PROJECT SUMMARY: Measurements, Models, and Simulation Scenarios for Internet Research

Networking researchers work from mental models of the Internet's important properties. The scenarios used in simulations and experiments (including our own) reveal aspects of these mental models, often including one or more of the following: Flows live for a long time and transfer a lot of data. Simple topologies, like the "dumbbell" topology, with one congested link, are sufficient to study many traffic properties. Flows on the congested link share a small range of round-trip times. Most traffic across the link is one-way; reverse-path traffic is rarely congested.

All of these models can deeply affect simulation and experimental results, and therefore our evaluations of research. But none of them are confirmed by measurement studies, and some are actively wrong. Some divergences from reality are unimportant, in that they don't affect the validity of simulation results, and simple models help us understand the underlying dynamics of our systems. However, as a community we do not yet understand which aspects of models affect fundamental system behavior, and which aspects can safely be ignored.

This project proposes to build a much richer understanding of the range of realistic models, and of the likely relevance of different model parameters to network performance. It is our belief that lack of good measurements, lack of tools for evaluating measurement results and applying their results to models, and lack of diverse and well-understood simulation scenarios based on these models, are holding back the field. Our program will directly address the models used by network researchers in simulations and experiments. We will build on the wealth of existing measurement results, generating new empirical results where needed. We propose to create high-quality tools to facilitate ongoing measurement by ourselves and the wider network research community, so that our results keep up with the Internet's high rate of change. We will attempt to synthesize the relevant body of measurement results to provide concise intuitions, and a range of simulation-ready models, for use by researchers. We will concretely illustrate how each aspect of the models can influence simulations and experiments.

This proposal will not create a universal model of the global Internet (an immense and unnecessary undertaking). Instead, we will focus on models used by researchers in specific research areas, motivated by addressing their pressing modeling concerns. Naturally, we will focus at first on our own areas of expertise, particularly modeling congestion-related mechanisms at a queue in a router. We will address other research areas, however, and we hope our approach will extend to help others improve their measurements and measurement tools, along with the resulting models and simulation scenarios, and improve the relationships between experiments, models, and our shared understanding of simulation results.

To illustrate the range of our proposed work, fundamental Internet properties that remain ill understood in practice, or that are poorly reflected in the models used in our research, include the following:

- Where is congestion on the Internet? What is the pattern of congestion at routers, for routers that experience congestion? Why does congestion occur?
- What range of round-trip times might one expect to see on a single link?
- How many congested links might a flow experience? Do congested flows often experience reverse-path congestion or network-induced burstiness?
- How do link properties such as the average levels of congestion, range of round-trip times, and so forth, differ on: home access links, server access links, backbone links, peering points, transoceanic links, wireless links, satellite links, links outside of North America?
- Should our simulation topologies take all of this into account? If so, then how?

Part C
PROJECT DESCRIPTION

1 Introduction

Modern Internet research is intricately bound to simulation. The Internet's great heterogeneity and rate of change [28] make it a deeply uncontrolled environment. Experiments are difficult to replicate, verify, or even understand without the stability and relative transparency provided by a simulator (such as ns [21]), emulator (such as the University of Utah's Emulab [33]), or self-contained experimental testbed. Simulators and emulators provide full control over the *network model*, meaning the full range of experimental parameters—network topology, end-node protocol behavior, queue drop policies, congestion levels, and so forth. Researchers can model unusual situations, or extrapolate current Internet trends to evaluate likely future behavior.

Researchers also model networks with little relationship to Internet reality, or with an unknown relationship to Internet reality. This isn't necessarily a problem. Some divergences between models and reality are unimportant, in that they don't affect the validity of simulation results, or useful, in that they clarify behavior in simple cases. Furthermore, some divergence is necessary in order to investigate the Internet of the future instead of the Internet of the past or present. However, the research community has not yet determined which divergences are acceptable and which are not. We simply don't know whether the models we use are valid. This basic question has led to difficulties both in our own research and in our evaluation of other work. There is a continuing frustration with poor quality of models in general in Internet research, and a need for a critical approach in evaluating the models that we use in simulations and experiments.

We need better models and better tools for evaluating our own and others' models. We need to know when a model might lead to bad results, and what those results might be. We need this work to be grounded in Internet measurement. Some of the relevant measurements have already been published; we will need to conduct new studies as well. The measurement tools we use must be publicly available and easy to use. This will facilitate ongoing measurement as the Internet evolves, so our models will keep up with reality.

The proposed work is as follows:

- We intend to begin with what modelers would or should want to know for specific research questions. One area of particular interest is congestion-related mechanisms at a queue in a router. This includes such research topics as differentiated services, active queue management, ECN, QoS, aggregate-based congestion control, fairness, and so forth, and touches on other issues, such as design of end-host protocols. We will branch into other research areas, such as unicast and multicast congestion control, routing lookups, and peer-to-peer systems, as time allows. Table 1, below, describes some relevant research questions in more detail.
- Analysis of the research questions will lead to a description of the experimental parameters relevant for constructing models.
- Simulation experiments will show how parameter settings affect the observed behavior of existing techniques. Relevant experiments will be based on published research in the area.
- For settings that do affect behavior, new measurement studies and analysis of the measurement literature will describe how the settings look on the real Internet.
- We will distill this work into a set of best practices for model construction. This may include ready-made simulation setups, papers, RFC-like documents, and so forth.

Of course, a new technique might expose the importance of a different parameter, and special network situations might require different values for the parameter. Our work will not determine the complete set of interesting simulations for a research area. Rather, it will point out the parameters that have proven important in the past, provide observations of their values on the Internet, and describe expected effects of other values.

- We will make the measurement programs we borrow, or create, available to the research community as high-quality, maintained software tools. This will make it easy for the community to keep the best-practice models up to date with changing Internet conditions.

To illustrate this program, here are some of our thoughts on modeling for active queue management and related research.

