Using SCTP to Improve QoS and Network Fault-Tolerance of DRE Systems

OOMWorks LLC	BBN Technologies	LM ATL
Metuchen, NJ	Cambridge, MA	Camden, NJ
Yamuna Krishnamurthy	Craig Rodrigues	Gautam Thaker
Irfan Pyarali	Prakash Manghwani	

Real-Time and Embedded Systems Workshop

July 12-15, 2004

Reston, Virginia, USA

BBN TECHNOLOGIES A Verizon Company

OOMWORKS

LOCKHEED MARTIN

Transport Protocol Woes in DRE Systems

UDP	TCP
Unreliable data delivery	Non-Deterministic Congestion control
Unordered data delivery	No control over key parameters like retransmission timeout
No Network Fault-Tolerance	No Network Fault-Tolerance

Stream Control Transport Protocol (SCTP)

- IP based transport protocol originally designed for telephony signaling
- SCTP supports features found to be useful in TCP or UDP
  - Reliable data transfer (TCP)
  - Congestion control (TCP)
  - Message boundary conservation (UDP)
  - Path MTU discovery and message fragmentation (TCP)
  - Ordered (TCP) and unordered (UDP) data delivery
- Additional features in SCTP
  - Multi-streaming: multiple independent data flows within one association
  - Multi-homed: single association runs across multiple network paths
  - Security and authentication: checksum, tagging and a security cookie mechanism to prevent SYN-flood attacks
- Multiple types of service
  - SOCK_SEQPACKET – message oriented, reliable, ordered/unordered
  - SOCK_STREAM – TCP like byte oriented, reliable, ordered
  - SOCK_RDM – UDP like message oriented, reliable, unordered
- Control over key parameters like retransmission timeout and number of retransmissions

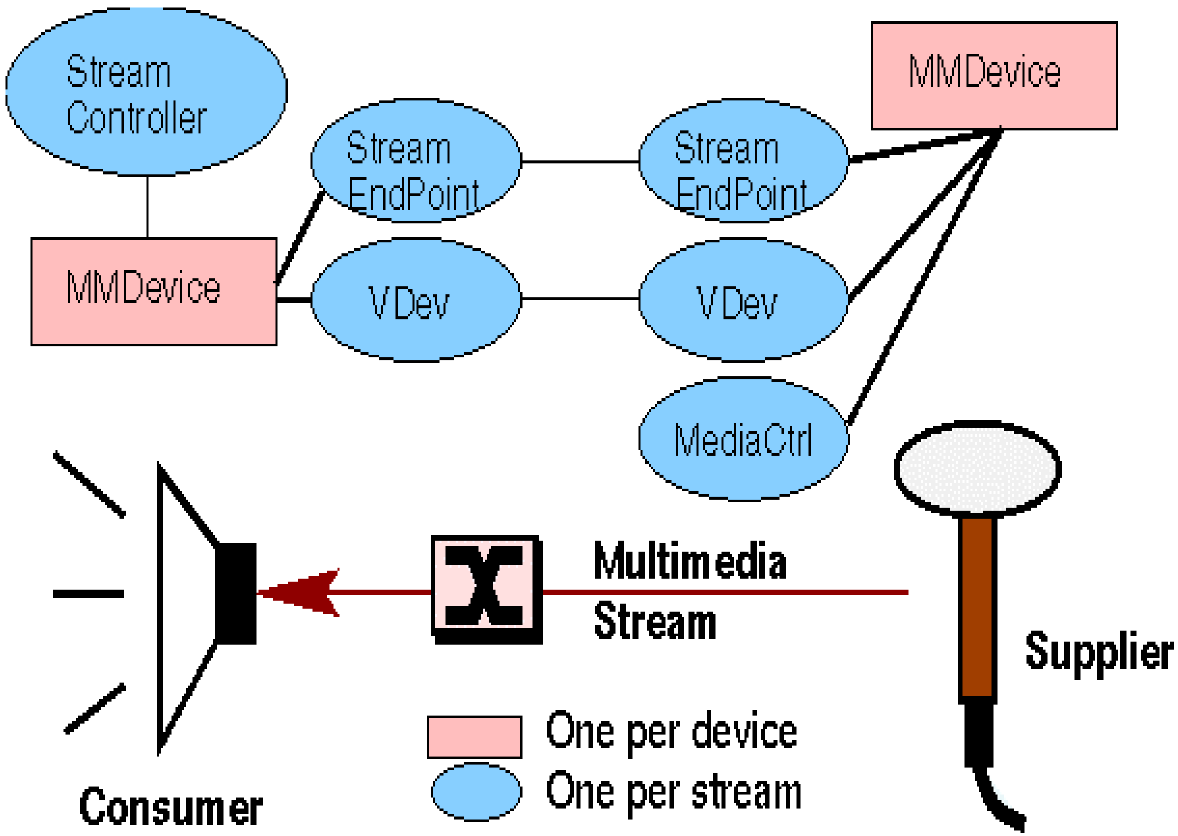
Integration of SCTP in DRE Systems

- Issues with integrating SCTP directly
  - Possible re-design of the system
  - Accidental and Incidental errors
  - Evolving API
  - Multiple SCTP implementations

- Solution
  - Integrate SCTP via COTS Middleware
  - SCTP was added as a Transport Protocol for GIOP messages (SCIOP)
  - *OMG TC Document mars/2003-05-03*
    - *Bound recovery time of CORBA objects after a network failure*
  - SCTP was added as a Transport Protocol to CORBA Audio/Video Streaming Service

- Proof of concept: Integration of SCTP in UAV Application
  - UAV is a reconnaissance image data streaming DRE application

CORBA AVStreams Overview



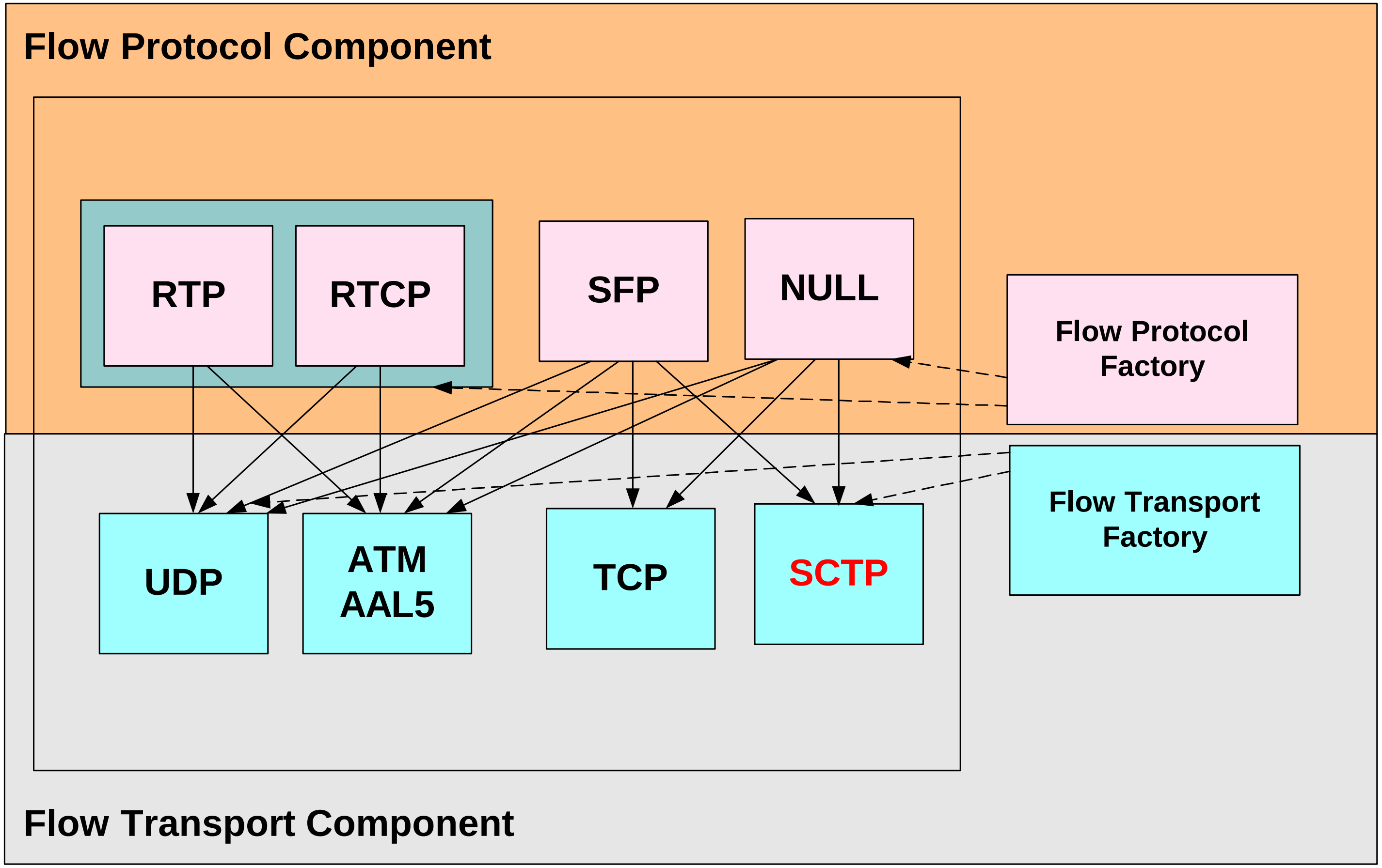
StreamCtrl
- controlling the stream

MMDevice
- interface for logical or physical device

StreamEndPoint
- network specific aspects of a stream

Vdev
- properties of the stream

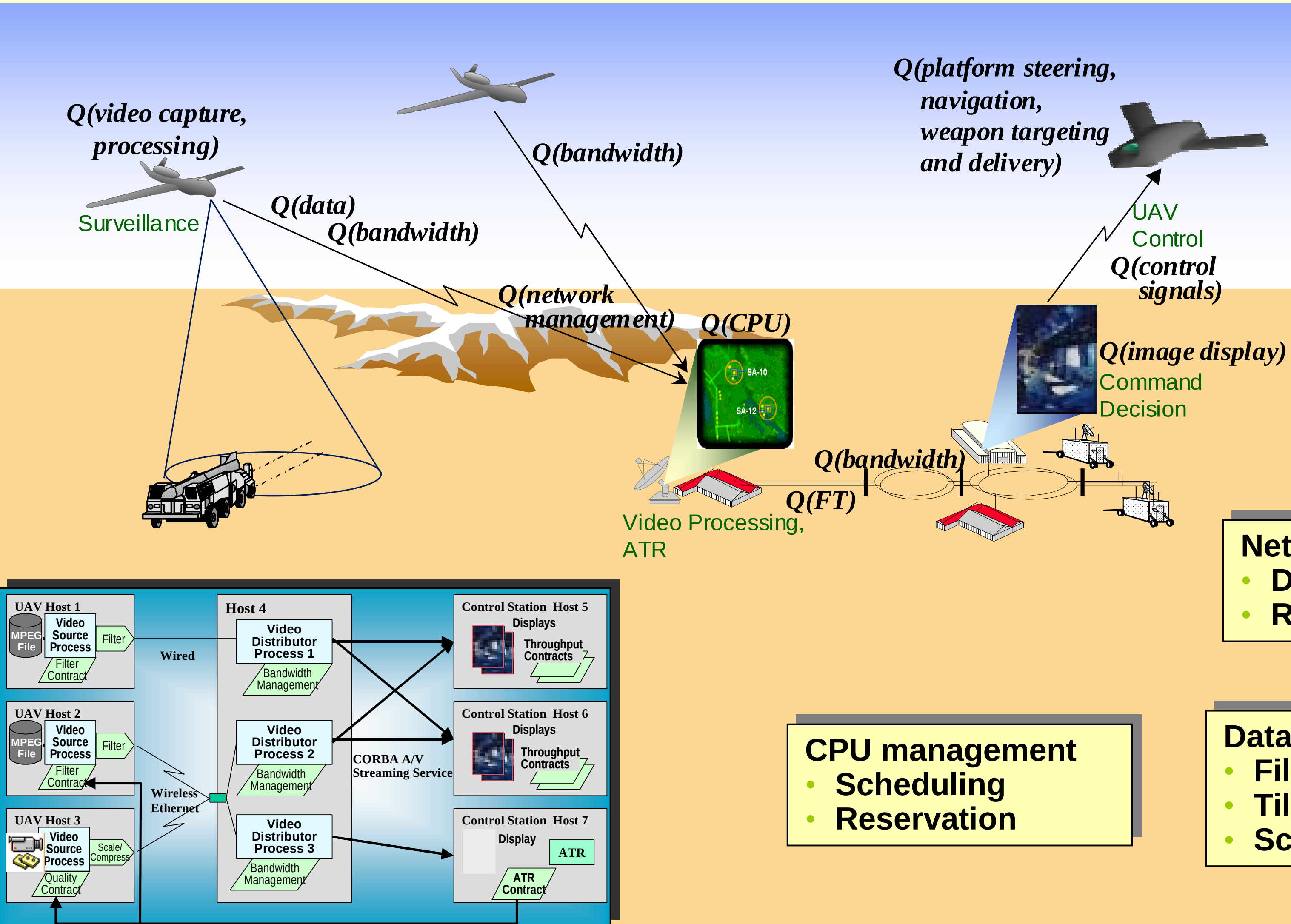
AVStreams Pluggable Protocol Framework



- SCTP added as a transport protocol

Distributed UAV Image Dissemination

Mission Requirements	Piloting	Surveillance	Command/Control
	• Requires an out-of-the-window view of imagery	• Must not lose any important imagery	• Require high fidelity imagery



Adaptation Strategies

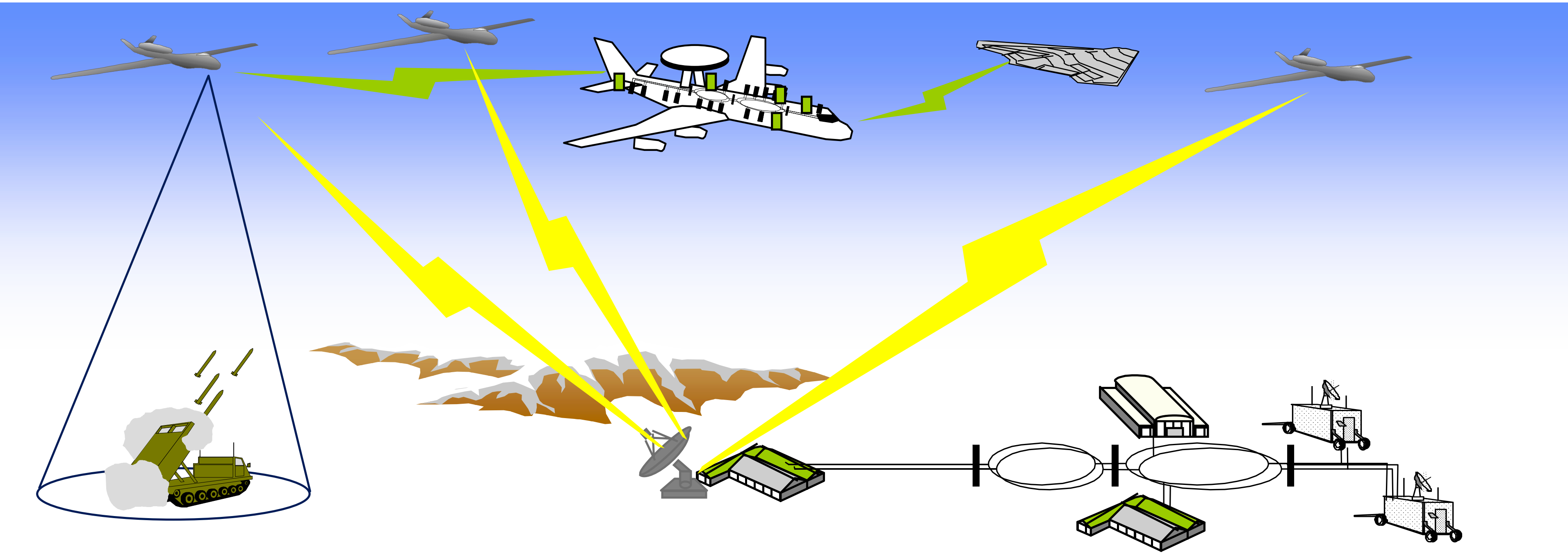
Load balancing
- Migrating tasks to less loaded hosts

