1 Preview and Motivation

We propose to explore economic mechanisms and protocols that create incentives for service and content providers to improve the network utility for users. We plan to study how these incentives might shape the industry. We want also to explore the implementation issues, including scalability, incremental deployment, and extensibility. Our main platform is the $100 \times 100$ architecture, but we want also to examine these ideas in the context of today’s Internet.

The $100 \times 100$ project, funded by NSF, investigates the architecture of a network that could provide 100 Mbps to 100 million homes in the U.S.A. Our involvement in this project concerns the retrospective analysis of the economics of the Internet and the study of the structure of the industry. In this proposal, we want to explore concrete mechanisms and protocols that provide the correct incentives. This study complements the investigation of structural issues that we conduct as part of $100 \times 100$. We need this complement because our part of the $100 \times 100$ funding is limited to two students and does not enable us to explore the topics of the present proposal.

Here is a quick preview of the questions we want to study. Do quality of service issues disappear as the network evolves? For instance, do delays and throughput become acceptable as one scales access speed, the number of users, and the applications? Does service differentiation help generate more revenues and improve the utility of users? What is the incentive for providers to remove bottlenecks? The current inter-domain routing protocol, BGP, creates incentives for some network providers not to advertise routes to limit traffic through their domain. Can mechanisms be introduced to correct this inefficiency? Do providers have an incentive to increase reliability, for instance by using diversity routing? Should we introduce routing choice for users? If economies of scale drive out competition, can one design suitable mechanisms that mitigate that effect? Could one design simple incentives for providers to be more effective at filtering out denial of service attacks?

We believe that these questions are essential. One of the lessons of the Internet is that, for historical reasons, it does not have mechanisms for market creation. That is, the best effort model of the Internet has no built-in mechanism for monitoring or controlling the services that the different segments of the network provide. Consequently, providers have limited incentives to improve their services and users cannot adjust their service demand to match the application requirements. The lack of such mechanisms is an impediment to the improvements of the Internet.

Now that the Internet has mutated from a government service to a commercial system, its evolution is driven by economic motivations more than by social welfare considerations. There is a disconnect between this reality and the basic mechanisms of the Internet. A simplified analogy of the current Internet is a postal system without stamps where you would place your packages in the mailbox and hope for the best.

Much work has been done on network protocols and their evaluation. Clever schemes such as ECN, DiffServ, MPLS with traffic engineering, diversity routing, Fast TCP, and many other such protocols came out of that work (see [CK95], [KMT98], [KS01], [JWL03], [F94], [ABKM01], [GK99]). At the present stage of networking, what is lacking is an understanding of the economics of networks and the interplay between protocols, the economic mechanisms they enable, and the economic incentives that result. We believe that we are particularly qualified to study this interplay because of our understanding of networking and our familiarity with economic models.
intuition, motivations, and techniques.

We plan to use combinations of modelling, optimization, game theory, and protocol designs. We will build on preliminary results on bilateral agreements, per flow pricing, wireless access pricing, and combinatorial auctions.

Section 2 provides the background of this work. In Section 3 we comment on the scaling properties of networks. Section 4 describes some preliminary ideas on service differentiation. Section 5 explains some models of revenue sharing mechanisms. Section 6 explores some pricing schemes for wireless access. Section 7 discusses the proposed research.

2 Background

The Internet was designed for providing connectivity. To that end, its protocols assume a minimum functionality from the networks it interconnects: the ability to deliver packets, possibly unreliably, between their nodes. Any additional capability, if needed by applications, should be provided by the end hosts. Although this initial end-to-end design philosophy has long been compromised, the basic best-effort service is still the underlying model.

As a result of this design approach, the Internet is peculiar in the lack of mechanisms for differentiating services and market segmentation. Cable and satellite television have sophisticated tariffs based on the number of channels the customer subscribes to and a growing menu of pay-per-view programs. Similarly, the power energy providers enter in complex tariffs with their main customers; some of these contracts have spot pricing provisions and some include the possibility of cutting some of the load when the system gets saturated. The Internet access providers have simple tariffs that are mostly flat rate or, at best, have a peak load component. There is no possibility to request better service for critical applications. Although service level agreements may have penalties for excessive outages, they do not guarantee any level of service in terms of bandwidth or delays. In fact, they could not since the customers are not able today to verify reliably and simply the bandwidth that their access provider makes available to them.