- Current models and simulation scenarios: Relevant model parameters include the level of statistical multiplexing, range of round-trip times on the link, traffic mix, TCP variants observed, the number of congested links a connection might traverse, the kinds of links that tend to be congested (for example, transoceanic links? access links? links at peering points?), the amount of data transferred by an average flow, traffic mix and congestion on the reverse path, and so forth.

  Published work in this area generally uses a dumbbell topology with a single congested link [9]. Frequently observed are single round-trip times, all long-lived flows ("elephants"), Reno TCP (an older TCP variant), and no traffic on the reverse path other than acknowledgements.

  The simplicity of these settings can be extremely valuable. Simplicity facilitates understanding, and we do not recommend indiscriminately replacing simple experiments with complex ones. Rather, we want to describe when simple experiments generalize well to Internet scales, and when they do not. For example, we believe that, for congestion-related queue mechanisms and unicast congestion control, experiments with dumbbell topologies generalize well. This assumption should be tested, however; and as described below, experiments with single round-trip times and all long-lived flows might generalize quite poorly.

- Determining which parameters matter: Parameters such as round-trip times and transfer sizes can affect the performance of queue management (see Section 2). For example, models with a narrow range of round-trip times and with long-lived transfers tend to induce queue oscillations much more than models with a wide range of round-trip times and transfer sizes.

- Measurement results: Measurement studies confirm that flows sharing a link on the Internet exhibit a wide range of round-trip times, typically varying over an order of magnitude. Measurement studies also characterize distributions for the range of transfer sizes, with most of data transfers being short-lived "mice", but with most of the bytes transferred coming from the lonter-lived "elephants". We do not know of studies characterizing where congestion occurs, or other studies that might indicate whether dumbbell topologies are suitable for modeling.

- Models: The best practice for simulations that evaluate queue management includes, at a minimum, a wide range of round-trip times and a wide distribution of transfer sizes.

Table 1 lists a range of topics in network research, along with the typical models used in simulations, experiments, or analysis in that area of research. The table also lists the supporting classes of measurements used in constructing these models. As the table suggests, our work starts with current active research topics and the current models used in that research, and considers the supporting measurement studies that need to be integrated to evaluate and improve those models.

Research Problems	Typical Models	Supporting Measurements
AQM, scheduling, differentiated services.	A dumbbell topology, with aggregate traffic.	Characteristics of congested links, Range of round-trip times, Traffic characterization (distribution of transfer sizes, etc.), Reverse-path traffic, Effects of congestion elsewhere.
Unicast congestion control.	A single path, with competing traffic.	Characteristics of links, Queue management along path, Packet-reordering behavior, Packet corruption on a link, Variability of delay, Bandwidth asymmetry.
Multicast congestion control.	A single multicast group in a large topology.	Router-level topologies, Loss patterns, Traffic generation by group members.
Routing protocols.	A large topology.	Router-level topologies, AS-level topologies, Loss patterns.
Routing lookups.	A lookup trace, or a model of the address space.	Ranges of addresses visible at a link.
Web caching and CDNs, Peer-to-peer systems.	Models of large topologies with application traffic.	Topologies, Application-level routing, Traffic patterns.
Controlling DDoS attacks.	Models of large topologies with aggregate traffic.	Topologies, Attack patterns.
Web cache performance.	A single cache with many clients and servers, as in Web Polygraph.	Detailed client behavior, server behavior.

Table 1: Some research problems, with typical models and required supporting measurements.

We aren't trying to construct a single model of the global Internet, with a single set of simulation scenarios, for use by all researchers. Each research area needs its own models, and supporting measurements for building and evaluating those models. Combining these into a single model would make people's simulations run slower without necessarily improving their precision, clarity, or applicability. We need to understand the factors relevant to each model, not duplicate the Internet in miniature.

In addition, we understand quite well the Internet's vast heterogeneity and rapid rate of change [28]. The Internet cannot be simply and accurately modeled in the same way that one might model a machine that one could hold in one's hand. Modeling the Internet, or even just modeling selected aspects of the Internet, is not a simple task. However, as researchers we are clearly not going to suspend our work until our models of the Internet have passed some critical threshold of accuracy and relevance; we continue with our work with the best models and simulation scenarios that we can find at the time. We believe that it is possible to make significant advances in the state of the art in providing models and simulation scenarios for the network research community.

Our concern with the broader impact of this work should be clear. Better models will make the Internet community's research efforts more effective. Lack of agreement over models complicates comparison and collaboration, and researchers risk expending valuable effort on dead ends caused by invalid models. Better models will therefore immediately improve the state of Internet research, and perhaps the Internet itself.

Our results will be disseminated in the forms most useful for the research community, both as best-practice documents and as infrastructure—for example, as high-quality measurement tools and simulation scenarios. Our focus on pragmatic measurement, grounded in the realities of the Internet, will prove valuable for the students that will assist us.

The rest of this proposal proceeds as follows. Section 2 illustrates the importance of the underlying model to network research, by showing how small model changes can have large effects in simulation results. Sections 3 and 4 discuss particular parameters relevant for router-queue and other interesting models. These sections present the relevant parameters, including pointers to related work, and describe supporting measurements for discovering their values in practice. The presentation is necessarily detailed here, as any working research model must be, but the goal should remain clear: improving the quality of models people use today. Section 5 describes why high-quality measurement tools are needed, and our plans for building them. The remaining sections list our current funding, summarize technological and societal impact, present our resarch plan, and conclude.

2 Why Models and Simulation Scenarios Matter

In this section we argue that there is a pressing need in the research community for assistance in evaluating and constructing models and simulation scenarios. If most of the models used today could be counted on to give similar results, and if the results could be counted upon to be relevant to the current and/or future Internet, then there would not be a problem. However, different models and different simulation scenarios do give different results, and have different degrees of relevance to the actual Internet. For example, some simulations demonstrate sensitive dependence on precise parameter settings, as shown below. That rich behavior is irrelevant to the modern Internet: it's an artifact of unrealistic aspects of the simulations, such as low levels of statistical multiplexing and no reverse-path traffic. We would like to discourage additional researchers from wasting their time investigating in depth the rich behavior of similarly unrealistic scenarios.

