



Ohio Supercomputer Center



## Measuring VVoIP QoE using the “Vperf” Tool

**Prasad Calyam** (Presenter)

*Ohio Supercomputer Center, The Ohio State University*

**Mark Haffner, Prof. Eylem Ekici**  
*The Ohio State University*

**Prof. Chang-Gun Lee**  
*Seoul National University*

*SC07, November 14<sup>th</sup> 2007*

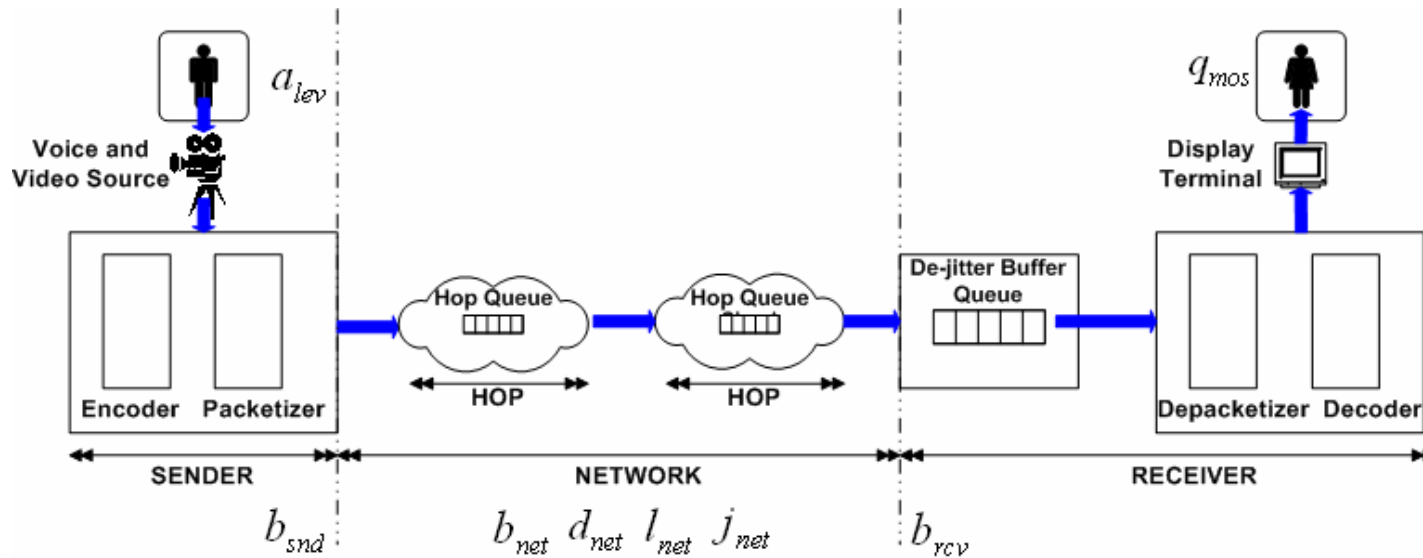
# Outline

- Background
  - Voice and Video over IP (VVoIP) Overview
  - Network QoS and End-user QoE in VVoIP
  - Streaming QoE versus Interaction QoE
- GAP-Model framework
  - Vperf tool implementation of GAP-Model
  - Performance evaluation
- Multi-Activity Packet Trains (MAPTs) methodology
  - Vperf tool implementation of MAPTs
  - Performance evaluation
- Concluding Remarks

# Voice and Video over IP (VVoIP) Overview

- Large-scale deployments of VVoIP are on the rise
  - Video streaming (one-way voice and video)
    - MySpace, Google Video, YouTube, IPTV, ...
  - Video conferencing (two-way voice and video)
    - Polycom, MSN Messenger, WebEx, Acrobat Connect, ...
- Challenges for large-scale VVoIP deployment
  - **Real-time** or online monitoring of end-user Quality of Experience (QoE)
    - Traditional network Quality of Service (QoS) monitoring not adequate
      - Network QoS metrics: bandwidth, delay, jitter, loss
  - Need **objective techniques** for automated network-wide monitoring
    - Cannot rely on end-users to provide subjective rankings – expensive and time consuming

# Network QoS and End-user QoE



- End-user QoE is mainly dependent on the combined impact of network factors
  - Device factors such as voice/video codecs, peak video bit rate (a.k.a. dialing speed) also matter

End-user QoE  $\rightarrow$   $q_{mos} = f(b_{net}, d_{net}, l_{net}, j_{net})$   $\leftarrow$  Network QoS

- Our study maps the network QoS to end-user QoE for a given set of commonly used device factors
  - H.263 video codec, G.711 voice codec, 256/384/768 Kbps dialing speeds


