Robust & Lightweight Networking

R. Guérin
University of Pennsylvania
Multimedia and Networking Laboratory

Acknowledgement

- This talk is based on joint work and discussions with many people
  - Former Ph.D. students
    - Y. Huang, A. Sridharan, S. Tao, E. Vergetis
  - And several colleagues
    - C. Diot, L. Gao, J. Kurose, S. Sarkar, D. Towsley, Z.-L. Zhang

whose contributions I gratefully acknowledge.
But don't blame them for errors or opinions you disagree with. These are all mine.

The Premises

- The networking "wars" are over – IP won!
  - Carries pretty much *any* application
  - Runs over pretty much *any* transmission technology
- But does it mean we have entered the "Golden Age" of the *simple* network, and that we can rest
  - In your dreams!
- Although we now have a "single" network, things have never been more complex
  - Lots of requirements, lots of options, lots of technologies, and to top it all a scale that keeps increasing

The Problem

- We now have ***one*** network that needs to be everything for everyone
  - Can it be done and how?
- Two (complementary) approaches
  1. Now that we have an answer (IP), lets see what we can do if we were allowed to start from scratch (the GENI perspective)
  2. Now that we have an answer (IP), lets see how far we can push it while preserving what allowed it to win in the first place (the hourglass perspective)

This Talk

- Very much focused on "option 2"
  - What can we build and how far can we go with what we have today
  - Motivational analogy: Networking is very much like politics – Don't underestimate the power of incumbency
    - Changes happen only if the incumbent is really bad and the contender really good, and/or we have a paradigm shift
- Some sample questions we'll investigate
  - Making IP routing intrinsically more robust
    - The *oblivious* routing perspective
  - Investigating the effect of scaling
    - What happens when traffic and network grow?
  - Leveraging the many faces of IP
    - Harnessing the power of diversity

Building Robustness into Routing

Robust Routing

- Goal: Configure IP routing to remain "near-optimal" across a broad range of conditions
  - Link and node failures, traffic fluctuations, etc.
- Lots of initial results pointing to feasibility of such an approach
  - Oblivious routing (primarily aimed at MPLS) - Sigcomm'03
  - Robust weight settings in traditional IP networks (Fortz & Thorup)
    - Robustness to traffic variations and failures
- Our focus: Robustness to all single link failures
  - Algorithm to compute link weights that produce near-optimal traffic distribution across all possible single link failure scenarios
  - Key challenge is computational efficiency

Approach & Results

- Two phase algorithm
  - Phase 1: Standard black-box traffic optimization used to gather statistics to identify *critical* links
    - That's the hard part!
  - Phase 2: Optimize for all failures of critical links only (much smaller set)
- Close to optimal across many network & traffic scenarios
- Computationally efficient
- Side benefit: Can help identify "extreme" scenarios
  - Intrinsically robust network
  - Intrinsically mis-matched network and traffic patterns

"Typical" POP-Level ISP Network

16 Nodes, 82 Links, Link Utilization of 0.7

Robustness to all possible single link failures

Heuristic	% Deviation from Optimal Re-routing									
	10	20	30	40	50	60	70	80	90	>100
Exhaustive	**9.1**	**81.8**	**9.1**	0	0	0	0	0	0	0
40 Critical Links	**13.3**	**73.3**	**6.67**	**6.67**	0	0	0	0	0	0
20 Critical Links	**10**	**75**	0	**5**	0	0	0	0	0	**10**
Plain OSPF	0	**3.33**	0	**16.67**	**3.33**	**6.67**	0	0	0	**70**

Large Network

150 Nodes, 432 Links, Link Utilization of 0.7

Robustness to all possible single link failures

Heuristic	% Deviation from Optimal Re-routing									
	10	20	30	40	50	60	70	80	90	100
10 Critical Links	**100**	0	0	0	0	0	0	0	0	0
Plain OSPF	0	0	0	0	0	0	0	0	0	**100**

Exhaustive search is impractical for this network size…

Note that things are getting better as network size grows (more on this later)

How Does Scale Affect Robustness?

Defining Robustness

- Assume network of size N nodes with link over-provisioning factor β (capacity of link ij is $l_{ij}=(1+\beta)E[f_{ij}]$)
  - β chosen to "accommodate" fluctuations in traffic f_{ij}
- Basic question: Does β ↘ as N ↗?
  - Lots of examples where scale helps
    - Trunking efficiency
    - Statistical multiplexing
- Networks are, however, complex beasts
  - Interactions between many parameters
    - Topology, routing, base traffic, traffic surges, etc.
  - Plus limited understanding of how things will grow
- Need to develop a "parametric" model to the extent possible

Network Model

- Two-level $\mathcal{G}(n,p)$ random graph
  - Two levels for intra-domain and inter-domain
    - Can grow domains and number of domains separately
  - Focus on core backbone networks
    - Usually mesh like and low degree
    - Impact of access networks (power-law component) can be captured through traffic matrix
- Independent and arbitrary *base* and *surge* traffic matrices
  - Vary total traffic intensity sourced by a node, distribution of destinations, growth as a function of network size, level of variability (random variables), etc.
  - Base traffic used to dimension the network
    - Link capacity is $(1+\beta)f_{ij}$,
  - Independent surge matrix captures possible traffic fluctuations

(Some) related works

- Aimed at estimating the load on the highest loaded link
  - Reached different conclusions because of their different choice of routing policy and traffic models
    - Gkantsidis, et al (ACM Sigmetrics 03), upper bound $O(n \log2 n)$
      - Assumptions: traffic proportional to src and dst degrees, optimal routing
    - Akella, et al (ACM PODC 03), lower bound $O(n1+1/\alpha)$
      - Assumptions: unit traffic between all node pairs, shortest path routing
- Key differences with these earlier efforts
  - Investigation of general traffic patterns.
    - Assume general traffic models: random traffic, base + surge
    - Explicitly quantify the contribution of traffic from individual src-dst pairs
  - Investigation of statistical properties of link load
    - Link load is a closed-form function of topology and the traffic parameters

