Network Distribution Capacity and Content-Pipe Gap

Mung Chiang, Princeton University

Abstract— **The growth of video content and diversification of content-sharing methods in the Internet lead to an exciting range of new problems in networking, communications, and signal processing. They range from fundamental theory and analytic models to practical design and industry deployment. This informal note briefly discusses some of the opportunities arising out of the "content-pipe" gap and presents some of the fundamental problems in distributing content over a network.**

I. Content-Pipe Gap

Networks are increasingly used for sharing content, especially video content, both copyrighted and user-generated ones. They shake many basic assumptions on how networks should be designed and used. For example, the asymmetry between uplink and downlink data rates is no longer appropriate, the "horizontal decomposition" into access-metro-core networks may be inefficient if both the content producers and consumers reside in the same access network, and the "vertical decomposition" into application layer and all the layers below it needs to be re-examined.

In this fast evolving landscape of sharing content by networking, there are those who generate and distribute content, e.g., media companies who own video and music, end-users who post video online, operators of content distribution network (CDN), and operators of peer-to-peer (P2P) sharing systems. Then there are those who provide and operate the network, e.g., Internet Service Providers (ISP), network infrastructure vendors, network management software vendors, and municipalities and enterprises running their own networks. On the one hand, the content-providers seek the best way to distribute content, through technologies including multimedia signal processing as well as content caching, relaying, and sharing. They often take the network as just a means of transportation. On the other hand, the pipe-providers seek the best way to meet end-user requirements, through technologies including those that manage resources on each link, between links, and end-to-end. They often take the content as just bits to transport between given nodes in the network. We believe that the interactions between the content and the pipes, whether by design or by default, is an important field to investigate.

Issues involving the content-pipe gap can be about both technology and economics. For example, ISP run traffic management protocols such as TCP congestion control and intra-AS routing, assuming that the traffic matrix is fixed and can be accurately estimated. Yet, 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 as induced by ISP's traffic management. This feedback loop is particularly challenging to ISP because, unlike the voice applications, the way videos are generated, shared, and viewed are quickly evolving through disruptive technologies and user-initiated protocols. Similarly, the 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, marking, and dropping. This gives rise to opportunities of jointly designing how video packets are coded and transported, e.g., assigning multiple streams of video packets coded differently for the same source on multiple paths. Another example is the heated debate on network neutrality: what kind of pricing structures by ISP over different content will be efficient, fair, and incentive-compatible to both content-providers and end-users, while allowing pipe-providers to move themselves out of the business of bit-transportation that is being increasingly commoditized? The mathematical languages of optimization theory, game theory, and control theory may play a role in structuring and clarifying the ongoing discussions on these issues [2], [3], [5], independent of the particular conclusion that a set of assumptions may lead to.

There are four general cases of interactions between content and pipe. First is the case where there is no cooperation, which is roughly the current state in the industry. The lack of cooperation can be due to both worries about network neutrality and lack of unilaterally-actionable, incrementally-deployable, and backward-compatible strategies for cooperation. Second is the case where an ISP receives information about the

content, either on its own, e.g., by deep packet inspection, or through explicit information passing from the content provider, and adapts its protocols in an intelligent way. Third is the case where a content provider and distributor receives information about the network, either on its own, e.g., by actively probing the network, or through explicit information passing from the ISP, and adapts its protocols in an intelligent way. The fourth possibility is a joint optimization by the ISP and content provider as if they were one entity owning all the degrees of freedom, a case that is likely to remain impractical and serves to establish benchmark for comparison. Clearly, cases two and three offer the more promising paths towards practical impact.

In the following we will present a basic model that leads to a set of questions on the fundamental limits of distributing content over a network, on architectures and algorithms that may approach the limits, and on the tradeoff with the cost and complexity in doing so. The general questions can be specialized in many ways, some of which have been answered in recent preprints. Most of the discussion will focus on wireline networks. Of course the introduction of wireless components, either in the air-interface of cellular systems or in all-wireless ad hoc networks, will bring further issues such as the impacts of mobility, shared medium, and time-varying links on content distribution.

Various notions of capacity have been used in well-established fields, e.g., capacities based on transportation of flows in a graph have been used in computer science, capacities based on the largest rate subject to vanishingly small decoding error probability in information theory, and capacities based on the largest set of arrivals subject to queue stability in queueing theory. The nature of Network Distribution Capacity (NDC) is a combination of these three and different from any particular one, since it involves both the combinatorial problems of overlay graph construction (emphasized by the word 'distribution' in NDC) and the communication network problems over various degrees of freedom (emphasized by the word 'network' in NDC).

