The Content-Pipe Divide

Mung Chiang
Electrical Engineering, Princeton University

ICME
June 22, 2009

A Brainstorming Talk

- Architectural decisions

CDN-ISP interaction

Content-aware networking

- Network economics

Two-sided pricing

- Fundamental limits

Content distribution capacity

P2P streaming capacity

Acknowledgement

Co-authors of papers behind this talk:

- Students and Postdocs: Prashanth Hande, Ying Li, Shao Liu, Wenjie Jiang, Rui Zhang-Shen
- Colleagues: Rob Calderbank, Gary Chan, Minghua Chen, Jennifer Rexford
- Industry: Phil Chou, Jin Li, Zhu Li, Sudipta Sengupta

Keith Cambron, Dah-Ming Chiu, John Lui, Raj Savoor, Steve Sposato

NSF, ONR, AT&T, Microsoft, Motorola

Content-Pipe Divide

Dominated by sharing content, especially video

- IPTV and triple play
- PPLive and BitTorrent
- YouTube and Internet TV
- Facebook and Web 2.0

The third wave of Internet usage

Shake many basic assumptions in network design:

- Uplink-downlink asymmetry, pricing model...
- "Horizontal decomposition": access-metro-core hierarchy...
- "Vertical decomposition": layering, socio-tech separation...

Leads to Content-Pipe Divide

Content Side

Those who generate and distribute content

- Media companies who own video and music
- End-users who post video online
- Operators of content distribution network (CDN)
- Operators of peer-to-peer (P2P) sharing systems

- Seek the best way to distribute content
- Through multimedia signal processing, caching, relaying, sharing...
- Take the network as just a means of transportation

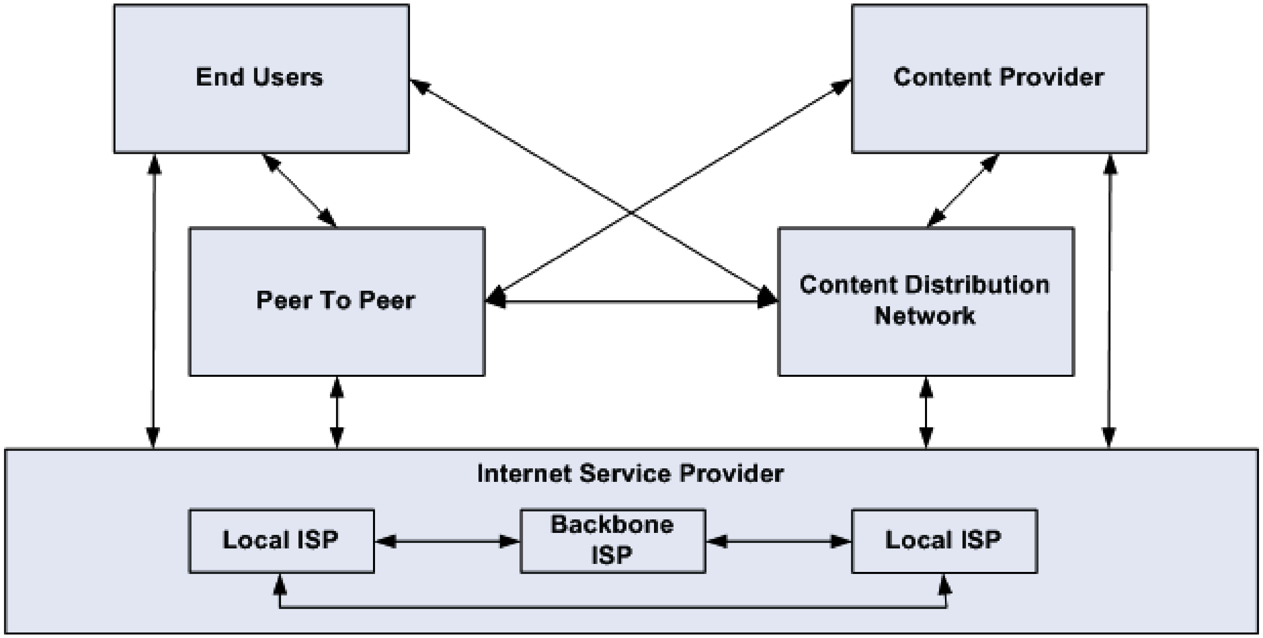
Pipe Side

Those who design and operate the network

- Internet Service Providers (ISP)
- Equipment vendors
- Network management software vendors
- Municipalities and enterprises running their own networks