Network management
- Diffserv
- Reservation

CPU management
- Scheduling
- Reservation

Data management
- Filtering
- Tiling, compression
- Scaling

Open Experimental Platform Scenario



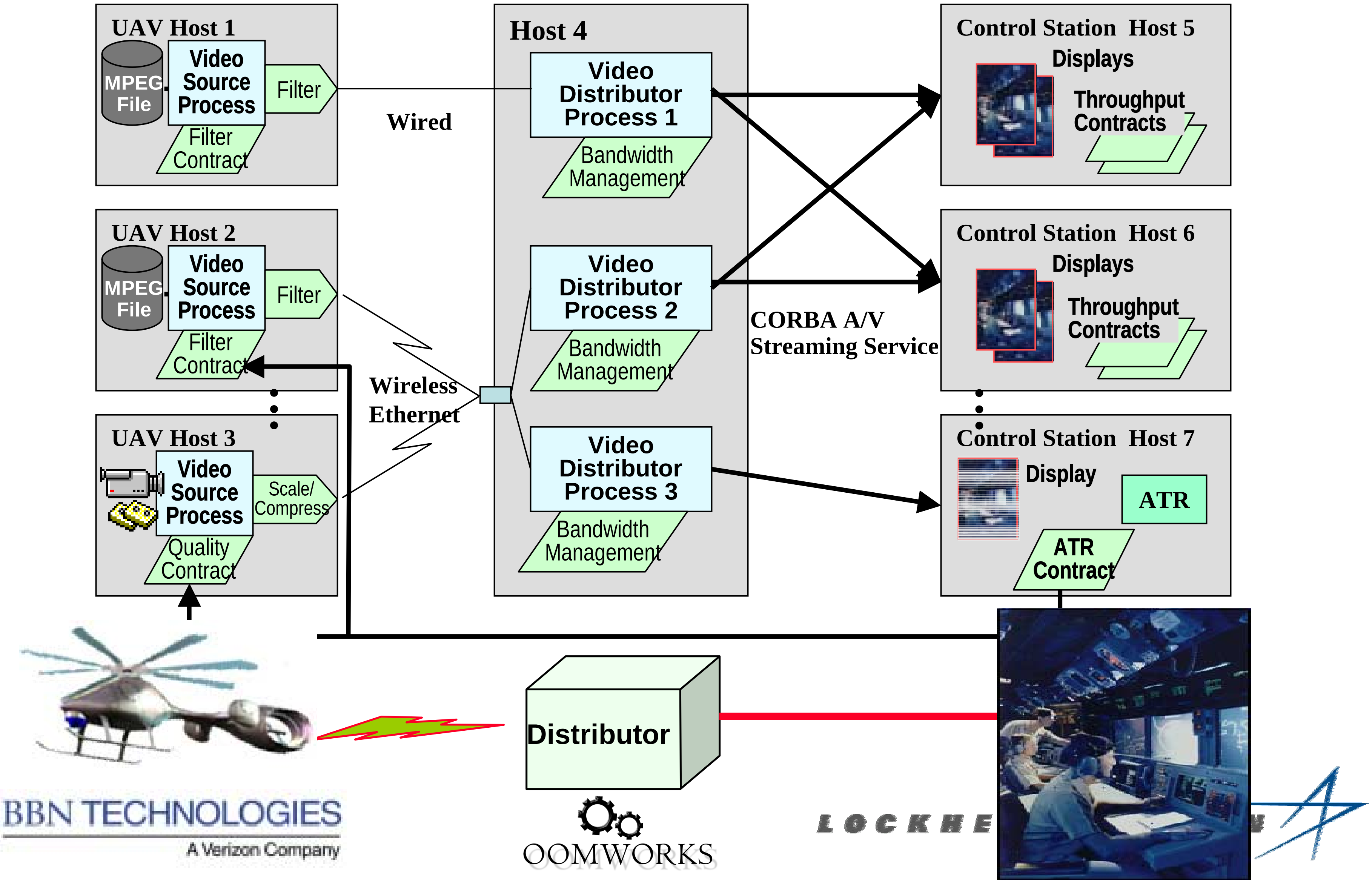
- End-to-end, delivery of video (and other forms of sensor) data, with control feeding back based upon data content and requirements
- Simultaneously, various forms of control stations receive real time video tailored to their needs, display it, process it, disseminate it, and
- Users act upon UAV information and interact with UAVs and other subsystems in real-time

BBN TECHNOLOGIES
A Verizon Company

OOMWORKS

LOCKHEED MARTIN

High Level View of UAV OEP Architecture

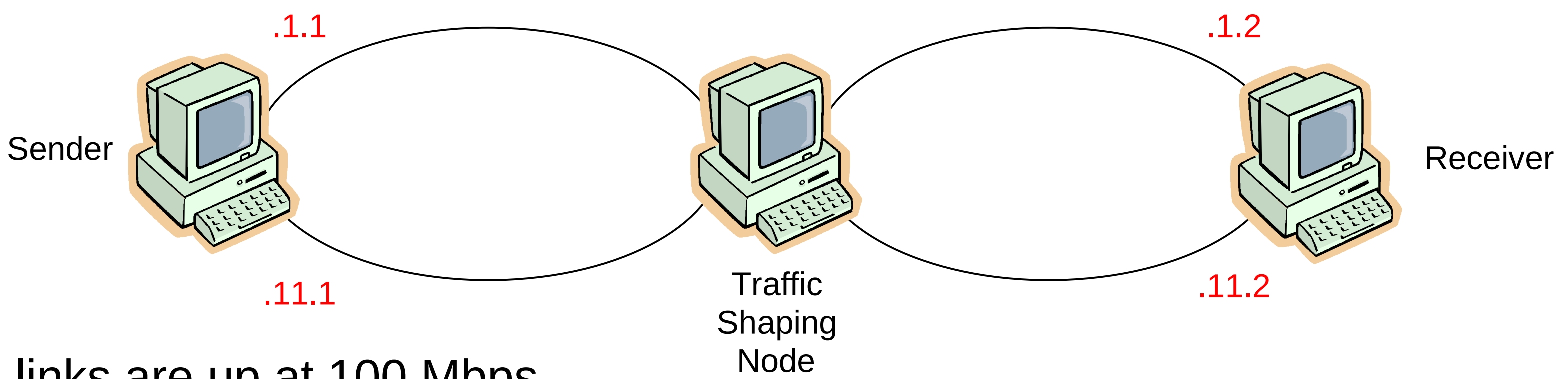


Experiments

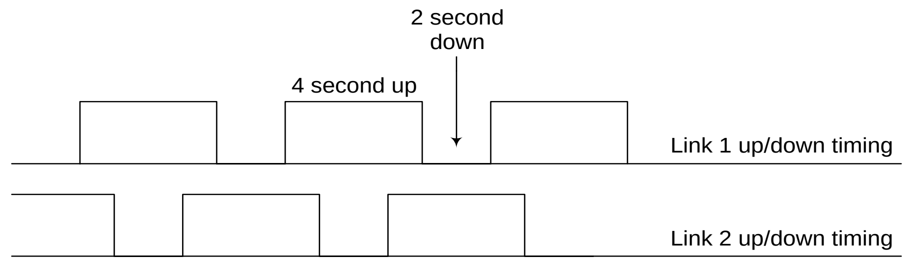
Protocol Analysis and Comparisons

- Several different experiments were performed
  - Paced invocations
  - Throughput tests
  - Latency tests
- Several different protocols were used
  - TCP based IIOP
    - `-ORBEndpoint iiop://host:port`
  - UDP based DIOP
    - `-ORBEndpoint diop://host:port`
  - SCTP based SCIOP
    - `-ORBEndpoint sciop://host1+host2:port`

Primary Configuration

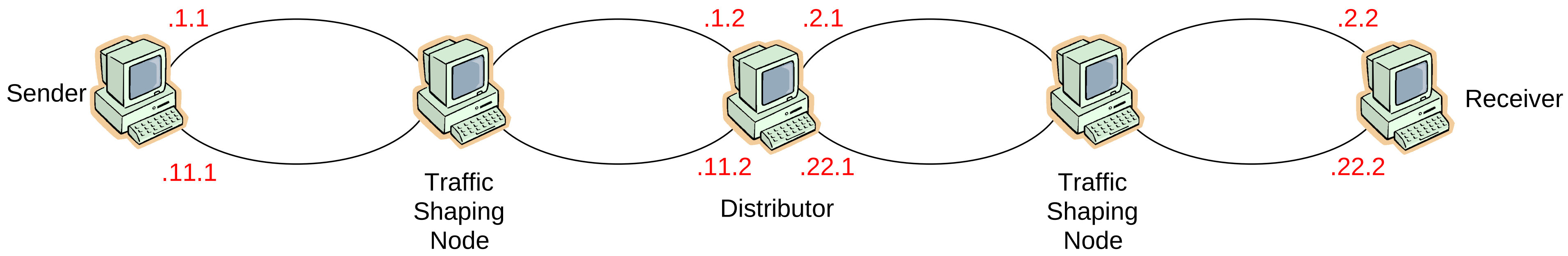


- Both links are up at 100 Mbps
- 1st link has 1% packet loss
- Systematic link failures
  - Up 4 seconds, down 2 seconds
  - One link is always available



- Host configuration
  - RedHat 9 with OpenSS7 SCTP 0.2.19 in kernel (BBN-RH9-SS7-8)
  - RedHat 9 with LKSCTP 2.6.3-2.1.196 in kernel (BBN-RH9-LKSCTP-3)
  - Pentium III, 850 MHz, 1 CPU, 512 MB RAM
  - ACE+TAO+CIAO version 5.4.1+1.4.1+0.4.1
  - gcc version 3.2.2

Secondary Configuration



- Results as expected:
  - Roundtrip latency about doubled
  - Throughput more or less the same
  - Little effect on paced invocations
- SCTP failures on Distributor
  - Maybe related to 4 network cards
  - Maybe related to the inability to specify addresses for local endpoints
- Sender, Distributor, and Receiver were normal CORBA applications
- Similar results were also obtained when AVStreaming was used to transfer data between these components

Modified SCTP Parameters

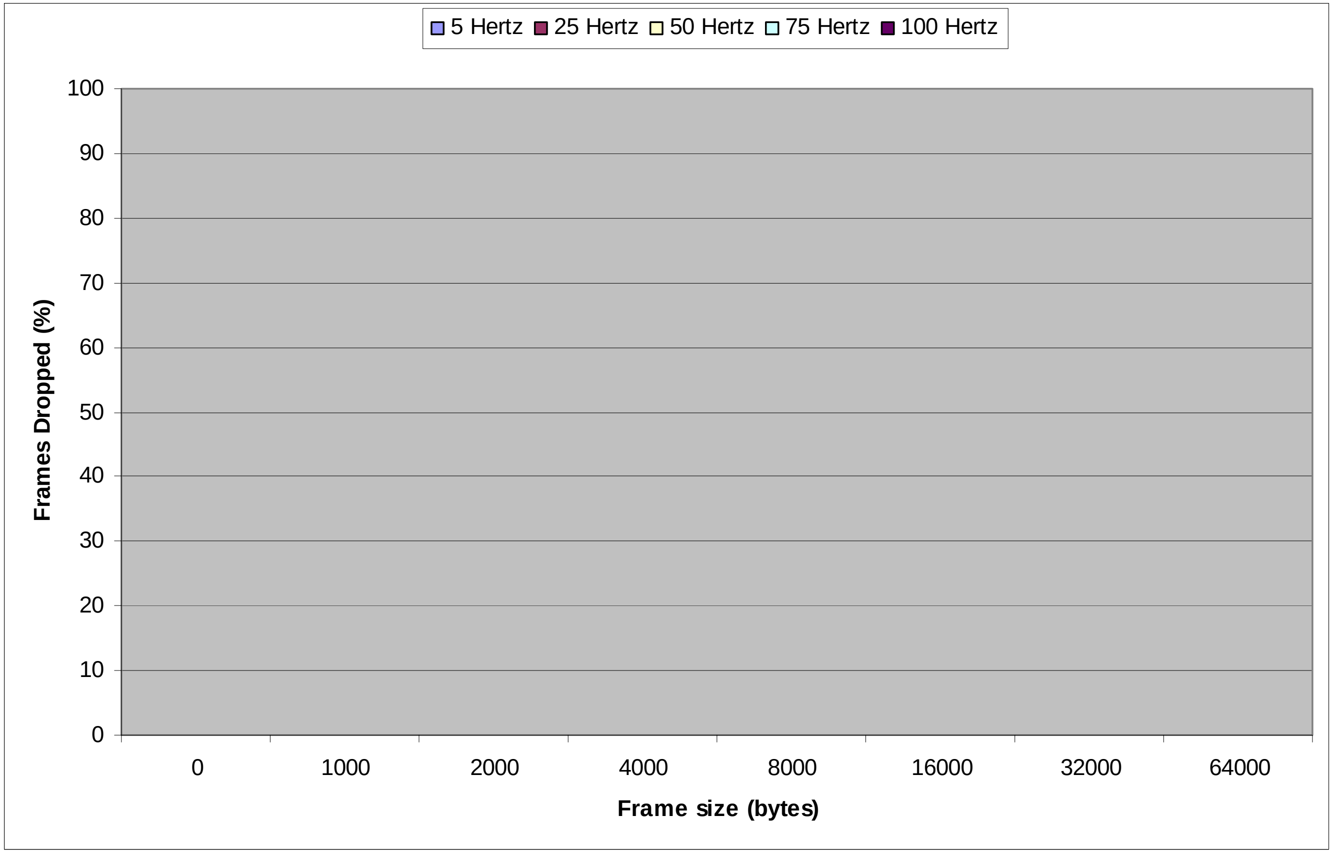
Parameter	Default	OpenSS7	LKSCTP
RTO initial	3000	0	10
RTO min	1000	0	10
RTO max	60000	0	10
Heartbeat interval	30	1	10
SACK delay max	200	0	?
Path retrans max	5	0	0
Association retrans max	10	25	25
Initial retries	8	25	25

- LKSCTP was less reliable, had higher latency and lower throughput relative to OpenSS7 in almost every test
- Also had errors when sending large frames