# Voice and Video Packet Streams

$$b_{snd} = b_{voice} + b_{video} = tps_{voice} \left( \frac{b_{codec}}{ps} \right)_{voice} + tps_{video} \left( \frac{b_{codec}}{ps} \right)_{video}$$

- Total packet size ( $tps$ ) – sum of payload ( $ps$ ), IP/UDP/RTP header (40 bytes), and Ethernet header (14 bytes)
- Dialing speed is  $\lceil b_{video} \rceil$ ;  $\lceil b_{voice} \rceil = 64$  Kbps fixed for G.711 voice codec
  - Voice has fixed packet sizes ( $tps_{voice} \leq 534$  bytes)
  - Video packet sizes are dependent on  $\alpha_{lev}$  in the content

# End-user QoE Types

- *Streaming QoE*
  - *End-user QoE affected just by voice and video impairments*
    - Video frame freezing
    - Voice drop-outs
    - Lack of lip sync between voice and video
- *Interaction QoE*
  - *End-user QoE also affected by additional interaction effort in a conversation*
    - “Can you repeat what you just said?”
    - “This line is noisy, lets hang-up and reconnect...”
- QoE is measured using “Mean Opinion Score” (MOS) rankings



<i>END-USER SATISFACTION</i>	<i>MOS</i>	<i>GRADE</i>
Very Satisfied	5.0	Good
Satisfied	4.3	
Some Users Dissatisfied	4.0	Acceptable
Many Users Dissatisfied	3.6	
Nearly All Users Dissatisfied	3.0	Poor
Not Recommended	2.6	
	1.0	

# Problem Summary

- **Given:**

- Video-on-demand (streaming) or Videoconferencing (interactive)
- Voice/video codec
- Dialing speed

- **Develop:**

- An objective technique that can estimate both streaming and interactive VVoIP QoE in terms of MOS rankings
- **Real-time** measurement without involving actual end-users, video sequences and VVoIP appliances
- An **active measurement tool** that can: (a) emulate VVoIP traffic on a network path, and (b) use the objective technique to produce VVoIP QoE measurements

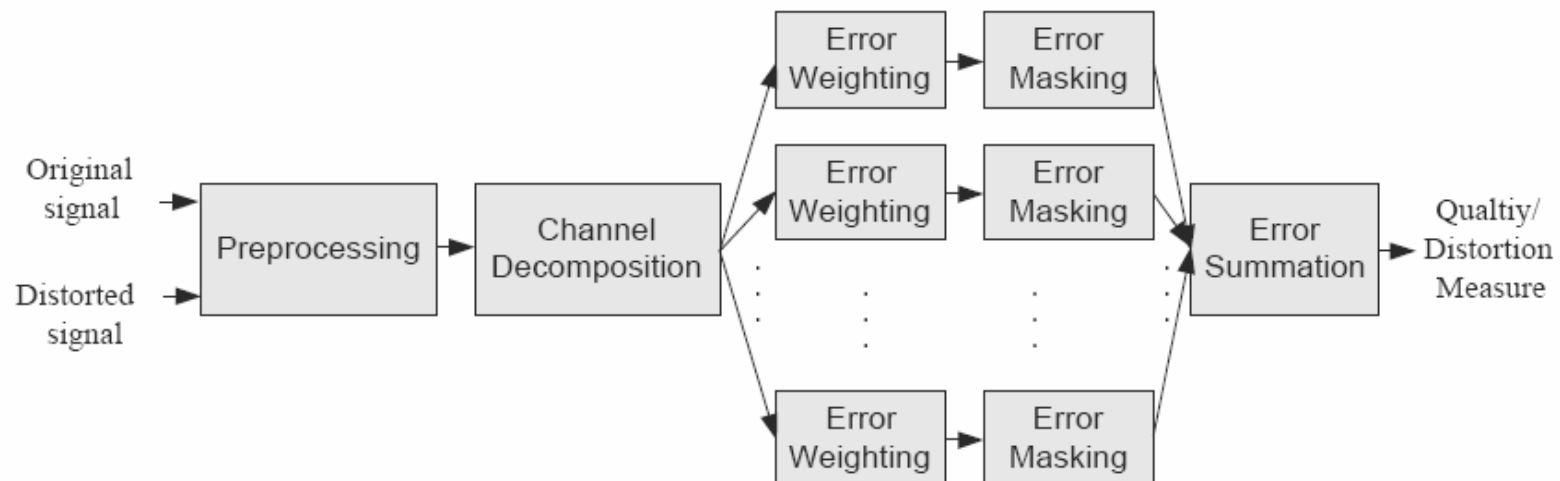
**Vperf Tool**



**NOTE:** *Vperf tool is a modified version of the Iperf tool; code extended from Vinay Chandrashekar's (NCSU) implementation of VBR Iperf*