One critical challenge facing the telecommunications industry is to increase the Internet’s profitability for network service providers. Currently, the network providers have limited economic incentives to invest in technology for value-added services. This situation results in an untapped potential of the network. Two ingredients are essential for improving the situation.

First, the network must implement mechanisms that enable providers to charge services differentially based on their characteristics. Second, the mechanisms must make it possible for the network providers to share the revenues fairly. If they can charge more for better services and collect a fair share of the resulting increased revenues, the network providers have the incentive to provide services that users value more.

3 Scaling

Riding Moore’s law and comparable progress in communication links and compression technologies, the Internet’s best effort service has provided an ever richer experience for its users over the last three decades. This continued progress may lead us to believe that there is no need to modify the basic protocols of the Internet and that the natural evolution of the technology will suffice to improve the network services and meet the desires of users.
Our first research topic is to study the scaling properties of networks. Is it true that performance improves as one increases bandwidth? Or, is it possible that as one increases access speed, the network becomes hopelessly congested?

South Korea and Japan have a much higher penetration of broadband access than the U.S. Not surprisingly, the backbone networks of these countries are becoming congested. One may argue that the backbones can be made faster to keep up with the access. That is particularly true if the backbones are controlled by a governmental agency that understands the value of improving the services. If the backbones are commercial, the incentive for improving them comes from increased fees that the tier-2 networks pay for a higher bandwidth.

Let us explore the technical aspects of scaling. If one could increase the speed of every link in the network by two orders of magnitude, to match an access speed of 100 Mbps, would everything go 100 times faster? We know that this is not the case. For instance, the propagation times do not change and we know that TCP does not work well with large bandwidth-delay paths. Moreover, applications will change to take advantage of the faster access. It is likely that file sizes will increase, so that their delivery time will not get smaller and the users will still experience congestion. The limiting factor may be the patience of users; if they are willing to wait ten seconds, then applications may be designed so that they have to wait ten seconds. (This is not unlike the booting time of operating systems that does not get shorter even though computers get much faster.) Finally, some technological limits prevent everything from going 100 times faster. The per-packet processing time will get smaller with specialized processors, but not 100 times smaller. Also, the 10-Gbps speed per wavelength is not likely to become 100 faster any time soon. As a consequence of these limits, the network will not be uniformly 100 times faster. The higher throughput will result from many parallel links instead of faster links. As we know, 100 parallel queues have the same average delay as a single queue, and they collectively have a 100 times larger backlog.

Summing up, it is not obvious to us how the network characteristics will scale and we propose to examine this question as part of this proposal. We will use simple models of the network and of plausible applications. The scaling question is important because its answer will shed light on the importance of service differentiation and quality of service. If quality of service improves dramatically as the network speed increases, then we should worry about reliability, security, privacy, mobility, ubiquitous computing, telepresence and other new applications. On the other hand, if quality of service remains a persistent problem, then careful designs are needed to avoid limiting the value of the network and the creation of these new applications.

4 Service Differentiation

Different applications require different service characteristics. Many applications work satisfactorily with best effort service. This is not surprising since these applications were designed to use the only service available. Other applications would benefit greatly from faster throughput. Yet others require low latency. A few applications need both low latency and fairly high throughput. Also, some customers value reliability much more than others.

We do not repeat the well understood argument that market segmentation generates more revenues. (See [CW03].) Instead, we comment on aspects of service differentiation that are specific to communication networks.

The first observation is that service differentiation may be simple to implement, at least techni-
Service differentiation matters only at the bottleneck links of the network. These links are mostly in the access network, such as the DSL links, the ATM cloud behind the DSLAMS, the cable that attaches the cable modems to the CMTS, the Wi-Fi or WiMAX access network, and other access links.