Paced Invocations

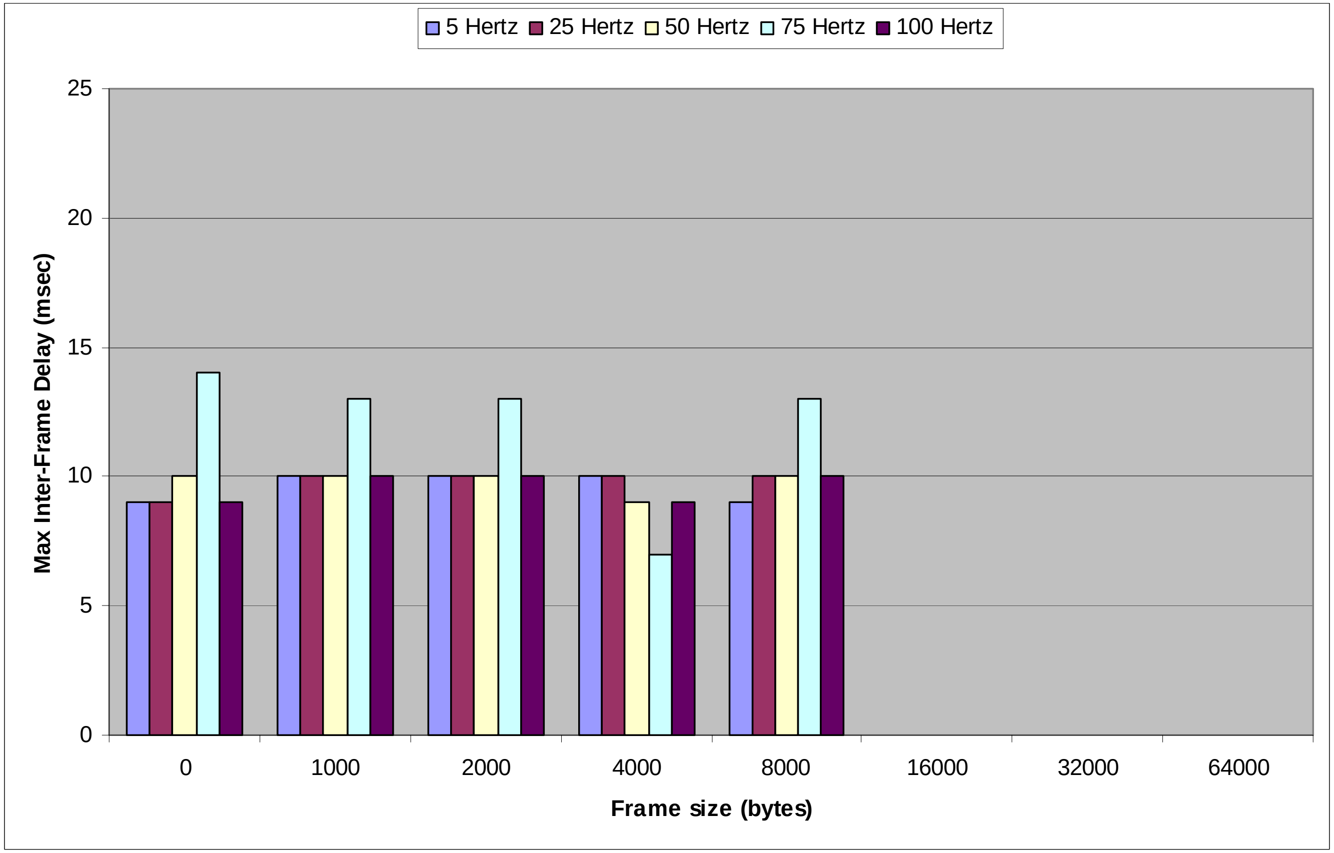
- Emulating rate monotonic systems and audio/visual applications
- Three protocols: IIOP, DIOP, SCIOP
- Frame size was varied from 0 to 64k bytes
  - DIOP frame size was limited to a maximum of 8k bytes
- Invocation Rate was varied from 5 to 100 Hertz
- IDL interface
  - **`one-way void method(in octets payload)`**
- Experiment measures
  - Maximum inter-frame delay at the server
  - Number of frames that were received at the server

Protocol = DIOP, Experiment = Paced Invocations
Network = Both links are up



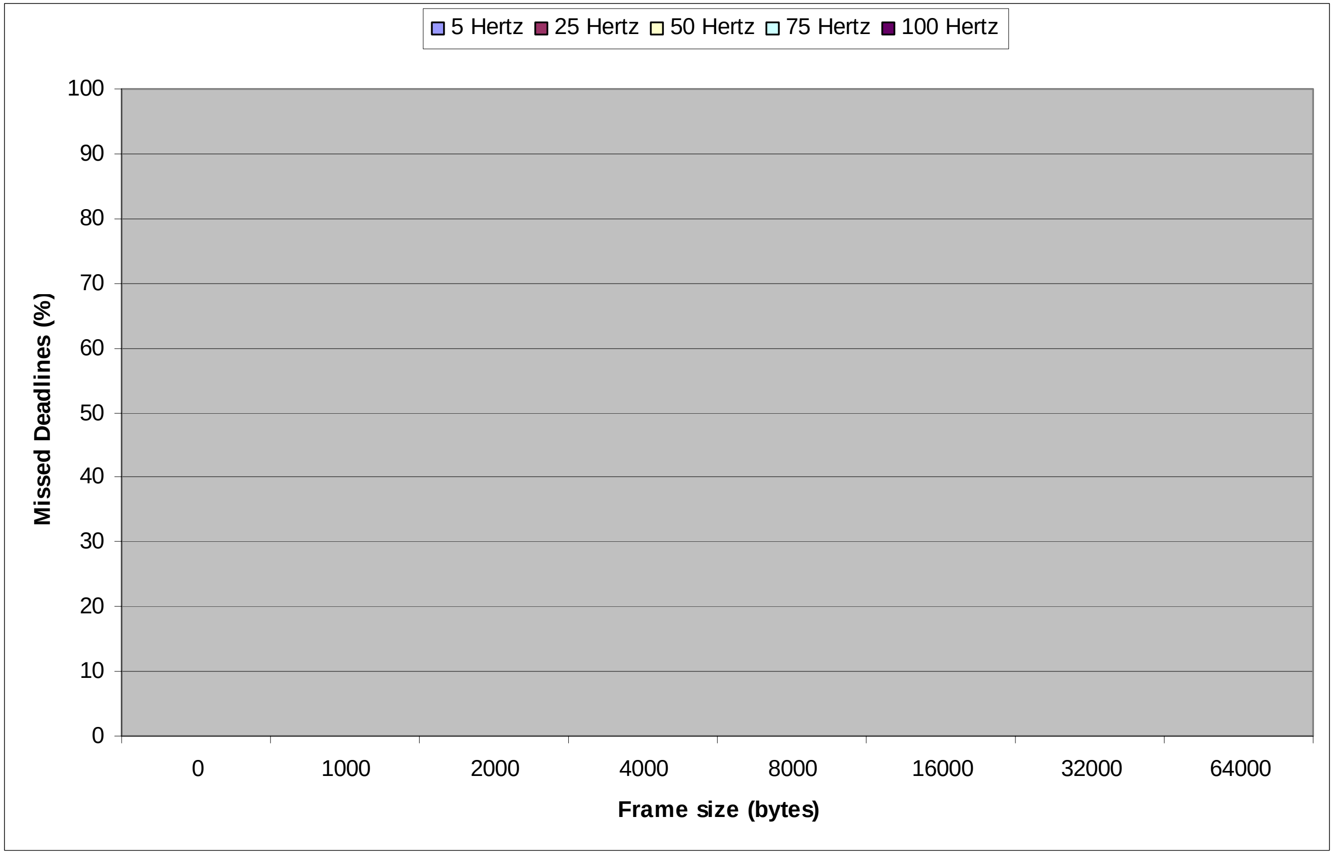
- Since network capacity was not exceeded, no DIOP packets were dropped
- Cannot measure missed deadlines when using DIOP because the client always succeeds in sending the frame, though the frame may not reach the server

Protocol = DIOP, Experiment = Paced Invocations
Network = Both links are up



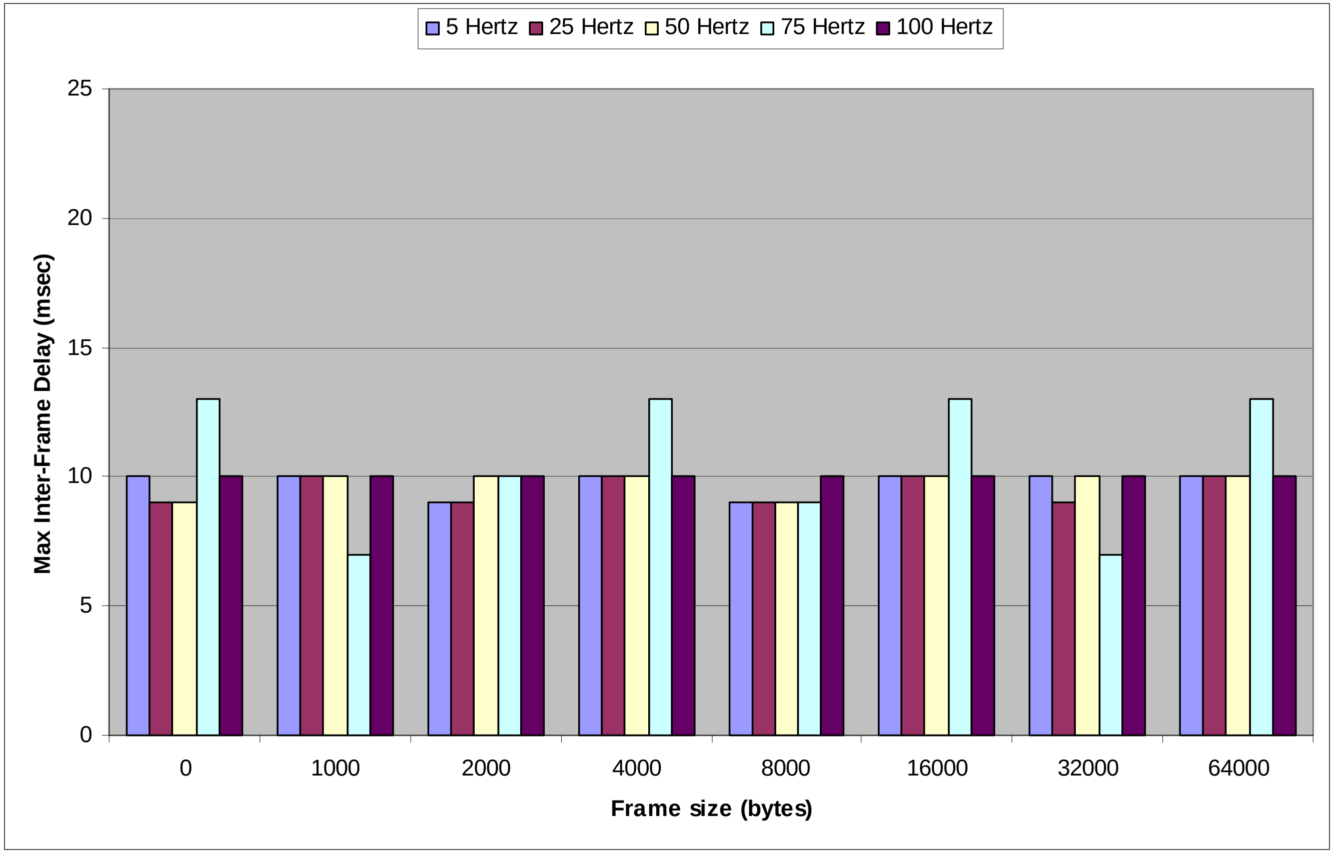
- Very low inter-frame delay
- DIOP good choice for reliable links or when low latency is more desirable then reliable delivery
- Drawback: Frame size limited to about 8k

Protocol = IIOP, Experiment = Paced Invocations
Network = Both links are up



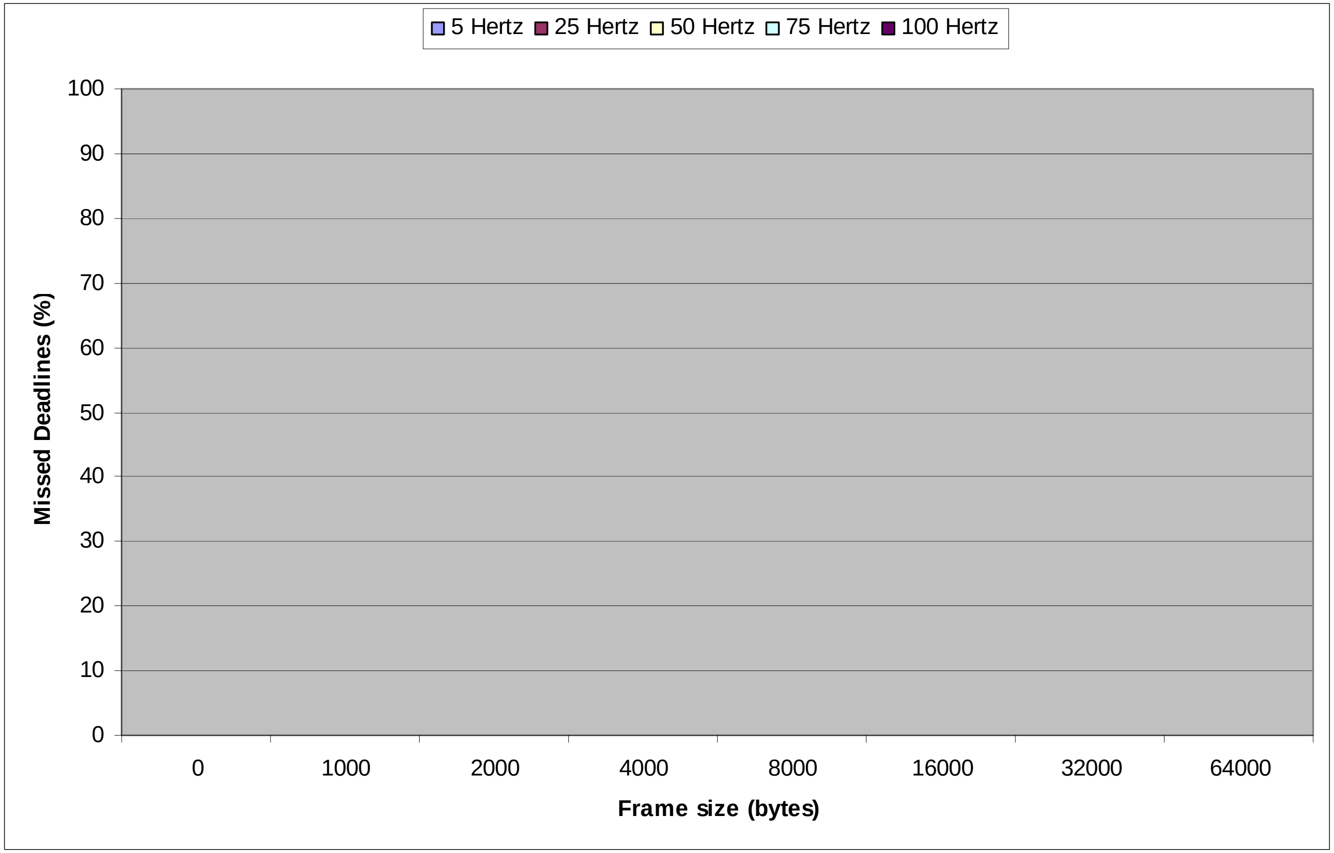
- Under normal conditions, no deadlines were missed