- Seek the best way to manage network infrastructure,
- Through resource allocation on each link, between links, and end-to-end
- Take the content as just bits to transport between given nodes

Content-Pipe Interactions



- Any pairwise interaction is interesting
- Triangle of interactions even more interesting
- So are multipaths in the interaction diagram

I. Architectural Decisions Revisited

- W. Jiang, R. Zhang-Shen, J. Rexford, and M. Chiang, "Cooperative content distribution and traffic engineering in an ISP network", *Proc. ACM Sigmetrics/Performance*, June 2009
- Y. Li, Z. Li, M. Chiang, and A. R. Calderbank, "Content aware distortion fair video streaming in networks", *IEEE Transactions on Multimedia*, 2009

Traffic Management

- ISP run traffic management protocols (TCP congestion control and intra-AS routing), assuming that the traffic matrix is fixed and can be accurately estimated
- On possibly different timescales, server selection by CDN (or peer selection by P2P) changes the traffic matrix by adapting in their own way to the user-perceived delay and throughput

A feedback loop is present

A variables-constants mirror image in two optimizations

CDN-ISP Interaction

- ISP Model: Minimize total link cost function subject to flow conservation by traffic engineering
- CDN Model: Minimize weighted user-perceived delay subject to content demand constraints by varying server selection and rate assignments

- Analysis: Model as non-cooperative game

Current practice: Nash equilibrium exists but Pareto-inefficient

CDN getting accurate information: still suboptimal

- Design: Joint control by Nash bargaining solution

Sharing Information

ISP shares topology, link capacities, routing information to CDN

- Good News

Existence of Nash equilibrium

Global optimality under same objectives and no background traffic

- Bad News

Pareto-inefficient and optimality gap arbitrarily large.

Paradox of extra information: more information could hurt CDN.

- From experiments: Limited improvement, ISP should be cautious on sharing information alone.

Sharing Control

Clean-slate design: control which content on which pipe, a broader design space revisited. Yet, requires,

Social optimality: Pareto-efficient

Fairness: tradeoff between network and user objectives

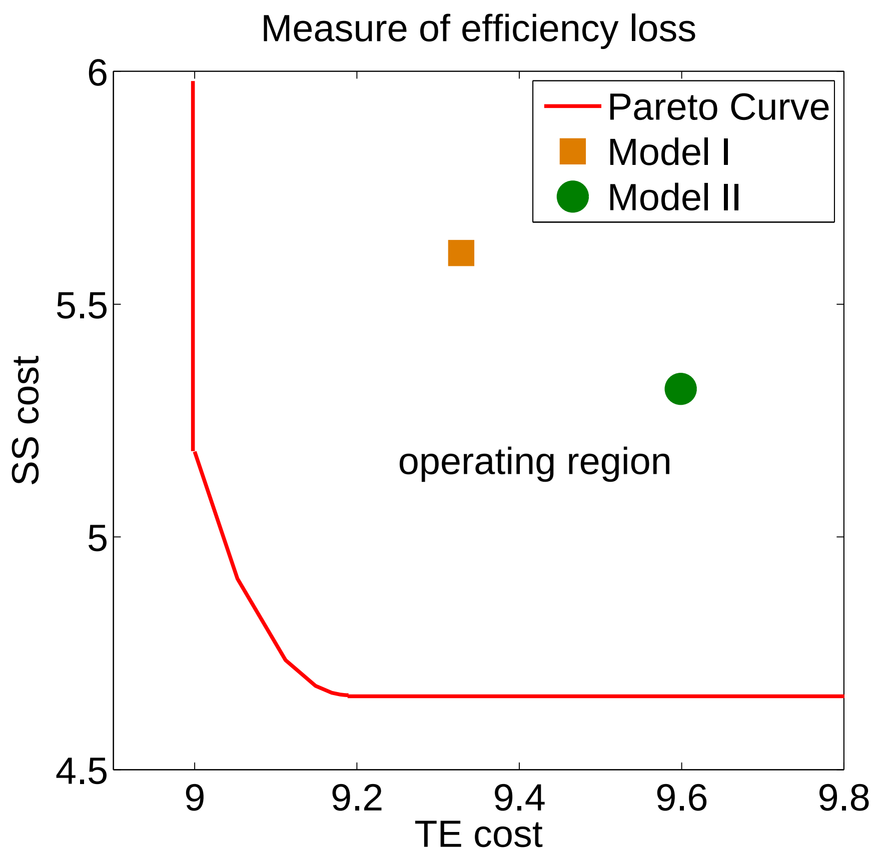
Modularity: keep separate TE and CDN functionalities