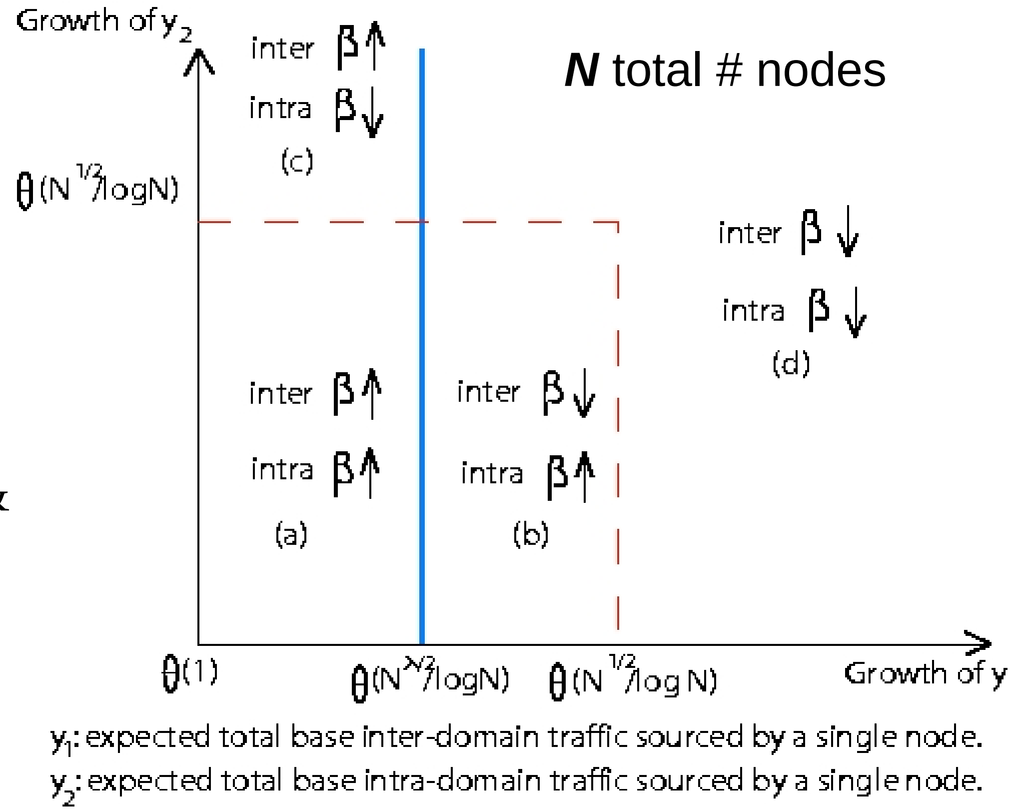
Scaling Properties of Over-Provisioning

- Back to our original question: How does a given β perform as network size grows (number of nodes per domain and/or number of domains) under different base and surge traffic growth models

- Basic approach:
  – Derive explicit expression for actual load on link *ij* as a function of network size, base and surge traffic, network topology, and routing
  – Use standard bounds, e.g., Chebyshev, to characterize the probability of link overload

Sample Representative Behavior

- Existence of clear thresholds that delineate regions of increasing and decreasing efficiency of over-provisioning
  - Traffic sourced by a node needs to grow faster than network size by a certain factor
- Some conclusions
  - Grow the network using faster & bigger routers, and not more small routers
  - Fewer large domains is better than many small domains
  - By and large things should get better for larger networks



N total # nodes

Base and surge traffic matrices are i.i.d
(similar results with other distributions)

A Quick Reality Check

Year	# Routers	Link Speed	# Links	Router Capacity	$N^{1/2}/\mathrm{Log}N$	Tput Increase
1988*	6	56kbps	4	~200kbps	3.15	-
1990*	15	1.544Mbps	8	~15Mbps	3.3	75
1995*	32	45Mbps	16	~100Mbps	3.8	500
2005†	600	10Gbps	64	~1Tbps (10^{12}bps)	8.8	5M

*** :** NSFNet progression

† : "Typical" large ISP backbone

We appear headed in the "right" direction

Taking Stock

- Will IP networks collapse under their own weight as they keep growing?
- Maybe, but likely for other reasons than raw performance or intrinsic instability
  - They should get better as they get bigger
    - Global QoS is even less of a requirement
  - No main reason for why basic IP routing should not be enough
    - Neither traffic engineering nor restoration capabilities seem enough of a motivator

➔ Plus we can (maybe) allow further improvements by letting *users* themselves take advantage of the scale/diversity of IP networks

On the Power of External Solutions

Broad Problem Setting

- An increasing number of (diverse) transmission options
  - Wireline and wireless
- Basic question
  - Can we take advantage of that *diversity* to improve MY network “performance”?
    - Diversity has helped a lot at the physical layer
      - MIMO, OFDM, etc.

Much Prior Work

- Maxemchuk “Diversity Routing”
  - 1975 Ph.D. thesis
- Some more “recent” investigations
  - Open-loop (diversity routing/coding)
    - Golubchik et al. (2002)
    - Abdouni et al. (2005)
    - Tsirigos & Haas (2004)
  - Closed-loop (path switching)
    - Chandra et al. (2004),
    - Miu et al. (2005),
    - Akella et al. (2005),
    - Tao et al. (2005)

Our Focus

- Packet-level solutions (more on this in a minute)
  - Portable across channel "types"
  - Minimal added complexity
- Closed-loop and Open-loop
  1. Closed-loop: Feedback based on simple channel monitoring
  2. Open-Loop: Long-term channel statistics and no feedback
- Goals
  - Better understanding of when and what performance benefits are achievable, and *how* to achieve them
  - Experimental validation (does is *really* work?)
    - Wide-area testbed of path-switching solutions for VoIP and video
    - 802.11 testbed of open-loop diversity (multiple frequency bands) solutions

Closed-Loop Solutions

Closed-Loop – Path Switching

- Approach
  1. Select multiple (uncorrelated) paths
  2. Monitor and predict path quality
  3. Send traffic on expected best path
- Basic issues
  1. How to select uncorrelated paths?
  2. How to monitor & predict path quality?
  3. When to switch from one path to another
     - Definition of "better" path is often application specific

