

Research Statement

Himabindu Pucha

Personal computers around the world are expected to touch the 2 billion mark in 2008; close to 1.5 billion people have some form of network access today. Current trends indicate that more than one IP-enabled device per user will be typical and that Internet access will soon hit the next billion people, while next-generation applications are waiting in the wings to ride this wave. My major research theme is building systems that improve the performance of data transfers or provide them with underlying support (e.g., routing) within and across these networked, distributed systems that span a range of scale and mobility.

My approach to enabling efficient data communication in these networked systems is by isolating their bottlenecks and proposing innovative techniques to improve upon the state of the art. In general, I focus on building robust and scalable solutions that can intelligently adapt to and benefit from underlying scenarios, while not straying from practicality. A common trait of my research is to marry algorithmic innovations with the construction of real systems. For example, handprinting [1], inspired by deterministic sampling in web search, significantly improved the scalability of locating peers for data download. Similarly, DPSR [2] improved the routing performance in small scale wireless networks by constructing a topology-aware overlay. Such overlays were first proposed in the context of millions of wired nodes.

My research style is a blend of comprehensive measurement studies and analysis coupled with hands-on system building to develop and refine solutions. I believe in demonstrating the performance and usability of a solution by building and deploying it. Such an approach gives a thorough understanding of the different aspects of a solution; it demonstrates the components that work well and highlights the design decisions that need to be revisited to accommodate real-world characteristics. There is no substitute to building a system—issues such as impact of CPU load, disk writes, fading in a wireless channel, and so on are uncovered only during live experimentation. At the same time, measurement studies are valuable in characterizing a problem, evaluating system design choices at scale, and understanding observed system performance so that the findings can be incorporated into the design process, resulting in a robust system design.

1 Dissertation Research: Improving Internet Data Transfer

Despite the increasing importance of bulk data transfers, such transfers frequently remain slow and inefficient. Many modern transfer systems achieve download throughputs of only tens of Kbit/s despite running on fast, asymmetric broadband connections. As bulk data transfers take center-stage to deliver content ranging from news and entertainment to critical services such as telemedicine and telepresence, there is a pressing need to improve their performance. This demand for efficient bulk data transfers motivated my dissertation—I investigated the performance limitations of current bulk data transfers and systematically improved their performance.

Enhance Performance via Innovative Use of Heterogeneous Resources. My dissertation addressed the challenges facing bulk transfers via three complimentary approaches: improving the performance of a single TCP transfer, improving multi-source transfers by enabling downloads from a larger number of peers, and improving file distribution by making aggressive use of files already resident at the receivers.

The foremost challenge is that TCP, the dominant transport protocol for the last twenty years, is inefficient in utilizing available bandwidth between a single sender-receiver pair in the presence of losses. Prior solutions that address this issue face significant deployment hurdles, hindering their wide-spread adoption. I built Slot [3], an incrementally deployable system, that addresses this challenge. Slot uses in-network resources from an overlay network to effectively shorten the end-to-end TCP feedback loop; it breaks the end-to-end loop into multiple pipelined sub-loops, each of which is more efficient than the end-to-end loop, via intermediary nodes carefully chosen from an overlay network. Slot's design includes scalable, efficient and practical path selection mechanisms to improve client throughput and load-balancing and admission control mechanisms to support multiple simultaneous clients using a common infrastructure.

Given that a single sender is typically unable to saturate the receiver's bandwidth, a popular approach to improve transfer performance is to download from multiple sources simultaneously (e.g., BitTorrent). The challenge, however, is that the current multi-source systems have a fundamental limitation—these systems may still not have enough

sources to saturate the receiver's bandwidth. I addressed this limitation by developing a novel technique to discover more sources for a receiver to draw from. The key idea of my approach is to exploit similarity among data objects to obtain additional sources [1]. These similar sources work in conjunction with sources that serve the desired data object (identical sources) to improve transfer performance. The proposed system, SET, efficiently locates and exploits such similar and identical sources using handprinting, a technique based on constant number of lookups and mappings in a global lookup table.

Current approaches have also attempted to leverage local resources to improve transfer performance (e.g., *rsync*). These approaches suffer from focusing on one particular resource strategy to the exclusion of others and are not adaptive across a wide-range of operating scenarios. To address this challenge, I proposed *dsync* [4], a system that harnesses both local and network resources intelligently. In particular, *dsync* includes a framework for locating relevant data on the local disk that might accelerate the transfer. *dsync* also adapts to diverse scenarios by determining at each step of the transfer which of the receiver's local resources can be used without introducing undue contention by monitoring queue back-pressure. For example, *dsync* uses queue information from its disk writer and network reader processes to infer disk availability.

