Aditya Akella **Research Statement, February 2009**

My broad research goal is to improve the *efficiency* and *robustness* of emerging networks and networked systems. My key research works within this general space fall into three categories:

1. A network-wide redundancy elimination service for improving effective network capacity and network efficiency [SIGCOMM08, SIGMETRICS09]

2. New approaches for mitigating the complexity of network management, thereby making networks more robust and less error-prone [NSDI09, Infocom07, Mobicom08-Poster]

3. Measurement-driven approaches to ensuring robust operation of networked systems under unexpected conditions such as large-scale attacks [USENIX08, INM07]

I try to ensure that my work is backed by sound theory and modeling. I rely mostly on simple tools from optimization, game theory, and graph theory. I also strive to convert ideas into real deployable systems and services. All my work involves several months of implementation followed by thorough testing in live networks and interaction with network operators.

Redundancy Elimination as a Network Service:

Different users on the Internet often access the same content, resulting in the same data being transferred repeatedly. Several redundancy elimination (RE) systems have explored how to suppress this redundant content and improve network efficiency. Examples include Web caches, content distribution networks, multicast routing protocols, P2P caches, peer-assisted transport systems (e.g. Bittorrent), and application-independent "WAN Optimizers" that can remove redundant strings from any traffic flow. A common constraint imposed by these systems is that they apply in a *localized* fashion to a specific system, protocol and/or network link. End-points wishing to derive the benefits of redundancy elimination must implement a specific mechanism that could in turn depend on the application-level protocols in use (HTTP vs Streaming vs P2P), the communication model (unicast vs multicast), the types of objects being requested (static, small objects vs dynamic content) and the network location of the end-points. Furthermore, the mechanism employed only benefits the end-points/applications using it but not others with similar needs.

My research explores a new network architecture where RE is a network-supported service that is accessible to all applications, protocols and flows. This service offers three architectural benefits: (1) the service frees communication end-points from having to re-implement point solutions for RE, and extends the benefits of redundancy elimination universally. (2) The inherent support simplifies the design of key applications and protocols. For instance, application layer multicast is vastly simplified in such a network because there is no longer a need to be careful about duplicate transmissions. (3) The service enables *new* applications and protocols. For instance, rather than use shortest path routing, networks (e.g. ISPs) could construct "minimum entropy routes" to place potentially redundant traffic on a common set of links to maximize the effectiveness of RE and improve overall network efficiency.

My research contributions on this topic include a proposal for an IP-layer redundancy elimination service and an exploration into its architectural benefits, a study of the properties of traffic redundancy to inform the design of the service, and algorithms for enabling RE in a variety of networks today:

(1) Supporting RE as an *IP-layer packet-level service*, and *new intra- and inter-domain routing protocols* that leverage this service [*SIGCOMM 09*]: My group devised a strawman framework for an IP-layer RE service. In this framework, a collection of network routers employ "packet caches" holding all packets that were forwarded in a certain time interval. Upstream routers use the caches to remove duplicate strings from packets, while routers immediately downstream reconstruct original packet payloads from the local cache.

We developed a new suite of routing algorithms that leverage the RE service. In these "redundancy-aware" approaches, networks compute estimates of the expected amounts of redundancy in network traffic between different network locations, and use them to construct routes that maximize the potential for RE (redundancy elimination is still applied in a hop-by-hop fashion along the constructed routes). Using synthetic and real packet traces, we find that IP-layer RE can reduce the overall link utilization in ISP networks by 10-50%, and redundancy-aware routing could bring it down by a further 10-25%. Thus, IP-layer RE helps ISP improve their effective switching capacity and exercise greater control over link loads.

(2) A large-scale *measurement study of redundancy* in network traffic [*SIGMETRICS 09*]: In order to understand the range of benefits from inherent support for RE and the best way to design packet caches, we conducted a large scale measurement study of redundancy in network traffic based on data collected at 12 network vantage points. We find that packet-level RE can deliver average bandwidth savings of 15-60% across enterprise and university access links. We also find that small packet caches (~200MB) can provide a bulk of the gain. We find that a significant fraction of savings in enterprise networks arises from redundancy within traffic between the same pair of end-points – this argues for end-host support for RE in addition to in-network support.