1. Selecting Alternate Paths

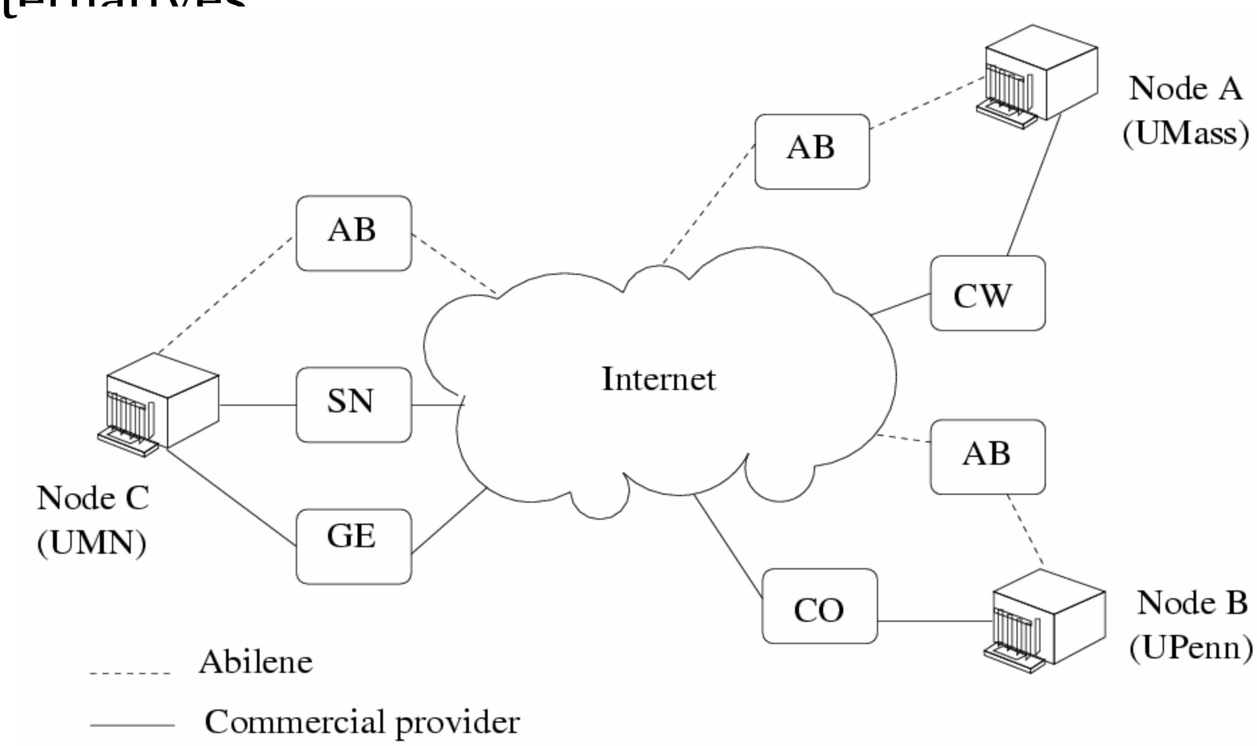
- Peer-to-peer systems are “ideal” for providing access to many alternate paths
- But, having many choices does not equate making a good choice
- Goal
  – Provide a lightweight solution for selecting peer nodes that provide “uncorrelated” alternate paths

Basic Strategy

- Use source specific AS path information
  - From source to destination
  - From source to peers
- Select overlay peer based on earliest divergence at the AS path level
  - Earliest Divergence Rule (EDR)
  - Earliest Branching Rule (EBR)
- Offers lightweight (minimal information gathering and updating) selection method that delivers good performance across a broad range of scenarios