Demonstrate Effectiveness through Deployment. I have extensively evaluated each of the above systems by deploying a working prototype on the Emulab emulation platform and on the PlanetLab and RON testbeds. Our experience shows that Slot improved the throughput of a single sender-receiver connection by 60-100%; SET improved the performance of multi-source transfers by 30-70% in real workloads; and *dsync* combined the benefits from using identical and similar sources over the network with local resources from a disk to outperform existing systems by a factor of 1.4 to 5 in one-to-one and one-to-many transfers. Building these systems was invaluable as we learned from the unexpected lessons. For instance, our experimentation with *dsync* showed that a robust *dsync* design must deal with practical system issues such as ensuring large sequential read/write operations, that can be hidden behind the content-based naming abstraction. All the systems in my work are available for users to experiment with [5].

While experimenting with a deployed prototype is important in systems research, it is a fundamental challenge to ensure that the reported performance is not an artifact of the deployment. My work [6] revealed the degree to which research network based testbeds (e.g., PlanetLab) can affect wide-area experiments; a deployment of overlay multicast, for example, exhibited a performance reduction of 40% when the underlying network connectivity was changed from research networks to commercial networks. My work also proposed guidelines for the research community on how to use testbeds like PlanetLab for experiments that better represent commercial Internet.

Guide Design with Comprehensive Measurement Studies. I have also successfully used comprehensive measurement based studies at various stages of my work, from motivating a solution to refining system design. Our 1.7 TBytes of multimedia files, for example, collected from the eDonkey and Overnet file-sharing networks for three months motivated the key idea behind SET. Analysis of this data showed that multimedia files are similar to other files and this similarity results in an increase in the total number of available data sources.

Another instance is the measurement infrastructure at the core of Slot's design, which continuously monitors network properties (e.g., RTT, loss) to perform overlay path selection. The design of this infrastructure required an understanding of the dynamics of network properties and their time scales—when and why do network properties change. The answers to these questions are also important in the design of other measurement-based systems such as VoIP and network positioning systems. Comprehensive measurements enabled studying these issues at scale using a large number of vantage points. Our work [7] used network delay as an example to understand its dynamics. It is well known that congestion on a routing path causes variation in the end-to-end delay between two hosts. My work is the first, however, to systematically analyze changes in network delays and jitter caused by routing path changes. Our key contribution was to study the predictability of the delay variations caused by these routing events by measuring properties such as the durations for which these paths last, the frequency at which path changes happen, the delay variations that can be caused due to these changes, and so on. My recent work [8] also investigated the predictability of one-way delay, important for applications like streaming and tree-based multicast. Both these studies highlight the symbiosis between comprehensive measurements and system design—they enable building systems with intelligent and efficient mechanisms for path selection and management by reacting to routing events rather than simply periodically probing the network.

2 Other Research

Wireless last mile access networks. Community wireless networks (mesh networks) are an appealing way to provide cheap, deployable, last mile Internet access to millions of homes, specially in rural and economically disadvantaged

zones. I have helped deploy MAP (Mesh@Purdue) [9], a 32 node multi-radio wireless mesh testbed, to understand the performance limitations of these mesh networks via a comprehensive measurement study [10]. Obtaining satisfactory throughput across several clients is a significant barrier in these wireless networks.

I proposed two systems, DMesh and MeshCache, to improve the throughput performance in mesh networks. DMesh [11] exploits cheap and simple directional antennas placed on multiple radios via intelligent directional channel assignment and routing algorithms. This provides spatial and frequency separation between competing transmitters, thereby reducing interference and improving throughput. MeshCache [12] is an application layer TCP relay and caching system for mesh networks. MeshCache enables data caching at each wireless mesh router by breaking long haul TCP connections (from clients to gateways) at every router. MeshCache then exploits locality in user traffic to fetch the data from peer routers instead of the Internet gateway. This reduces the hop count of the data transfers and mitigates congestion at the gateway, thereby improving throughput performance.

DMesh and MeshCache leverage intelligent algorithms and efficient protocol design to work well in a resource-constrained wireless environment. Our evaluation also exemplifies the advantages of hands-on system building. For example, we learned that a conservative channel assignment algorithm is the most suitable choice in the presence of interference from side and back lobes as observed in deployed directional antennas.

Small-scale Mobile Systems. I built Carawan, a testbed with 5 laptops and 15 PDAs, to experiment with multi-hop wireless networks formed in an ad-hoc fashion. Such networks are useful in settings such as stadiums and conferences. Although such networks are typically small in scale (about 100 nodes), the possible mobility of nodes and their churn creates performance hurdles—these networks suffer from limited scalability of routing protocols and lack of service discovery mechanisms.

I developed Dynamic P2P Source Routing (DPSR) [2], a new routing protocol for mobile wireless ad hoc networks (MANETs) that exploits the synergy between topology-aware structured P2P overlays and MANETs for enhanced scalability. In a network of size N , when each of the N nodes communicates with the remaining $N-1$ nodes directly, DPSR introduces a virtual address space for routing through overlay peers, effectively reducing the number of links (and hence routes) required for all pairwise communication from $O(N^2)$ to $O(N \cdot \log N)$. Like its Internet counterpart, an efficient DHT substrate in MANETs can greatly ease the construction of distributed applications and services. My work [13] implemented the first DHT substrate in wireless networks, Ekta, by adopting Internet DHTs and then adapting them to leverage wireless network characteristics (e.g., wireless broadcast advantage). I then built a service discovery mechanism on top of Ekta which was more efficient than an approach directly built on top of physical layer broadcast.

