Network Simulation and Testing

Polly Huang
Department of Electrical Engineering
National Taiwan University
http://cc.ee.ntu.edu.tw/~phuang
phuang@cc.ee.ntu.edu.tw

P2P Papers

- Salman A. Baset and Henning Schulzrinne. An Analysis of the Skype Peer-to-Peer Internet Telephony Protocol. In Proceedings of IEEE Infocom 2006.
- Kuan-Ta Chen, Chun-Ying Huang, Polly Huang, and Chin-Laung Lei. Quantifying Skype User Satisfaction. In Proceedings of ACM SIGCOMM 2006.
- X. Zhang, J. Liu, B. Li, and T.-S. P. Yum. CoolStreaming/DONet: A Data-driven Overlay Network for Efficient Live Media Streaming. In Proceedings of IEEE INFOCOM 2005
- Meng-Ting Lu, Jui-Chieh Wu, Kuan-Jen Peng, Polly Huang, Jason J. Yao, Homer H. Chen. Design and Evaluation of a P2P IPTV System for Heterogeneous Networks. IEEE Transactions on Multimedia, Vol. 9, No. 8, pp. 1568-1579, Dec. 2007

Multimedia P2P

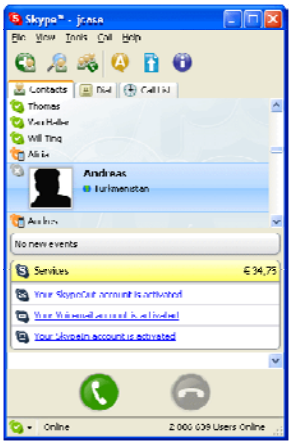
Skype

CoolStreaming

HotStreaming

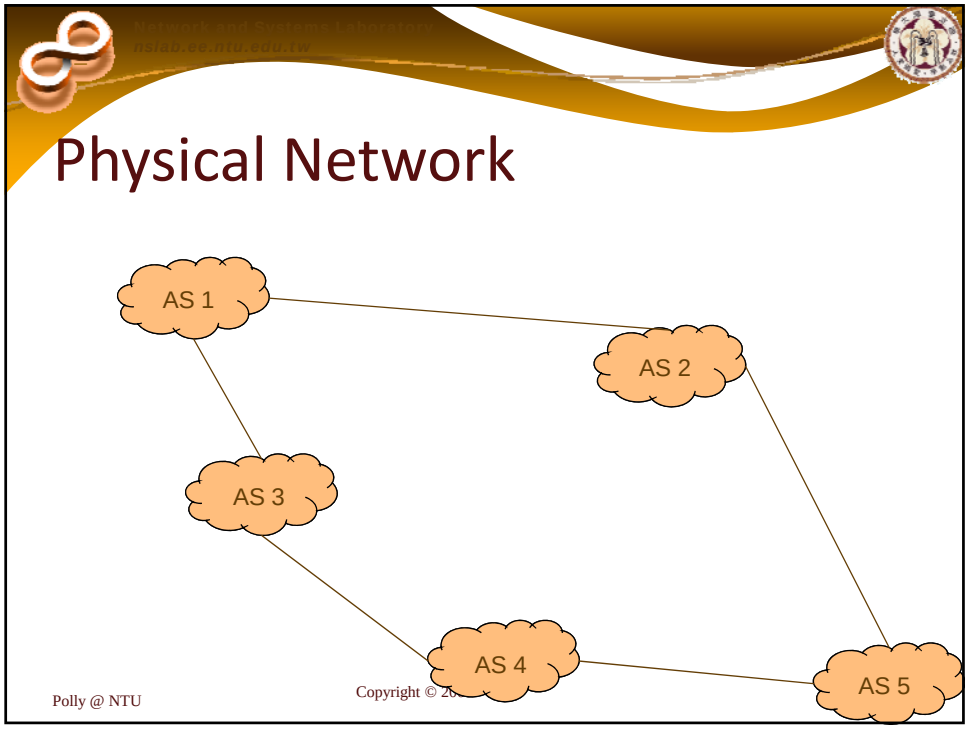
What is it?

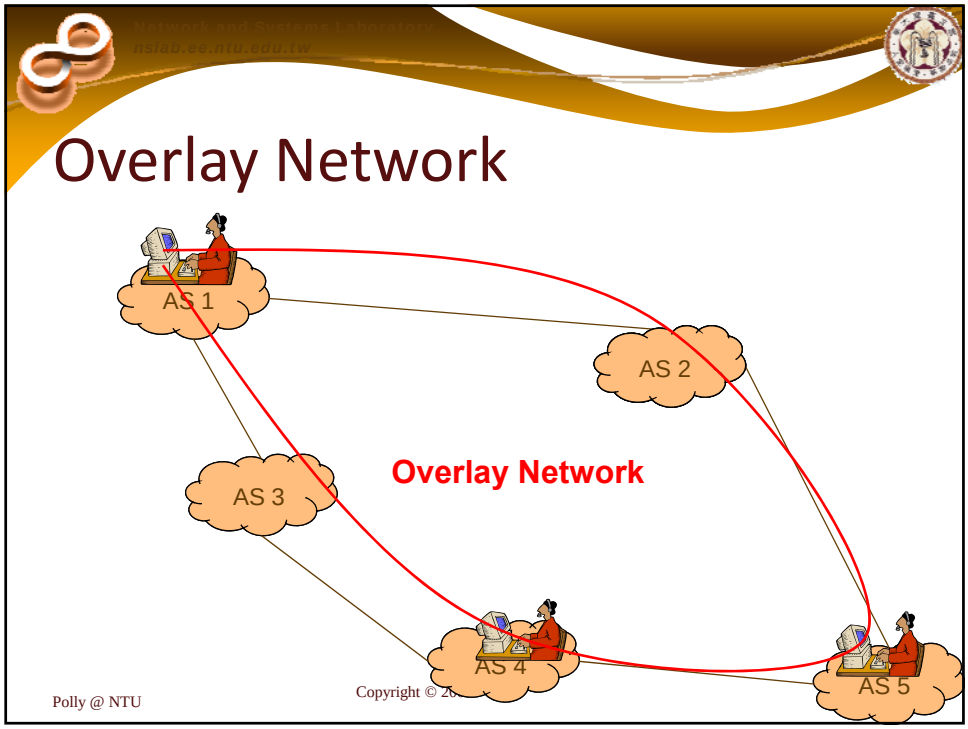
- A peer-to-peer (P2P) overlay network for VoIP
- By the KaZaA developers
- To an average user, it is an Instant Messaging system that supports instant voice