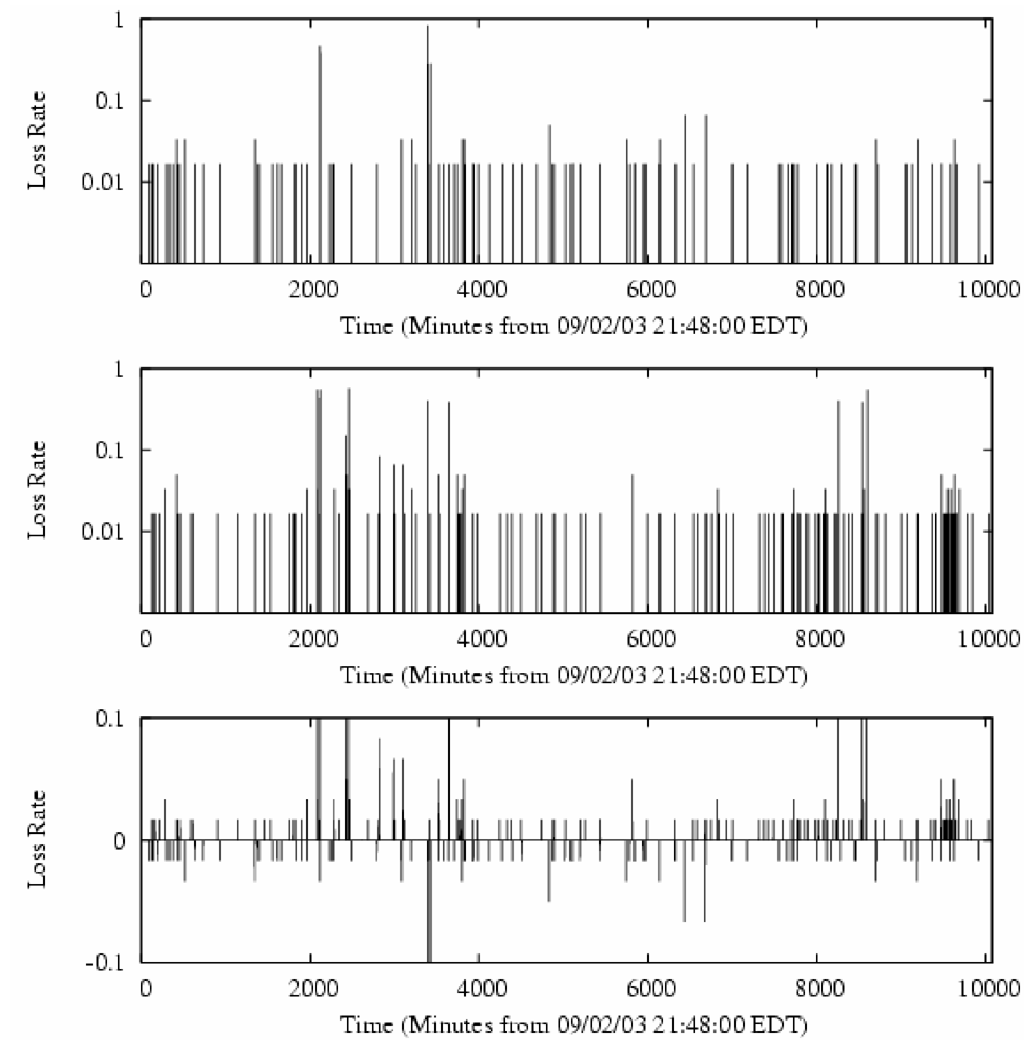
2. Monitoring Paths

- What type of path performance predictors?
  - Testbed involving multiple providers and overlay alternatives



Basic Conclusions

- Path performance monitoring can be carried at relatively coarse time-scale (10’s of secs)
- Simple path prediction based on previous state performs as well as more complex models
- Switching to the predicted better path can yield meaningful improvements



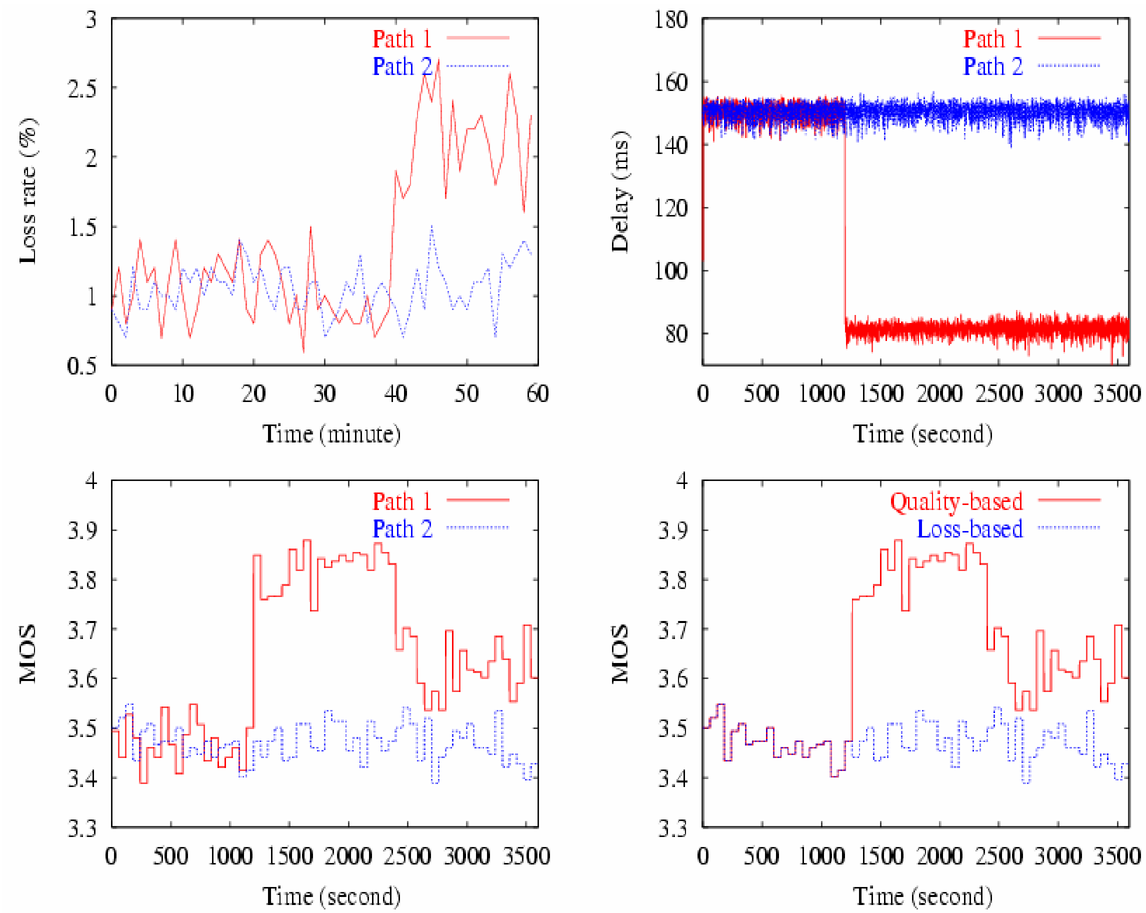
3. Switching to the Best Path

- How to define the "best" paths
  - Many different metrics
    - Loss rate, loss burstiness, delay variations, etc.
- Definition of "best" is very much application specific
  - VoIP quality depends on loss rate, loss patterns, delay, codec, etc.
  - Even worse for video, as the *type* of video also plays a role
- Conclusion: Maybe it's best to let the application decide!
  - Is it feasible and how?
  - Does it matter?

- Note: This is an *old* problem that in some way has plagued earlier approaches in deploying *network* QoS

The VoIP Scenario

- Yes it is feasible
  - Models exist that "easily" map network performance to voice quality, e.g., E-model
- And yes, it can make a difference
  - Two paths with different loss and delay characteristics