One form of service differentiation might be a preferential scheduling at a DSLAM for some users. The provider could offer two services: premium and regular. Premium DSL would be served with a larger average bandwidth per port than regular DSL. This service could be activated on demand. This class of service does not require complex provisioning since it does not make explicit guarantees on the service characteristics. Similar schemes can be used to differentiate services with other access technologies.

The second observation is that the impact of service differentiation on the network needs to be evaluated with care. To illustrate that point consider two users $A$ and $B$ that face a choice of two service classes, say high and low priority. When the two users choose the same service class, their traffic experiences the same average delay $T$. If user $A$ chooses the high-priority and user $B$ the low-priority service, then user $A$ faces a delay $T_1$ and user $B$ a delay $T_2$ with $T_1 < T_0 < T_2$. The situation is symmetric if user $B$ chooses the high-priority and user $A$ the low-priority service.

Assume first that the utility of user $A$ (respectively, $B$) is $f_A(T)$ (respectively, $f_B(T)$) when the delay is $T$. Assume also that the price of high-priority service is $p_1$ and that of the low-priority service is $p_2$ with $p_1 > p_2$. The question is what service will the users choose? (See also [SV03] and [JK02] for related studies of service differentiation.)

We can view the users as players in a game with the following reward matrix (see [FT91]):

<table>
<thead>
<tr>
<th>A;B</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$f_A(T_0) - p_1; f_B(T_0) - p_1$</td>
<td>$f_A(T_1) - p_1; f_B(T_2) - p_2$</td>
</tr>
<tr>
<td>2</td>
<td>$f_A(T_2) - p_2; f_B(T_1) - p_1$</td>
<td>$f_A(T_0) - p_2; f_B(T_0) - p_2$</td>
</tr>
</tbody>
</table>

This matrix shows the rewards of players $A$ and $B$ that correspond to the four possible pair of choices of service class. For instance, the entry $(1, 1)$ of the matrix indicates that when both users choose the high-priority class, user $A$ gets a net reward $f_A(T_0) - p_1$ and user $B$ a net reward $f_B(T_0) - p_1$. The net reward of user $A$ is his utility $f_A(T_0)$ when he experiences a delay equal to $T_0$ minus the price $p_1$ of the service. Similarly, entry $(1, 2)$ corresponds to user $A$ choosing the high-priority service and user $B$ the low-priority one. In that case, user $A$ has a reward $f_A(T_1) - p_1$ because he experiences a delay $T_1$; user $B$ has a reward $f_B(T_2) - p_2$.

Under reasonable assumptions about the delays and the utility functions $f_A$ and $f_B$, surprising situations may occur. For instance, a pure Nash equilibrium may not exist. That is, given any pair of choices, one of the two users is better off by changing its choice. It is also possible that the only Nash equilibrium of this game corresponds to both users choosing the high-priority service. This is surprising because both users then experience the same delay as if they both chose the less expensive low-priority service. This situation corresponds to a prisoner’s dilemma. It occurs when the difference in costs of the two services is not large enough to compensate the increase in utility of a smaller delay.

As you may expect, these conclusions hold for models with a large population of users. We have analyzed a model with a continuum of users with a distribution of utilities. In such a model we find that an unstable desirable equilibrium exists and the stable equilibrium corresponds to an undesirable prisoner’s dilemma.

The conclusion of this analysis is that simple service differentiation mechanisms such as priority...
scheduling exist but, when implemented, they lead to undesirable user behavior. Of course, these conclusions assume that users are strategic. We plan to study likely behaviors of users facing such choices.

Other service differentiation mechanisms do not present the undesirable characteristics discussed above. For instance, one can avoid them by separating the service classes strictly. Thus, if the network reserves a fixed fraction of its bandwidth for premium service and the rest for regular service, these peculiar user behavior will not happen. Unfortunately, such a strict separation is not efficient. (See [098].)

We believe that an economic analysis is a necessary complement to the usual performance evaluation of protocols. We propose to study different schemes for service differentiation and their impact on the user behavior and the resulting profitability of the services. We elaborate on the propose research in Section D.8.