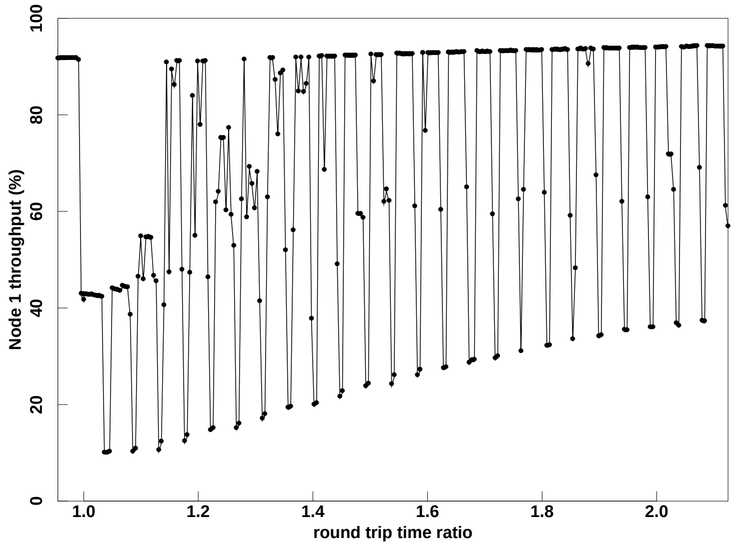

Figure 1: Flow 1's throughput as a function of the ratio of the two flows' round-trip times.

Figure 1 (borrowed from [10]) illustrates phase effects, where a small change in the propagation delay of a single link completely changes the fraction of link bandwidth received by one of two flows sharing a Drop-Tail queue. Each dot on the graph represents the result of a single simulation; the *y*-axis shows the throughput of flow 1 in that simulation. The simulation topology is a simple dumbbell. When the propagation delays of the two competing flows' access links are equal, then both flows have the same round-trip time and receive the same fraction of the link bandwidth. However, as the propagation delay of one of the access links changes slightly, flow 1 can shift to receiving almost all of the link bandwidth to receiving very little of the link bandwidth, depending on the exact propagation delays of the two access links. In real networks, of course, the traffic mix contains many short-lived flows and core links have a much higher degree of statistical multiplexing, making phase effects much less likely. The lesson is not that phase effects are a significant or important dynamic to address in current networks, but rather, that simulations can be very tricky, and that the combination in a simulation scenario of DropTail queue management with one-way long-lived traffic can be deadly indeed.

In simulations of RED active queue management, we have shown how oscillations in the average queue size can be strongly affected by moderate changes in the traffic mix [9, Section 5.1]. In particular, we showed a simulation scenario based on other simulations in the literature, where all of the flows have the same long (250 ms) round-trip time, with most of the traffic consisting of one-way traffic of long-lived flows. In the simulation with RED, there are undesirable oscillations in the average queue size. A small change in the traffic mix, with the addition of a modest amount of web traffic and reverse-path traffic, is sufficient to significantly reduce oscillations.

In other work in progress on evaluating active queue management mechanisms [24], we have many simulation sets that show similar performance from a range of queue management mechanisms. However, we also have simulation scenarios with long round-trip times and mostly long-lived flows that show oscillations in the queue size with RED [11] or with Adaptive RED [9]. We have simulation scenarios with mostly web traffic, or with changes in the level of congestion over time, that show poor performance with the Proportional-Integral Controller (PI) [12] or with Random Early Marking (REM) [2]. We have simulation scenarios with Drop-Tail or with Adaptive Virtual Queue (AVQ) [18] that give competitive performance in terms of delay-throughput tradeoffs, but that also give high packet drop rates. It would be helpful to have more grounding in deciding which models and simulation scenarios were critical to explore, and which are edge cases that were less likely to occur in practice. It is unsettling to feel that one could construct a simulation to show almost anything that one wanted, and that there is so little agreement within the research community about why one chooses to explore one set of simulation scenarios rather than another.

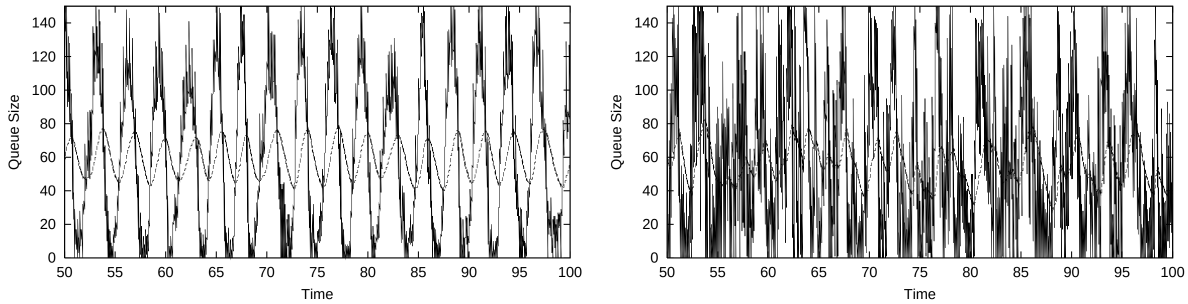

Figure 2: Scenarios with different traffic mixes and distributions of round-trip times. Left: mostly long-lived traffic, all with the same round-trip time. Right: mostly web-traffic, with a range of round-trip times.