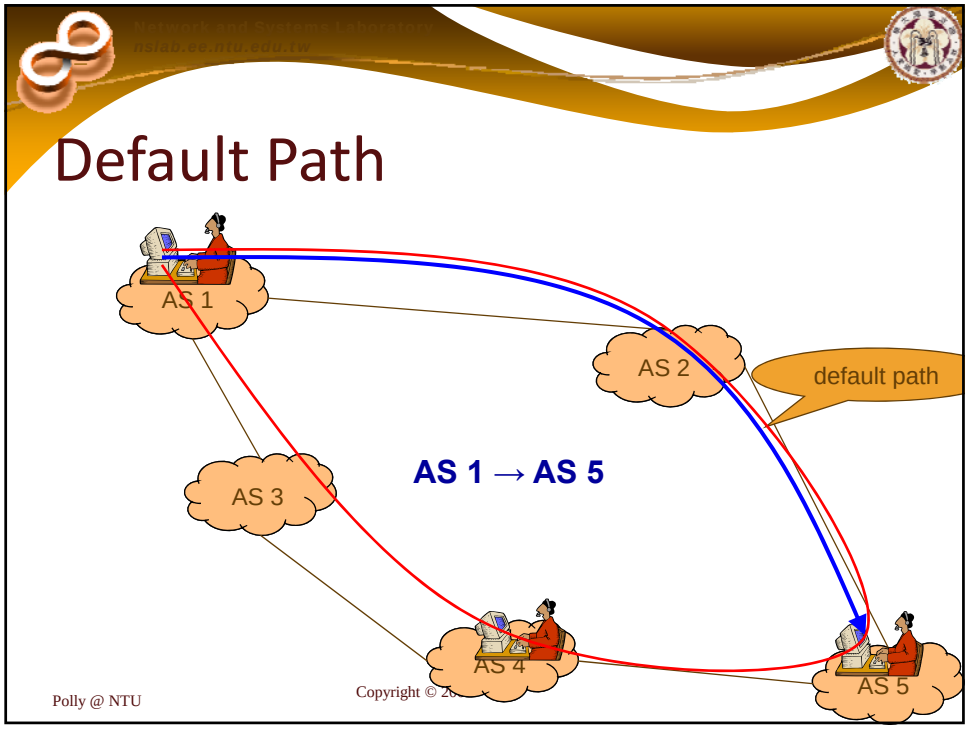


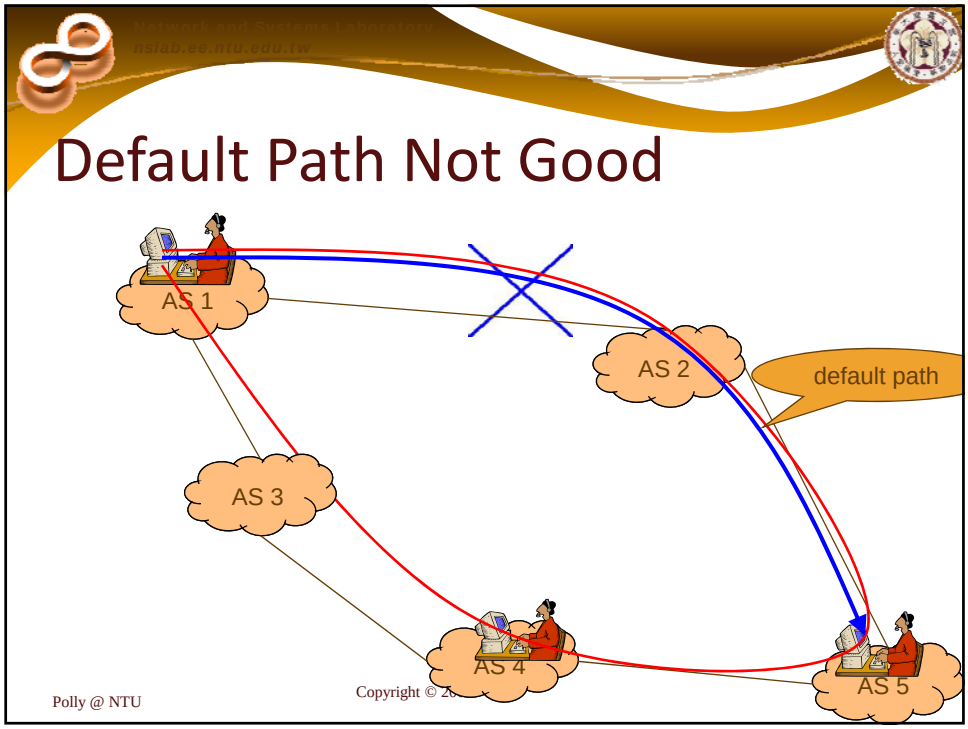
Secret of Success

- Codec (provided by Global IP Sound)
  - 67 bytes packet payload
  - 24 to 120 kbps
- NAT and firewall traversal
- **Alternative paths!**

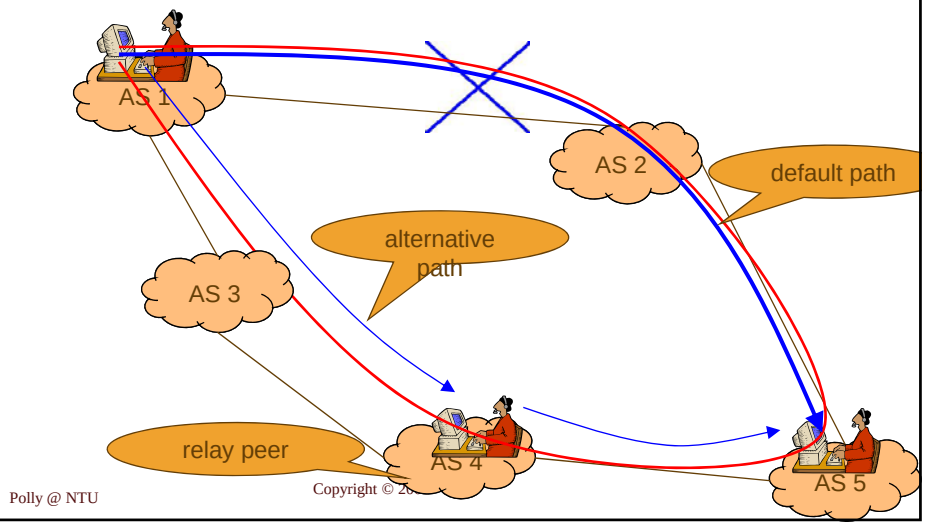

Network and Systems Laboratory
nslab.ee.ntu.edu.tw
Physical Network
AS 1
AS 2
AS 3
AS 4
AS 5
Polly @ NTU
Copyright © 20