Protocol = IIOP, Experiment = Paced Invocations
Network = Both links are up



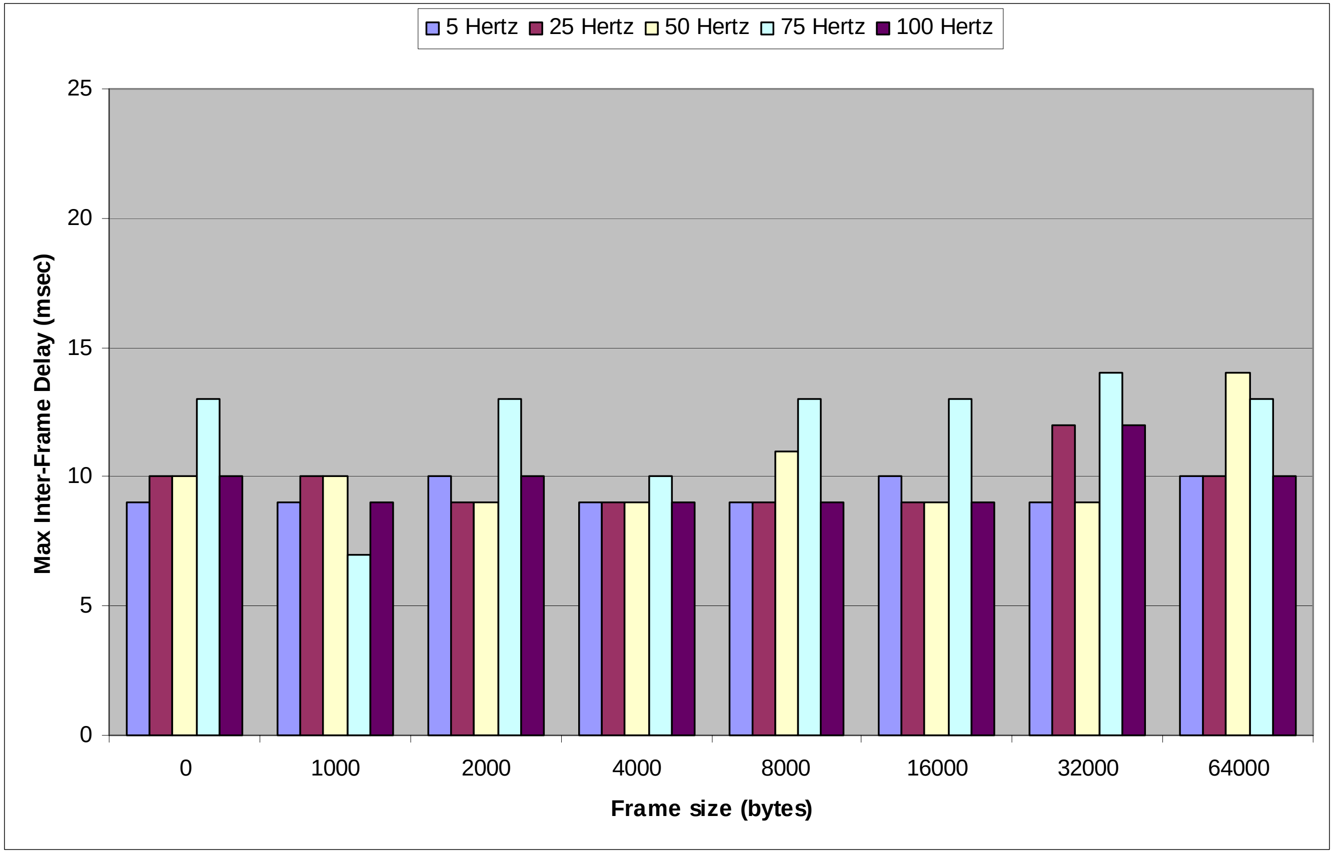
- Very low inter-frame delay
- IIOP good choice in most normal situations

Protocol = SCIOP, Experiment = Paced Invocations
Network = Both links are up



- Under normal conditions, very comparable to DIOP and IIOP

Protocol = SCIOP, Experiment = Paced Invocations
Network = Both links are up



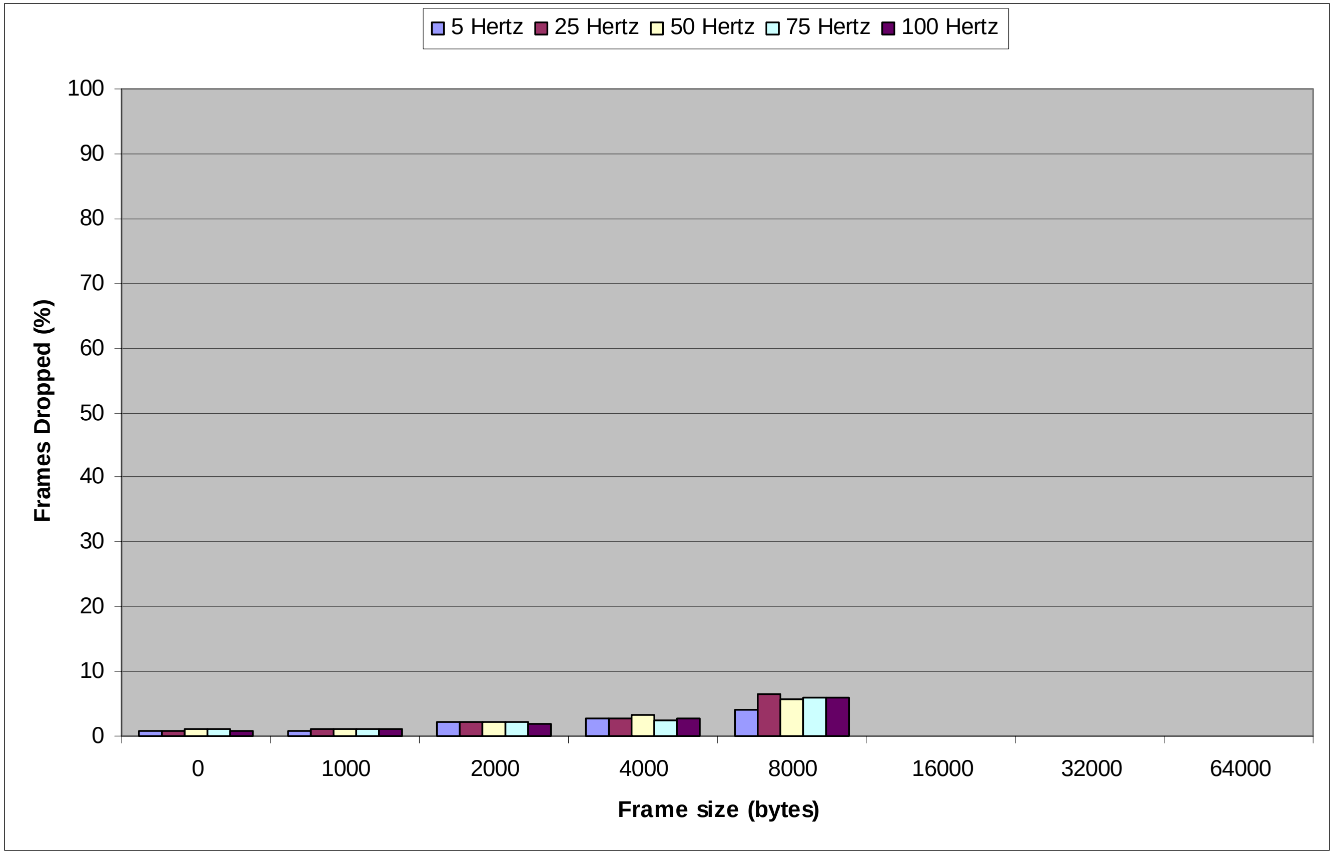
- Under normal conditions, very comparable to DIOP and IIOP

BBN TECHNOLOGIES A Verizon Company · OOMWORKS · LOCKHEED MARTIN

Summary of Experiments under Normal Network Conditions

- Under normal conditions, performance of DIOP, IIOP, and SCIOP is quite similar
- No disadvantage of using SCIOP under normal conditions

Protocol = DIOP, Experiment = Paced Invocations
Network = 1% packet loss on 1st link



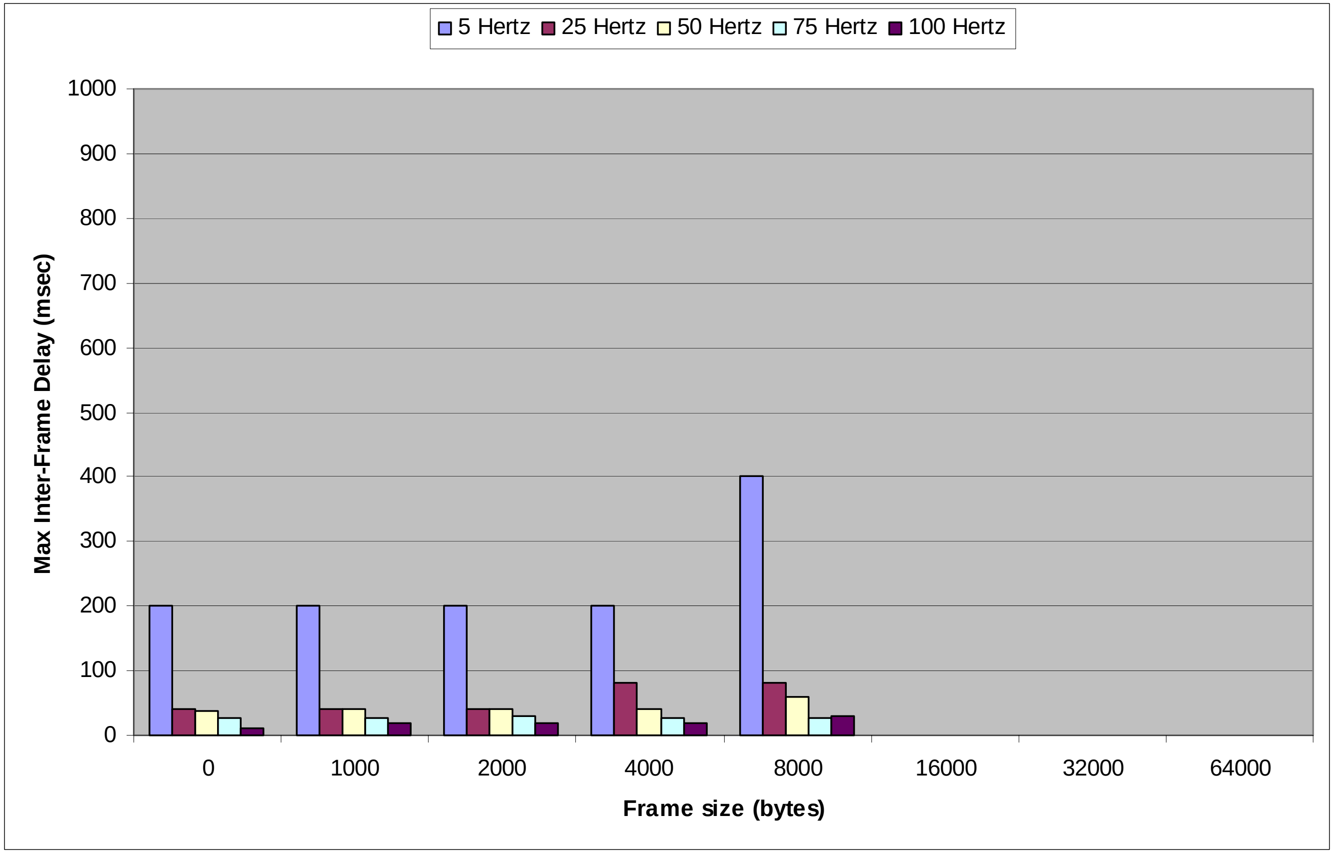
- Packet loss introduced at Traffic Shaping Node causes frames to be dropped
- 1% to 7% frames were dropped – higher loss for bigger frames

BBN TECHNOLOGIES
A Verizon Company

OOMWORKS

LOCKHEED MARTIN

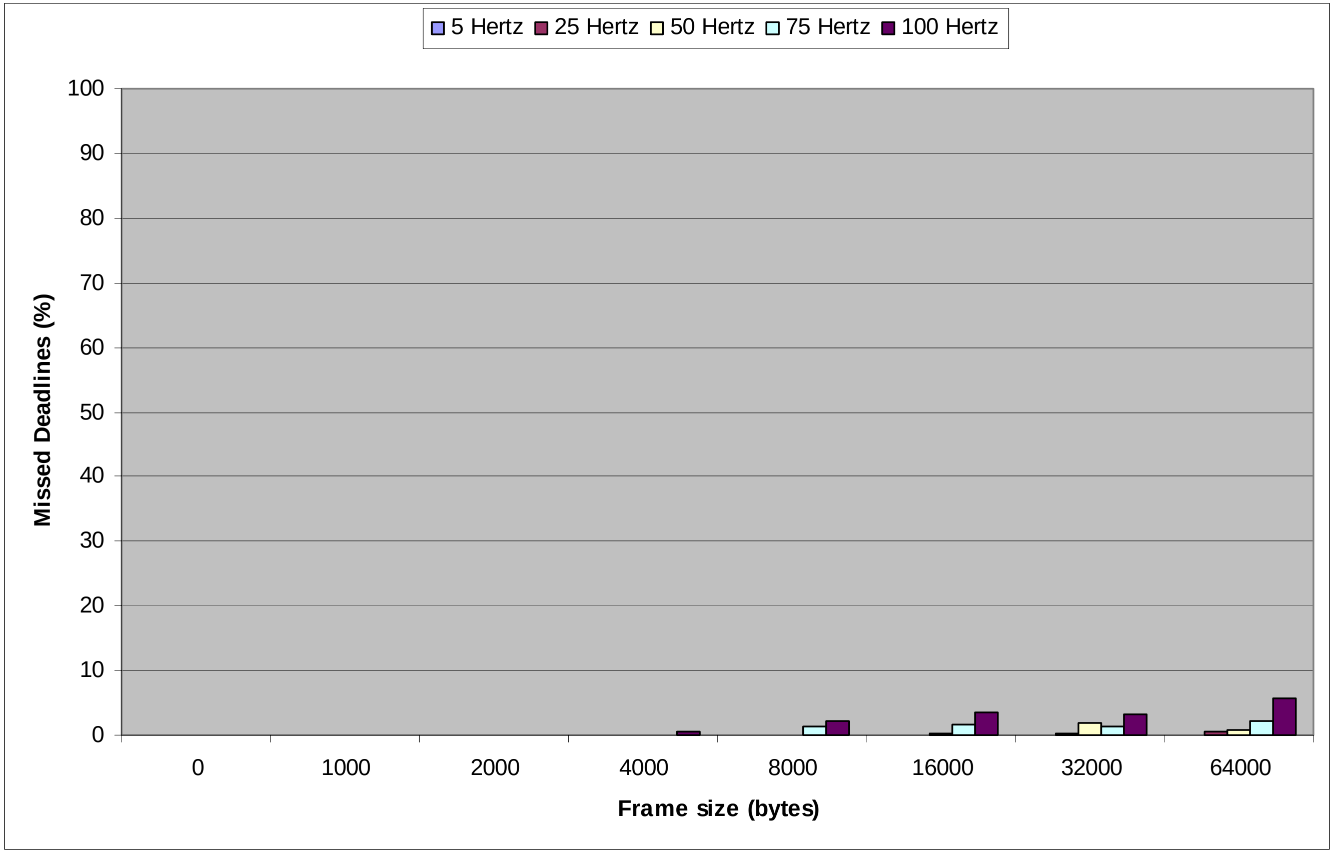
Protocol = DIOP , Experiment = Paced Invocations Network = 1% packet loss on 1st link



- Increase in inter-frame delay due to lost frames
- Slowest invocation rate has highest inter-frame delay