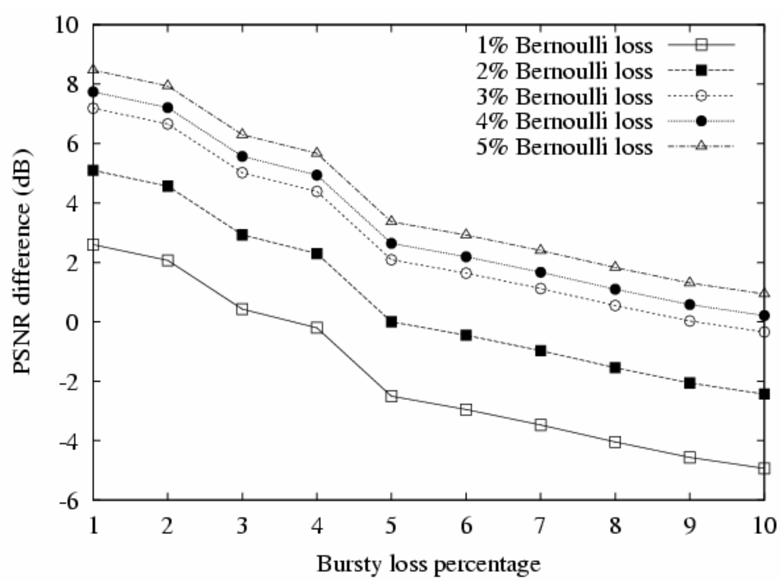


The Video Scenario

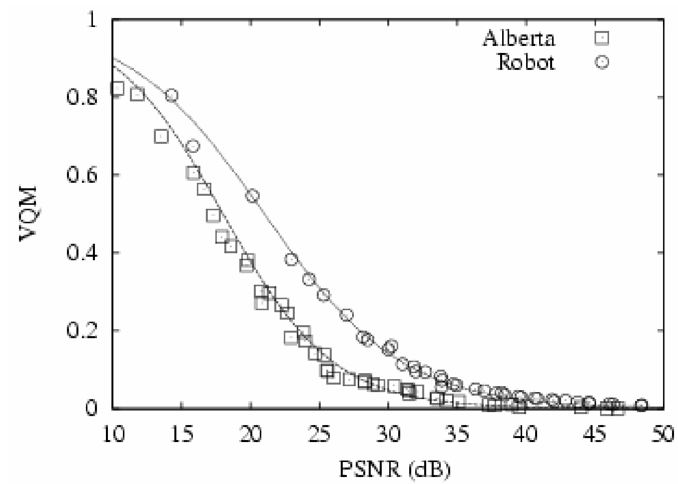
- A more complex situation
  - Most quality assessment models rely either on *comparing* received and original videos, or on performing complex feature extraction in received video
  - Diversity of videos, codecs, and error concealment strategies, further complicates matters

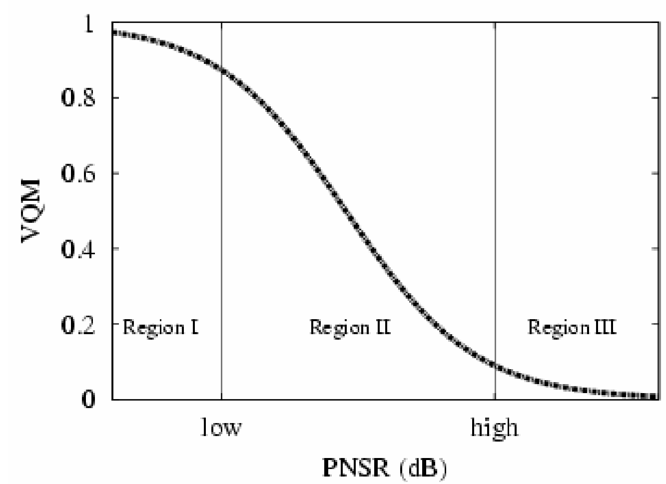
Network Quality ≠ Video Quality

- Bernoulli and bursty losses have different impact



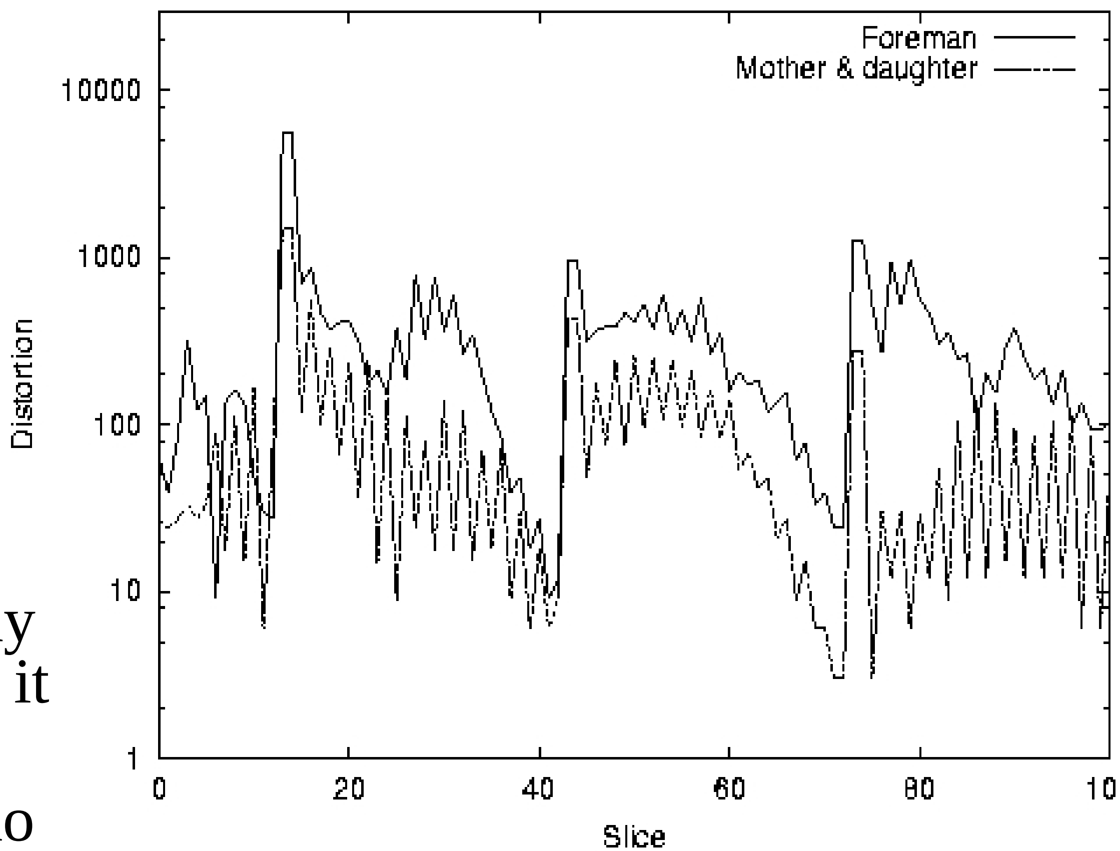
- And the impact varies **non-linearly!**





Network Quality ≠ Video Quality

- Things also depends on
  - What frame/slice is lost within a group of picture
  - The content of the video itself
    - High vs low motion
    - Scene changes
  - The frame format and packetization scheme



- In short – There is no way the "network" can figure it out
- But can the application do it?