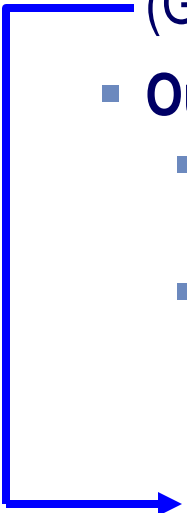
# Existing Objective Techniques

- **ITU-T E-Model** is a success story for VoIP QoE estimation
  - OSC'S H.323 Beacon tool has E-Model implementation
  - It does not apply for VVoIP QoE estimation
    - Designed for CBR voice traffic and handles only voice related impairments
    - Does not address the VBR video traffic and impairments such as video frame freezing
- **ITU-T J.144** (NTIA VQM tool) developed for VVoIP QoE estimation
  - “**PSNR-based MOS**” – PSNR calculation requires original and reconstructed video frames for frame-by-frame comparisons
  - Not suitable for online monitoring
    - PSNR calculation is a time consuming and computationally intensive process



# GAP-Model Framework

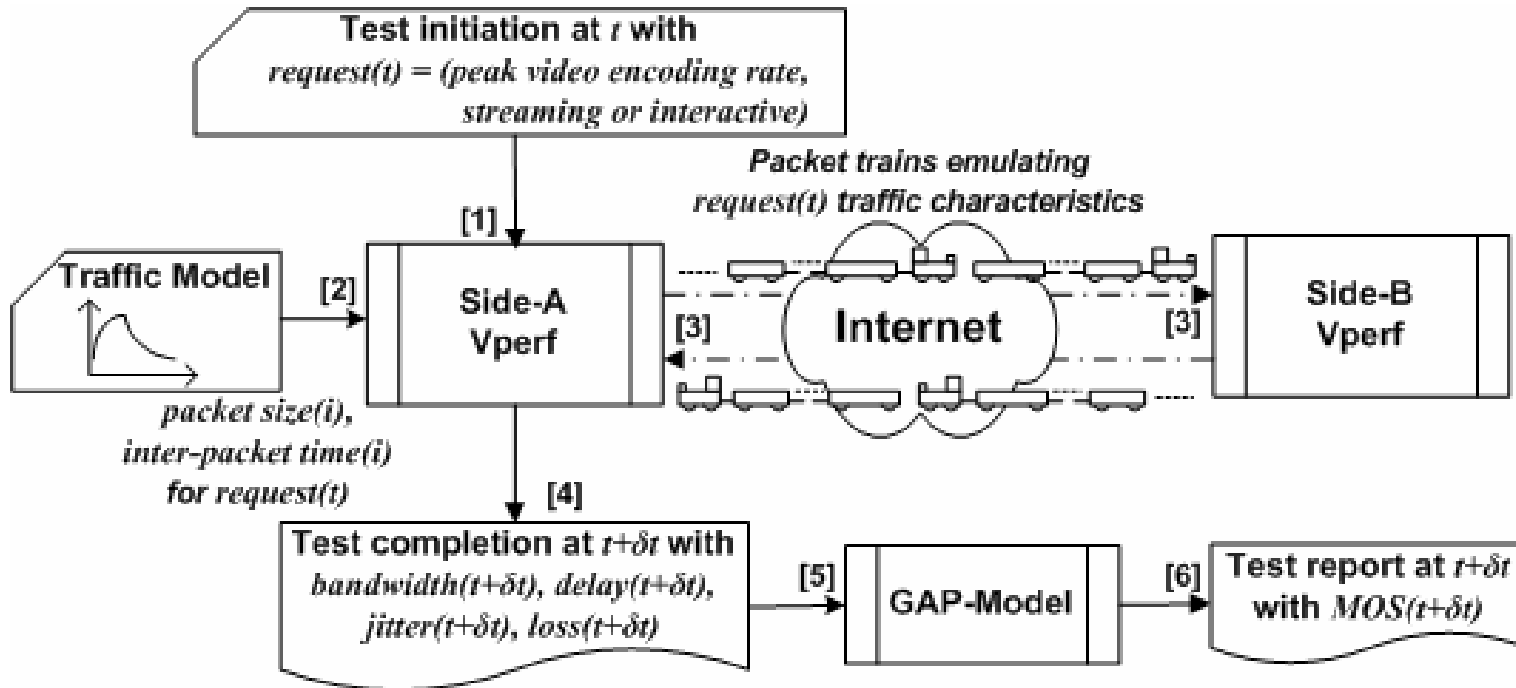
- Earlier studies estimate QoE affected by QoS metrics in isolation
  - E.g. impact due to only bandwidth/delay/loss/jitter
- We consider network health as a combination of different *levels* of bandwidth, delay, jitter and loss – hence more realistic
- The *levels* are quantified by well-known “Good”, “Acceptable” and “Poor” (GAP) performance levels for QoS metrics
- **Our strategy**
  - Derive “closed-form expressions” for modeling MOS using offline human subject studies under different network health conditions
  - Leverage the GAP-Model in Vperf tool for online QoE estimation for a measured set of statistically stable network QoS metrics