As a concrete example of the effect of the simulation scenario, Figure 2 shows two simulations with quite different results in terms of the queue dynamics at the congested link.[1] The left-hand simulation in Figure 2 shows the instaneous queue size over the second half of a 100-second simulation. The simulation scenario is of a dumbbell topology with a 15 Mbps, 20 ms congested link, with all of the round-trip times

[1]The scripts that created these simulations are available at `http://www.icir.org/models/nsf_sims.html`.

of 180 ms. The traffic consists mostly of 80 long-lived flows, each with a receiver's advertised window of 20 MSS-sized segments. The packet drop rate is 2.6%, and the throughput is 96%, but the oscillations in the instantaneous queue are quite pronounced. The dashed line shows the estimated average queue size used by RED in determining the packet drop rate.

The right-hand simulation in Figure 2 shows a simulation scenario in which two changes have been made. First, although the average round-trip time is still 180 ms, the round-trip times vary from 40 to 320 ms. The second and more critical change is that the traffic has been changed to a mix consisting mostly of traffic from the web traffic generator in NS, along with a small number (15) of long-lived flows with a large receiver's advertised window. The demand from the web traffic generator was chosen to give the same packet drop rate of 2.6%. The throughput is also relatively high, at 93%, but the queue dynamics and the distribution of queueing delay are rather different. Which is more important to explore, the more pronounced oscillations in a scenario with long-lived flows all with the same round-trip time, or the variability of demand over shorter time scales that comes from a traffic mix closer to that observed on real links, with flows with a distribution of transfer sizes? It is not necessarily the case that the mechanisms proposed to address the oscillations in the left-hand graph also perform well in scenarios with more diverse traffic, as in the right-hand graph.

For some issues, such as the distributions of transfer sizes for web traffic, there is a substantial body of analysis, measurement work and synthesis of that work, including a well-established web traffic generator in NS itself. What is needed is additional work in providing guidance, and concrete simulation scenarios, to help researchers to use the existing traffic generators. In other cases, such as typical values for the range of round-trip times, some measurement work has been done, but tools are needed to expand and integrate the measurement work into tools that can be used in generating and evaluating simulation scenarios. However, for other important modeling issues, we have very little information about the underlying characteristics of the Internet. One such issue concerns the presence (or absence) of congestion along the reverse-path.

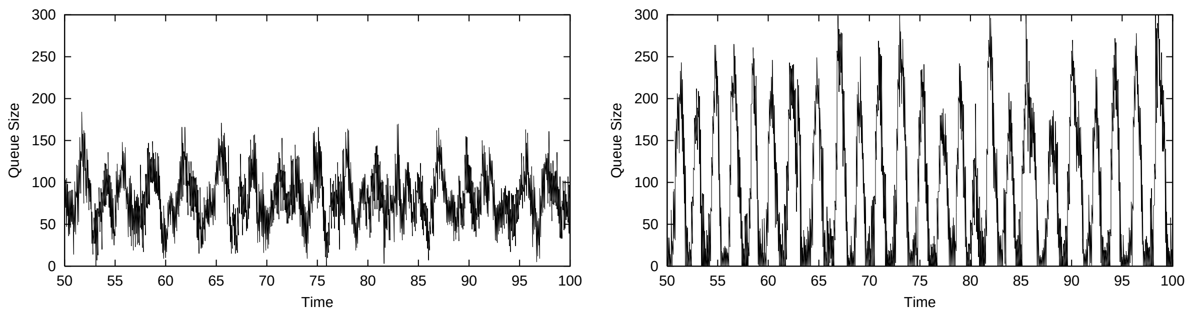

Figure 3: Scenario without (left) and with (right) reverse-path congestion.

Figure 3 shows two simulations, one without congestion on the reverse path, and one with such congestion. The left-hand simulation, without reverse-path congestion, contains a range of round-trip times and a rich traffic mix of web traffic along with long-lived traffic, and shows high throughput (97%), moderate drops (2.4%), and low delay. The right-hand simulation differs only in that it contains more long-lived TCP connections along the reverse path (although the reverse-path queue is less congested than the forward-path queue). As the simulations show, the presence of congestion on the reverse-path strongly influences the queue dynamics for the forward-path queue, significantly increasing oscillations, and resulting in higher delay and lower average throughput.

Thus, the ability of the queue management module to achieve high throughput and low delay can be heavily affected by the presence or absence of reverse-path congestion in the queue, but we have no idea if reverse-path congestion is or is not a common occurrence for traffic in the Internet. When acknowledgements encounter queues on the reverse path, this results in changes in the spacing between ACK packets (ACK compression), and drops of ACK packets; these in turn increase the burstiness of traffic in the forward path. In addition, the presence of congestion along the reverse path increasing queueing delay, increasing round-trip times. However, we know very little about whether most traffic that encounters congestion in one direction is likely to also encounter queues and congestion in the reverse direction or not.

3 Parameters for Queuing Models

In this section we discuss in more detail the models for aggregate traffic at a queue needed for analysis and simulations in queue management, scheduling, differentiated services, aggregate-based congestion control, and similar topics. In all of these areas, researchers are concerned about mechanisms for act on a single queue in a router. The models used often consist of a dumbbell topology, sometimes with reverse-path traffic and a range of round-trip times, and sometimes with a range of traffic include web traffic as well as long-lived file transfers.

The previous section has already discussed briefly the relevance of the range of round-trip times, the traffic mix, and the presence (or absence) of reverse-path congestion, for the specific research topic of active queue management. In this section we discuss the various measurement topics in more detail.

3.1 Characterizing the congested links

For research on scheduling and queue management mechanisms, the only links that are of interest are those links that sometimes experience congestion. As a first question, we don't even know where the congested links are, and what are the patterns of congestion. The prevailing assumption is that congestion rarely occurs within backbone networks in North America, but that congestion occurs in edge networks, access links to the home, links to peering points, and transoceanic links. Do we have sufficient data to back this up? How well can we characterize these links, if at all, in terms of typical bandwidths and propagation delays, the level of statistical multiplexing, the packet drop rates, patterns of congestion, and so on? And how does congestion look outside of North America? Backbone networks in Asia and Europe differ structurally from those in North America, and may have different congestion characteristics, but often network research focuses exclusively on the North American situation. We want to take an explicitly global viewpoint, collecting information from our colleagues in other countries. And if we had information on congested links on the current Internet, how well could we extrapolate from this to models for future congestion in the Internet? Should we be building separate models for separate classes of congested links, e.g., for access links, campus links, transoceanic links, links to public or private peering points, etc.? We will begin to address these issues over the summer, including an investigation into our ability to infer the answers to some of these questions from packet traces.