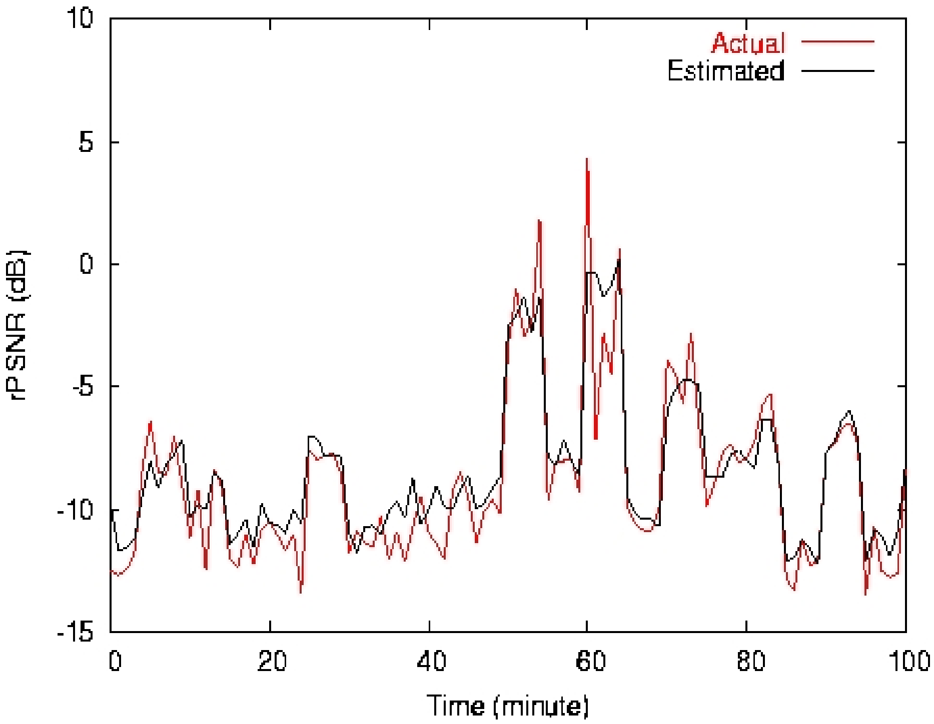
Lightweight Application Quality Estimation

- *Relative* quality metric, *rPNSR*
  - Quality difference between received video and video that *would have been received* on a baseline path
  - Eliminates most of the video specific components
- Enables relative path quality estimation based *only* on network performance measurements
- Challenges
  - Choosing the baseline path
  - Better does not mean good, and worse does not mean bad!

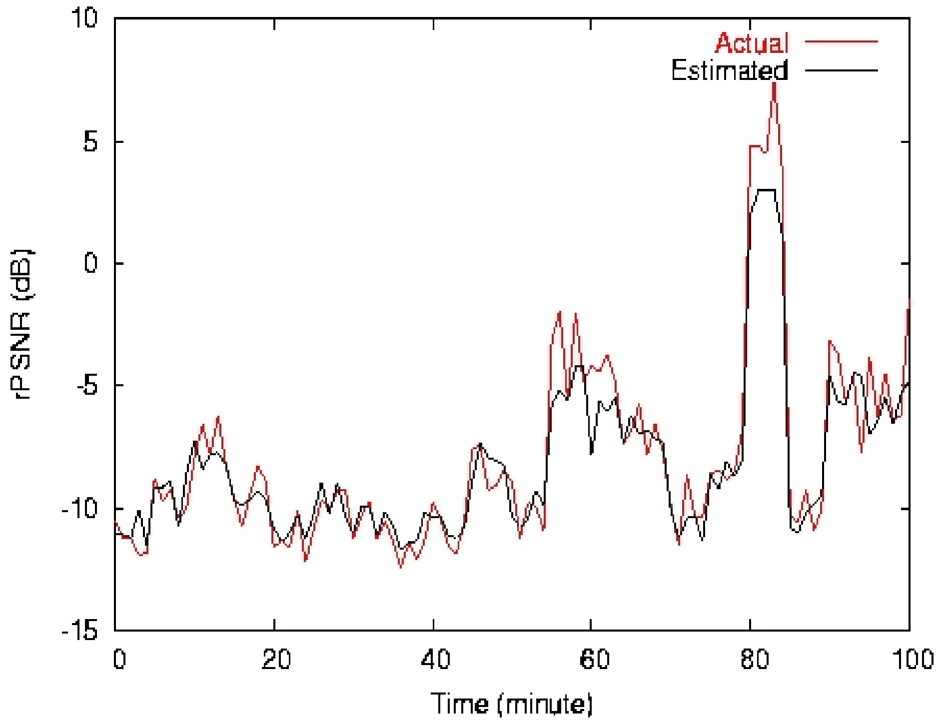
Experimental Validation

- Video transmission on the path between UMN and UPenn
  - Estimation interval of ~1 minute