Network and Systems Laboratory
nslab.ee.ntu.edu.tw
Overlay Network
AS 1
AS 2
AS 3
Overlay Network
AS 4
AS 5
Polly @ NTU
Copyright © 20


Network and Systems Laboratory
nslab.ee.ntu.edu.tw
Default Path
AS 1
AS 2
default path
AS 3
AS 1 → AS 5
AS 4
AS 5
Polly @ NTU
Copyright © 2

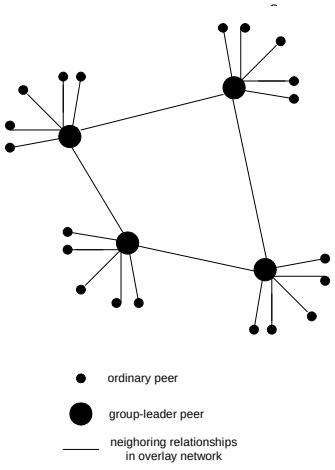

Network and Systems Laboratory
nslab.ee.ntu.edu.tw
Default Path Not Good
AS 1
AS 2
default path
AS 3
AS 4
AS 5
Polly @ NTU
Copyright © 2

Find an Alternative Path



Based on KaZaA

- A hybrid solution
  - Some are group leaders
    - Super nodes or SN
  - Most are ordinary peers
    - Ordinary nodes or ON
- 1st level
  - Between a group leader and its ordinary peers
  - Napster-like
- 2nd level
  - Between group leaders
  - Gnutella-like



Phase I: Login

- Send 'I'm alive' to other peers
  - Showing in your friends' Friend lists
- Determines the type of NAT and firewall it is behind
  - More details if time allow
- **Discover online SNs**

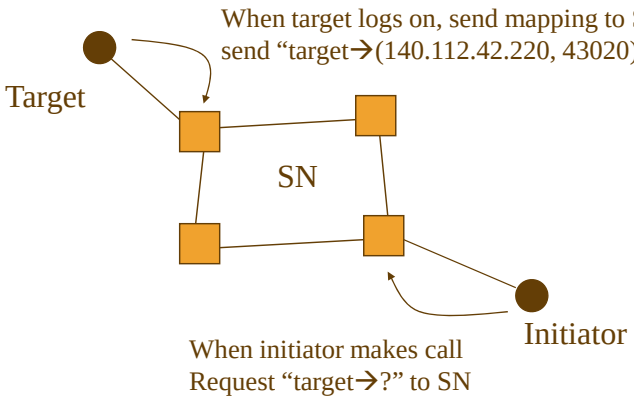
Where are the SNs?

- 1st-time use of Skype
  - Connect first to a **bootstrap** server to acquire a candidate SN list
  - Select one from the list
  - Report to the SN meta data about user/files to be shared
  - Acquire from the SN the updated list of candidate SN
  - Cache the candidate SN list
- 2nd-time+ use of Skype
  - Select one from the cached candidate SN list
  - The rest is the same

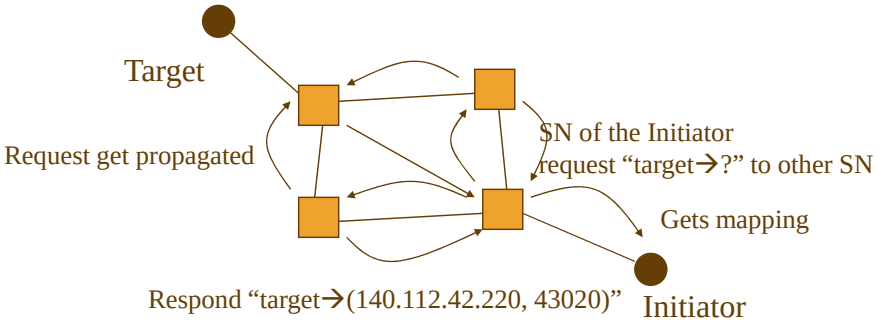
Phase II: Address Lookup

- Initiator calling target
  - Calling pollyhuang…
  - Where is pollyhuang at?
- Username → (IP, port) address mapping
  - The meta data about users stored in SNs

Address Lookup Service



P2P Style Search

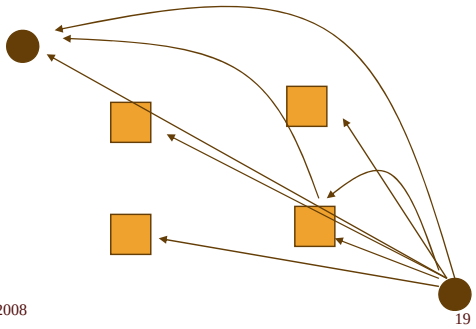


Phase III: Call Establishment

- Two directions
  - Initiator connects to target first
  - Target connects back to initiator
- Operation for each direction is identical

The Operation

- Probe a list of nodes
  - Target, candidate relay nodes (RN)
- Pick the best to connect
  - If target is picked
    - Connect to target
  - If RN is picked
    - Connect to RN
    - RN connect to target



Candidate Relay Nodes (RN)

- Pre-defined and change dynamically
- Simply the candidate SNs in the host cache (HC)
- Probes can be a priori in low frequency
  - Shorten the delay in call establishment

Who should be the relay peer?

such that the alternative path
is **better** than the default
one…

Skype's Solution

Round-Trip Delay & Loss Rate (Probably)

[ICDCS 06] S. Ren, L. Guo, and X. Zhang, ``ASAP: an AS-Aware Peer-relay protocol for high quality VoIP with low overhead", *Proceedings of the 26th International Conference on Distributed Computing Systems* (ICDCS'06), Lisbon, Portugal, July 4-7, 2006

A Fundamental Question

Which is the right path selection criteria?

Source Rate?
Congestion Level?
Delay?
Combination of the above?

User Satisfaction

$$USI = 2.15 * \log(\text{bit rate}) - 1.55 * \log(\text{jitter}) - 0.36 * RTT$$

bit rate:	data rate of voice packets
jitter:	level of network **congestion**
RTT:	round-trip times between two parties

[SIGCOMM 06] Kuan-Ta Chen, Chun-Ying Huang, Polly Huang, Chin-Laung Lei, "Quantifying Skype User Satisfaction," *ACM SIGCOMM* 2006

Secret of Success

- Codec (provided by Global IP Sound)
  - 67 bytes packet payload
  - 24 to 120 kbps
- **NAT and firewall traversal**
- Alternative paths!