We note that there is a great deal of information about packet drop rates in the Internet [5], from the Internet Traffic Report [22], the Internet Weather Report [20], the Internet End-to-end Performance Monitoring Group [13], Telcordia's NetSizer [32], and other sites. However, most of these measurements are based on ping data, or other end-to-end path properties, and do not identify the characteristics of the congested link. Some research has begun on characterizing network anomalies such as network operation anomalies, flash crowd anomalies, and network abuse anomalies that can cause congestion [4], but again, little is known about general patterns of congestion at occasionally-congested links. In the absence of more information, network researchers (including ourselves) often tend to pick some arbitrary but plausible packet drop rate, and assume a relatively unchanging pattern of congestion over time.

3.2 The range of round-trip times

The range of round-trip times can have a clear affect on the results of network simulations, particularly in simulations involving end-to-end congestion control. The question can be phrased as follows: For packets on a particular link, assuming that each packet could be assigned the round-trip time of its associated connection, what would be the per-packet distribution of round-trip times? Measurements of the range of round-trip times for traffic on a link go back at least as far as the 1996 Olympic Web server [3], and continue to the present [1]. Techniques for estimating the round-trip times from packet traces are discussed by Jiang and Dovrolis [14].

Questions that remain to be addressed include the following:

- What additional measurements and additional measurement tools are needed, if any, for examining the range of round-trip times at different congested links in the Internet?
- Given known results about the range of round-trip times on a link, how do we translate in a simulation scenario to assigning propagation delays to links in a dumbbell topology? (The distribution of end-to-end propagation delays in a simulation topology is not the same as the distribution of round-trip times of the TCP traffic on the congested link.)
- How do measurements of round-trip times match to the underlying propagation delay, and how much of the round-trip times are due to other factors? (Andre Broido has some preliminary measurements that address this issue.)
- Are there likely to be changes in the ranges of round-trip times over time, as the Internet evolves with caching infrastructures, content distribution networks, and the like? Do we need tools to help track these changes?
- It is clear that theoretical analysis—using, for example, queuing theory—is much more tractable for scenarios of connections that all have the same round-trip times. How can we evaluate the applicability of this analysis for more complex and more realistic scenarios?

Our plans are to begin to address all of these issues.

3.3 Characterizing and generating traffic

There has been a wealth of measurement studies, analysis, and simulation tools for characterizing aggregate traffic, and for traffic generation for simulations and experiments. Questions that have been investigated, and some that might need to be investigated further, concern the breakdown of aggregate traffic as TCP and UDP, the distribution of flow sizes, and the identification of TCP variants.

Transport protocols:

What fraction of the link traffic is TCP? Are the results sensitive to the variants of TCP represented in the aggregate traffic (e.g., Reno, NewReno, and SACK TCP; one-way or two-way TCP; with or without ECN)? What about the subset of the traffic that is not using end-to-end congestion control?

Recent measurements tend to show 90–95% of the bytes on a link come from TCP packets [6, 30].

TBIT, the TCP Behavior Inference Tool, has been used to identify the characteristics of TCP implementations in current web servers [23]. As an example, the findings were that the web servers were essentially all using conformant end-to-end congestion control mechanisms, and that Reno TCP is no longer the dominant TCP variant. It would be useful to have similar measurement results of the TCP behavior of web clients as well.

Transfer sizes:

What is the distribution of flow sizes, ranging from "mice" (short web transfers) to "elephants" (long-lived sessions)? Several studies have shown that, while most of the connections represented in a trace are mice, most of the bytes or packets in a trace generally come from the elephants [27, 31]. The distribution of connection sizes is often modeled by a log-normal distribution for the body of the distribution, and with a heavy tail (e.g., a Pareto distribution with shape parameter less than two) [28].

NS includes a web traffic generator, based in part on the earlier SURGE web traffic generator. The simulation script can specify parameters such as the distributions for the number of objects per web page and the number of packets per object. In [8], it is shown that the aggregate traffic generated by the web traffic generator has self-similarity properties similar to those observed in actual Internet traffic.

There is also some initial work on characterizing file sizes for peer-to-peer traffic; large video files dominate much of the peer-to-peer traffic.

Session start times:

It has generally been shown that sessions are well-modeled as having independent start times [28].

Will it become important to add traffic from flash crowds and from DDoS attacks to simulation scenarios?

Future work:

Because so much work has already been done in this area, the work for us to do is largely to synthesize some of the existing work, to identify some of the research topics where the choice of traffic characteristics can have a key impact, and to provide simulation scenarios that incorporate rich mixes of traffic. It is possible that additional measurement studies will be needed for things such as tracking the deployment of ECN, or for adding to existing work on characterizing large UDP flows not using end-to-end congestion control.

3.4 Small-scale burstiness

As Section 2 has illustrated, the performance of queue management can be affected by the bandwidth limitations, queueing, and packet drops experienced by data and acknowledgement packets elsewhere in the network. Feldman et al. show that burstiness of aggregate traffic over small time scales can be impacted by TCP-related factors, such as ACK compression and the small-scale dynamics of end-to-end congestion control [8]. Sarvotham et al. suggest that bursts also arise from high-volume "alpha" traffic, with the "beta" traffic consisting of traffic constrained by low-bandwidth links elsewhere in the network [29]. In our own simulations, we have found that the dynamics of the aggregate traffic on a link can be strongly affected simply by the details of the queue management on the reverse path (e.g., RED or Drop-Tail, or by the buffer size).