<b>Network Factor</b>	<b>Good Grade</b>	<b>Acceptable Grade</b>	<b>Poor Grade</b>
$b_{net}$	(>922] Kbps	(576-922) Kbps	[0-576) Kbps
$d_{net}$	[0-150) ms	(150-300) ms	(>300] ms
$l_{net}$	[0-0.5) %	(0.5-1.5) %	(>1.5] %
$j_{net}$	[0-20) ms	(20-50) ms	(>50] ms

P. Calyam, M. Sridharan, W. Mandrawa, P. Schopis “Performance Measurement and Analysis of H.323 Traffic”, *Passive and Active Measurement Workshop (PAM)*, Proceedings in Springer-Verlag LNCS, 2004.

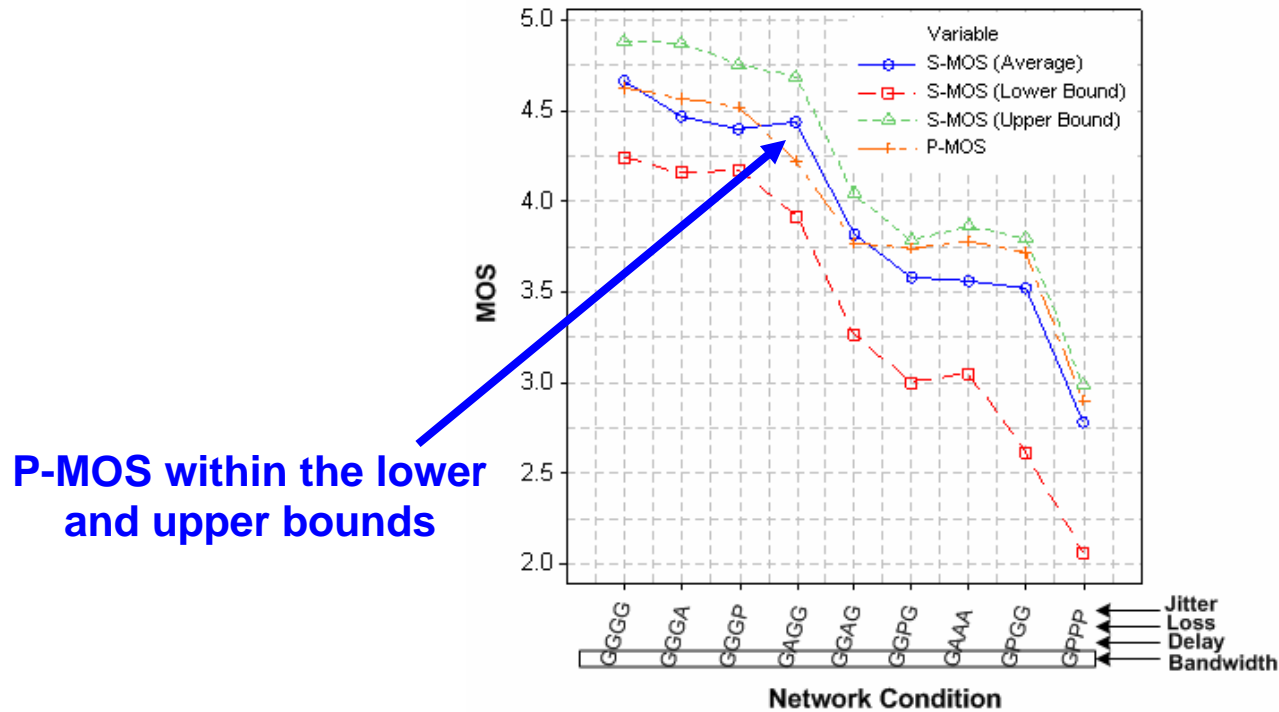
# Vperf Tool Implementation of GAP-Model



- After test duration  $\delta t$ , a set of statistically stable network QoS measurements are obtained
- When input to GAP-Model, online VVoIP QoE estimates are instantly produced