Special Ports

- Skype client listening on a randomly selected port
- If blocked by firewall, use
  - 80 -- HTTP
  - 443 -- HTTPS

TCP Connection

- Skype client listening using UDP
- If blocked by firewall, use
  - TCP

NAT Traversal

- STUN
  - Simple Traversal of UDP through NATs
  - **More details if time allows**
- TURN
  - Traversal Using Relay NAT
  - I.e., the relay node approach

Secret of Success

- **Codec** (provided by Global IP Sound)
  - 67 bytes packet payload
  - 10 to 120 kbps
- NAT and firewall traversal
- Alternative paths!

Codecs

- iLBC
  - 13.3 kbps (30 ms frames) 15.2 kbps (20 m frames)
  - Better than G.729A and G.723.1
- iSAC
  - 10-120 kbps (**adaptive and variable**)
  - Wideband communication
- No silence suppression

Multimedia P2P

Skype

CoolStreaming

HotStreaming

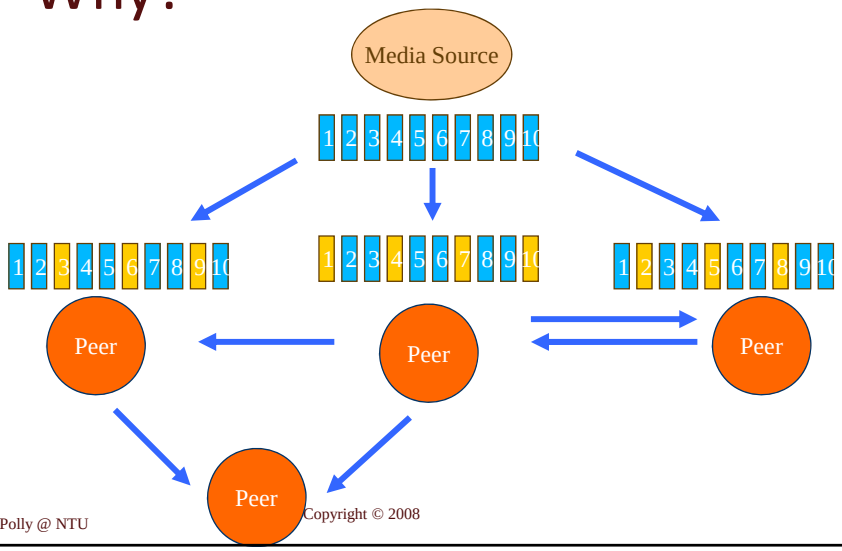
CoolStreaming → PPLive

- A peer-to-peer (P2P) overlay network for IPTV
- Hongkong-China based
- To an average user, it's free live sports event broadcast service
- 10,000s of subscribers

Secret of the CoolStreaming's Success

- Free content
- Smooth and quality video
- The more users, the better the quality

Why?



Major Technical Components

	CoolStreaming
P2P Network Formation	Pure random
Frame Scheduling	Fewest copies first Dead-line oriented
Codec	Mpeg-4, windows WMA

Outline

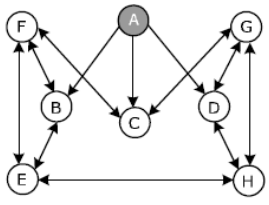
- **Design of DONet**
- Performance Evaluation
- CoolStreaming
- Conclusion

DONet

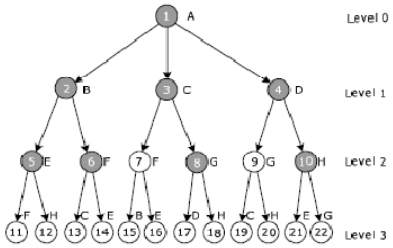
(Data-driven Overlay Network)

- Concept
  - Form a gossip P2P network
    - Essentially, a peer connect to other peers **randomly**
  - P2P network dynamically updated
    - Essentially, which peer connect to which peers is changed **periodically**
- Characteristics
  - Easy to implement
  - Robust and resilient
  - Efficient

P2P Network Topology



Mesh



As opposed to a tree

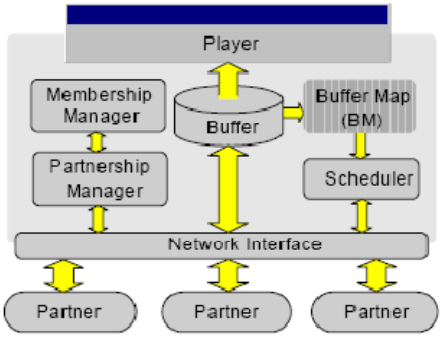
The rest of the network can still function with a node fails!

Network Comparison

	DONet	CoopNet
Base	mesh	tree
Load	balanced	unbalanced
Supplier/Client	not specified*	specified
Structure	distributed	centralized
Dynamic switch	robust	vulnerable

Design of DONet

- 3 Key modules
  - Membership Manager
  - Partnership Manager
  - Buffer Map & Scheduler
- Terminology
  - Member → active nodes on t overlay network
  - Partner → nodes can exchan data (**a subset of members**)



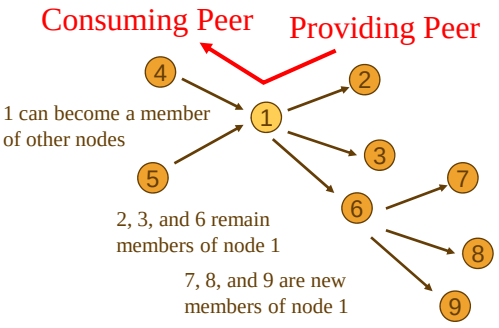
Membership Management

- SCAMP (Scalable Gossip Membership protocol)
- Join
  - A newly joined node contacts first a bootstrap node
    - Redirected to a deputy node
    - Randomly selected from the bootstrap node's mCache*.
  - Deputy node gives the new node its mCache
- Maintain
  - **Periodical membership messages to update mCache**
  - Nodes will be removed from the list after a fixed amount of time

*The membership cache contains a partial list of active nodes in DONet.