Decomposability: distributed protocols and information

Use Nash Bargaining solution

Distributed implementation: pass information, but hide objectives

Illustration



CP-ISP Interactions

	ISP no change	ISP changes
CP no change	Current practice	
CP changes		

- Collaboration III: Sharing control
- Collaboration II: Sharing various types of information
- Collaboration I: Anticipate reaction by the other party

Bottleneck: Lack of unilaterally-actionable, incrementally-deployable, and backward-compatible strategies for cooperation

Content Aware Networking

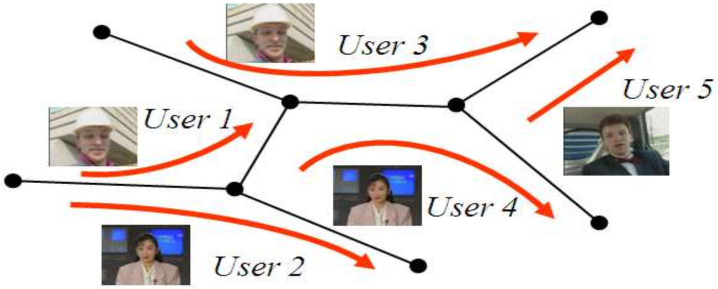
Generation and processing of multimedia signals have traditionally been designed in separation from the way the resulting packets are treated inside the network, e.g., shaping, queueing, marking, and dropping

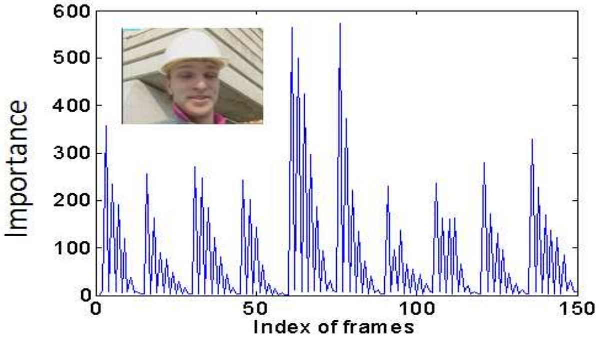
Content Aware Networking (and Network Aware Content Generation):

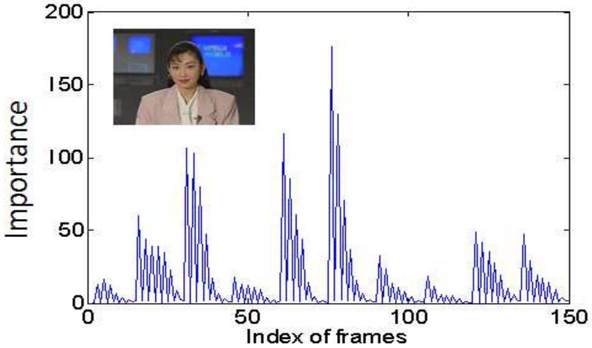
Jointly designing how video packets are coded and transported

- Dropping packets by frame-utility (e.g., I, P vs. B)
- Assigning multiple streams of video packets coded differently for the same source on multiple paths
- Joint summarization and resource allocation by Layering as Decomposition

An Example



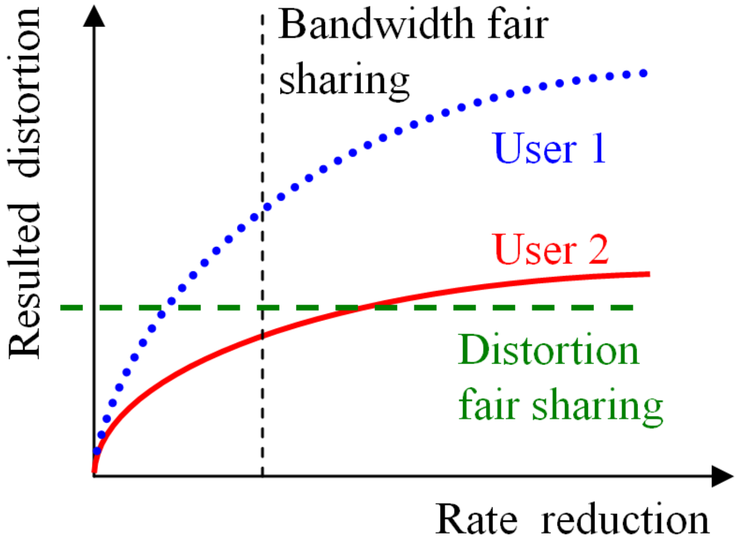




- Frames have different importance based on video content and frame dependency

Content-aware Packet Dropping during Congestion

- Distortion-fair, rather than rate-fair
- Slow motion video compensate for fast motion video