BBN TECHNOLOGIES A Verizon Company OOMWORKS LOCKHEED MARTIN

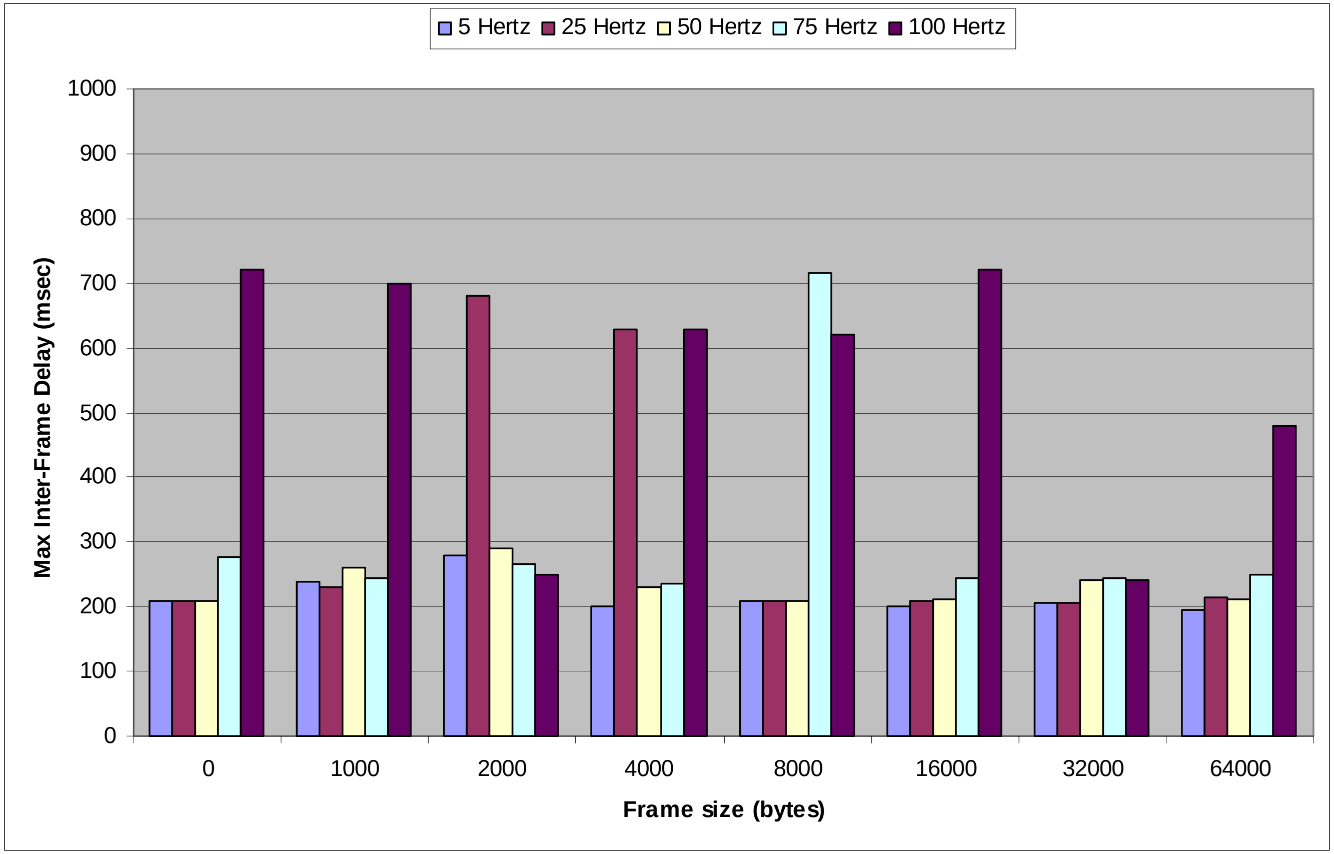
Protocol = IIOP, Experiment = Paced Invocations
Network = 1% packet loss on 1st link



- Packet loss did not have significant impact on smaller frames or slower invocation rates since there is enough time for the lost packets to be retransmitted
- Packet loss did impact larger frames, specially the ones being transmitted at faster rates – 6% of 64k bytes frames at 100 Hz missed their deadlines

BBN TECHNOLOGIES A Verizon Company OOMWORKS LOCKHEED MARTIN

Protocol = IIOP, Experiment = Paced Invocations
Network = 1% packet loss on 1st link



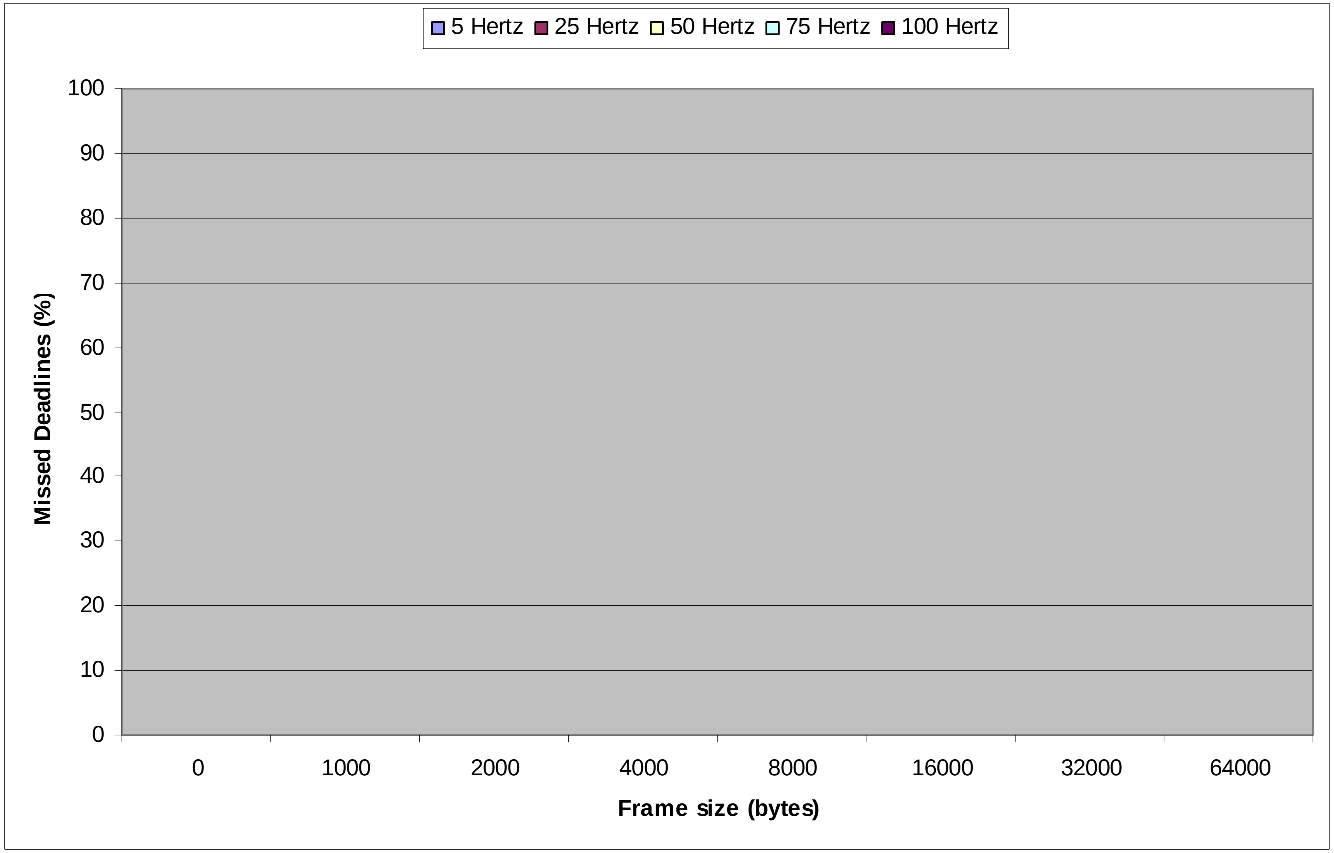
- IIOP does not recover well from lost packets
- 720 msec delay for some frames (only 13 msec under normal conditions)

BBN TECHNOLOGIES A Verizon Company

OOMWORKS

LOCKHEED MARTIN

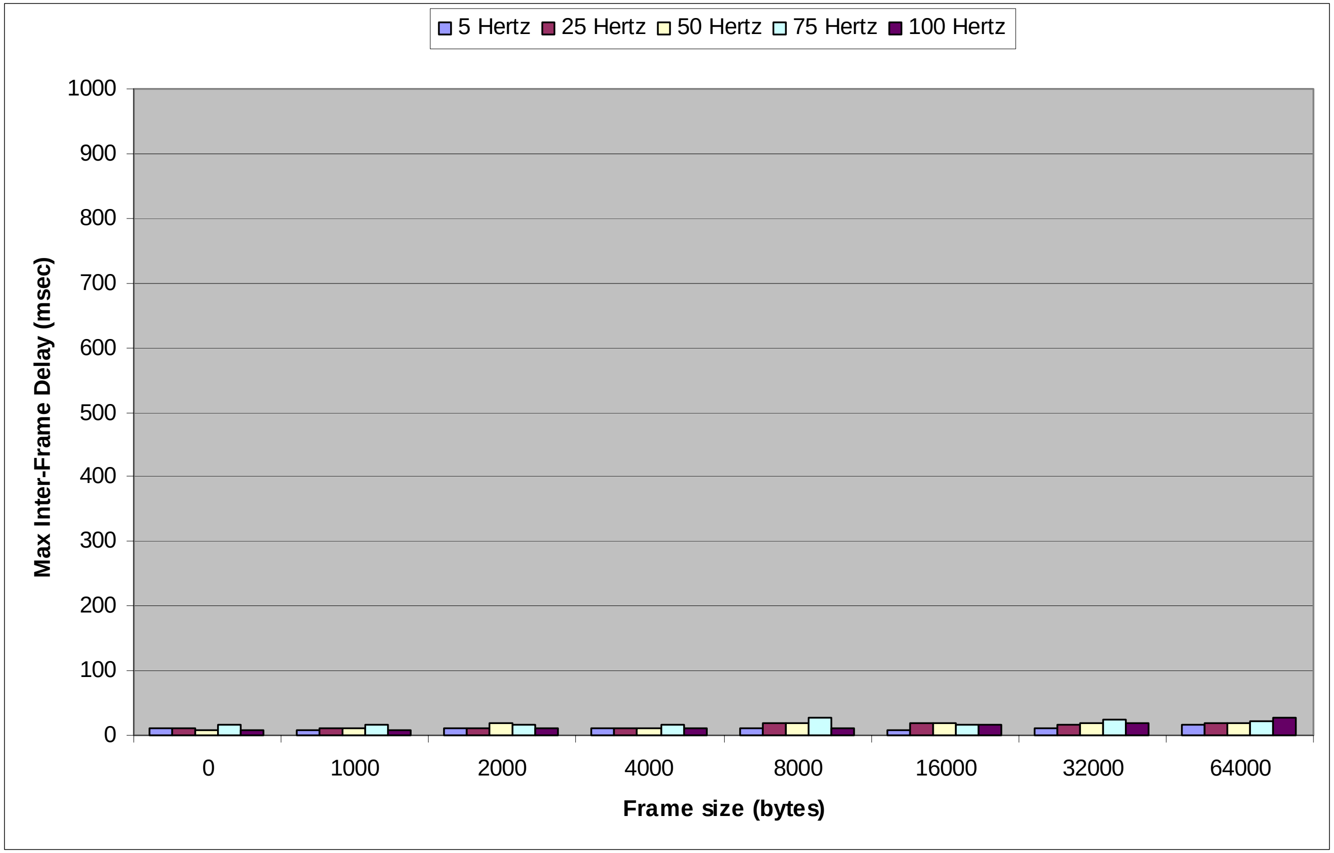
Protocol = SCIOP, Experiment = Paced Invocations
Network = 1% packet loss on 1st link



- SCIOP was able to use the redundant link during packet loss on the primary link
- No deadlines were missed

BBN TECHNOLOGIES A Verizon Company   OOMWORKS   LOCKHEED MARTIN

Protocol = SCIOP, Experiment = Paced Invocations
Network = 1% packet loss on 1st link



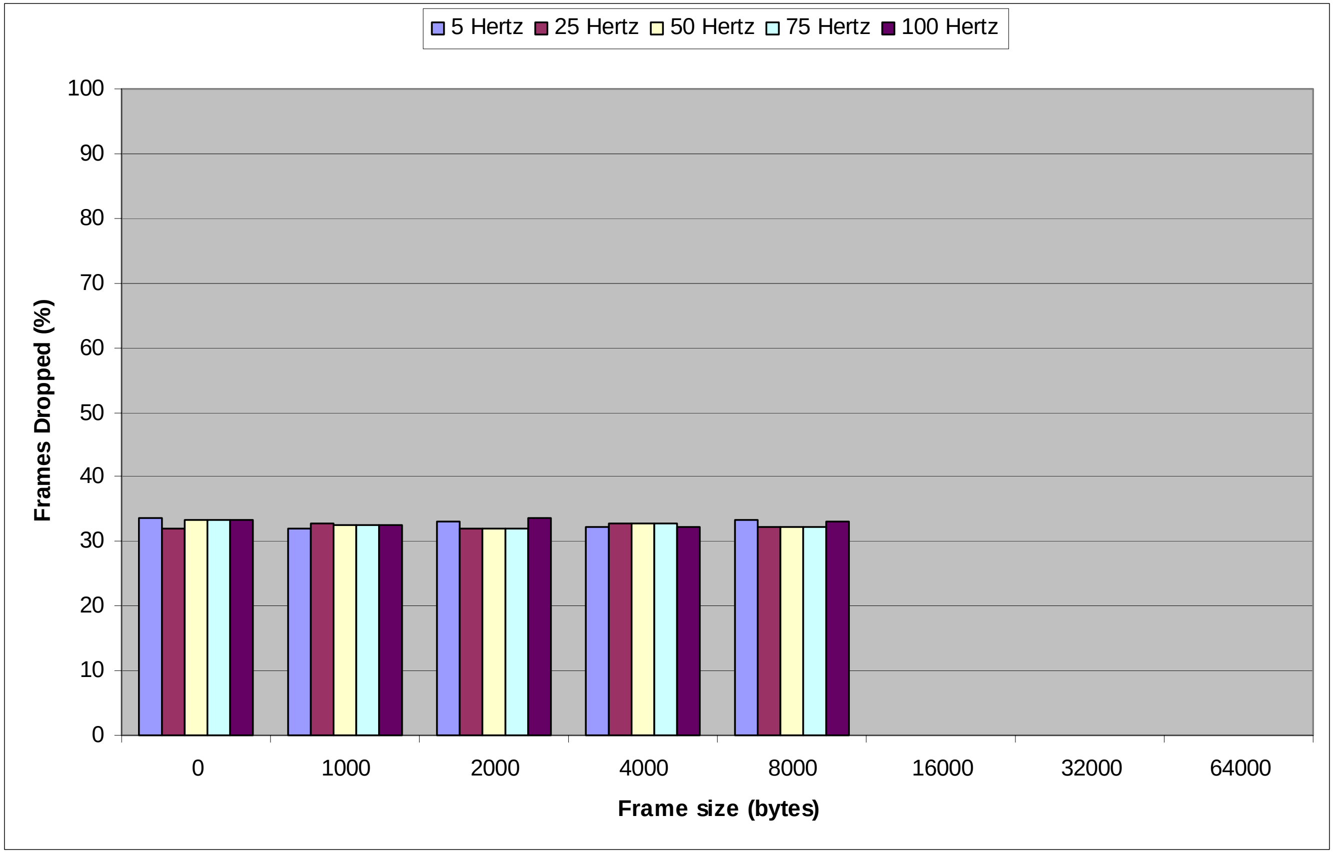
- Inter-frame delay in this experiment is very comparable to the inter-frame delay under normal conditions (26 msec vs. 14 msec)