Implementation

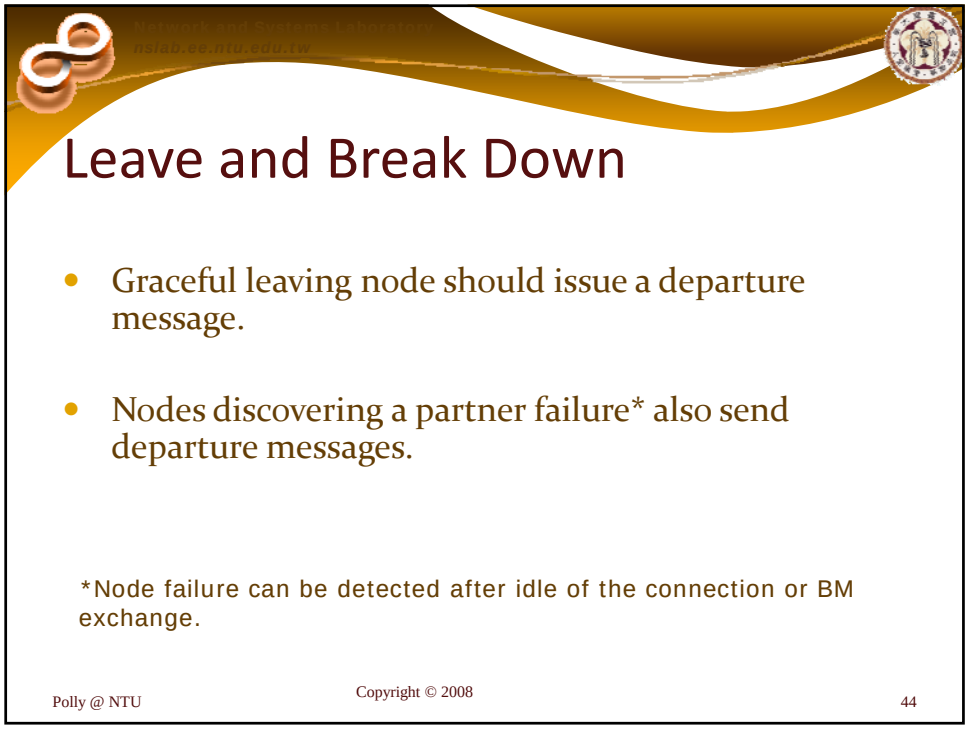
Membership Relationship



Every peer:
1. Periodically send
2. An advertisement message (ADV)

Upon receiving the ADV:
1. Roll a dice
2. If win, select the ADV sender to provide video segments


Network and Systems Laboratory
nslab.ee.ntu.edu.tw
Partnership Management
Periodically establish new partnerships with nodes randomly selected from mCache
Reject the lowest rating partner
Polly @ NTU
Copyright © 2008
43


Network and Systems Laboratory
nslab.ee.ntu.edu.tw
Leave and Break Down
Graceful leaving node should issue a departure message.
Nodes discovering a partner failure* also send departure messages.
*Node failure can be detected after idle of the connection or BM exchange.
Polly @ NTU
Copyright © 2008
44

Buffer Map Management

- Buffer Map
  - Semi-synchronized playback progresses.
  - A **sliding window of 120 segments** can effectively represent the buffer map. Segments outside the window is ignored.
  - Use **120 bits** to record BM, with bit 1 indicating that a segment is available.
  - The **sequence number of the first segment** in the sliding window is record by another **2 bytes**.
- Each node continuously exchanges its BM with the partners.

Scheduling Algorithm

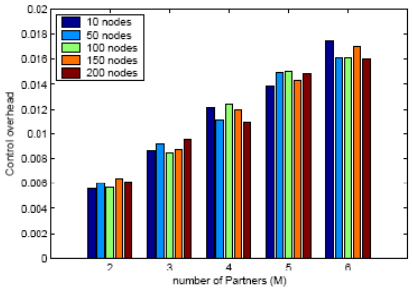
- Two constrains
  - Playback deadline for each segment
  - Heterogeneous streaming bandwidth
- Minimize segments missing deadlines.
- Heuristic algorithm
  - Calculate the number of the potential suppliers for each segments.
  - Less suppliers means higher risks of missing the deadlines.
  - Determine the supplier of each segment starting from the one with **lowest potential suppliers**.
  - Among the supplier candidates, pick the one **with higher BW** and enough available time.

Outline

- Design of DONet
- **Performance Evaluation**
- CoolStreaming
- Conclusion

Performance Evaluation*

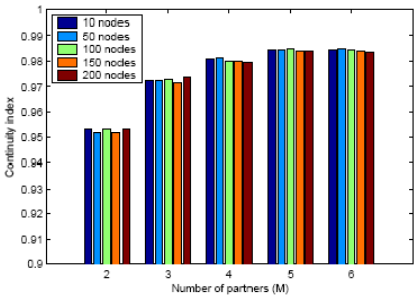
- Control overhead
  - Control traffic volume /Video traffic volume
- Continuity index
  - Number of segments that arrive in time over the total number of segments
- On/Off period
  - Average of node join/leave-period, which is exponentially distributed
- Stable Environment



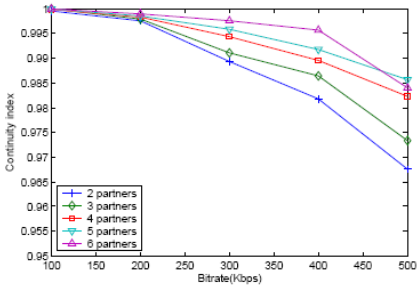
Control overhead vs. Partner #

*Experiments are conducted in PlanetLab.