An online video demo:
http://www.princeton.edu/∼yingli/research/CAF.htm

II. Network Economics

- P. Hande, M. Chiang, A. R. Calderbank, and S. Ragan, “Network rate allocation with content provider participation”, *Proc. IEEE INFOCOM*, April 2009

Challenge 1: ISP Service Exposure

ISP's current revenue models drying up

The need for new interfaces between ISP and what it supports:

- Bandwidth meter-reading for users
- Incentive-compatible sockets for application developers
- Profit-margin-sharing for ISP

Challenge 2: Net Neutrality

- Many definitions of “net neutrality”
- Many parties with competing interests
- Many conflicting objectives

What kind of pricing structures by ISP over different content will be

- Efficient
- Fair
- Incentive-compatible
- Innovation-enabling
- Enforceable

Example 1: Two-Sided Pricing

ISP price the usage of bandwidth to end users and content distributors

- Flat rate, volume based, mixture

Different degrees of neutrality means different constraints on prices

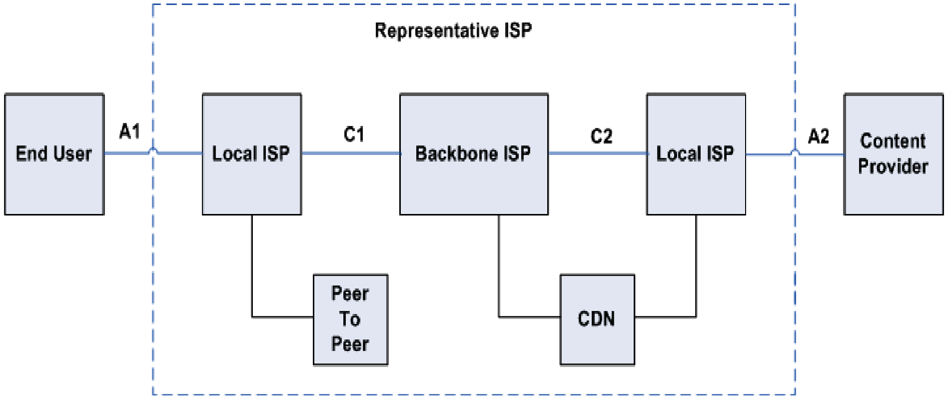
Lead to different

- Monopoly power
- Total surplus and distribution of surplus
- Efficiency-fairness tradeoff

Content Provider getting charged in two-sided market

- When (monopoly power and elasticity) is that good for everyone?

A Basic Model



- Data rate x (Mbps) between end-user and content-provider through representative ISP
- End-user utility $u(x)$ and content-provider utility $v(x)$, price takers
- μ is unit cost of capacity, end-user pays p and content-provider pays q per unit data-rate
- End-user demand $y(p) = u'^{-1}(p)$, and content-provider demand $z(q) = v'^{-1}(q)$

Scenarios

Competitive ISP Market

- ISP is also price-taker, resulting in a three way game between ISP, end-user and content-provider
- Resulting equilibrium:

$$\begin{aligned} p + q &= \mu \\ y(p) &= z(q) \end{aligned}$$

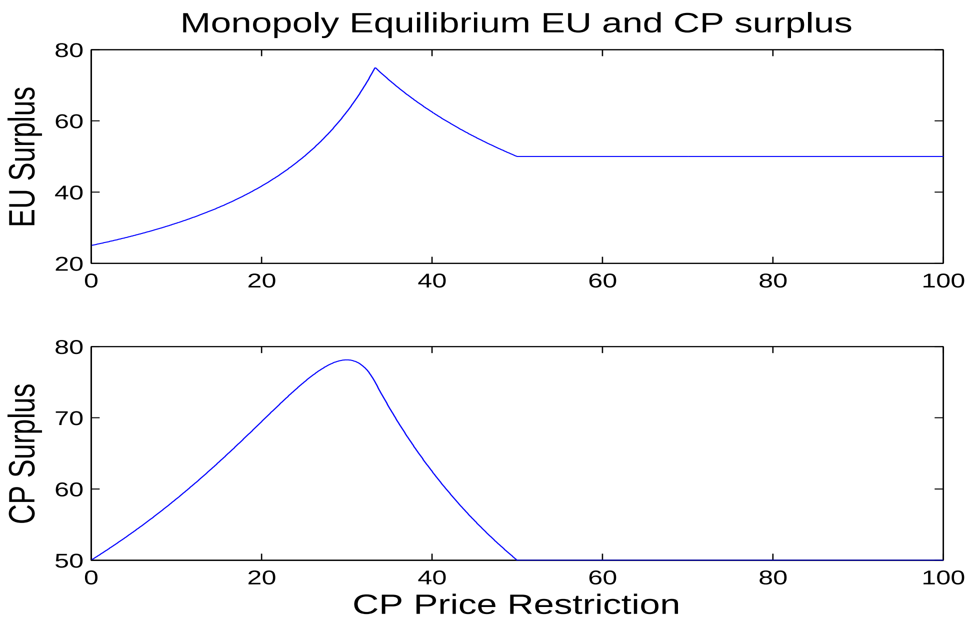
Monopoly ISP Market

- ISP determines price to maximize profits
- Analyze equilibrium in terms of end-user (η^E) and content-provider (η^C) demand elasticities
- Resulting equilibrium:

$$p(1 - 1/\eta^E) + q(1 - 1/\eta^C) = \mu$$

$$x = y(p) = z(q)$$

Effect of Price Restriction: Monopoly ISP

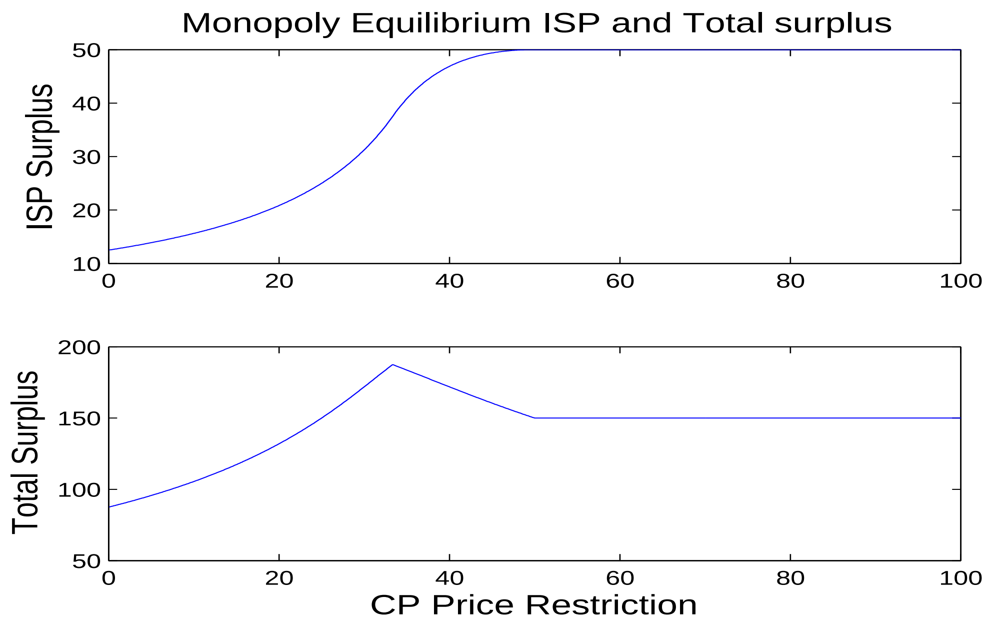


- Content-provider price restricted to $\bar{q}$, analyze effect through measure of economic surplus: net-utility or net profit

- End-user benefits from relaxation of $\bar{q}$ under both ISP competition and monopoly

- Content-provider benefits from relaxation of $\bar{q}$ provided end-user elasticity η^E is high and capacity cost μ is low

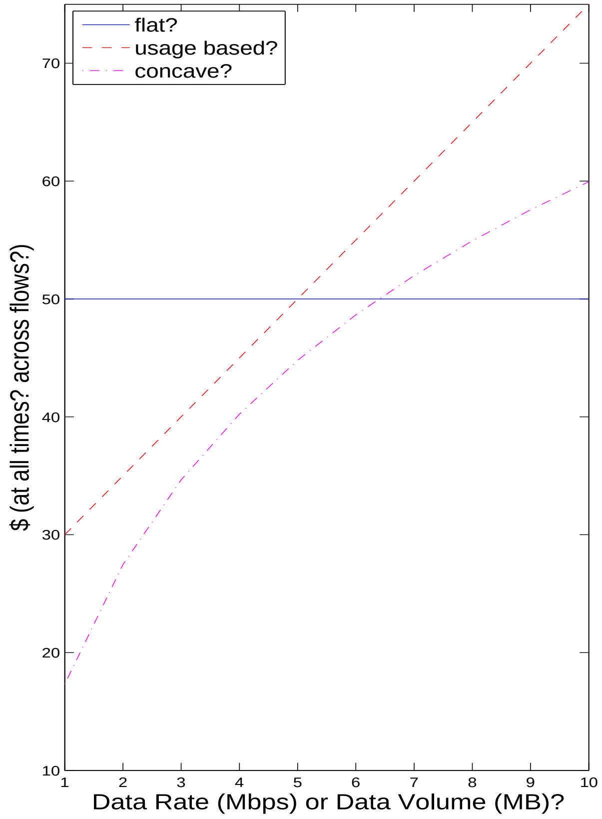
Effect of Price Restriction: Monopoly ISP