Both DPSR and Ekta highlight the innovative and successful use of algorithmic solutions, i.e., topology-aware overlay routing proposed in the context of millions of wired nodes, in a small-scale mobile wireless scenario. This work was among the very first to propose ideas in this interdisciplinary topic and has been cited extensively [5].

Large-scale Mobile Systems. Another area of my research is large-scale mobile systems which have important applications in military command and control, emergency services, and sensing networks. Scalable routing, however, in such large networks in the presence of mobility is a challenging problem due to the frequent topology changes and the need for global route discovery. I proposed practical and scalable unicast [14] and multicast routing protocols [15] for these networks. My work demonstrated the importance of accounting for practical challenges such as mobility when designing solutions; our simple geographic-hashing-based protocols were more robust and scaled better (upto $8\times$ lower overhead) than complex asymptotically scalable solutions in practical scenarios.

3 Future Directions

I plan to continue my research in the area of networking and distributed systems. In the short term, I intend to build on my recent work in agile transfer systems. My long term interests lie in the broader scope of exploring the challenges and the opportunities enabled by the next generation of ubiquitous computing, specifically at the intersection of wired and wireless networks. Please send email to hpucha@cs.cmu.edu for further details.

References

- [1] H. Pucha, D. G. Andersen, and M. Kaminsky, "Exploiting Similarity for Multi-Source Downloads using File Handprints," in *Proceedings of USENIX Symposium on Networked Systems Design and Implementation (NSDI)*, 2007.
- [2] H. Pucha, S. M. Das, and Y. C. Hu, "Imposing Route Reuse in Mobile Ad Hoc Network Routing Protocols using Structured Peer-to-Peer Overlay Routing," *IEEE Transactions on Parallel and Distributed Systems (TPDS)*, vol. 17, pp. 1452–1467, December 2006. Also appeared in HotOS 2003.
- [3] H. Pucha, Z. Zhang, and Y. C. Hu, "Buffering in the middle: Overcoming Internet Data Transport Limitations," in *Submission, USENIX Annual Technical Conference (ATC)*, 2008. Poster in SIGCOMM 2005.
- [4] H. Pucha, M. Kaminsky, D. G. Andersen, and M. A. Kozuch, "Adaptive File Transfers for Diverse Environments," in *Submission, USENIX Annual Technical Conference (ATC)*, 2008. WIP in SOSP 2007.

- [5] Project Descriptions. <http://web.ics.purdue.edu/~hpucha/projects.htm>.
- [6] H. Pucha, Y. C. Hu, and Z. M. Mao, "On the Representativeness of Wide Area Internet Testbed Experiments," in *Proc. of ACM SIGCOMM/USENIX Internet Measurement Conference (IMC)*, 2006.
- [7] H. Pucha, Y. Zhang, Z. M. Mao, and Y. C. Hu, "Understanding Network Delay Changes Caused by Routing Events," in *Proc. of ACM International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS)*, 2007.
- [8] A. Pathak, H. Pucha, Y. Zhang, Z. M. Mao, and Y. C. Hu, "A Measurement Study of Internet Delay Asymmetry," in *Proc. of Passive and Active Measurement Conference (PAM)*, 2008.
- [9] MAP. <http://www.engineering.purdue.edu/MESH>.
- [10] S. M. Das, H. Pucha, K. Papagiannaki, and Y. C. Hu, "Studying Wireless Routing Link Dynamics," in *Proc. of ACM SIGCOMM/USENIX Internet Measurement Conference (IMC)*, 2007.
- [11] S. M. Das, H. Pucha, D. Koutsonikolas, Y. C. Hu, and D. Peroulis, "DMesh: Incorporating Practical Directional Antennas in Multichannel Wireless Mesh Networks," *IEEE Journal on Selected Areas in Communications (JSAC)*, vol. 24, pp. 2028–2039, November 2006.
- [12] S. M. Das, H. Pucha, and Y. C. Hu, "Mitigating the Gateway Bottleneck via Transparent Cooperative Caching in Wireless Mesh Networks," *Ad Hoc Networks*, vol. 5, pp. 680–703, August 2007.
- [13] H. Pucha, S. M. Das, and Y. C. Hu, "Ekta: An Efficient DHT Substrate for Distributed Applications in Mobile Ad Hoc Networks," in *Proc. of IEEE Workshop on Mobile Computing Systems and Applications (WMCSA/HOTMOBILE)*, 2004.
- [14] S. M. Das, H. Pucha, and Y. C. Hu, "On the Scalability of Location Services for Geographic Ad Hoc Routing," *Computer Networks (COMNET)*, vol. 51, pp. 3693–3714, September 2007. Also appeared in INFOCOM 2005.
- [15] S. M. Das, H. Pucha, and Y. C. Hu, "Distributed Hashing for Scalable Multicast in Wireless Ad Hoc Networks," *IEEE Transactions on Parallel and Distributed Systems (TPDS)*, vol. 19, March 2008.