Stable Environment

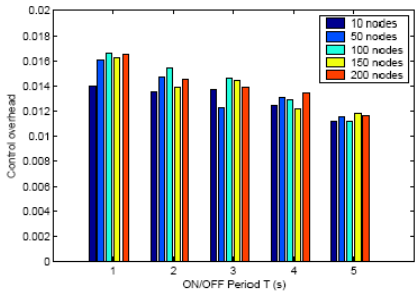


Continuity index vs. Partner #

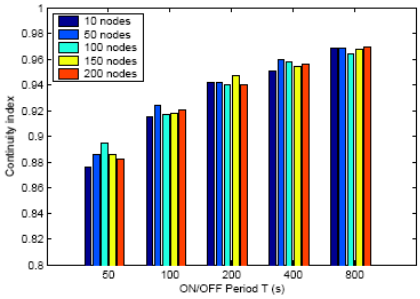


Continuity index vs. Bitrate

Dynamic Environment

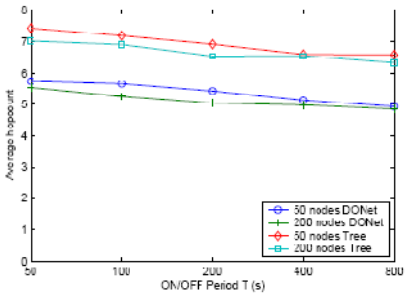


Control overhead vs. On/Off period

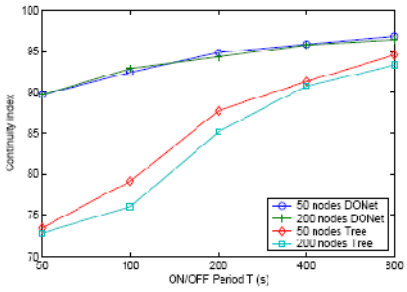


Continuity index vs. On/Off period

Compared with Tree Overlay

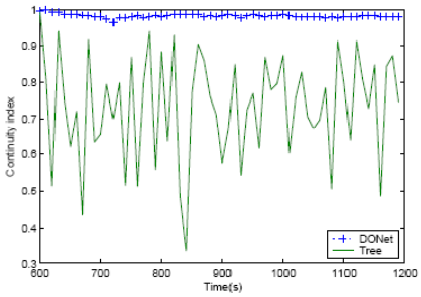


Average hopcount vs. On/Off period



Continuity index vs. On/Off period

Fluctuation Over Time



Continuity index vs. Time

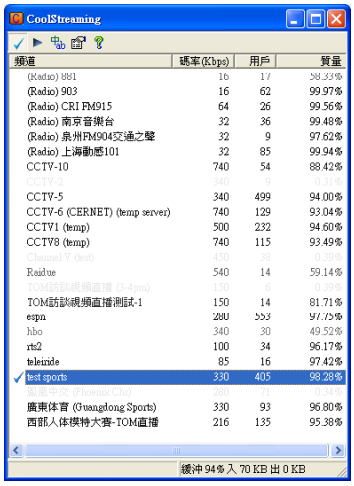
- DONet summary
  - Low control overhead (1% of the video traffic)
  - Better playback continuity and smaller end-to-end delay, especially under dynamic environments
  - Scalable both in overlay size and streaming rate
  - Stable quality

CoolStreaming

- TFRC (TCP friendly rate control) protocol
- About 5~10 minutes to settle down the

	ESPN	Test Sports	Channel V
Bitrate (kbps)	280	330	450
Online user	5000	2500	67
Server quality	95%	50~60%	75%
Buffer level	80%+	50%	90%
Video quality	good	poor	Mediocre

*CoolStreaming v 0.0.42 released at March 1, 2005



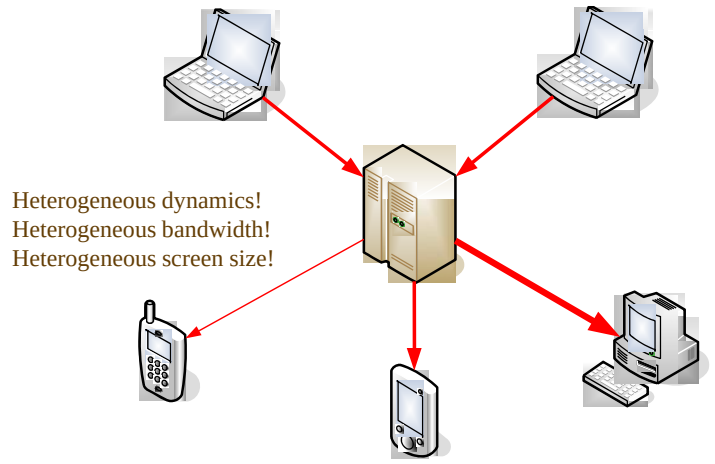
Summary

- DONet is a **capacity amplifier** between a content provider and clients.
  - The larger the data-driven overlay is, the better the streaming quality it delivers.
- DONet doesn't maintain an explicit overlay structure
  - A mesh-based overlay network
  - Scalable and robust
- The problem of user behind Firewall can be solved by TCP connection.
  - Statistical results show that more than 95% of the nodes can become relay nodes.

Multimedia P2P

Skype
CoolStreaming
HotStreaming

Mobile IPTV



Differences

	CoolStreaming	HotStreaming
Partnership Formation	Pure random	Preferential random (rotate)
Frame Request Scheduling	Heuristic: (fewest copies first Earliest Deadline first)	Optimal: (min loss under bw constraint)
Codec	Mpeg-4, windows WMA	Mpeg-4, interleaving MDC (bw heterogeneity)

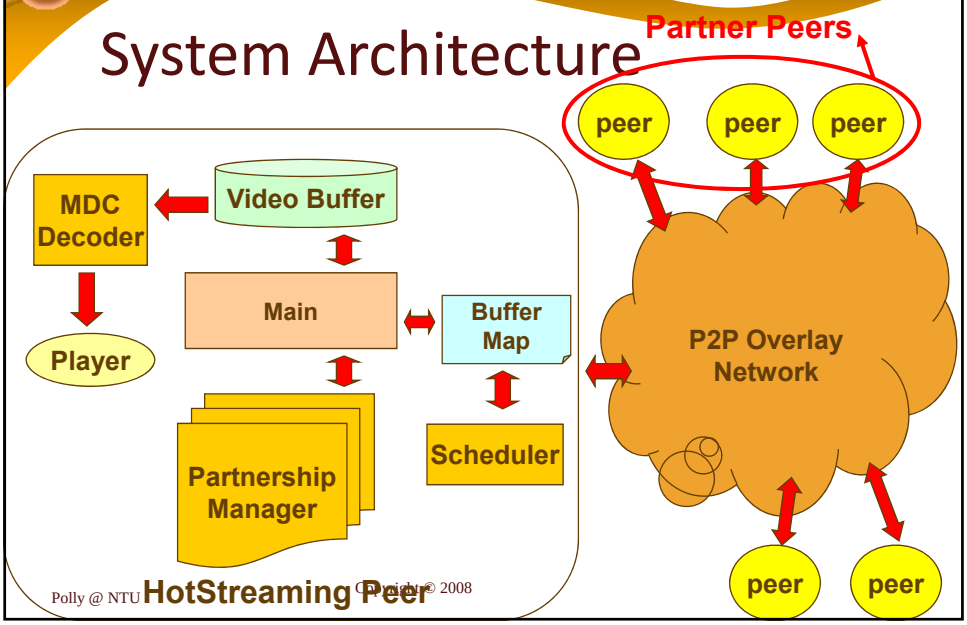
System Advantages

	HotStreaming	Advantages
Partnership Formation	Preferential random (rotate)	Stability
Frame Request Scheduling	Optimal: (min loss under bw constraint)	Minimum loss
Codec	Mpeg-4, interleaving MDC (bw heterogeneity)	Sustainable quality for heterogeneous users