BBN TECHNOLOGIES A Verizon Company · OOMWORKS · LOCKHEED MARTIN

Summary of Experiments under 1% packet loss on 1st link

- Client was able to meet all of its invocation deadlines when using SCIOP
- DIOP dropped up to 7% of frames
- IIOP missed up to 6% of deadlines

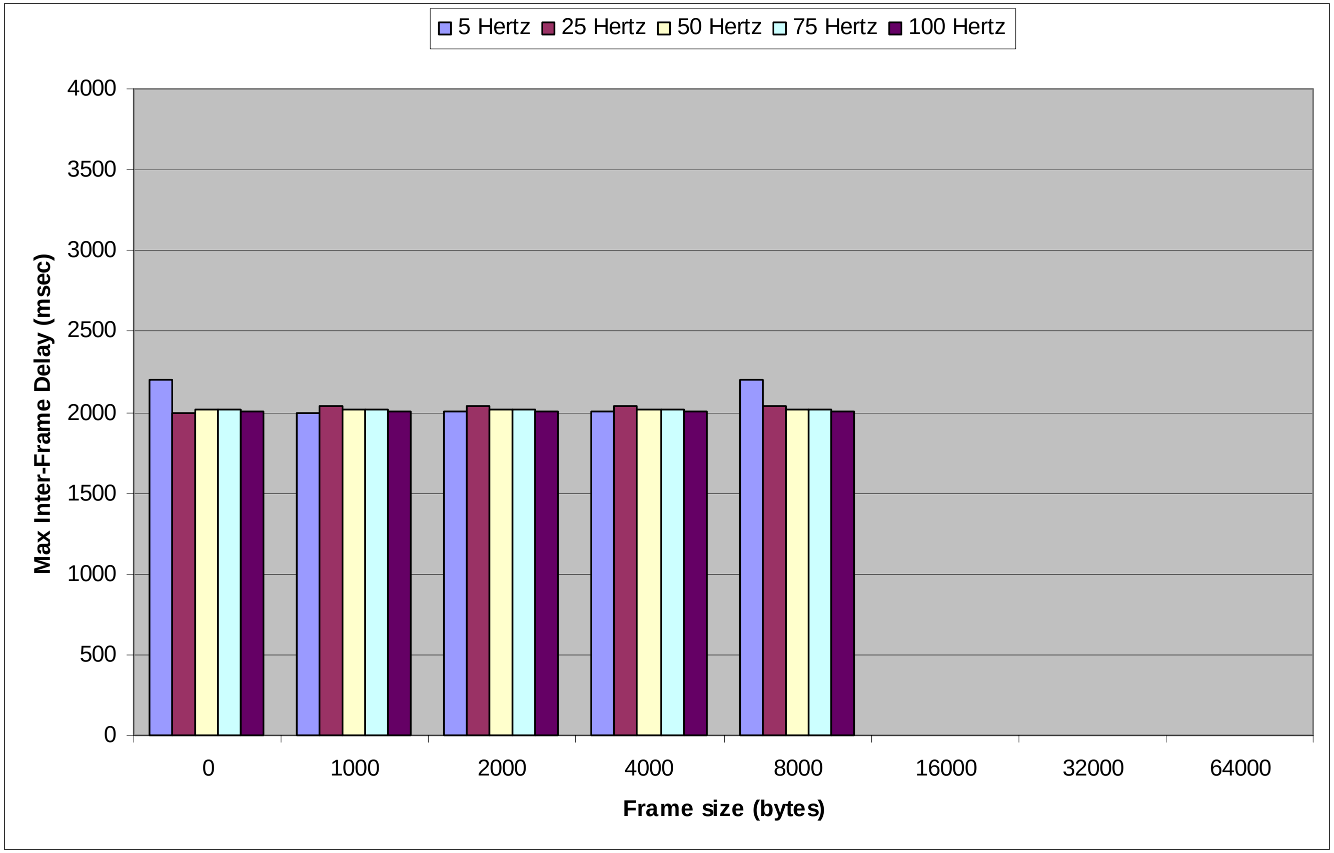
Protocol = DIOP, Experiment = Paced Invocations
Network = Systemic link failure



- Link was down for 33% of the time
- 33% of frames were dropped

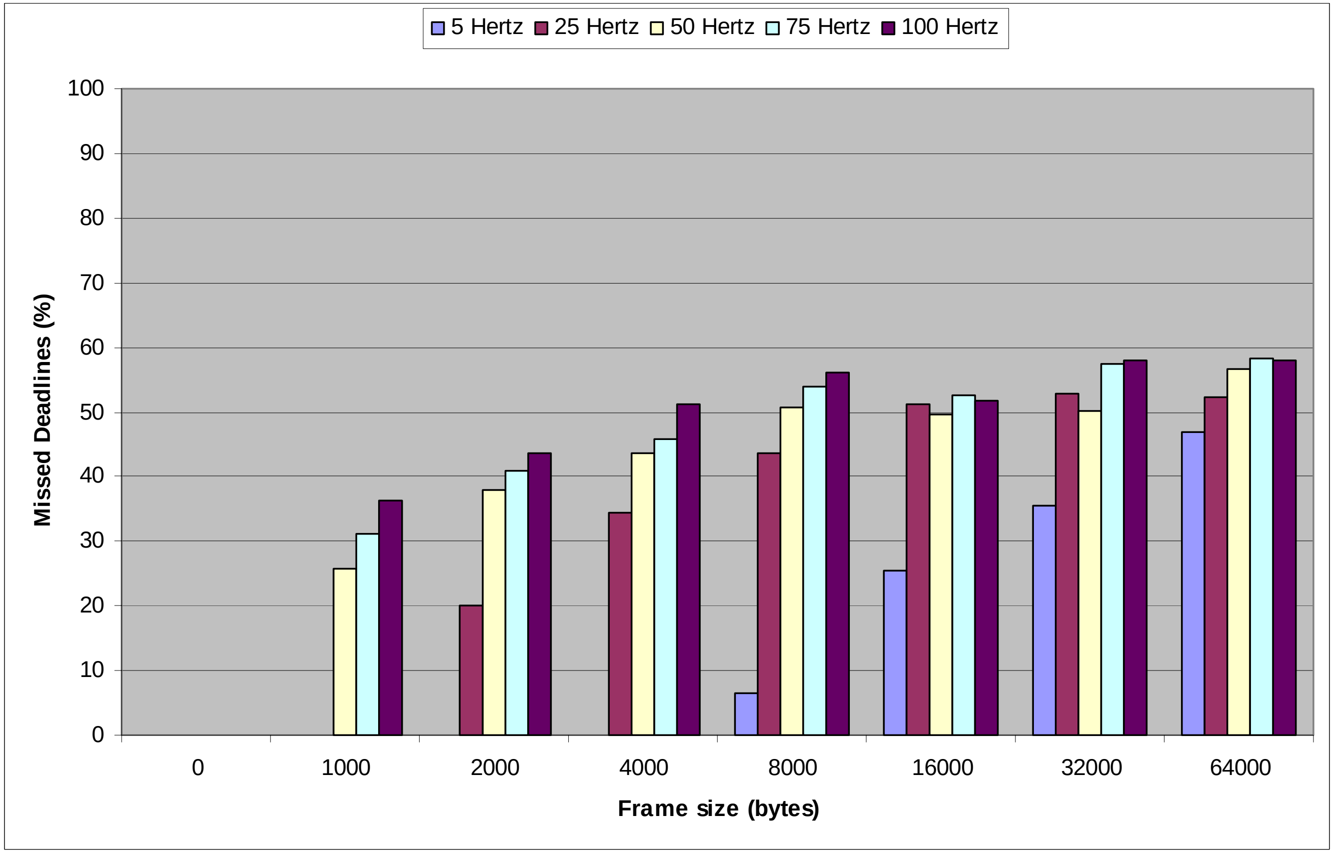
BBN TECHNOLOGIES A Verizon Company OOMWORKS LOCKHEED MARTIN

Protocol = DIOP, Experiment = Paced Invocations
Network = Systemic link failure



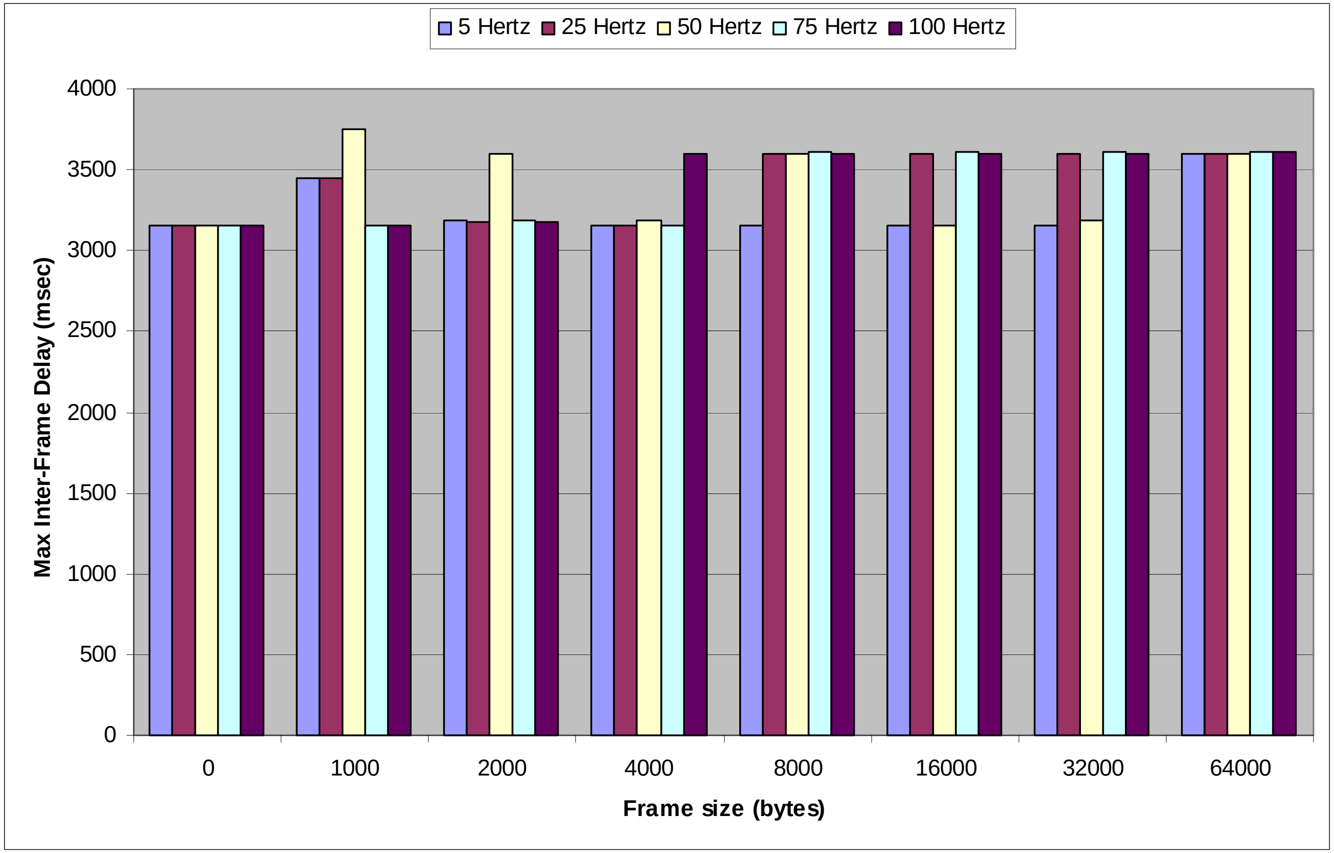
- Link was down for 2 seconds
- Inter-frame delay was about 2 seconds

Protocol = IIOP, Experiment = Paced Invocations
Network = Systemic link failure



- Link failure has significant impact on all invocation rates and frame sizes
- Impact is less visible for smaller frames at slower rates because IIOP is able to buffer packets thus allowing the client application to make progress
- Up to 58% deadlines are missed for larger frames at faster rates

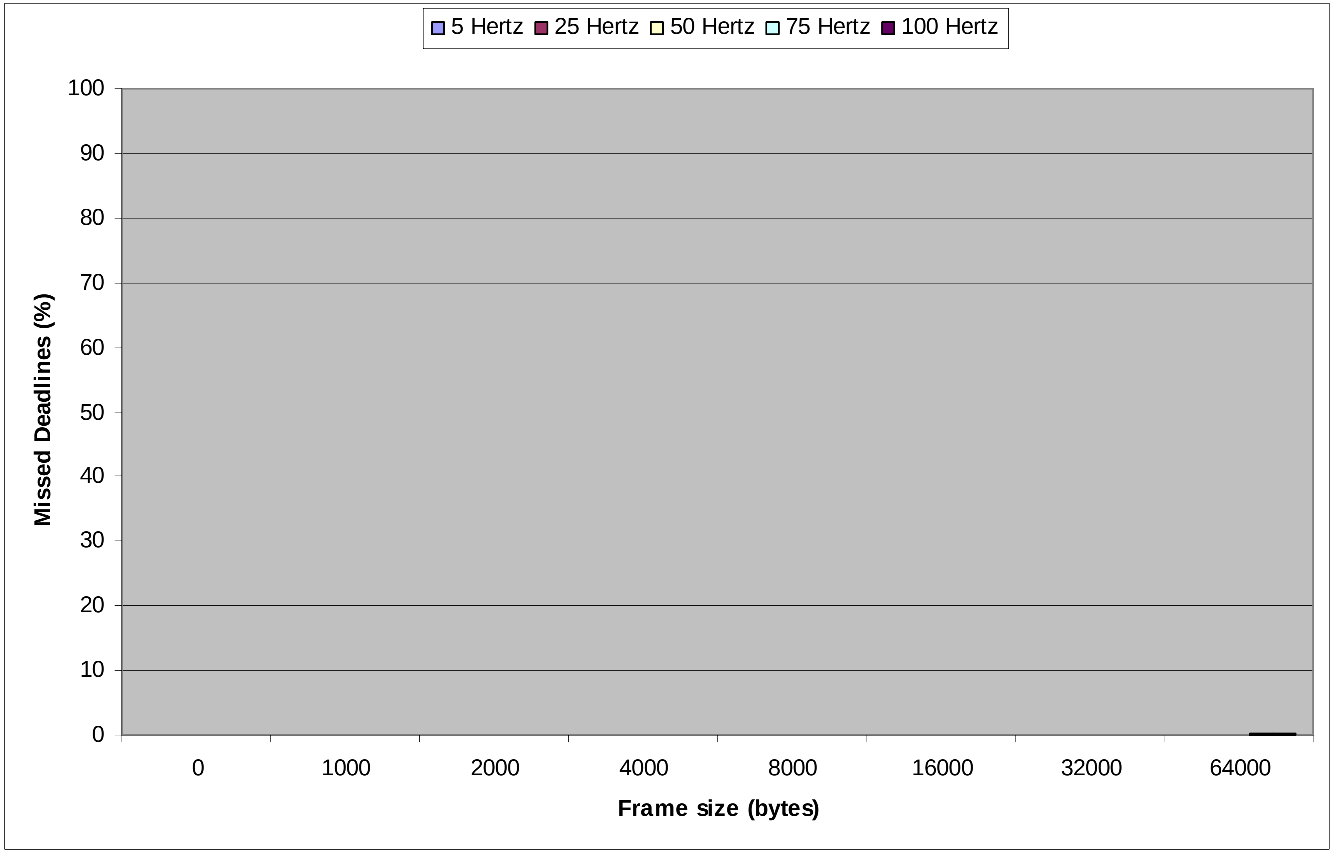
Protocol = IIOP, Experiment = Paced Invocations
Network = Systemic link failure