5 Revenue Sharing

Most end-to-end services involve a number of network providers. The general structure of the Internet consists of tier-1 networks (e.g., AT&T, UUNET, Sprint) that span the country and connect a number of tier-2 networks (e.g., SBC, BBN) to which customer networks are attached. A typical connection goes across a customer network (or link), a tier-2 network, a tier-1 network, another tier-2 network, and finally another customer network or link. Roughly, tier-1 networks are backbones and tier-2 networks are access networks. The contracts are between customers and access networks and between access and backbones. The actual situation is complicated by peering agreements between access networks that circumvent payments to backbones (see [BW98], [GH00]). Some customer networks also build peering links, such as the campuses of the University of California, to reduce payments to access networks. (See [CRT00], [LMPRT01], [CW03].)

A first order approximation is that backbones are cheap (per unit bandwidth) compared to access networks. Today the contracts between customers and access and between access and backbone are primitive. They do not support multiple classes of service and have elementary considerations for availability.

We propose to study the economics of revenue sharing among providers and explore whether more sophisticated contracts might create incentives for better services.

Pricing can be used as signaling to regulate the sharing of bottleneck links. In such a scheme, routers mark packets when they become congested and the acknowledgments of marked packets carry that mark back to the sender. The applications are charged a (fictitious) price for each mark. Applications that do not have a sufficient utility for the traffic slow down instead of paying for the marks. Other applications keep on sending packets. A number of studies show that this approach can maximize the sum of user utilities, provided that the connections are long lasting. The key observation is that the shadow price is the same for all the connection on any given link because the inequality constraints are on the sum of the rates. (See [KMT98] and [CW03].)

In our research, we use pricing differently. Instead of being a congestion signaling mechanism, we use pricing as a way to establish a market for the network resources. The prices generate revenues and regulate the user demand. To appreciate the difference, note that even when the network bandwidth is large enough so that there is no bottleneck the prices are non-zero in our
protocols to compensate the network providers for their operating costs. The research questions concern the selection of the prices and the distribution of the resulting revenues.

A good pricing scheme must have a number of properties. First, it should promote market differentiation to increase network revenues. Some users are willing to pay quite a bit more for services that have only a small additional cost (see [JK02], [O98], and [SV03]). Second, the scheme should provide an incentive to all the network providers to supply good services. That is, a provider’s revenues increase when it provides better services. The scheme should penalize cheating. Third, the pricing should encourage network upgrades. For instance, a provider should derive more revenues by increasing the capabilities of its network. Fourth, the scheme must be scalable. It should require a small amount of signaling and few states to be kept by the routers. Fifth, if the mechanism solicits information from users and providers, it should be incentive-compatible.

To date we have explored two different models that we review below.

5.1 Bilateral Agreements

In the first model we focus on a service similar to Voice over IP or Video over IP. In this model users make connections that transport a fixed rate packet stream. The users have a latent demand for such connections. The providers enter in bilateral contracts with users and other providers. The study of this model shows that if providers charge for bandwidth, then the economic market behaves poorly. Technically, it has degenerate equilibriums that may correspond to incorrect incentives and may result in undesirable network characteristics. The intuition is that each provider has a short-term incentive to reduce its downstream bandwidth costs. On the other hand, if the contracts specify quality of service (here in the form of a call blocking probability), then the market is well-behaved. Specifically, there is a unique equilibrium and it has good properties. The intuition for these desirable properties is that a provider suffers a loss of revenues if it reduces its quality of service. Consequently, each provider has an incentive to improve its quality of service, which also increases the revenues of other providers.

Mathematically, we explored a number of version of the game between users and players. In the Nash equilibrium, each provider maximizes its profits by adjusting the bandwidth it contracts to its own providers. We also studied the repeated version of this game. In the Stackelberg equilibrium, the downstream providers select their bandwidth by anticipating the optimal response of the upstream providers. For details, see [HW02].