Three Components

- Partnership formation
- Multiple description coding
- Video segment scheduling

System Architecture



Qualitative Comparison

- SCAMP
  - Forwarding of ADV
    - To all partners
  - Selection of partner
    - Depending on current #partners
- TYPHOON
  - Forwarding of ADV
    - To partners who have few partners
    - To avoid peers being isolated → better stability
  - Selection of partner
    - Depending on current #partners
    - Bounded by the max #partners → better load balance

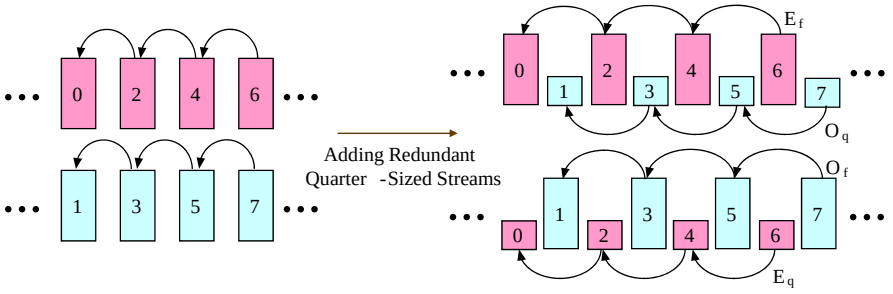
Quantitative Comparison

	SCAMP	TYPHOON
Connected Time (sec)	59180	74320
Disconnected Time (sec)	15820	680
Disconnected Node (N)	305	18
Instability Index	51700	2580

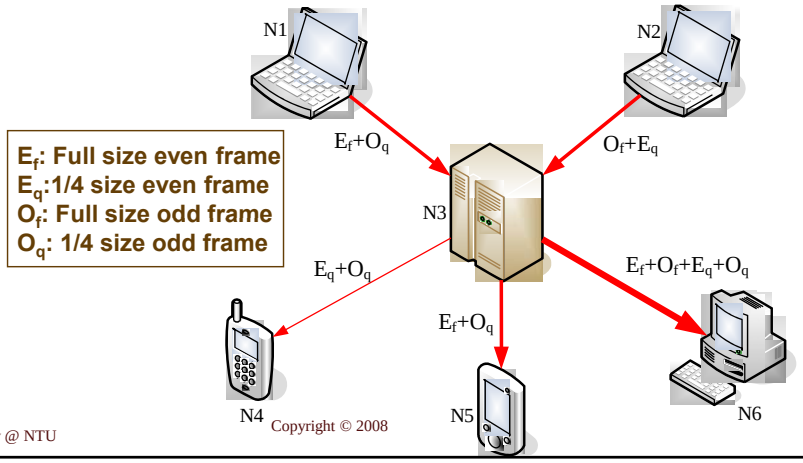
Stability Improvement ~ 20 fold

MDC: Idea

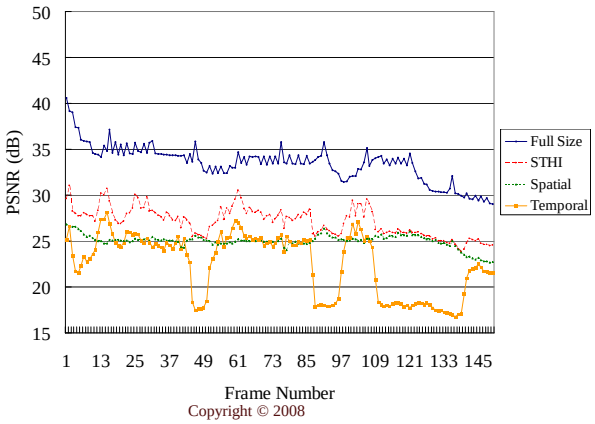
MDC-STHI:
MDC with Spatial-Temopral Hybrid Interpolation



Very Key to Mobile IPTV



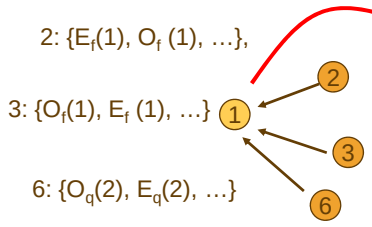
PSNR: Stefan (CIF)



Segment Request Scheduling

- From a peer's partners
  - Available segment type (e.g. E_f, O_q) and size
- The peer calculates
  - Maximize the video quality
    - Score of full-sized segments > quarter-sized ones
  - Under 2 constraints
    - Estimated time for segment to arrive < playback time
    - Total segments to request < estimated available bandwidth
- A linear programming problem
  - Efficient approximation algorithm
  - Computation < 10s milliseconds

Segment Request: Illustrated



Variables

x_i^j : a schedule that requires MDC type j for segment i

Maximize

$$\sum_{i=1}^{K}\sum_{j=1}^{L} x_i^j U_i^j$$

Subject to

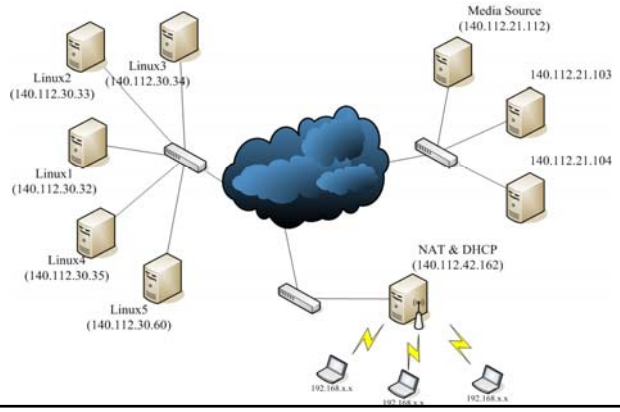
$$T_i = T_{i-1} + 1 - \sum_{j=1}^{L} \frac{S_i^j}{B} x_i^j > 0, 1 \le i \le K$$

$$0 \le \sum_{j=1}^{L} x_i^j \le 1, 1 \le i \le K$$

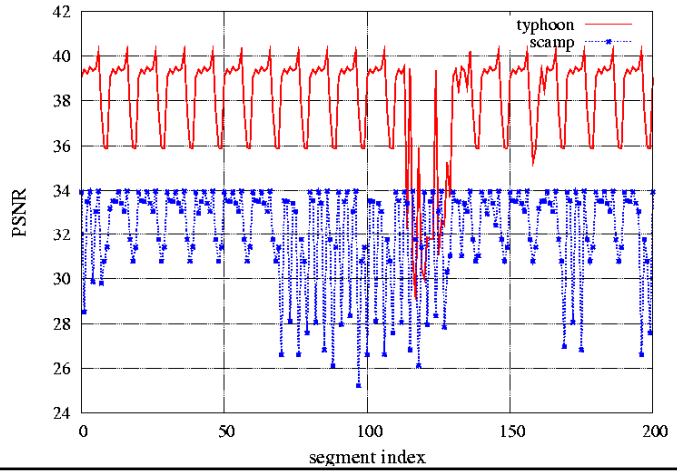
$$x_i^j \in \{0,1\}, 0 \le x_i^j \le 1, 1 \le i \le K, 1 \le j \le L$$

