

Cristian Lumezanu

RESEARCH STATEMENT

My research interests lie in the areas of networking and distributed systems, with emphasis on designing, building and analyzing systems that enhance the functionality of the Internet as a communication medium for its users.

In my dissertation, I showed how to use the properties of the Internet delay space to achieve fair, scalable and fast end-to-end communication. I designed and built PeerWise, a latency-reducing routing overlay network, and demonstrated its ability to find and use faster paths under realistic network conditions. I have also worked on improving communication through efficient resource allocation for distributed applications that send data on fixed paths. In the future, I plan to continue exploring ways to enhance both our understanding and the functionality of the Internet: a social map of the Internet and a route avoidance mechanism through which users can divert their traffic away from malicious or contentious nodes.

My approach to research is from the perspective of both a scientist and an engineer. This duality enables me to develop real systems that are based both on solid mathematical reasoning and on measurement. As a scientist, I conceptualize simple ideas, formulate hypotheses and validate them. Because my science is motivated by practical problems, I have found that, to validate my hypotheses, it is often necessary to engage in realistic and rigorous experimentation.

PeerWise Overlay Networks

The Internet was designed as a best-effort communication medium for end users, limited to a most basic role: providing connectivity. Paths are chosen by Internet Service Providers (ISPs) based on their own cost, policies or past performance. This often results in suboptimal routes for end users: packets may take longer or more congested paths than necessary, they may be delayed by slow reaction to failures, they may even fail to find a usable path between users. With thousands of users joining the network every day and new applications being constantly deployed, the diversity of interests and requirements becomes increasingly important. Finding *any* path is not sufficient anymore, users and applications need *optimal* paths.

I believe that a major goal for the future Internet is the transition to a more flexible communication medium, where end users have more leeway in negotiating and choosing paths based on their own performance. My dissertation, in which I explore how to improve communication using routing overlay networks, is a step towards this goal. I have designed and built PeerWise, a latency-reducing routing overlay that employs the properties of the Internet delay space to make end-to-end communication fast, fair and scalable.

PeerWise offers several properties desired in a practical communication infrastructure. It preserves the *connectivity* and *scalability* of the underlying routing layer. In addition, it offers increased *end-to-end performance* to users by finding low latency paths. Finally, it provides *fairness* through an incentive mechanism that encourages participation of honest users and discourages freeloaders and adversaries. Next I present how real world observations, measurement and experimentation lead me to the design and implementation of PeerWise. Publications describing my work appear in the proceedings of HotNets 2007, PAM 2009 and NSDI 2009.

First, I observed that triangle inequality violations (TIVs) in the Internet can be used to predict detours. A detour between two nodes is a shorter-than-default alternate one-hop path. Three nodes form a TIV when the round-trip time (RTT) between two of them (the long side of the TIV) is greater than the sum of RTTs to the third node (the short sides). Pairs of nodes that are long sides in TIVs may benefit from detours; pairs that are short sides may be part of detours. To pursue this further, I found it necessary to better understand TIVs. I examined TIVs by collecting and analyzing latency measurements between thousands of Internet hosts over the course of several weeks. My measurement study showed that TIVs are prevalent, lasting features of the Internet and that they offer significant latency reduction.

Measuring all possible latencies to discover TIVs would limit the scalability of a latency-reducing overlay. Instead, it would be better to detect TIVs and use them to predict good detours. I use network coordinates to find TIVs scalably. A network coordinate system associates nodes with points in a metric space such that the distance between points estimates the real latency between nodes. Since TIVs are not allowed in metric spaces by definition, estimated distances between nodes in a TIV may have high errors. I observed, and subsequently confirmed through experiments, that when the estimated distance between two nodes is much smaller than the real distance, the nodes have a higher chance of benefiting from a detour; conversely, when the estimated distance is much larger than the real distance, the nodes are more likely to be part of a detour.

The final missing piece from PeerWise was fairness. I wanted to provide an incentive mechanism that encourages honest users to participate and discourages freeloaders. I introduced mutual advantage as a fundamental design principle: overlay edges exist only between hosts that can help each other find detours. Mutual advantage induces better cooperation among nodes and helps avoid the tragedy of the commons, when only a few, well connected nodes provide transit. Of course, it would all be for naught if mutual advantage would severely limit the number of detours that a node can find. Using real-world latencies, I showed that this is not the case: the mutual advantage requirement reduces the number of destinations reachable to approximately half, yet even popular web servers using content distribution are reachable.

Translating the design of PeerWise into a real, working system required a number of engineering mechanisms, devised after analyzing many measurement results and taking several trips back and forth to the drawing board. One challenge was how to use PeerWise to scalably discover detours to non-participating nodes, such as web servers. Because these destinations do not maintain network coordinates, my second key insight could not be easily applied. With colleagues, I implemented a virtual coordinate system through which a PeerWise node can become responsible and compute coordinates for any host in the Internet. We also proposed and evaluated several policies that nodes can use to predict quickly the best detours to any specific destination. I deployed PeerWise on the PlanetLab testbed and showed that nodes quickly find detours to popular destinations, that these detours are stable and that they offer significant latency reductions. I then confirmed that user-level applications such as web transfers can benefit from the network-level detours of PeerWise.

Distributed Optimization for Information Dissemination Applications

Recent years have witnessed the emergence of distributed information dissemination and processing applications such as stock tickers, program trading, medical alerting, environmental monitoring, and airline ticket pricing. These applications are generally deployed on fixed overlay topologies, sometimes with several applications sharing the same infrastructure. They disseminate data using predetermined paths; thus, improving end-to-end communication by choosing the nodes on the

data path is not an option. Instead, I focused on how to optimize the allocation of resources on the path (network bandwidth for the links and CPU share for the nodes). While visiting IBM Research as an intern, I formulated the resource allocation as a constrained optimization problem. I used the concept of utility functions to measure the benefit of an application and developed two distributed optimization algorithms, LRGP and LLA, that optimize the overall utility while keeping the resources of the underlying infrastructure uncongested. This work appeared in ICDCS 2006 and ICDCS 2008.