5.2 Path Pricing of Data Services

In our second model, we study a service similar to TCP with a specified end-to-end delay. The providers achieve this delay objective by limiting the utilization of their links. In this model, the users pay per unit bandwidth and they have some demand that decreases with the unit price. We study a number of models for how the providers can share the network revenues and we show that many of these schemes result in poor behavior. For instance, if the individual providers individually select the price of the packets they transport, then they get more revenues if they are a bottleneck. Accordingly, the providers have no incentive to improve the network and they even have an incentive to artificially limit their throughput. Fortunately, we identify a revenue-sharing scheme that provides the correct incentives. Under this scheme, the providers share the revenues per path equally and the total price is selected to maximize revenues. In addition, we
describe a packet-marking protocol to implement this scheme and we show that it converges to the correct prices.

Mathematically, we studied the Nash, Stackelberg, and cooperative versions of this game. For details, see [HW03].

6 Wireless Access

Two approaches are emerging for ubiquitous wireless access to the Internet. In the first approach, access is via WiFi access points that companies deploy in public places and, possibly, that individual users make available. The second approach is based on the IEEE 802.16, or WiMAX, standard. In this approach, some corporations set up a WiMAX infrastructure that consists of antennas with a range of a few miles that form a cellular-like arrangement. At present it is hard to predict which approach will be ultimately successful or whether a combination of the two might be used.

We propose to explore the economic aspects of these alternative approaches. We describe some preliminary results on the WiFi approach in the sections below. These results concern economic incentives for individual users of WiFi access points to make them available. The idea is that a mobile user could connect to an access point for a fee. The access point would charge micro payments, say per minute, instead of requiring a subscription to some provider. Techniques have been proposed in the literature to make such micro payment schemes secure and easy to implement, so we do not focus on that aspect. Instead we examine how the prices could be calculated. We study that question by modelling the relationship between the mobile user and the access point as a game. An interesting aspect of this game is the asymmetry of information between the players: the mobile user knows more about his utility for the service than the access point. For instance, the user may be interested in browsing the web for fun or he may be using the web to carry out an important business transaction. The access point is unaware of the user’s utility of the transaction. We examine two different situations. In the first, the user is browsing the web and is gaining a utility proportional to the duration of the session. In the second, the user is downloading a file and derives a utility only upon the completion of the download.

6.1 Web Browsing Model

Assume that the user is browsing the web. His intended session duration is \( \tau \) time units. Assume that the actual duration of the connection is equal to \( T \leq \tau \) time units. The access point charges the user \( p_t \) in time slot \( t \). The user’s utility is equal to

\[
UT = (p_1 + \cdots + p_T).
\]

In this expression, the quantity \( U \) is the user’s utility per unit time; it depends on what the user does with the session. We model \( U \) as a random variable that the user knows but the access point knows only its distribution. Similarly, we assume that the user knows \( \tau \) but the the access point knows only its distribution. Finally, we assume that \( \tau \) and \( U \) are independent.

The user and the access point want to maximize their payoff. The strategy of the user is to accept or reject the price \( p_t \) that the access point charges at time step \( t \). In our model, the user walks away the first time that he refuses a price. The strategy of the access point is to calculate prices that maximize the expected value of \( p_1 + \cdots + p_T \).
One might be tempted to guess that the access point learns from the payments that the user is willing to make. Accordingly, one expects that the access point may keep on increasing its prices as it learns that the user is willing to pay increasing amounts. Surprisingly, the analysis shows the unique equilibrium of this game is for the access point to charge a fixed price $p^*$ and for the user to accept the price if it is smaller than $U$ and to walk away otherwise. This pair of strategies is a sub-game perfect equilibrium in the sense discussed earlier, that the strategy-pair remains an equilibrium in every possible continuation of the game, but now with the additional feature that the access point maintains and refines the conditional probability distributions of the random variables that describe the client.

6.2 File Transfer Model

Consider now a variation of the model where the user wants to download a file, which would require a session duration equal to $\tau$. As before, the access point charges $p_t$ during time step $t$. The net utility of the user, if the connection lasts $T$ time units, is equal to

$$\tau 1\{T = \tau\} - (p_1 + \cdots + p_T).$$

In this expression, $1\{T = \tau\}$ takes the value 1 if $T = \tau$ and the value 0 otherwise. That is, the user derives a utility proportional to the duration of the file, as long as the file is completely downloaded.