II. Model for Network Distribution Capacity

The primary goal of the following models and problems is to understand how to distribute content from one set of points to another set in a network. The network physical (but not logical) topology and the content demand-supply locations are given, everything else are up to engineering design, subject to economic, technology, and policy constraints. What are the fundamental limits of these designs for the end-users of content?

Given a directed graph $G = (V, E)$, a set of content demands $D = (D_1, \ldots, D_N\}$, 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. Special cases arise as $|T_i|$ or $|S_i|$ is 1. A node can be a source but not a destination, like a server, a destination but not a source, like a client, or both, like a peer, or neither, like a router. We do not differentiate which source has which piece of the content at this point, nor the time-variation of the nodes. These issues of content chunk availability and peer churn may be added to the model later.

Consider the following as design freedoms:

- For each content i and destination T_{it}, a subset of S_i, denoted by S_{it}, that serve node T_{it}. To complete the notation, for each source S_{is}, there is a set T_{is} of destinations served by it. These subsets (S_{it}, T_{is}) are variables, while the original source and destination sets (S_i, T_i) are constants.
- Transmission rate of each node $x_v \geq 0$, and queue management policy in each router.
- Routing matrix A, which in turn depends on the variable of load balancing matrix H and the constant of physical topology matrix W: $A = HW$, $H_{vp} \in [0, 1]$ is the fraction of traffic from node v on path p, and $W_{pl} \in \{0, 1\}$ is the boolean indicator of whether link $l \in E$ is on path p or not.
- Link capacity $c_l \geq 0$, which in turn depends on some variable vector b representing the cost of building a link with this much capacity (e.g., transmit power in fast-timescale or trenching cost in long-timescale), with a total cost constraint $\sum_l b_l \leq B$ for a given constant B.

We have assumed that scheduling on the links is a constant factor. This can be modified in the model, especially when queueing delay is key component of overall delay objective function. We have also assumed that the cost of storing content in a source is zero, and storage buffer is infinite in each node, which can be modified in extensions of the basic model.

The variables are obviously constrained with each other, e.g., $Ax \leq c(b)$. This reflects a single-AS, wireline network constraint. For wireless networks, more sophisticated constraints would be necessary. For multiple ASes, further policy-based constraints on routes need to be added. In contrast to the standard problem of joint optimization and decomposition across the protocol stack, here even the source and destination sets are

changing, and the construction of content distribution topology over time is a design variable.

Consider two types of performance 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 completion time Q_{it} of content distribution for each of the receivers T_{it}? We can also raise the question of throughput $R_{it}(M)$ averaged over time here.

Both types of metrics above obviously are functions of the given constants (G, D, B), where clearly G and D have much more complicated data structures than constant B. Sum of the metrics over all (i, j) may be easier to study than metrics for individual (i, j).

We have assumed that links are perfectly reliable both for information decoding and service availability. Otherwise, reliability and availability are also metrics. Other utility or cost metrics in standard works of content distribution or network engineering, such as utility function based on user-perceived video quality, may also interact with these capacity metrics.

The above model is neither complete nor tractable. Adding details to the formulations and taking asymptotic limits in appropriate dimensions will be necessary. Even then, only special cases (holding some degrees of freedom as constants) over special (G, D, B) will be readily solvable.

III. General Case

A. Question 1: Network Distribution Capacity

What are the best achieavable R_{ij}, denoted as R^*_{ij}, and Q_{ij}, denoted as Q^*_{ij}?

Throughput and delay are two of the possible variants as shown above. One may also consider, for example, stability capacity as the arrival of D that would keep Q finite over time. A subquestion is only on the converse part: upper bound $\bar{R}$ and $\bar{Q}$.

For special graph G (a less desirable restriction) and special demands D (a more interesting restriction), closed form solutions may be possible. In general, either asymptotic results in terms of the order of growth of R or decay of Q, or an efficient computation of R and Q are more likely than closed-form solutions. It is also possible that computing the answers to the question is NP-hard for sufficiently general G and D.

B. Question 2: Achieving Capacity

What R_{ij} and Q_{ij} are achieved by the current practice of network providers and content distributors? What can be achieved by alternative practice like the other three cases discussed in Section 1?

By "practice", we mean a set of architectures and protocols of controlling an entity's own variables and of interacting with other entities. For example, ISP runs OSPF at the timescale of hours, TCP Reno at the timescale of round trip time, and bandwidth allocation at the timescale of WDM wavelength assignment, P2P runs BitTorrent with tit-for-tat and opportunistic unlocking for peer selection, CDN runs locality-based, user-delay-minimization for content caching, and their interaction is only based on each entity's measurements without any explicit message passing.