- IIOP does not recover well from temporary link loss
- Maximum inter-frame delay approaching 4 seconds

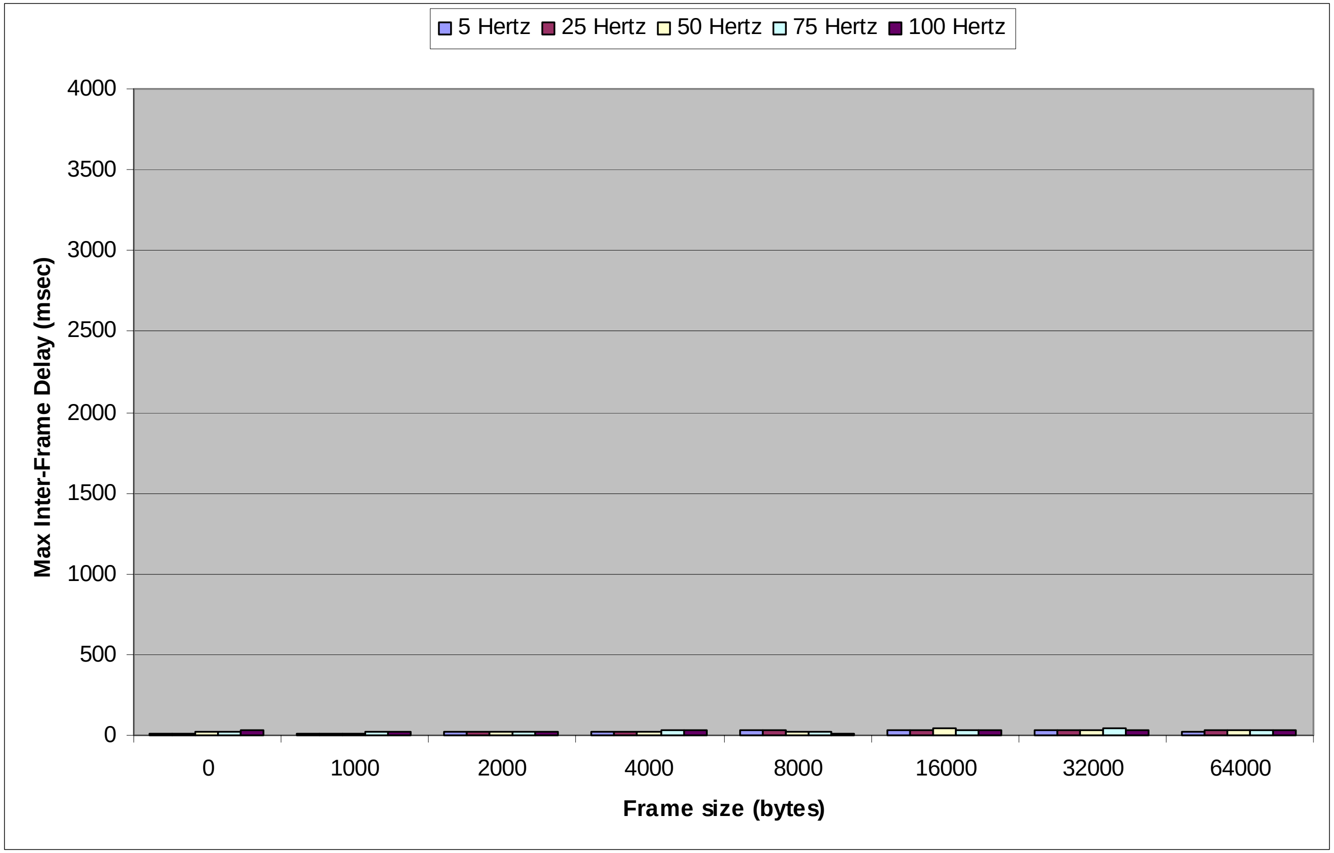
BBN TECHNOLOGIES A Verizon Company   OOMWORKS   LOCKHEED MARTIN

Protocol = SCIOP, Experiment = Paced Invocations Network = Systemic link failure



- SCIOP was able to use the redundant link during link failure
- No deadlines were missed

Protocol = SCIOP, Experiment = Paced Invocations
Network = Systemic link failure



- Inter-frame delay never exceeded 40 msec

BBN TECHNOLOGIES A Verizon Company OOMWORKS LOCKHEED MARTIN

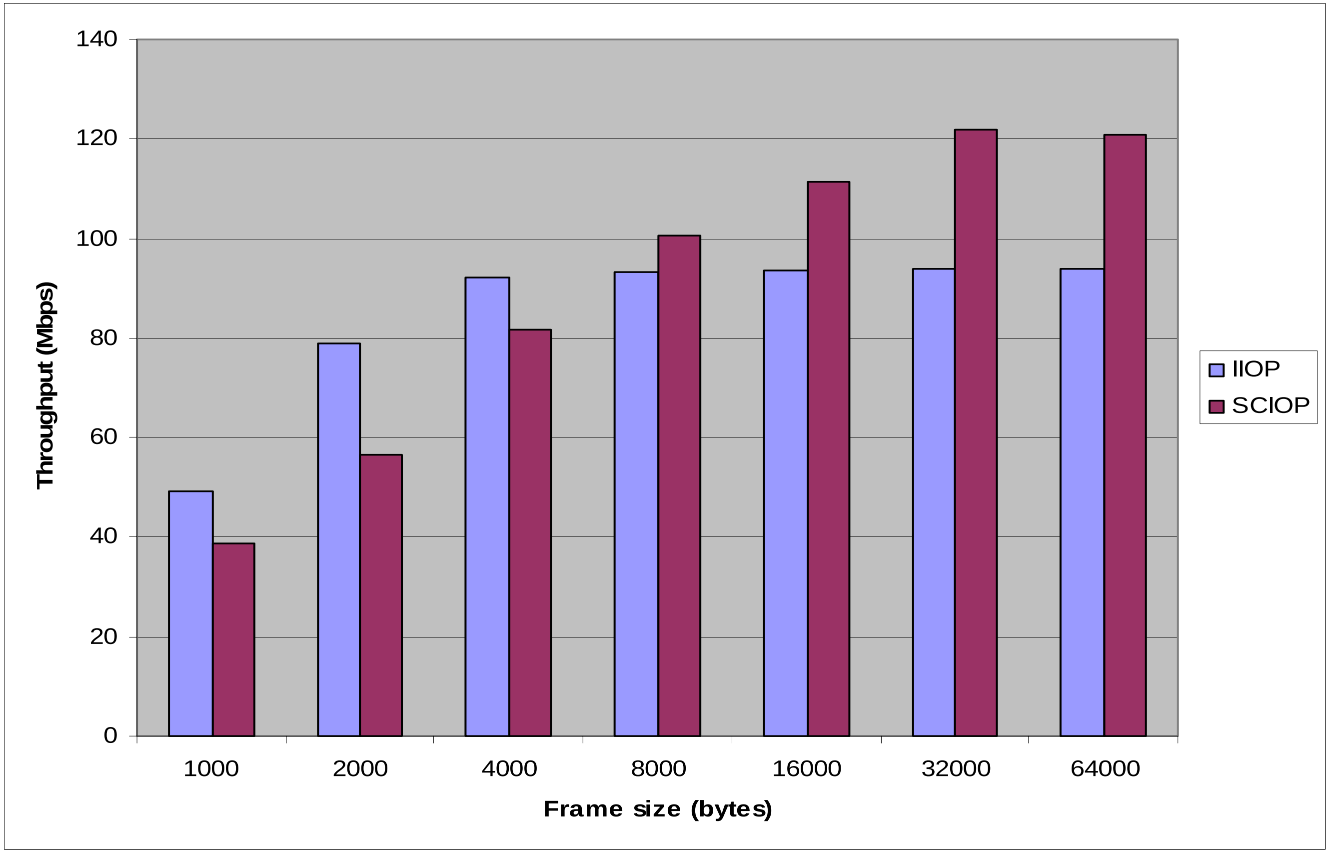
Summary of Experiments under Systemic link failure

- Client was able to meet all of its invocation deadlines when using SCIOP
- DIOP dropped up to 33% of frames
- IIOP missed up to 58% of deadlines

Throughput Tests

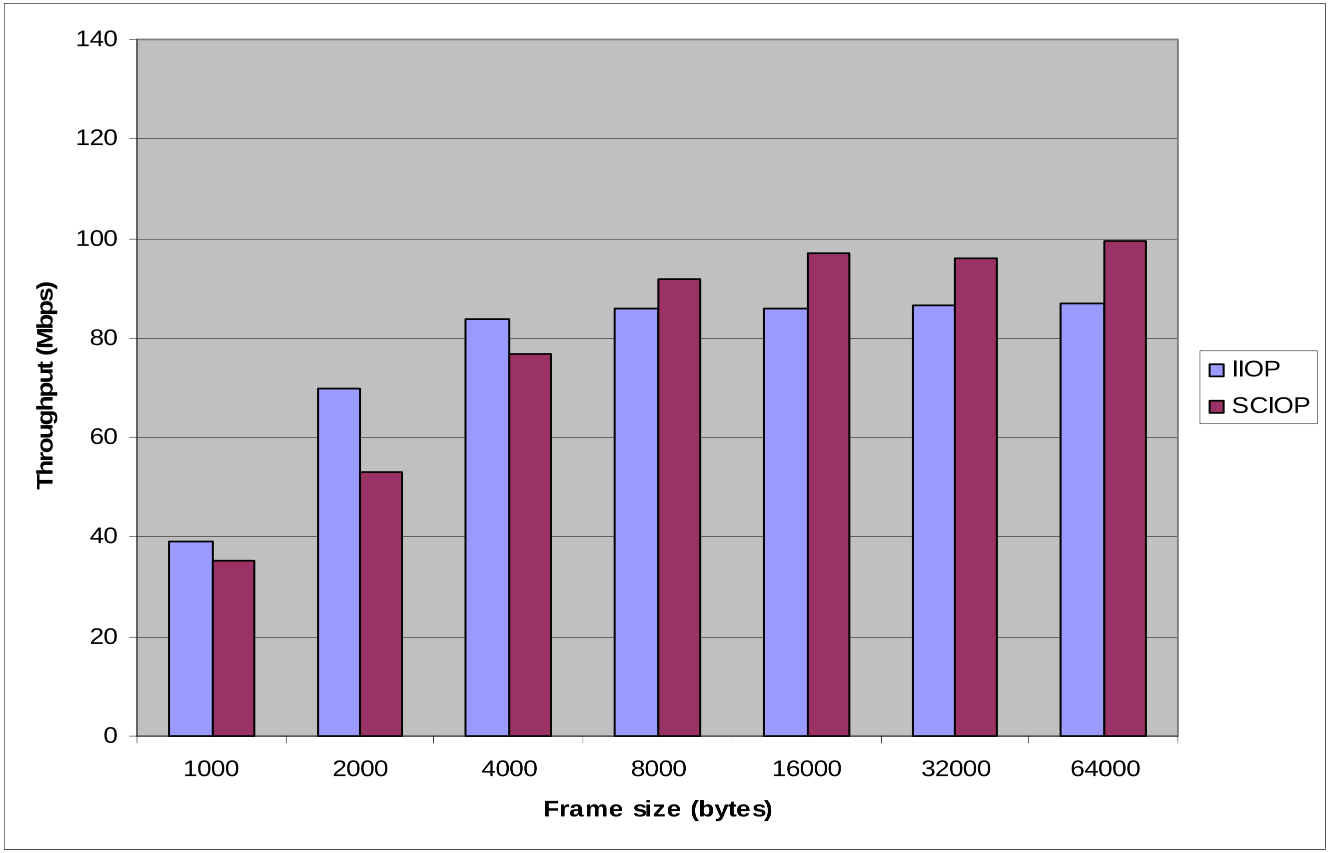
- Emulating applications that want to get bulk data from one machine to another as quickly as possible
- Two protocols: IIOP, SCIOP
  - DIOP not included because it is unreliable
- Frame size was varied from 1 to 64k bytes
- Client was sending data continuously
- IDL interface
  - **`one-way void method(in octets payload)`**
  - **`void twoway_sync()`**
- Experiment measures
  - Time required by client to send large amount of data to server

Experiment = Throughput
Network = Both links are up



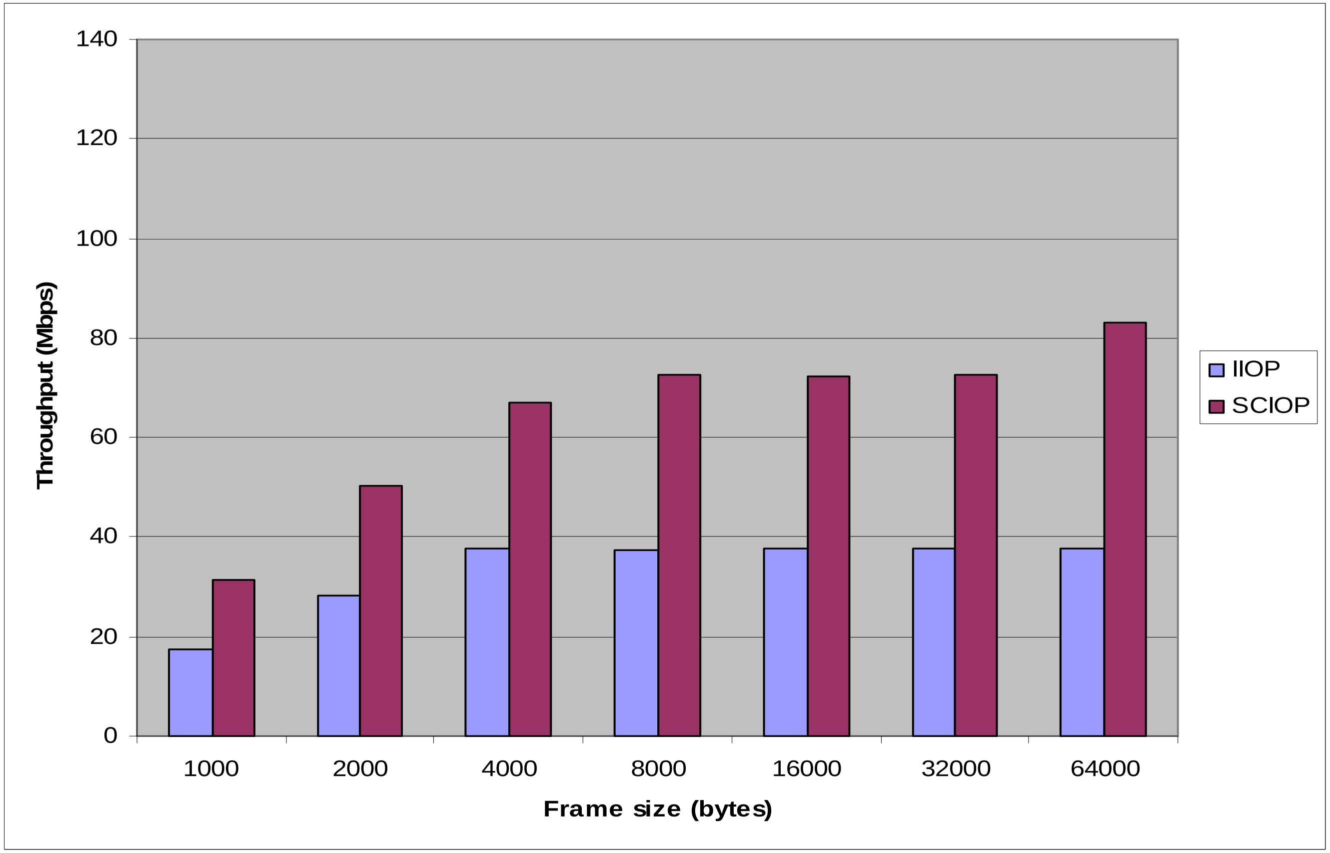
- IIOP peaks around 94 Mbps
- SCIOP is up to 28% slower for smaller frames
- SCIOP is able to utilize both links for a combined throughput up to 122 Mbps