(3) Supporting RE in *resource-constrained environments* [under submission]: RE involves several resource-intensive tasks such as indexing packet content, looking up content matches and compressing data, and reconstructing original data. We have developed a new caching framework to support these operations effectively while operating under tight resource constraints at network devices; Example of constraints include limited amount of memory available to store content, bounds on number of memory lookups possible per second, power consumption etc. Our insight is that the different RE actions (e.g. compression and reconstruction) can be decoupled and allocated in an intelligent fashion across a collection of devices in accordance with their local resource constraints. We use an *implicit coordination scheme* based on consistent hashing to achieve decentralized control over the RE actions of the devices. This framework applies to many RE scenarios such as supporting IP-layer RE in core routers, employing RE inside data centers to alleviate congestion, and improving cooperative caching schemes in multi-hop wireless networks. Our framework allows the RE service to be introduced in an incremental, backward-compatible fashion. We have built Click software prototypes of the coordinated caching framework. Our software prototypes can perform RE actions at OC-192 rates.

New Approaches to Network Management:

Experience has shown that the high complexity underlying the design and configuration of today's networks generally leads to significant manual intervention when managing them. While hard data implicating complexity in network outages is difficult to come by, both anecdotal evidence and operator interviews suggest that more complex networks are more prone to failures, and are difficult to upgrade and manage. Complexity can also have a "ripple effect" where errors made by human operators in one network seriously impact neighboring networks.

My vision is *to make today's networks simple, robust (especially to human errors), easy to manage and easy to modify*. Ultimately, my goal is to take the "human out of the loop" and enable fully automated management and upgrades of networks today. Two of my key research contributions in this direction are as follows:

(1) Modeling the *complexity* of network design and management [*NSDI09*]: As a first step toward realizing the above vision, my group has developed a family of *complexity models and metrics* that describe the complexity of the design of an *enterprise network* in a succinct fashion, abstracting away all the details of the underlying configuration language. Given the frequency with which configuration errors are responsible for major outages, we argue that creating techniques to quantify systematically the complexity of a network's configuration is an important first step to *reducing* that complexity. Developing such metrics is difficult as they must be automatically computable yet still enable a direct comparison between networks that may be very different in terms of their size and design. In databases and software engineering, metrics

and benchmarks have driven the direction of the field by defining what is desirable. In proposing these metrics, we hope to start a similar conversation for network design.

We have designed three measures of complexity of an enterprise network's routing design: (i) the complexity behind configuring network routers accurately, (ii) the complexity arising from identifying and defining distinct roles for routers in implementing a network's policy, and (iii) the inherent complexity of the policies themselves. Our models and metrics are designed to have the following characteristics: (i) They align with the complexity of the mental model operators use when reasoning about their network – networks with higher complexity scores are harder for operators to manage, evolve or reason about correctly. (ii) They can be derived automatically from the configuration files that define a network's design. This means that automatic configuration tools can use the metrics to choose between alternative designs when, as frequently is the case, there are several ways of implementing any given policy.

A key part of this work was the empirical validation of the metrics and a study of the complexity of existing live networks. *We applied the metrics to 7 live networks, including 5 universities and 2 enterprises. For 5 of the 7 networks, we systematically interviewed the operators to identify the relevance of our metrics.* We devised tests where the operators were asked to perform a standardized set of configuration-related tasks. The operators' actions were timed by measuring the number of steps taken for each task. In all cases, we found the metrics to be *predictive* of the difficulty operators face in conducting management tasks. Our models also discovered bugs in network designs, where changes that the operators had made to network configurations did not have the intended global effect. This exercise served both to validate the effectiveness of our approach while also providing a direct avenue for impacting today's operational practices.

A final part of this work was identifying the *causes* of complexity. The operator interviews highlighted several important causes, including network evolution over time, special cases for network policies, and, most interestingly, cost constraints. It turns out that in some cases network providers choose a seemingly complex network design because it is more cost-effective than simpler alternatives. We believe that these insights can inform the design of future clean-slate architectures such as SANE, Ethane and ConMAN.

(2) Building blocks for *coordinated management* of networks [*Infocom07, Mobicom08-Poster*]: In today's interconnected world, actions taken by different networks are tightly coupled and inter-dependent. For example, the routing decisions taken by one network can influence the ability of its neighbors to conduct traffic engineering effectively. Recent studies have shown that unilateral myopic decisions of networks could collectively lead to reduced network reliability and performance.