# GAP-Model Validation

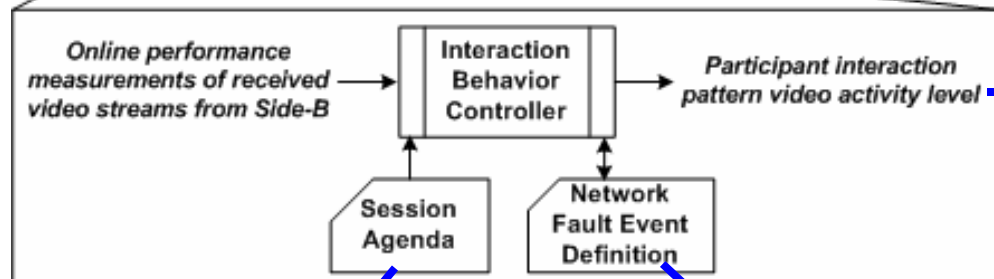
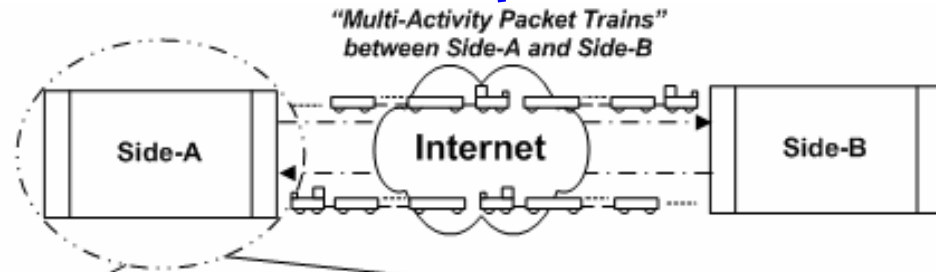
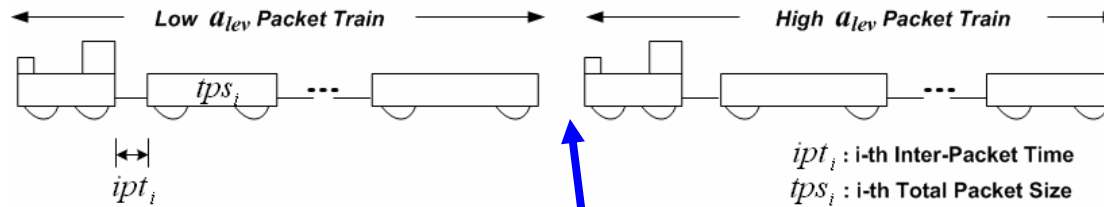
- GAP-Model validation with ITU-T J.144 estimates (P-MOS) and network conditions not tested during model formulation



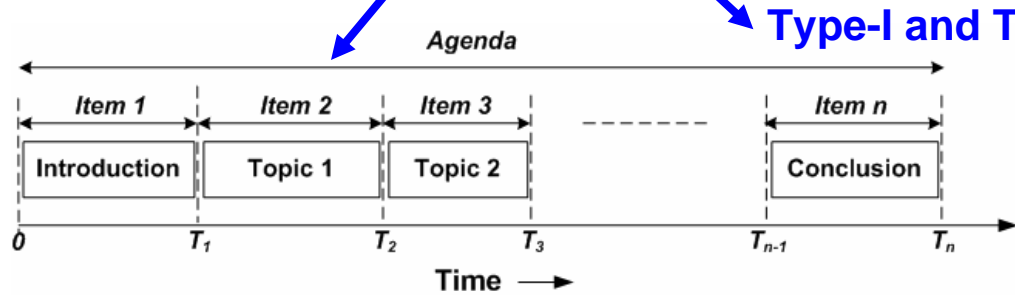
# MAPTs Methodology

- “Multi-Activity Packet Trains” (MAPTs) measure Interaction QoE in an automated manner
  - They mimic participant interaction patterns and video activity levels as affected by network fault events
  - Given a **session-agenda**, excessive talking than normal due to unwanted participant interaction patterns impacts Interaction QoE
  - “Unwanted Agenda-bandwidth” measurement and compare with baseline (consumption during normal conditions)
    - Higher values indicate poor interaction QoE and caution about potential increase in Internet traffic congestion levels
    - Measurements serve as an input for ISPs to improve network performance using suitable traffic engineering techniques