H.264



MPEG-2



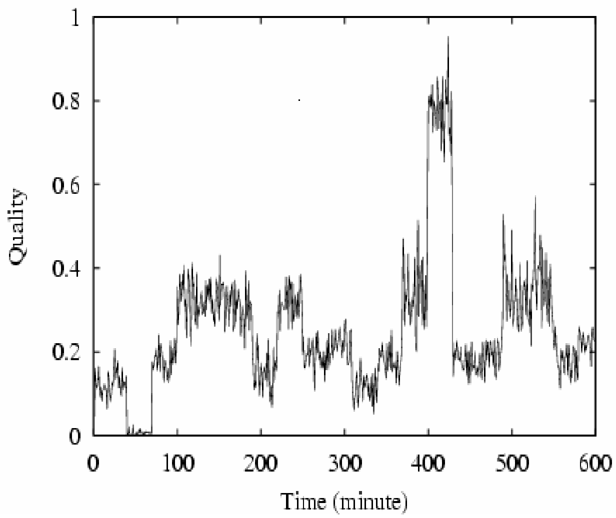
Application to Path Switching

Table 1: Quality improvement from path switching

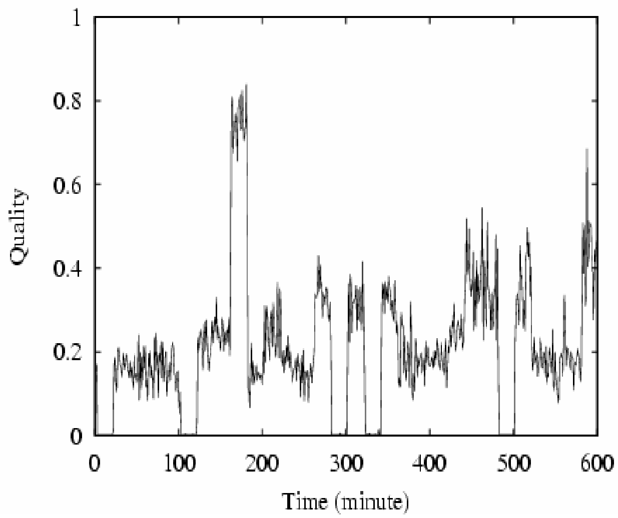
	Path 1	Path 2	Path switching
Overall quality	0.251	0.214	0.165
Quality variation	0.158	0.176	0.108

(0 is best quality)

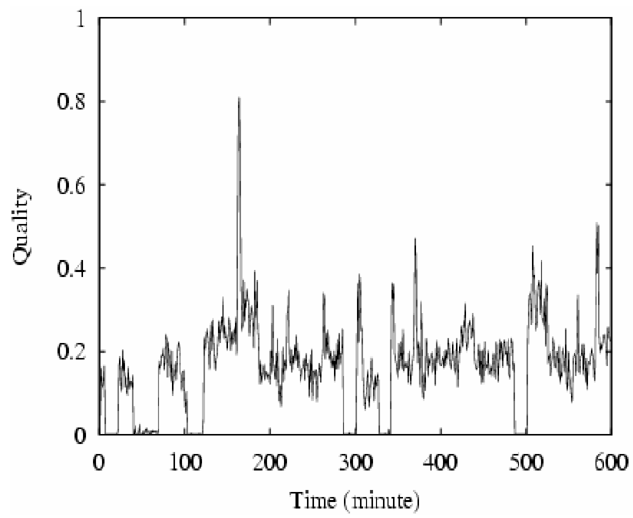
Path 1



Path 2



Path Switching



Open-Loop Solutions

Open-Loop Diversity – The Theory

- User can choose from *C* channels with "known" statistics
  - Long-term error rate (LTER), expected burst length (EBL)
  - User transmissions do not affect channel statistics
- User distributes packet transmissions across all *C* channels according to some policy
  - Deterministic and probabilistic policies
- User wants to maximize performance
  - Highest possible message (consisting of *k* packets) delivery rate that meets a certain reliability target $P_{\min}$
- Design knobs
  - Transmission policy
    - What set of channels to use and how?
  - Code selection
    - What (N,k) code to choose (smallest N that achieves $P_{\min}$)?

Open-Loop Diversity – The Results

- Under certain assumptions
  - Channel independence
  - Stationary, Markovian (Gilbert-Elliot) channels
  - No overhead in switching transmissions from one channel to another

- In scenarios where diversity "helps," a simple, round-robin policy is close to optimal, and "usually" wont hurt
  - Higher effective message rate (*ER*) and relative insensitivity to errors in channel statistics

Open-Loop Diversity – The Intuition

- Channel diversity is useful because/when
  - It allows breaking-up error bursts
  - It avoids being "stuck" with a bad channel
- Deterministic, round-robin policy works well because
  - It spaces out successive use of a given channel (minimizes the odds of coming back early to an ongoing error burst)
- Distributing packet transmissions across multiple channels yields
  - The ability to use a smaller code length N to satisfy Pmin
  - And/or a higher probability of successful message delivery
  - Most of the gains from diversity are through reducing N

Open-Loop Diversity – The Questions

- How well do the assumptions hold in practice?
  - Independent, stationary channels, with known statistics
  - No impact of user transmissions on channel statistics
  - No channel switching overhead