As in the previous model, we assume that the user knows the value of the random variable $\tau$ whereas the access point only knows its distribution.

The analysis of this game shows that an equilibrium is as follows. The access point gambles that the file download should last a precise duration $t^*$ and charges the amount $t^*$ in time step $t^*$ and nothing prior to that time. The user refuses to pay any amount prior to the actual duration $\tau$ and walks away if the access point charges anything at any step prior to $\tau$. As a result, the access point gets paid only if $\tau = t^*$ and the user gets the file. Otherwise, if $\tau \neq t^*$, the access point and the user both get nothing.

This result is disappointing because it shows that the expected gain of the access point is small. As a result, the owner of the access point may not be willing to take part in this game.

We have explored mixed situations where the user has some probability $\pi$ of using the access to browse the web and to download a file otherwise. Not surprisingly, the equilibrium depends on $\pi$. We plan to explore more realistic scenarios that involve repeat customers, competition among access points, and reputation effects.

7 Proposed Research

7.1 Scaling Properties of Networks

We propose to study models of networks and applications and explore how backlogs, delays, and throughput are likely to change. The models will reflect the technology limitations on per-packet processing times and the rate of optical links. We will study the performance bottlenecks and examine the need for service differentiation as the penetration and speed of broadband access increases and as the applications take advantage of that progress. We believe that the user characteristics, such as willingness to wait for downloads, must be included in the models.
7.2 Behavior of Users Facing Service Choice

As our preliminary investigations show, strategic users may be caught in an undesirable equilibrium. We propose to study service differentiation mechanisms that prevent undesirable equilibriums and yet make an efficient use of the network resources. We also propose to examine how actual users might behave and whether they would be rational as the game theoretic models assume or would act with bounded rationality. We plan to conduct simple experiments to compare the prediction of the theory with actual behavior.

7.3 Behavior of Providers Offering Service Choice

Imagine providers competing for users and offering multiple services. Will some providers specialize in gold service and others in basic service or will they all compete across all service classes?

7.4 Bilateral Agreements Between Providers

Our preliminary investigation of revenue sharing through bilateral agreements indicates the importance of well-designed schemes. Our initial study explored limited models and a number of essential questions require further analysis.

So far we have considered only one class of traffic. This class can be viewed as the gold class, everything else being best effort and treated with lower priority. We have not studied multi-class versions of this scheme, but this is an obvious next step.

Algorithms that converge to a suitable equilibrium are not known yet for general network topologies. We suspect that some side payments may be needed to force convergence. We only have very preliminary results on the adjustment of prices.

We have only studied a simple VoIP model. We plan to explore models of data transfers.

7.5 Path Pricing

We propose to explore scalable pricing protocols. Consider a large network with many providers. Assume we want to divide the price $p$ per packet equally among the $N$ providers that the packet goes through. Here is a possible protocol that avoids the need for pricing tables in the provider routers. The edge router (the client’s provider) keeps a list of ongoing connections. When a packet of a new connection shows up, it marks that packet as new and sets a "provider count" field to one and notes that the new connection ID. Whenever a new packet crosses a new provider, a router of that provider increments the provider count field by one. As the packet reaches the exit provider, the exit router sends back a message to the ingress provider with the connection ID and the provider count. From the time on, subsequent packets are marked as old and the provider count field becomes a price per provider field, where the price is the total price divided by the provider count. The ingress provider adjusts that price to maximize the revenues. Only the ingress routers need to keep track of the connection ID and price information. They have to do so anyway to charge the customers.

We plan to study protocols that support multiple classes of service and combine our preliminary ideas on path pricing with those on service differentiation.
7.6 Competition Among Providers

Consider two independent service providers offering the same kind of Internet access service to a large number of users, except that they may charge different prices. Users choose which provider to use based on their price alone, i.e. the one with the lower price. For simplicity, let us assume that each provider has sufficient capacity to carry all the potential load, but has a per-unit-bandwidth cost \( c_i \) in carrying traffic. The providers choose their prices \( p_i \) to maximize their revenues. Under these assumptions, it is fairly easy to show that the provider with lower cost seizes all the traffic, at a price equal to \( \max\{c_1, c_2\} \). The profit it makes equals \( \max\{c_1, c_2\} - \min\{c_1, c_2\} \), while the other gets nothing.