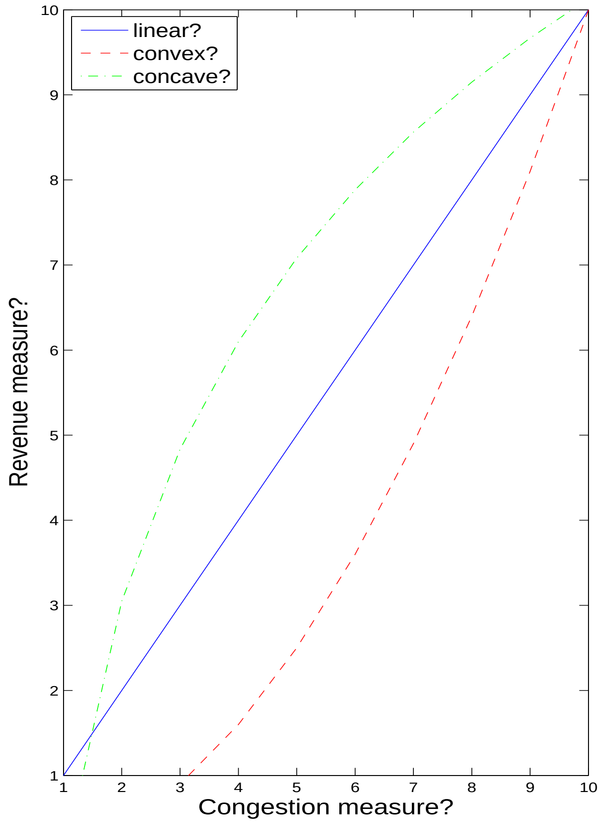


- ISP benefits from relaxation of $\bar{q}$ with monopoly power

- Total-surplus increases from relaxation of $\bar{q}$ under ISP competition, attains a peak at $\bar{q} = q^t$ under monopoly

Example 2: Adaptive Access Charging





Many More Examples in Two-Way Interaction

- Economics shapes technology evolution
- Technology's potential changes economics decision making

III. Fundamental Limits

- S. Liu, R. Zhang-Shen, W. Jiang, J. Rexford, and M. Chiang, “Performance bounds for peer-assisted live streaming”, *Proc. ACM Sigmetrics*, June 2008
- S. Sengupta, S. Liu, M. Chen, M. Chiang, J. Li, and P. A. Chou, “P2P streaming capacity”, *Submitted to IEEE Transactions on Information Theory*, 2009

Network Distribution Capacity

Various existing notions of capacity:

- Transportation of flows in a graph
- Largest rate subject to vanishingly small decoding error probability
- Largest set of arrivals subject to queue stability

Network Distribution Capacity (NDC) is a combination of these:

- Combinatorial problems of overlay graph construction ("distribution" in NDC)
- Communication network problems over various degrees of freedom ("network" in NDC)

Introduction of wireless components will bring further issues:

- Mobile nodes, time-varying links, shared medium

Constants

Given

- A directed graph $G = (V, E)$
- A set of contents $D = (D_1, \ldots, D_N\}$, where each D_i consists of 3-tuples: the size of the content M_i, a set T_i of destinations $T_{it} \in V,\ t = 1, 2, \ldots, |T_i|$, who demand the content, and a set S_i of sources $S_{is} \in V,\ s = 1, 2, \ldots, |S_i|$, who can supply the content

- The set of sources of content can become larger after more nodes $v \in V$ obtain the content
- A node can be a source but not a destination (server), a destination but not a source (client), or both (peer), or neither (router)
- Extensions: content chunk availability and peer churn

Some Variables

- Who serves whom
- Video codecs
- Capacity of links and nodes
- Transmission rate
- Routing: topology part and load balancing part
- Resource allocation: queue, schedule, power...