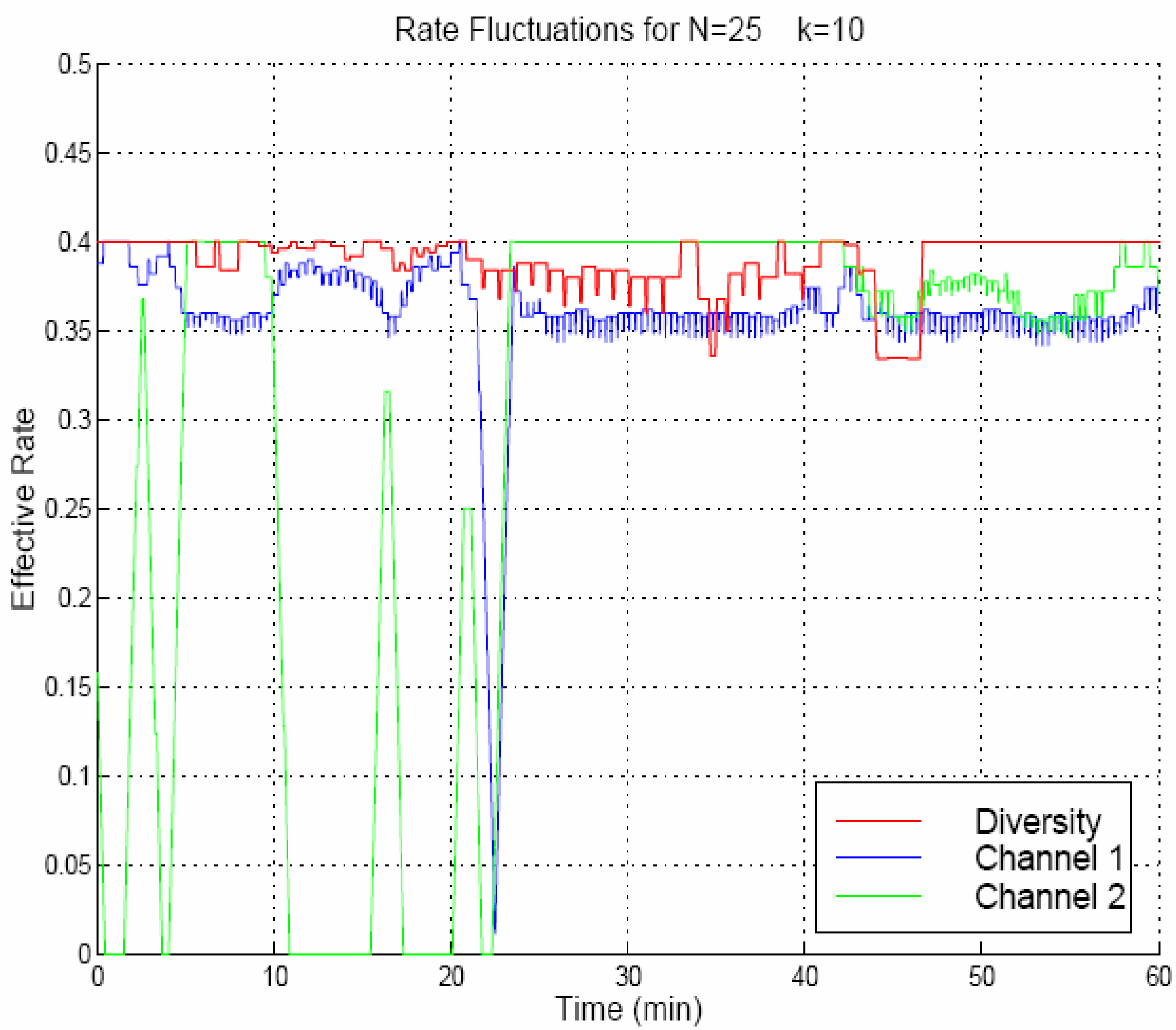
- What can actually be realized?
  - 802.11b environment
  - Standard end-systems (PCs) without precise control of transmission timings

Experimental Setup

- Two 802.11b Access Points (APs)
  - Intel StarEast board, with one miniPCI NIC each
  - External omni-directional antennas
  - Assigned “non-overlapping” frequency bands
  - Located ~1m from each other
  - Logging of all incoming packets without performance degradation
  - Within reach of other APs interfering in all 11 frequency bands
- Sender
  - Standard laptop with *two* NICs
    - One external PCMCIA NIC, and one internal miniPCI NIC
    - Linux operating system
    - Transmission speed set at 2Mbps
  - Located between 2m and 10m away from the two APs
    - Maintains association with both APs
    - Line-of-sight (LoS) as well as non-LoS (indoor wall) transmissions

Diversity As a Performance Stabilizer

- Two channels:
  - $LTER_1 \sim 11.4\%$
  - $EBL_1 \sim 10$ pkts
  - $LTER_2 \sim 29.2\%$
  - $EBL_2 \sim 11$ pkts
- *ER* measured over a 200 messages sliding window
  - Mean value improves by 6%/30%
  - Variance decreases by 60%/90%



Wrap-Up

- Clearly we should not stop dreaming up new networking technologies and architectures!
- But, we need to be aware that new ≠ better
  - There are many, many different ways to make current IP networks better
  - The hourglass design paradigm is still valid and powerful

- In the end, Darwin tells us that we need both to ensure that we "evolve" towards better networks

Related References (2004)

1. S. Tao and R. Guerin, "On-line Estimation of Internet Path Performance: An Application Perspective." In Proc. IEEE INFOCOM'2004, Hong Kong, March 2004.
2. S. Tao, K. Xu, Y. Xu, T. Fei, L. Gao, R. Guerin, J. Kurose, D. Towsley, and Z.-L. Zhang, "Exploring the Performance Benefits of End-to-End Path Switching." In Proc, ICNP 2004, October 5-8, 2004, Berlin, Germany.
3. S. Tao and R. Guerin, "Application-Specific Path Switching: A Case Study for Streaming Video." In Proc. ACM Multimedia 2004, October 10-16, 2004, New York, NY.

Related References (2005)

4. S. Tao, K. Xu, A. Estepa, T. Fei, L. Gao, R. Guerin, J. Kurose, D. Towsley, and Z.-L. Zhang, “Improving VoIP Quality Through Path Switching.” In Proc. IEEE INFOCOM'2005, Miami, FL, March 2005.
5. A. Sridharan and R. Guerin, “Making IGP Routing Robust to Link Failures.” In Proc. Networking'2005, University of Waterloo, Ontario, Canada, May 2005.
6. S. Tao, J. Apostolopoulos, and R. Guerin, “Real-Time Monitoring of Video Quality in IP Networks.” In Proc. ACM NOSSDAV'05, Skamania, OR, June 2005.
7. E. Vergetis, R. Guerin, and S. Sarkar, “Improving Performance Through Channel Diversity in the Presence of Bursty Losses.” In Proc. ITC'19, Beijing, China, August 2005.
8. E. Vergetis, R. Guerin, and S. Sarkar, “Realizing the Benefits of User-Level Channel Diversity.” ACM SIGCOMM Computer Communication Review, Vol. 35, No. 4, October 2005.
9. Y. Huang and R. Guerin, “Does Over-Provisioning Become More or Less Efficient as Networks Grow Larger.” In Proc. ICNP 2005, Boston, MA, November 2005.

Related References (2006)

10. T. Fei, S. Tao, L. Gao, and R. Guerin, "How to Select a Good Alternate Path in Large Peer-to-Peer Systems?" In Proc. IEEE INFOCOM'06, Barcelona, Spain, April 2006.
11. T. Fei, S. Tao, L. Gao, R. Guerin, and Z.-L. Zhang, "Light-Weight Overlay Path Selection in a Peer-to-Peer Environment." In Proc. Global Internet Symposium, Barcelona, Spain, April 2006.
12. E. Vergetis, E. Pierce, M. Blanco, and R. Guerin, "Packet-Level Diversity – From Theory to Practice: An 802.11-based Experimental Investigation." In Proc. Mobicom 2006, Los Angeles, CA, September 2006.