The first algorithm, LRGP, is intended for applications that follow the publish/subscribe paradigm, such as stock tickers or sports scoreboards. In such applications, producers (publishers) periodically inject messages into the system, which then are delivered to consumers (subscribers) that register interest using subscriptions. Message flows may be transformed along the path according to filters specified by consumers. The benefit of these applications increases with the consumer population and the data delivery rates to them. However, because the utility depends on both the number of consumers and on the flow rates, the objective function in the optimization problem is not concave; therefore, there is no global maximum. My key idea was to partition the optimization into two parts: admission control and rate control. The admission control uses a greedy approach to admit consumers based on the benefit that each brings. The rate control uses Lagrangian decomposition to compute the flow rates that keep all resources uncongested. The optimization algorithm runs continuously, iterating between the two parts and enacting new allocations periodically or when significant changes occur. LRGP is completely distributed and uses the concept of price to indicate how congested each resource is at every step. I tested LRGP on several workloads and showed that it converges quickly, scales with the number of flows and consumers, and achieves close to optimal utility even though it is distributed.

The success of LRGP was impetus to investigate how to use a similar approach for other types of applications. A significant number of information dissemination applications are required to respond rapidly to real-world events, or to continuously analyze data in real-time to build models for prediction or prevention. These applications place diverse real-time requirements on the underlying infrastructure: different deadlines for data delivery and different levels of importance for each deadline. With colleagues from IBM Research, I developed a framework for modeling applications with flexible real-time requirements and diverse levels of importance. In the framework, we divided each application into tasks and subtasks, assigned to different resources for execution. We expressed the utility of an application as a non-increasing function of the application latency. We designed LLA, a distributed algorithm based on Lagrangian multiplier theory, that assigns resource shares to each subtask such that the utility is maximized while the deadline of each task is met. LLA runs continuously using feedback about the level of congestion of each resource to compute new shares and establish the latency of the application on each resource. LLA demonstrated fast convergence and scalability both in simulation and under realistic network conditions.

Future Work

My future research agenda is dominated by two goals. I want to continue understanding the complex mechanisms that make the Internet function by performing new measurement studies. I also want to build systems that improve the functionality of the Internet and are based on principles derived from experimentation. I describe next two specific objectives: a route avoidance mechanism and a social map of the Internet.

Route Avoidance. In recent years, ISPs, enterprises or even governments have been increasingly interfering with end-to-end user traffic that they help carry. For example, ISPs such as Comcast and Bell Canada throttle peer-to-peer traffic and countries like China and Australia enforce strict Internet censorship laws. These contentious actions are not likely to stop here. AT&T announced its intention to filter content in the future. Such behavior emphasizes the necessity of a mechanism that allows users to send traffic on paths that avoid certain parts of the Internet.

I plan to study the effects of introducing route avoidance as a user policy in the Internet and build a routing overlay system that implements it. Users will be able to avoid regions of the Internet by smartly forwarding their traffic through other nodes in the overlay. The success of such a system hinges on the answer to several important questions. First, *is route avoidance feasible?* Would enough nodes be able to benefit from it without significant penalties in performance (*e.g.*, longer or more congested paths)? It is known that ISPs do not choose default paths with end user performance in mind. By avoiding these routes, users would not necessarily suffer. It remains to be seen how one should choose overlay paths such that the loss in performance is minimum and how would route avoidance compare with simply tunneling and encrypting traffic. Second, *is route avoidance verifiable?* Would users be able to verify that their traffic did indeed avoid the boycotted region? My idea is to choose the overlay path in such a way that I can unambiguously use network-level latency probes to identify misbehaving nodes. Finally, *what effects would the adoption of a route avoidance policy have on the Internet?* If communication patterns change drastically, so would traffic engineering and business agreements. Would this direction democratize end-to-end communication and foster innovation or would it lead to more chaos and selfish attitudes?

Social Map of the Internet. Understanding the structure of the Internet is paramount. With precise network maps, the effectiveness of many applications and protocols would be easier to evaluate and the consequences of failures would be easier to predict. Most efforts in mapping the Internet focus exclusively on physical connectivity: identifying all nodes (routers) and connections between them. However, physical links are sometimes insufficient for predicting performance; how nodes interact with each other is a better predictor than whether they share a wire or not. A different vision is to study the Internet as a *social* medium of communication. I will attempt a new structural decomposition of the network and identify relationships between nodes that transcend physical connectivity. I have considered two types of relationships so far, symbiosis and dependency, but I believe that many more exist, each leading to a different map. Symbiosis captures the mutual ability of two nodes to benefit from each other's position or resources, while dependency occurs when one node's performance relies exclusively on the other's. An example of symbiotic relationship appears in my dissertation work: pairs of nodes are able to offer each other lower latency to other regions of the network. Such a relationship cannot be captured with a simple topology map.

A social map of the network is exciting because it would change the way researchers view the Internet. It would create more complete models that, coupled with topology maps, would better predict the performance of new protocols. Social network maps could help overlay network construction, routing protocol design or measurement techniques by revealing new patterns of dependency or redundancy between nodes. They would also benefit the day-to-day operation and running of the network (*e.g.*, by identifying which links or nodes to overprovision, where to establish peering agreements). Studying the social structure of the Internet at the network level requires leveraging existing measurement techniques and studies, while devising new ones that capture social interplays between nodes. My experience in measuring the Internet and in designing and building applications that use the measurements puts me in an excellent position to achieve this goal.