Understanding how to characterize aggregate traffic on a link in terms of the congestion experienced by that traffic elsewhere on the forward and reverse path remains an open question, and, to the best of our knowledge, very little has been done in terms of integrating this aspect into our models and simulation scenarios. Further, it would be nice to be able to translate the effect of congestion elsewhere in the network into our simulation scenarios as concisely and simply as possible. Our goal is not to construct models and simulation scenarios that include the entire range of behaviors in the Internet, but to look for the simplest models and scenarios sufficient to investigate the issues at hand.

4 Parameters for Other Models

This section briefly addresses models for research topics other than router queue behavior—in particular, for routing lookup and for unicast congestion control. This demonstrates how our methodology applies to other domains. We will continue the measurement and modeling studies described here, and may investigate models and simulation scenarios for other research topics listed in Table 1 as time permits.

4.1 Modeling for research in router lookup

The basic operation performed by a router is to look up a destination IP address in a routing table. Any model for routing lookups, therefore, must address both the set of routes in force at a given point in the network, and the addresses visible on packets at that point. These factors combine with the routing lookup algorithm in force to determine average-case lookup performance. For example, the size and rate of change of the set of visible addresses affects the potential effectiveness of caching strategies. Also, the measurements needed for this research problem—visible addresses for routing lookup models—are also needed for models for other research problems, such as aggregate-based congestion control. We hope to apply our measurements to multiple research problems in this way whenever possible.

Several freely available sets of routes exist, such as Oregon RouteViews, and AS and router topology have already been extensively addressed in the existing literature [7]. However, we are not aware of previous work characterizing the structure of addresses visible on packets. Our study (with Li, Karp, Paxson, and Shenker) of several widely-varying packet traces indicates that addresses follow a multifractal distribution [16]. This has several pragmatic consequences. For example, many properties of address aggregates, such as the number of active addresses in a given /8 network, demonstrate high variability. We also found that address replacement—the way in which the set of visible addresses changes—is well represented by a model with two kinds of addresses, "long-lived" addresses that survive any time interval with a probability p_l and "short-lived" addresses that survive with a lower probability p_s. This model could aid in the construction of realistic traffic generators for testing address-related algorithms.

As future work, we will further analyze requirements for routing lookup models and see more carefully how our findings fit in, and release the tools developed and used for the initial phase. We may also produce more measurements using these and other tools.

4.2 Modeling for research in unicast congestion control

For constructing models used in research in unicast congestion control—for example, research into end-to-end TCP behavior and improvements—the researcher has to specify at least the characteristics of the end-to-end path in terms of delay and link bandwidth, any competing traffic, and the level of statistical multiplexing to be explored. A number of other important factors, which have sometimes received less attention, include the queue management and buffer configuration for the congested links along the path, the presence or absence of ECN, and "corner cases" such as packet reordering, packet corruption, or variable delay from unreliable wireless links.

Measurement studies have examined some of these factors in practice, such as packet reordering [25]. In other cases, such as characterizing the queue management and configured buffer sizes for routers, we know very little about how current routers are configured, and any experimental data, either from ISPs or from end-to-end packet traces, could be quite helpful.

Congestion control models and router queuing models are particularly interlinked. Many measurements, such as those of flow round-trip times, are relevant to both models. Furthermore, research into router mechanisms such as AQM schemes depends on end-to-end congestion control behavior, and vice versa. Researchers using more realistic models of end-to-end traffic would perhaps make a convincing case that smaller configured buffers were sufficient at the routers, or that larger configured buffers were a convincing need, resulting in changes in the actual practice of configured buffer sizes in the Internet. Or take packet reordering: Although reordering is rare on today's Internet, technological changes (particularly the increasing speed demanded of routers) make reordering more and more likely in the future. TCP's bad reaction to reordering is delaying the deployment of these important technological changes. Unicast congestion control research that improves TCP's robustness to reordering, then, would cause changes to the Internet's router mechanisms by a ripple effect.

As part of this proposal, we will begin work on some of the issues in models for unicast congestion control, with a particular eye towards reusing measurements and tools developed for the queuing models described in Section 3.

5 Tool Building

This section turns to the measurement tools we plan to build. We explain why tool-building will require engineering effort, list the useful properties tools can have, and describe one way to achieve those properties.

Any measurement project needs both an algorithm describing how to take the measurement and a tool that implements that algorithm. Measurement algorithms are significant contributions in and of themselves, and turning a local measurement tool into a robust piece of software easily usable by others is a significant engineering undertaking. For example, a locally-developed tool might run slowly, or fail easily, or require onerous conditions on its input (such as that a trace contain only a single TCP flow), or come without documentation. This presents little difficulty in the context of a focused project, where the tool author is around to fix anything that goes wrong. In the global context we intend, however—where the measurement tools are used widely and over a long period of time—these problems are unacceptable.

We plan, therefore, to consolidate measurement tools from others, to add our own as necessary, and to develop these tools into a suite of robust, solid, documented software easily usable by researchers. For example, we might take a round-trip time estimator for TCP flows, such as that described by Jiang and Dovrolis [14], or an estimator for TCP loss rate, and produce tools working on arbitrary traces that output combined loss-rate and RTT information.

We have expected at least the following properties from tools used in prior measurement studies, and will build them into our measurement tools.

- **Input flexibility.** The tool should read all common trace input formats, including tcpdump's pcap format, the University of Waikato's DAG format, and various ASCII formats, as well as reading packets directly from the wire. It might also be useful to read from simulator traces, both to characterize simulation results and for tool testing.

- **Selection flexibility.** It should allow selection of subsets of the input traffic using various criteria: classification by protocol or address, say, or random sampling, or a limited time range.

- **Anonymization.** Any tool working with IP addresses and/or packet data should be able to anonymize both, facilitating public data sharing.

- **Speed.** Processing huge trace files will always be slow, but the tool should add no overhead.

- **Output flexibility.** Output files should be easily processable by user scripts.

- **Usability.** The tool should have a consistent interface, inform the user of its progress when processing large files, and everything else we expect of quality software.