# MAPTs Methodology (2)

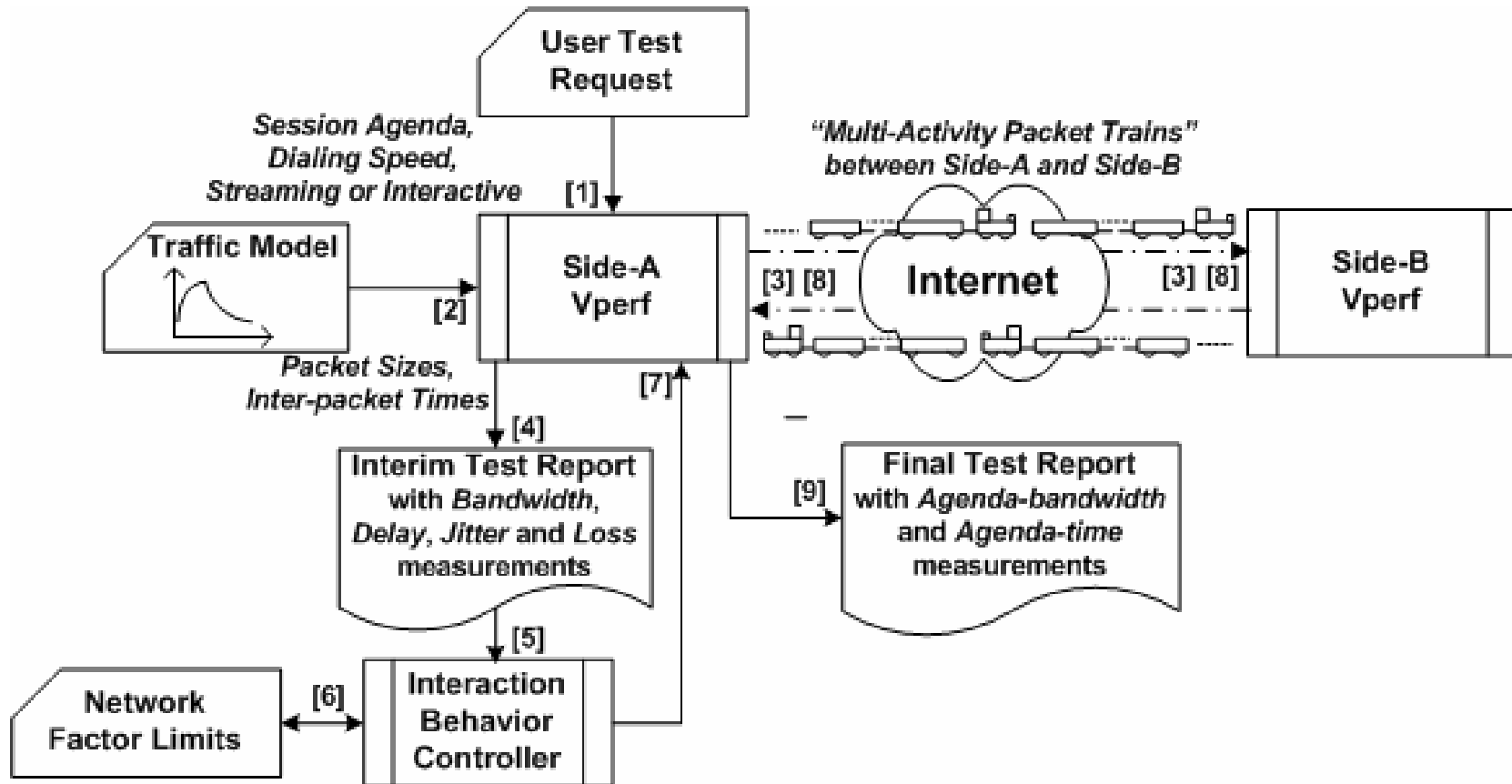


'repeat'  
 'disconnect'  
 'reconnect'  
 'reorient'