- The variables are obviously constrained with each other
- Source and destination sets are changing, and the construction of content distribution topology over time is a design variable

Some Metrics

- When $\{M_i\}$ are infinite, what are the time-averaged throughputs R_{it} of content distribution for each of the receivers T_{it}?
- For finite $\{M_i\}$ arriving at the system according to some pattern, what are the startup time or completion times Q_{it} of content distribution for each of the receivers T_{it}?

Other possibilities:

- Utility function based on user-perceived video quality
- ISP cost functions
- Robustness metrics

NDC

Model so far is neither complete nor tractable

- Adding details to the formulations
- Hold some degrees of freedom as constants
- Special case of (G, D)
- Take asymptotic limits along some dimension

What are the best achieavable R_{it} and Q_{it}?

- Inner bound
- Outer bound

A Tiny Corner for Now

	Technology	Economics and Policy
Content distribution alone	⋆	
Interact with transportation		
Interact with video processing		
Interact with social networks		

Many under-explored areas in this emerging field

P2P Streaming Capacity

Metrics:

- Throughput (maximum rate for all receivers)

Degrees of freedom:

- Overlay topology and peering relationship:

Tree (single or multi) or mesh or hybrid, Pull or push, Locality based, Clustered architectures for scalability

- Streaming rate (at application layer)

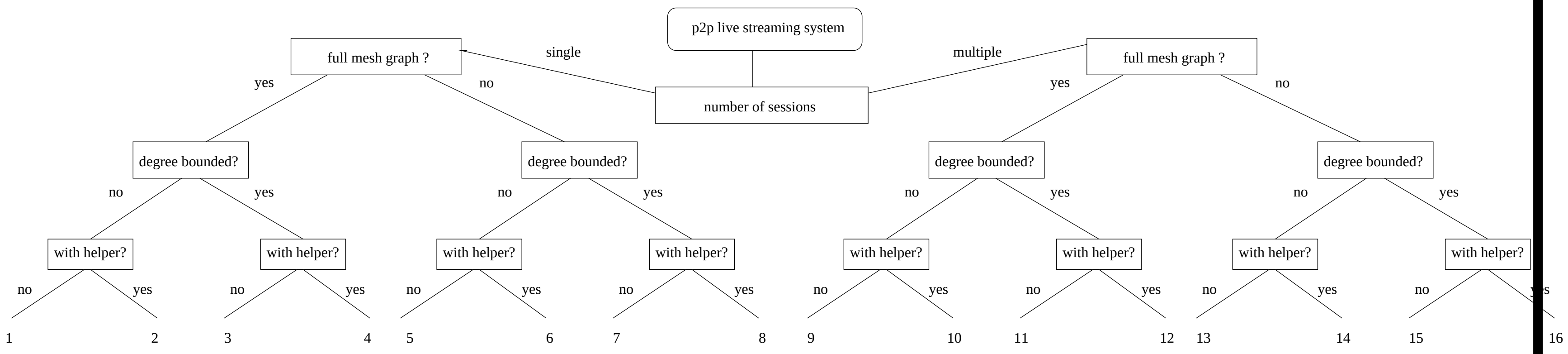
Motivation: Peer-Assisted Streaming (IPTV, video conferencing...)

Use: Benchmarking, Theory-inspired meshing, Impact of degree bound, capacity heterogeneity, presence of helper...

Assumptions

- No peer churns
- No chunk availability issues
- No repetitive downloading
- Uplink bandwidth is the only bottleneck

Taxonomy of 16 Problems



p2p live streaming system

number of sessions

single

multiple

full mesh graph ?

yes

no

degree bounded?

no

yes

with helper?

no

yes

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

Two Approaches

Tree construction for exact answer:

- Li Chou Zhang 2005, Kumar Liu Ross 2007: Cases 1 and 2
- Liu, Zhang-Shen, Jiang, Rexford, Chiang 2008: Case 3
- Liu, Chiang, Sengupta, Li, Chou 2008: Case 3 with hetero capacity and degree bound
- Liu, Chen, Chiang, Li, Chou 2009: Case 3 with total degree bound

PD outer loop with smallest-price-tree inner loop for approx. answer:

- Chen, Ponec, Sengupta, Chou, Li 2008: $1-\epsilon$ for Cases 9 and 10
- Liu, Sengupta, Chiang, Li, Chou 2008: $1-\epsilon$ for Cases 4, 5, 11, 12, 13, $1/\log|R|-\epsilon$ for Cases 6, 14, $1/4-\epsilon$ for Cases 7, 15
- Open: Cases 8, 16