When appropriate, we will achieve these engineering goals using the Click modular router [17]. Click is well-suited both for packet processing, such as trace processing, and for software component reuse. For example, several projects have considered IP address anonymization [19, 34]. However, the available tcpdpriv tool [19] works only on tcpdump traces (or wire packets). In Click, a component "*AnonymizeIPAddr*" implements address anonymization; it may be included independent of the trace format. We have built several tools using this methodology, including tools for analyzing address structure in a packet trace. Even a very simple tool, the ipsumdump program [15], continues to be useful for both us and others, proving the practical value of careful tool-building.

6 Relation to Ongoing Work

Sally Floyd is one of the two co-founders of the *ns* simulator, and she has been continuously involved with its development. She also has a long history of interest in how we choose the models that we use in our research, and how we apply the results of this research back to the real world. Her other ongoing work has mostly been on the understanding and development of transport protocols, and on router mechanisms for active queue management, explicit congestion notification, detecting and policing misbehaving flows, and aggregate-based congestion control. The further development of all of this work will contribute to, and build upon, our understanding of models and simulation scenarios, and how they affect our ongoing research.

Since 1999, Sally has been at ICSI's Center for Internet Research, funded by AT&T; this funding from AT&T terminates at the end of 2003. She is also one of the PIs for an NSF proposal on "Addressing Fundamental Issues for Robust Internet Performance" that was recently funded. Her work in that proposal involves the design of congestion control mechanisms for increased robustness in a wide range of environments; that work should be complementary to the model-based work in this proposal.

Eddie Kohler also works at ICSI's Center for Internet Research, on router forwarding paths, protocol design, and measurement projects. He developed the Click modular router for his doctoral work at MIT, and values, and has experience in, producing high-quality, well-documented software tools for the research community; Click's distributions are used at many universities and companies. His ongoing work on Click's development is supported by Xorp, the eXtensible Open Router Project at the ICSI Center for Open Networking (funded by Intel and AT&T/ICSI), and by Mazu Networks, Inc., a startup at which he is Chief Scientist. He has received no prior NSF funding.

7 Technological and Societal Impact

This proposal was designed to serve the Internet research community and, by extension, society at large. Better models will tie research more firmly to the global Internet. It will become easier to do research that focuses on the Internet's actual problems, and hopefully, to fix those problems, rather than expending effort on cul-de-sacs from inappropriate models. It will also become easier to do research when one can readily build upon a publically-avaliable set of simulation scenarios, along with guidelines to the relevance of different parameters of the underlying models. Just as the publically-available network simulator NS has been a powerful tool for the network research community, a publically-available set of models and accompanying simulation scenarios can be a powerful tool as well. The result should be a better Internet for everyone.

Our proposal should result in tools, models, and simulation scenarios that are of widespread use to the research community. We will distribute our results widely, and in the forms most immediately useful for use and development by others. For example, we will publish NS simulation scenarios corresponding to the models we develop, and maintain repositories for simulation scenarios developed by other researchers. We will produce papers clearly evaluating underlying models, clarifying which model parameters are well understood or appear to be stable invariant properties, and which are not well understood or are in a state of rapid change. We will distribute our measurement tools in high-quality, maintained distributions. All of this work will be freely available.

We plan to execute our plan with the help of summer interns and other graduate students. Through measurement studies and tool building, these students will get their hands dirty in the real Internet, and learn the importance of valid models and assumptions for their own research.

8 Research Plan

This section summarizes our current plan for how we will accomplish the goals of the proposal. Of course, the actual order of execution depends on factors unknown to us now, such as the preferences of particular students that express interest in working with us.

- **Year 1: Consolidation.** Collect relevant measurement studies from the literature and summarize their conclusions. Develop initial versions of measurement tools—for example, effective tools for measuring TCP loss rates from packet traces. Investigate the first set of research questions, including particularly congestion-related mechanisms at router queues, and determine the experimental parameters relevant for building models. Collect and make available simulation scenarios representative of the current state of the art.

- **Year 2: Analysis.**

  Conduct measurements using the measurement tools created in Year 1 and some combination of public traces, the National Internet Measurement Infrastructure [26], and private traces. Combine measurement studies with new measurements to determine experimental values for the relevant parameters. Integrate results from recent measurements (our own, and those of others) into simulation scenarios, and experiment with simulations to determine how experimental parameters can affect simulation results.

- **Year 3: Solidity and Extension.** Digest the preferred settings for experimental parameters into simulation scripts and papers. Prepare high-quality distributions of these scripts and the measurement tools. If time allows, extend this work into other research questions.

9 Conclusions

In this proposal we have outlined a plan to improve the models and simulation scenarios used in network research. Specifically, starting in those research areas in which we have expertise, we plan to do the following:

- Understand the existing measurement studies and perform new measurement studies where needed.
- Create tools as needed to facilitate ongoing measurement by ourselves and the wider network research community.
- Understand existing models, and build upon and improve these models as needed.
- Understand existing simulation scenarios, and provide a range of diverse simulation scenarios for researchers to build upon.
- Publish papers outlining the methodology in going from measurements to models to simulation scenarios.

- Publish papers clarifying, to the extent possible, the general impact of different aspects of the simulation scenarios, and of the underlying models, on network performance.

If this effort is successful, we would explore the possibility of creating a repository for the network community to share tools, experimental results, models and simulation scenarios, and to build up a body of work that rigorously and systematically investigates the impact of different models and different simulation scenarios on aspects of network performance.

Part D
REFERENCES CITED

[1] Mark Allman. A Web server's view of the transport layer. *ACM Computer Communication Review*, 30(5), October 2000.

[2] Sanjeewa Athuraliya, Steven Low, and David Lapsley. Random Early Marking. In *Proc. 1st International Workshop on Quality of Future Internet Services (QofIS 2000)*, September 2000.