Type-I and Type-II fault detection

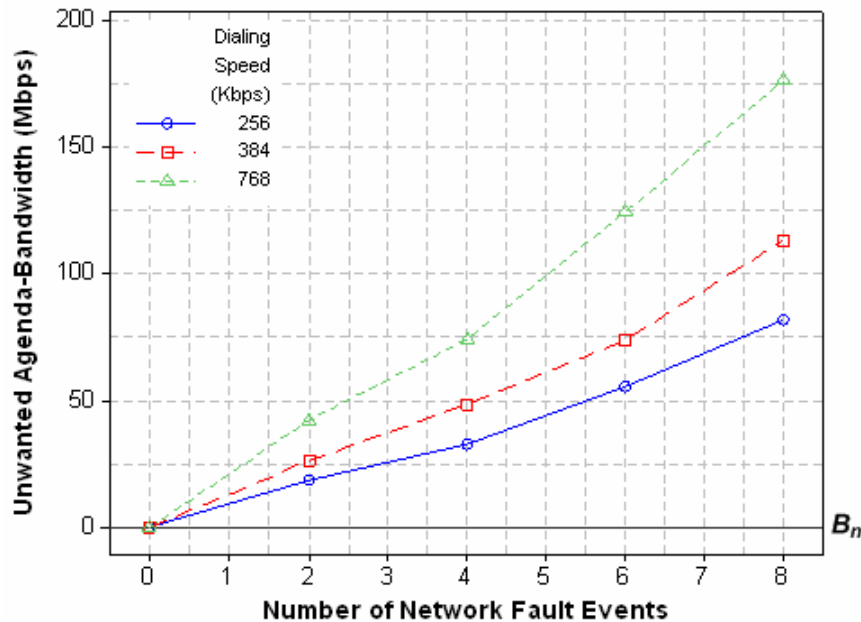
# Vperf Tool Implementation of MAPTs



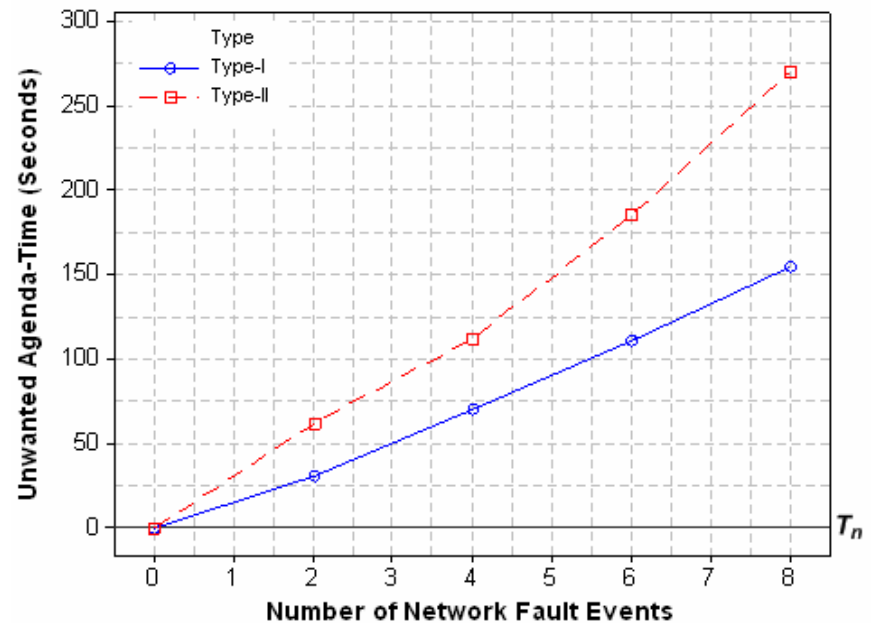
- Per-second frequency of “Interim Test Report” generation
- Interaction QoE reported by Vperf tool - based on the progress of the session-agenda

# MAPTs Measurements Evaluation

- Increased the number of Type-I and Type-II network fault events in a controlled LAN testbed for a fixed session-agenda
  - NISTnet network emulator for network fault generation
- Recorded Unwanted Agenda-Bandwidth and Unwanted Agenda-Time measured by Vperf tool



