  - Discussion form
  - Bug tracking, New feature requests, documentation,…

- Invitation to join the development
  - Add new features, Fix bugs, …

Questions?

INTERNET 2®

www.internet2.edu