Alternative, ISP on the one hand and P2P or CDN on the other hand can run different protocols and also interact by providing certain information to each other at a certain timescale and granularity, e.g., a subset of the following or their limited version: topology, traffic matrix, traffic mixture, link utilization, delay, buffer size, window size, peer selection, exit point, and link weights.

Each detailed system model would provide a possibly different achieveable R and Q. In addition, properties of the trajectories of such interactions are important to characterize, including convergence, convergence time, and invariance. And tradeoff among individual entities' own objective (e.g., maximization of revenue or minimization of link utilization by ISP, and maximization of the number of active users or the total content distribution volume by P2P) are also interesting to understand as content distribution games.

C. Question 3: Capacity-Cost Tradeoff

What is the Pareto-optimal capacity-cost tradeoff, both upper bounds and achievable lower bounds?

Cost here refers not just to B but more generally to the cost of cooperation between the entities, including the cost of measuring, transferring, and leaking information like those in the last subsection. Some of the costs can be measured in time, processing power, or monetary terms, and others in terms of security and privacy.

IV. Special Case: P2P Streaming Capacity

In the case of infinite-backlog, throughput problem in Question 1 can be specialized to a problem where only the content provider's degrees of freedom are optimized, keeping those from the ISP as constants and assuming upload capacity in the access is the only bottleneck. Even just in this branch of formulations, there is at least a taxonomy of 16 formulations depending on the overlay topology allowed, number of peers allowed, presence of relay routers, and the number of streaming sessions.

While this set of formulations does not investigate the content-pipe gap, it provides fundamental limit to the best achievable throughput of peer-assisted content distribution systems, especially P2P streaming [6], [9] where chunk availability is not as much of a concern as in P2P file sharing. It also sheds light on how much can P2P help meet the bandwidth demands on the links of a network, in addition to alleviating the server loads on the nodes. This question has very recently been studied in [1], [4], [8], and in [7] approximation solutions to both computing and achieving this "streaming capacity" are provided for most of the 16 formulations, and exact solutions for a subset of them.

These recent results represent small steps towards the goal of understanding content, its distribution and sharing over a network, and the interactions between content and pipes.

ACKNOWLEDGEMENT

This paper has been inspired by joint work on the subject with Rob Calderbank, Phil Chou, Prashanth Hande, Wenjie Jiang, Jin Li, Ying Li, Zhu Li, Shao Liu, Jennifer Rexford, Sudipta Sengupta, and Rui Zhang-Shen, by discussions with Keith Cambron, Gary Chan, Xing Jin, Raj Savoor, and Steve Sposato, and by industry collaborations with AT&T, Microsoft, and Motorola.

REFERENCES

[1] M. Chen, M. Ponec, S. Sengupta, J. Li, and P. A. Chou, "Utility maximization in peer-to-peer systems", *Proc. ACM Sigmetrics*, June 2008.

[2] P. Hande, M. Chiang, and A. R. Calderbank, "A mathematical model for network neutrality", *Preprint*, 2008.

[3] W. Jiang, R. Zhang-Shen, M. Chiang, and J. Rexford, "Interplay of content distribution and traffic engineering", *Preprint*, 2008.

[4] X. Jin, W. P. Yiu, S. H. Chan, and Y. Wang, "On maximizing tree bandwidth for topology-aware peer-to-peer streaming", *IEEE Trans. Multimedia*, vol. 9, no. 8, pp. 1580-1592, December 2007.

[5] G. Kesidis and G. de Veciana, "On flat-rate and usage-based pricing for tiered commodity Internet services", *Proc. CISS*, March 2008.

[6] R. Kumar, Y. Liu, and K. Ross, "Stochastic fluid theory for P2P streaming systems", *Proc. IEEE Infocom*, May 2007.

[7] S. Liu, S. Sengupta, M. Chiang, J. Li, and P. A. Chou, "Achieving streaming capacity in P2P", *Preprint*, 2008.

[8] S. Liu, R. Zhang, J. Jiang, J. Rexford, and M. Chiang, "Performance bound in peer-assisted live streaming", *Proc. ACM Sigmetrics*, June 2008.

[9] K. Suh, C. Diot, J. Kurose, L. Massoulie, C. Neumann, D. Towsley, and M. Valleo, "Push-to-peer video-on-demand system: Design and evaluation", *IEEE J. Sel. Area Comm.*, 2007.