(a) Impact of Type-I Network Fault Events on Unwanted Agenda-Bandwidth



(b) Impact of Type-I and Type-II Network Fault Events on Unwanted Agenda-Time

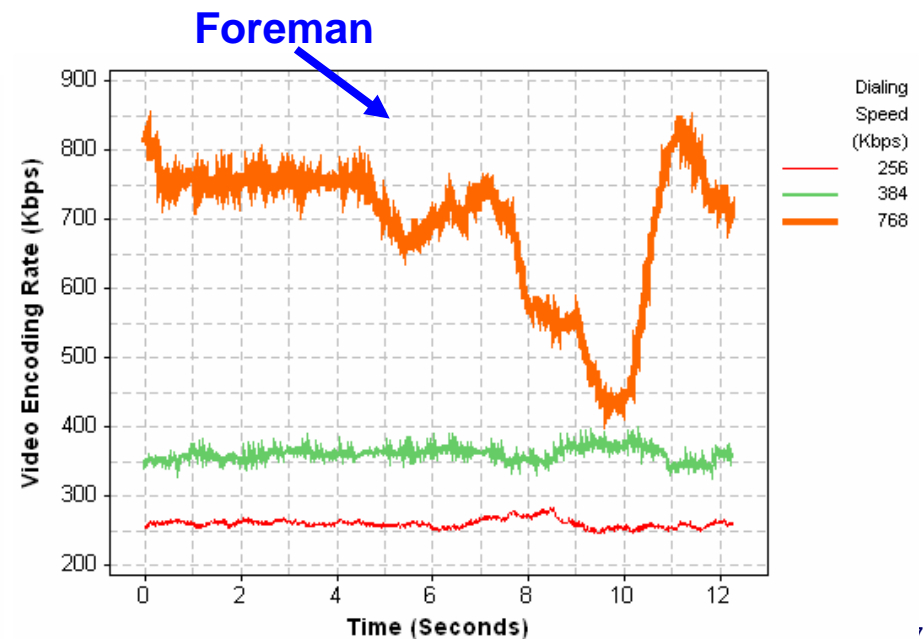
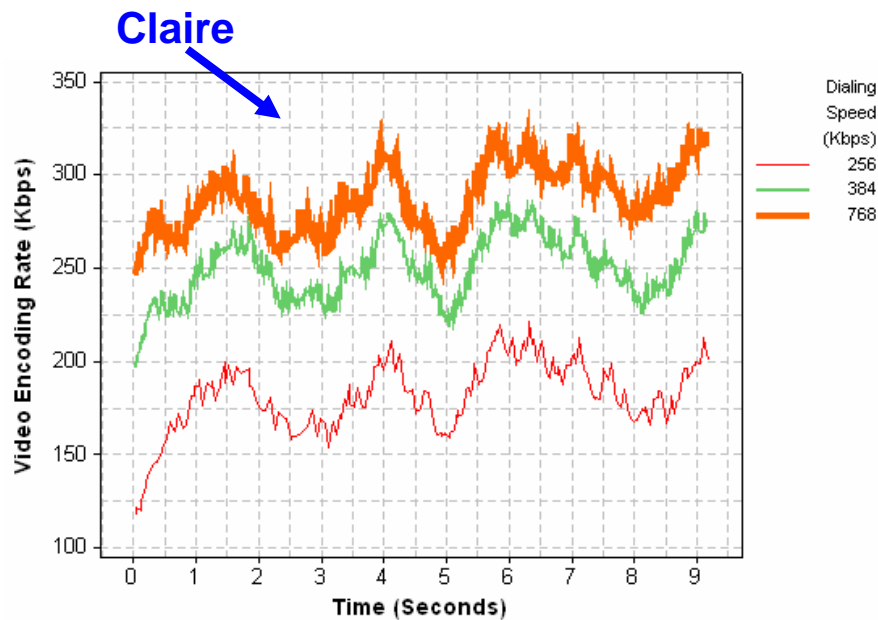
**Thank you for your attention!**



**Any Questions?**

# Video $a_{lev}$

- Low  $a_{lev}$ 
  - Slow body movements and constant background; E.g. *Claire* video sequence
- High  $a_{lev}$ 
  - Rapid body movements and/or quick scene changes; E.g. *Foreman* video sequence
- ‘Listening’ versus ‘Talking’
  - Talking video  $a_{lev}$  (i.e., High) consumes more bandwidth than Listening video  $a_{lev}$  (i.e., Low)



# Example – Session Agenda and Network Factor Limits File

<i>PIP-Type</i>	Side-A Train ( <i>a<sub>lev</sub></i> )	Side-B Train ( <i>a<sub>lev</sub></i> )	Train Duration (Seconds)
<i>PIP<sub>1</sub></i>	<i>L</i>	<i>H</i>	30
	<i>H</i>	<i>L</i>	10
	<i>L</i>	<i>H</i>	30
	<i>H</i>	<i>L</i>	20
	<i>L</i>	<i>H</i>	30
	<i>H</i>	<i>L</i>	20
	<i>L</i>	<i>H</i>	20
	<i>H</i>	<i>L</i>	20
<i>PIP<sub>2</sub></i>	<i>H</i>	<i>H</i>	5 ← <i>Repeat</i>
<i>PIP<sub>3</sub></i>	<i>H</i>	<i>H</i>	5 ← <i>Disconnect</i>
	<i>N</i>	<i>N</i>	15 ← <i>Reconnect</i>
	<i>H</i>	<i>H</i>	5 ← <i>Reorient</i>

# Traffic Model for MAPTs Emulation

- Traffic Model for probing packet trains obtained from trace-analysis
  - Combine popularly used low and high  $a_{lev}$  video sequences and model them at 256/384/768 Kbps dialing speeds for H.263 video codec
  - Low – *Grandma, Kelly, Claire, Mother/Daughter, Salesman*
  - High – *Foreman, Car Phone, Tempete, Mobile, Park Run*
- Modeling
  - Video Encoding Rates ( $b_{snd}$ ) time series
  - Packet Size ( $tps$ ) distribution
  - Derived instantaneous inter-packet times ( $tps$ ) by dividing instantaneous packet sizes by video encoding rates