My goal is to develop building blocks that allow networks with conflicting goals to reconcile each other's decisions and cooperate systematically when conducting their management tasks. A key requirement in such cooperation-based management approaches is *fairness*: all participating entities should observe equitable improvements from cooperation; otherwise some entities will refuse to cooperate to the start with. A second requirement is to *accommodate heterogeneity*: different entities may have diverse local goals that must all be honored within the cooperation-based framework. I have identified two generic building blocks in designing coordinated management protocols with the above salient properties: *Nash bargaining* (a key concept in cooperative game theory) and *dual decomposition*. In my work so far, I have applied these building blocks to designing protocols for coordinated inter-domain traffic engineering and managing dense deployments of heterogeneous wireless networks.

In this framework, participating entities with distinct local goals use an iterative procedure to jointly optimize a abstract social cost function, referred to as the Nash product; in the context of traffic engineering, for example, the Nash product is simply the product of the local objectives of the participating ISPs. By optimizing the Nash product, the entities in essence are negotiating their operating parameters, but in a manner that results in a globally optimal and locally fair outcome. Using tools for multi-criteria optimization, we show how to decompose the global optimization problem into sub-problems that can then

be solved independently and in a decentralized manner by the individual entities. Our approaches do not require the entities to share any sensitive internal information (such as network topology or user locations).

Measurement-driven Approaches to Improving the Robustness of Networked Systems:

As Web-based and data-center based applications on the Internet grow in popularity, their providers face the key challenge of determining how to configure server-side resources to provide consistently good response time and high availability to users. Ideally, these resources, such as processing and memory capacity, database and storage, and access bandwidth, should be provisioned to deliver satisfactory performance under a broad range of operating conditions. Since an operator's ability to predict the volume and mix of requests is limited, this can be difficult. My research asks if it is possible to develop a light-weight flexible framework for automatically inferring resource constraints of a live Internet application, and use the inferences to drive further provisioning decisions.

I have developed a novel wide-area profiling service that helps Web server operators better understand the ability of their Internet applications to withstand increased request load [*USENIX08, INM07*]. Our *mini-flash crowd* (MFC) mechanism reveals bottlenecks in the application infrastructure by quantifying the number and type of simultaneous requests that are likely to affect response time by taxing specific resources. Using the service, an application provider could, for example, compare the impact of an increase in database-intensive requests versus an increase in bandwidth-intensive requests. The operator could then make better decisions about how to manage additional provisioning, or take other actions.

The MFC technique is based on a phased set of simple controlled probes in which an increasing number of clients distributed across the wide-area Internet make synchronized requests to a remote application server. These requests attempt to exercise a particular part of the infrastructure such as network access sub-system, storage sub-system, or back-end data processing subsystem. As the number of synchronized clients increases, one or more of these resources may become stressed, leading to a small, but discernible and persistent, rise in the response time. We deployed MFC across ~100 PlanetLab sites.

We have applied the MFC service to a top-50 ranked commercial site and three university Web servers with the active cooperation of the site operators. The operators confirmed MFC's non-intrusive nature and its usefulness in uncovering problems with their site infrastructure, some of which the operators did not anticipate at all. Several other Web site operators have also signed up to run MFC against their set-ups.

Summary of My Research Approach:

As the above descriptions show, I try to ensure that my research ideas are vetted by solid implementation, practical deployments, testing on live networks, and interaction with operators. Interacting with operators and finding the right contacts is non-trivial and time-consuming. Nevertheless, I try to make this is a key priority because it ensures high impact. Most of my research has already had direct significant impact by the time it appears in print.

I particularly enjoy using theoretical insights and contributions from other diverse areas as the basis for the systems and protocols I build. I find such "cross-cutting" work exciting (and educational). Students who join my group quickly diversify and become adept at a variety of skills including system building, modeling and theory.

I enjoy working on deep, challenging and multi-faceted problems that take a few years to address thoroughly. Rather than follow someone else's lead, my goal is to *define* the next hot area and make the first contributions in it.

I expect future work from my research group to be driven by similar trends.