[3] Hari Balakrishnan, Venkata N. Padmanabhan, Srinivasan Seshan, Mark Stemm, and Randy H. Katz. TCP behavior of a busy Internet server: Analysis and improvements. In *Proc. IEEE INFOCOM 1998*, pages 252–262, March 1998.

[4] Paul Barford and David Plonka. Characteristics of network traffic flow anomalies. In *Proc. ACM SIGCOMM Internet Measurement Workshop 2001*, November 2001. `http://www.icir.org/vern/imw-2001/proceedings.html`.

[5] CAIDA. Internet Measurement Infrastructure. `http://www.caida.org/analysis/performance/measinfra/`.

[6] CAIDA. Traffic workload overview (Web page). `http://www.caida.org/outreach/resources/learn/trafficworkload/tcpudp.xml`.

[7] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. On power-law relationships in the Internet topology. In *Proc. ACM SIGCOMM 1999 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*, pages 251–262, August 1999.

[8] Anja Feldmann, Anna C. Gilbert, Polly Huang, and Walter Willinger. Dynamics of ip traffic: A study of the role of variability and the impact of control. In *Proc. ACM SIGCOMM 1999 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*, August 1999. `http://www.acm.org/sigcomm/sigcomm99/papers/session8-3.html`.

[9] Sally Floyd, Ramakrishna Gummadi, and Scott Shenker. Adaptive RED: An algorithm for increasing the robustness of RED's active queue management, August 2001. `http://www.icir.org/floyd/papers/adaptiveRed.pdf`, submitted for publication.

[10] Sally Floyd and Van Jacobson. On traffic phase effects in packet-switched gateways. *Internetworking: Research and Experience*, 3(3), September 1992.

[11] Sally Floyd and Van Jacobson. Random Early Detection Gateways for Congestion Avoidance. *IEEE/ACM Transactions on Networking*, 1(4):397–413, August 1993.

[12] C. Hollot, Vishal Misra, Don Towsley, and Wei Bo Gong. On designing improved controllers for AQM routers supporting TCP flows. In *Proc. IEEE INFOCOM 2001*, April 2001.

[13] IEPM Group, Stanford Linear Accelerator Center. Internet End-to-End Performance Monitoring home page (Web site). `http://www-iepm.slac.stanford.edu/`.

[14] Hao Jiang and Constantinos Dovrolis. Passive estimation of TCP round-trip times, 2002. http://www.cis.udel.edu/~dovrolis/Papers/rtt.ps.gz.

[15] Eddie Kohler. Ipsumdump (Web page). http://www.icir.org/kohler/ipsumdump/.

[16] Eddie Kohler, Jinyang Li, Vern Paxson, and Scott Shenker. Observed structure of addresses in IP traffic. Work in progress, citation for acknowledgement purposes only.

[17] Eddie Kohler, Robert Morris, Benjie Chen, John Jannotti, and M. Frans Kaashoek. The Click modular router. *ACM Transactions on Computer Systems*, 18(3):263–297, August 2000.

[18] Srisankar Kunniyur and R. Srikant. Analysis and design of an adaptive virtual queue (AVQ) algorithm for active queue management. In *Proc. ACM SIGCOMM 2001 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*, pages 123–34, August 2001.

[19] Greg Minshall. Tcpdpriv: Program for eliminating confidential information from traces. http://ita.ee.lbl.gov/html/contrib/tcpdpriv.html.

[20] Matrix NetSystems. The internet weather report (web site). http://www.mids.org/weather/.

[21] NS project. The network simulator—ns-2 (Web site). http://www.isi.edu/nsnam/ns/.

[22] Opnix, Inc. Internet Traffic Report (Web site). http://www.internettrafficreport.com/.

[23] Jitendra Padhye and Sally Floyd. On inferring TCP behavior. In *Proc. ACM SIGCOMM 2001 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*, pages 287–298, August 2001.

[24] Jitendra Padhye, Sally Floyd, and Scott Shenker. Simulation-based comparison of AQM schemes. Work in progress, citation for acknowledgement purposes only.

[25] Vern Paxson. End-to-end routing behavior in the internet. *IEEE/ACM Transactions on Networking*, 5 (5):601–615, October 1997.

[26] Vern Paxson, Andrew Adams, and Matt Mathis. Experiences with NIMI. In *Proc. Passive and Active Measurement Workshop*, April 2000.

[27] Vern Paxson and Sally Floyd. Wide-area traffic: The failure of Poisson modeling. *IEEE/ACM Transactions on Networking*, 3(3):266–244, June 1995.

[28] Vern Paxson and Sally Floyd. Difficulties in simulating the Internet. *IEEE/ACM Transactions on Networking*, 9(4):392–403, August 2001.

[29] Shriram Sarvotham, Rudolf Riedi, and Richard Baraniuk. Connection-level analysis and modeling of network traffic. In *Proc. ACM SIGCOMM Internet Measurement Workshop 2001*, November 2001. http://www.icir.org/vern/imw-2001/proceedings.html.

[30] Henning Schulzrinne. Internet technical resources: Long-term traffic statistics (Web page). http://www.cs.columbia.edu/~hgs/internet/traffic.html. This page summarizes and links to other studies.

[31] Anees Shaikh, Jennifer Rexford, and Kang G. Shin. Load-sensitive routing of long-lived IP flows. In *Proc. ACM SIGCOMM 1999 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*, pages 215–226, August 1999.

[32] Telcordia, Inc. Internet quality of service assessment, daily update (Web site). `http://www.netsizer.com/daily/quality_today.html`.

[33] Utah Emulab project. Emulab.net, the Utah network testbed (Web site). `http://www.emulab.net/`.

[34] Jun Xu, Jinliang Fan, Mostafa Ammar, and Sue B. Moon. On the design and performance of prefix-preserving IP traffic trace anonymization. In *Proc. ACM SIGCOMM Internet Measurement Workshop 2001*, November 2001. `http://www.icir.org/vern/imw-2001/proceedings.html`.