Notation

Given a set of nodes V

K sessions: indexed by k

s^k: source

R^k: set of receivers

H^k: set of helpers

r^k: supported rate

T^k: set of allowed subtrees

$t \in T^k$: a subtree, with rate y_t

$r^k = \sum_{t \in T^k} y_t$

$m_{v,t}$: Outgoing degree of node v in tree t

$M(v)$: Outgoing degree bound of node v in each tree

Notation

x_{vu}: Streaming rate from node v to node u

U_v: Total uplink rate: $U_v = \sum_{u \in V} x_{vu} = \sum_{t \in T} m_{v,t} y_t$

C_v: Uplink capacity: $U_v \leq C_v$

p_v: price of node v

$Q(t, p) = \sum_{v \in V} m_{v,t} p_v$

$\alpha(p) = \min_{t \in T} Q(t, p)$

I_t: Set of internal nodes in tree t

Key Idea

- Turn a combinatorial problem into a continuous optimization
- Too many trees to search through
- Primal-dual iterative outer loop to guide tree search by price
- Combinatorial tree construction for inner loop

Present the case of single session

Single Session: Primal and Dual Problem

maximize	$r = \sum_{t \in T} y_t$
subject to	$\sum_{t \in T} m_{v,t} y_t \leq C_v \,, \quad \forall \; v \in V$
	$y_t \geq 0 \quad \forall \; t \in T$
variables	$(t, m_{v,t}), y_t, r$
constants	T, C_v

minimize	$\sum_{v \in V} C_v p_v$
subject to	$\sum_{v \in V} m_{v,t} p_v \geq 1 \,, \quad \forall \; t \in T,$
	$p_v \geq 0 \quad \forall \; v \in V$

Single Session: Primal-Dual Outer Loop

```
$p_v \leftarrow \delta,\ Y \leftarrow 0,\ flow(v) \leftarrow 0, \quad \forall\ v \in V$

while $\alpha(p) < 1$
    Pick tree $t \in T$ with the smallest $Q(t, p)$
    $y \leftarrow \min_{v \in I_t} C_v / m_{v,t}$
    Assign rate $y$ to tree $t$
    $flow(v) \leftarrow flow(v) + y m_{v,t}, \forall v \in t$
    $Y \leftarrow Y + y$
    $p_v \leftarrow p_v(1 + \epsilon \frac{m_{v,t} y}{C_v})$
    Compute $\alpha(p)$ from the updated $p$
end while
Compute scaling factor $\mathfrak{L} \leftarrow \max_{v \in V} \frac{flow(v)}{C(v)}$ ;
Output capacity $r^* \leftarrow Y / \mathfrak{L}$ ;
```

Optimality and Complexity

- Approximation accuracy: $\epsilon_{tree} - \epsilon$

For appropriately chosen δ

- Time complexity: $O(\frac{N \log N}{\epsilon^2} T_{tree})$

Use Garg and Konemann 1998

Can we find smallest-price-tree algorithms with small T_{tree} and $\epsilon_{tree} = 1$?

Combinatorial Inner Loop: SPT

- Several innovations on combinatorial and graph theoretic constructions

1. Direct construction
2. Snowball Algorithm
3. Translation to:

Shortest arboresence of directed graph, Chu Liu 1965

Min cost undirected group Steiner tree, Charikar et al 1998

Degree constrained survivable network, Lau et al 2007

- Sometimes can get $\epsilon_{tree} = 1$ in polynomial time
- Sometimes only get $\epsilon_{tree} = 1/4$ or $1/\log |R|$

Challenging Extensions

Additional Metrics:

- Delay
- Robustness (traffic fluctuation, peer churn, flash crowd)
- ISP-friendliness (revenue, congestion, server rate)

More Degrees of freedom:

- Delivery schemes (P2P)
- Location-aware peering (P2P)
- Construction of caching nodes (ISP)
- Underlay traffic engineering and congestion control (ISP)

Tradeoff surfaces in 4-dimensional metric space?

More Fun Questions

- Theory-inspired implementation

An ongoing global test

- Where an analogy with packet switching breaks

Is tree richer than path?

Summary

To design content sharing over a network, we need to understand, leverage, or bridge the gaps between content and pipe

- Need to work with sociologist, economists, psychologists, and political scientists
- Need new models with concise language and predictive power
- Need deployment experience and data to close the loops

Contacts

chiangm@princeton.edu
www.princeton.edu/∼chiangm