Experiment = Throughput
Network = 1% packet loss on 1st link



- 1% packet loss causes maximum IIOP bandwidth to reduce to 87 Mbps (8% drop)
- IIOP outperforms SCIOP for smaller frames
- SCIOP maintains high throughput for larger frames, maxing out at 100 Mbps

Experiment = Throughput
Network = Systemic link failure



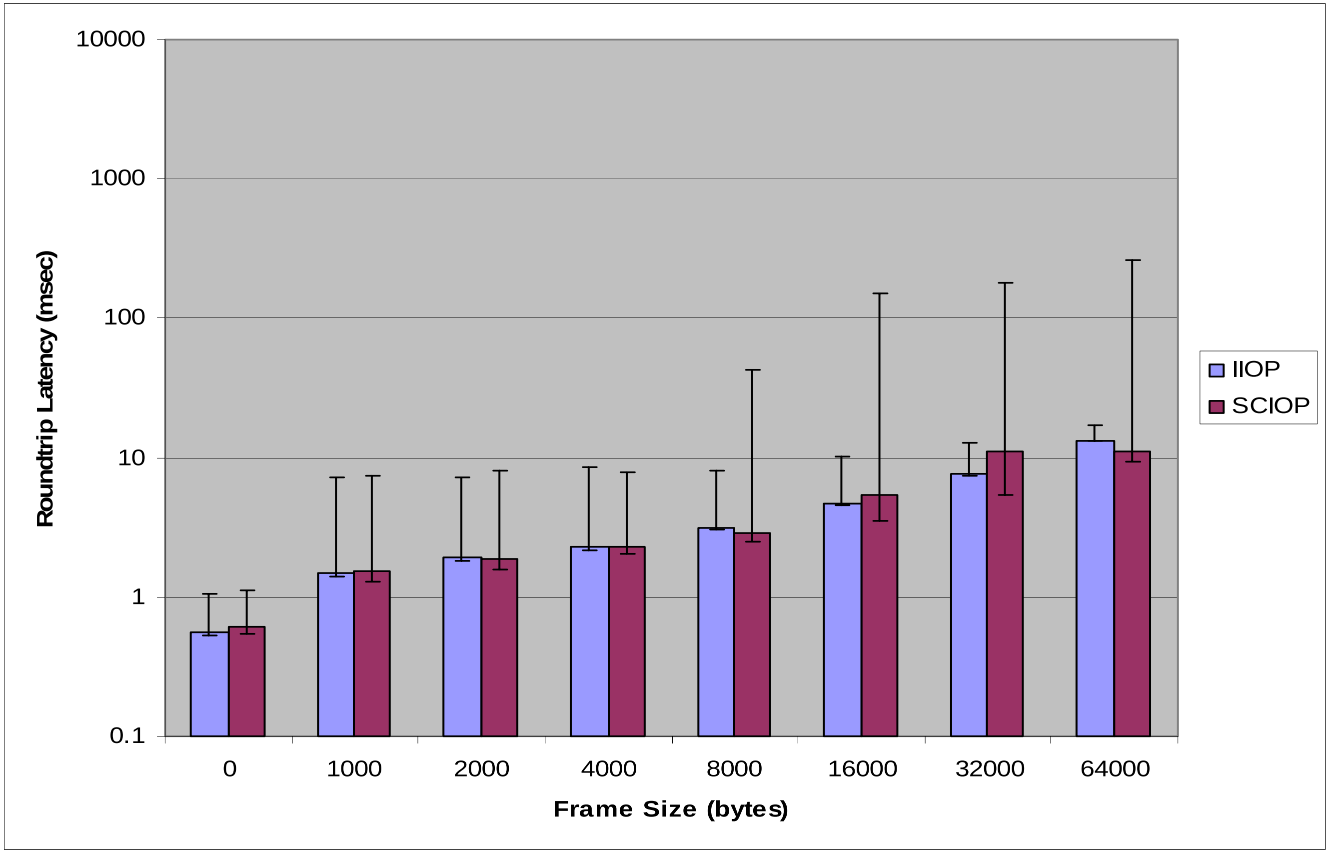
- Link failure causes maximum IIOP throughput to drop to 38 Mbps (60% drop)
- SCIOP outperforms IIOP for all frame sizes
- SCIOP maxes out at 83 Mbps

BBN TECHNOLOGIES A Verizon Company OOMWORKS LOCKHEED MARTIN

Latency Tests

- Emulating applications that want to send a message and get a reply as quickly as possible
- Two protocols: IIOP, SCIOP
  - DIOP not included because it is unreliable
- Frame size was varied from 0 to 64k bytes
- Client sends data and waits for reply
- IDL interface
  - **`void method(inout octets payload)`**
- Experiment measures
  - Time required by client to send to and receive a frame from the server

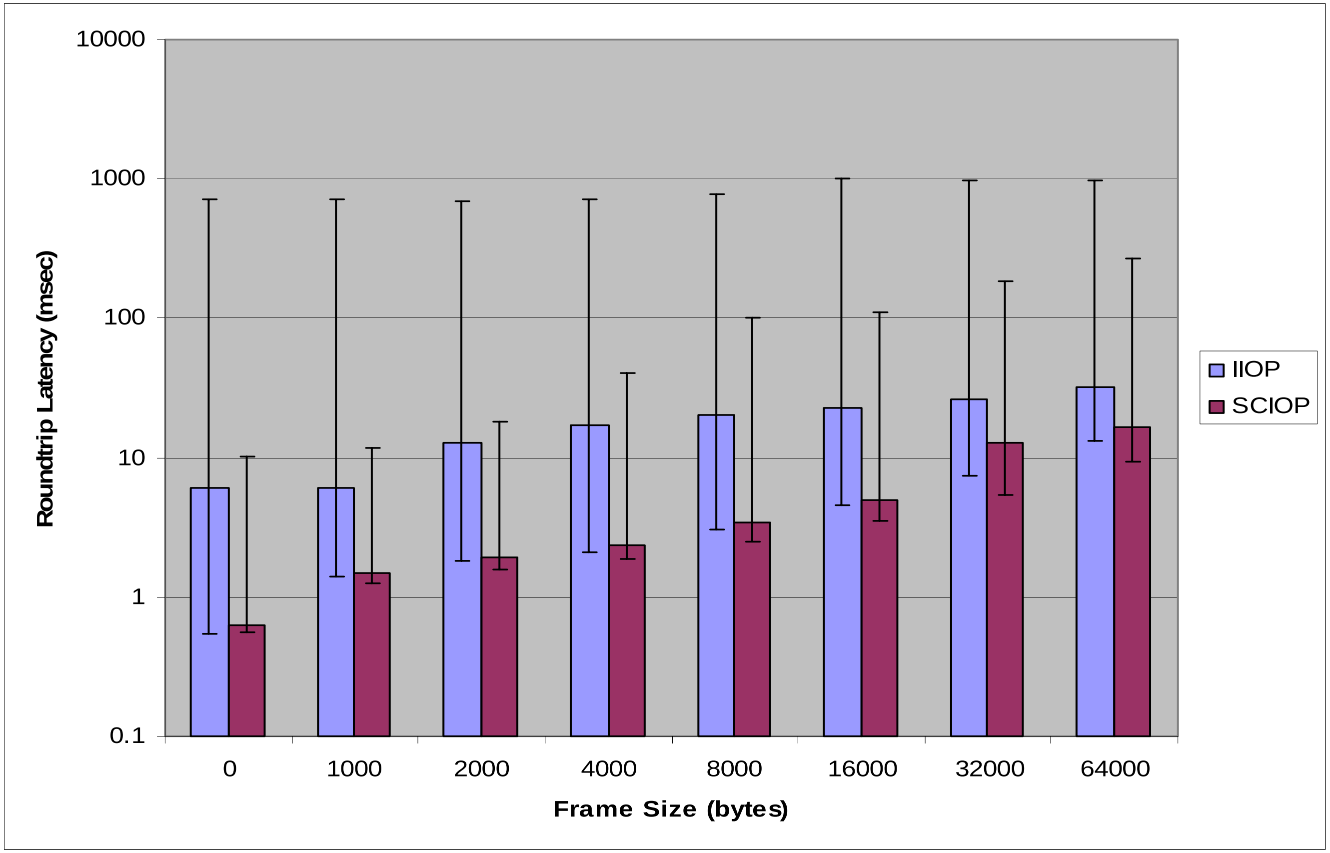
Experiment = Latency
Network = Both links are up



- Mean IIOP latency comparable to SCIOP
- For larger frames, maximum latency for SCIOP is 15 times maximum latency for IIOP

BBN TECHNOLOGIES
A Verizon Company
OOMWORKS
LOCKHEED MARTIN

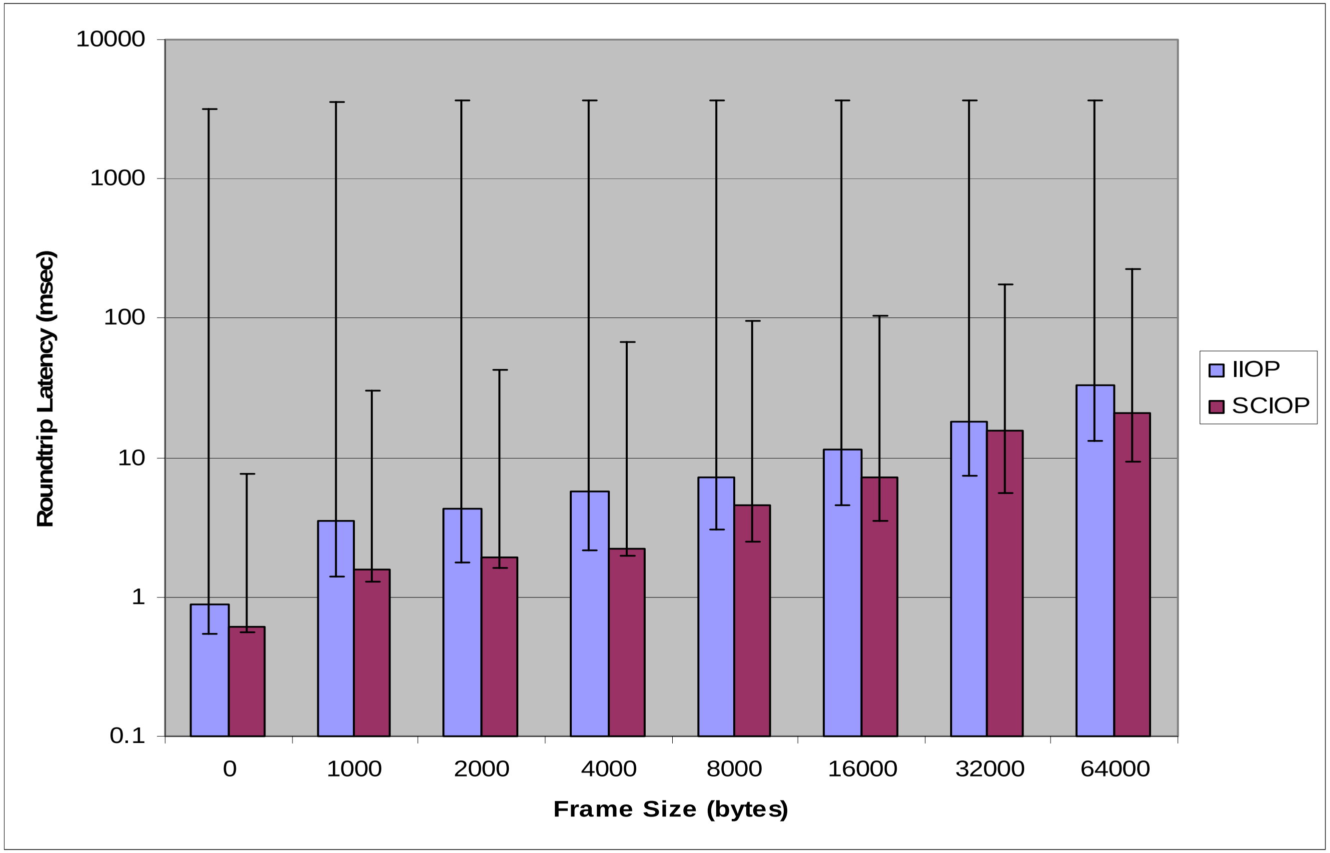
Experiment = Latency
Network = 1% packet loss on 1st link



- 1% packet loss causes maximum IIOP latency to reach about 1 second
- SCIOP outperforms IIOP for both average and maximum latencies for all frame sizes

BBN TECHNOLOGIES
A Verizon Company
OOMWORKS
LOCKHEED MARTIN

Experiment = Latency
Network = Systemic link failure



- Link failure causes maximum IIOP latency to reach about 4 seconds
- SCIOP outperforms IIOP for both average and maximum latencies for all frame sizes

Experiments Summary

PACED INVOCATIONS	DIOP		IIOP		SCIOP	
	Maximum Delay (msec)	Dropped Frames (%)	Maximum Delay (msec)	Missed Deadlines (%)	Maximum Delay (msec)	Missed Deadlines (%)
Normal Conditions	14	0	13	0	14	0
1% Packet Loss	400	7	720	6	26	0
Link Failure	2000	33	4000	58	40	0

THROUGHPUT	DIOP	IIOP	SCIOP
Normal Conditions		94	122
1% Packet Loss		87	100
Link Failure		38	83

LATENCY	DIOP		IIOP		SCIOP	
	Average	Max	Average	Max	Average	Max
Normal Conditions			0.6	1.1	0.6	1.1
1% Packet Loss			5.4	717	0.6	10.1
Link Failure			5.8	3641	0.6	7.5

Conclusions

- SCTP combines best features of TCP and UDP and adds several new features
- SCTP can be used to improve network fault tolerance and improve QoS
- Under normal network conditions, SCTP compares well with TCP and UDP
  - In addition, it can utilize redundant links to provide higher effective throughput
- Under packet loss and link failures, SCTP provides automatic failover to redundant links, providing superior latency and bandwidth relative to TCP and UDP
- Integrating SCTP as a pluggable protocol into middleware allows effortless and seamless integration for DRE applications
- SCTP is available when using ACE, TAO, CIAO and AVStreaming
- Continue to use other network QoS mechanisms such as DiffServ and IntServ with SCTP
- Both OpenSS7 and (specially) LKSCTP implementations need improvement
  - Crashes during failover
  - Differences in conformance to SCTP specification
  - Limited technical support
- Emulab provides an excellent environment for testing SCTP
- Future work
  - Load-sharing in SCTP – Concurrent multipath data transfer
  - Adaptive Failover
  - Ongoing research at Protocol Engineering Laboratory, University of Delaware