Testbed Experiments

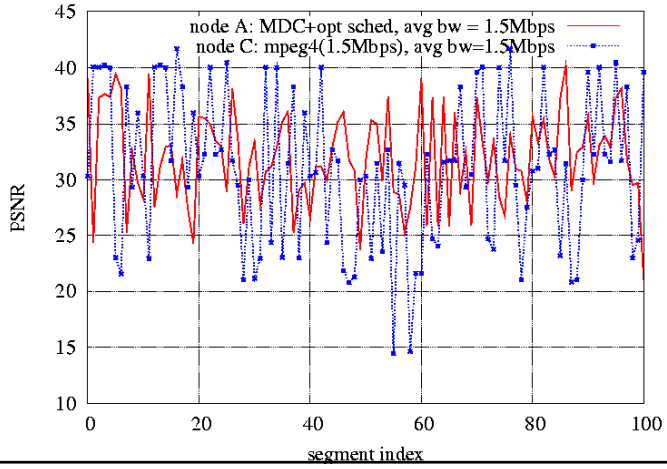
- All three components integrated



Impact of TYPHOON



Impact of MDC-STHI+ Opt Schedule



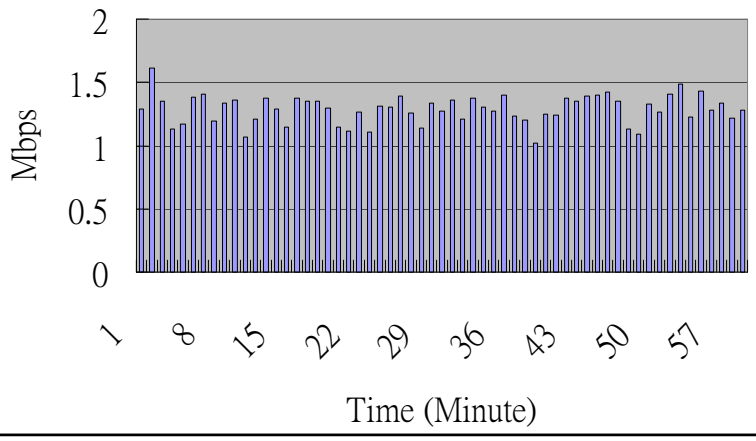
Summary

	HotStreaming	Advantages
Partnership Formation	Preferential random (rotate)	Stability
Frame Request Scheduling	Optimal: (min loss under bw constraint)	Minimum loss
Codec	Mpeg-4, interleaving MDC (bw heterogeneity)	Sustainable quality for heterogeneous users

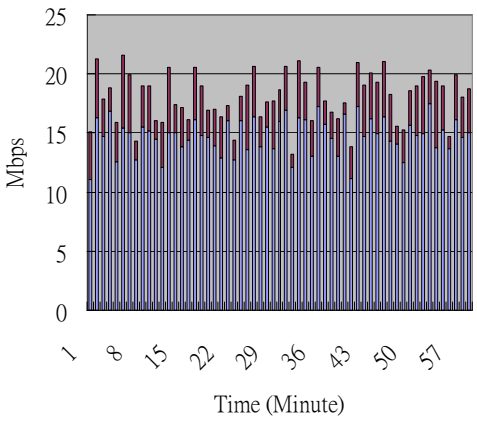
PlanetLab Results

- Small-scale testing on Campus Network
  - 15-node scale testing
  - 1 server, 14 peer users
  - 1 Mbps video source for an hour
- Network centric measurement
  - Scalability
  - Packet loss rate (arrival rate within deadline)
  - Control message overhead

Video Receiving Rate: per Peer



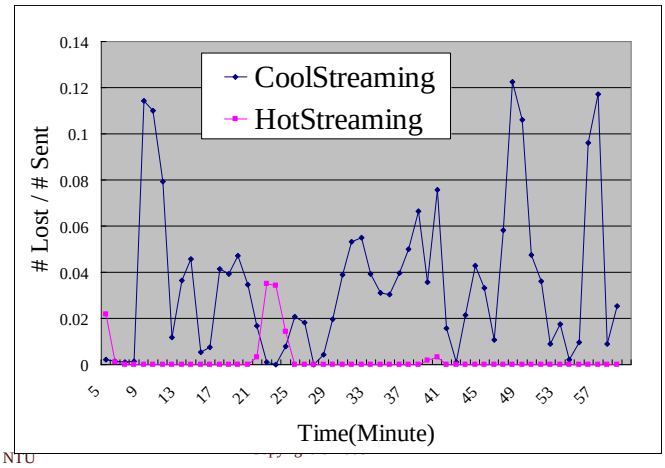
Video Sending Rate: Total



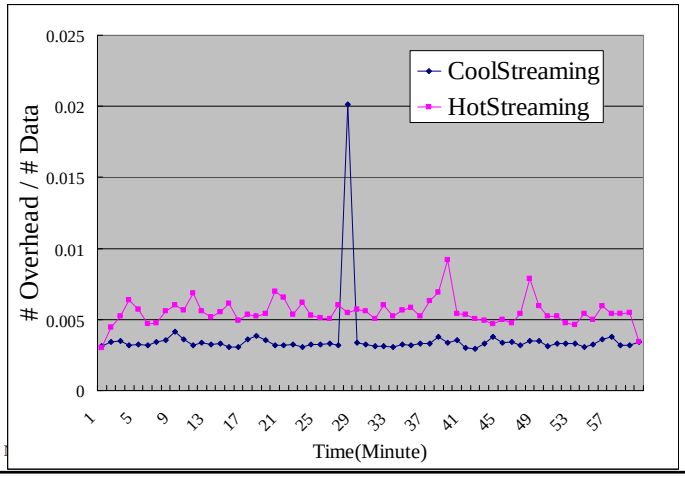
Scalability (server load)
MOD ~ stream_size * #user
P2P ~ stream_size * in_degree

→ For non-live-programs, no user limit on HotStreaming!

Playback Continuity



Overhead




Network and Systems Laboratory
nslab.ee.ntu.edu.tw
Questions?
Polly @ NTU
Copyright © 2008
77