Although one may argue that this outcome is socially efficient for the users, because it assigns all the traffic to the least-cost route, it is questionable if it is good for the benefits of the service providers. First, notice that the profit at equilibrium depends on the cost difference only and has nothing to do with the demand. So even if the demand is strong, neither provider would be able to exploit that. Moreover, over the long term, one may expect that to reduce cost \( c_i \), this "price war" between the providers could turn into competition in building capacity, so that eventually neither provider is able to make much profit to justify all the investment they make.

A number of schemes could be used to correct this undesirable outcome. One approach, which is well-known in economics, is to have providers offering different tiers of services and segment the market. The question of interest here is whether it is possible to design pricing schemes that can help providers identify and move into their respective "best" market segment, based on their private cost information. Another approach is to design pricing schemes that would allow providers to split the traffic in a mutually beneficial way. For instance, the providers may agree to charge their prices as an increasing function of their load. Since users are free to choose the most economical provider to use, this pricing policy helps the providers reach a common price without coordination. The question to be answered is: 1. what is a fair allocation acceptable to both providers; 2. how to choose the pricing function to reach that equilibrium, preferably without having the providers reveal their private cost information.

7.7 Inter-Domain Routing

Currently there are several backbone providers in the Internet. Since no one has the full connectivity to the entire Internet, they rely on each other to forward the traffic. To facilitate the exchange of traffic, providers of compatible size often forward each other’s traffic for free. However, due to the strategic interest of the providers, such an agreement would lead to possible inefficient routes between the providers, known as "hot-potato routing." This happens because a provider wants to route any transit traffic out of its network as soon as possible, even if that route may not be the most efficient one to take from an end-to-end perspective. It has been shown in [JT03] that the increase in the routing cost due to this selfish forwarding behavior can be as high as three times that of a shortest-path route.

This inefficiency is clearly due to the fact that how traffic is routed downstream has no impact on the routing decision upstream. To rectify this inefficiency, one obvious approach is to apply some form of cost-sharing: the routing cost incurred downstream should be somehow included in the upstream provider’s routing decision. There are two challenges in designing such a cost-sharing policy: 1. what is a fair allocation of revenue between the providers, given their costs? 2. since routing cost is private information to each provider, how to design a mechanism which can
incentivize the providers to reveal their true cost information, in order to allocate the revenue fairly?

7.8 Strategies for Ubiquitous Wireless Access

Making an access point available to other users makes possible to derive direct revenues. More importantly, if this approach is widely adopted, it creates an ubiquitous access to the Internet. One may argue that an access provider should not allow such open access. On the other hand, such open access would increase the network usage and appropriate pricing schemes could guarantee that everyone is better off, including the providers.

We propose to study more complete models of such access schemes, including competition for access, arriving and departing customers, and the effects of reputation and of longer term contracts. We also plan to study the economics of WiMAX access as it may compete with wired access and offer mobility benefits.

7.9 Economics of Reliability

A number of companies provide simple schemes for alternate routing to increase the reliability of Internet services. As can be expected, the increased reliability that users of such schemes experience comes at the cost of excess congestion for other users. Accordingly, it might make sense to price differently services with an increased reliability and to reduce the price of unreliable services.

The resulting situation is similar to that of service differentiation and it may happen that the resulting market has poor equilibriums or even no equilibrium. We propose to model the economic incentives of services with increased reliability.

7.10 Lessons for 100 × 100 Architecture

The ‘green field’ study of 100 × 100 provides a rare opportunity to incorporate economic mechanisms in the network protocols without the constraints of backward compatibility. We will examine the scaling effects and the economics aspects of service differentiation and revenue sharing for the proposed architecture of that network. We will compare bilateral agreements and path pricing for that architectures.

The architecture of 100 × 100 has the advantage of being highly symmetric and ‘clean.’ For instance, the congestion control mechanisms of the access network introduce constraints on the rates that suggest specific pricing schemes. Also, some early versions of that architecture are simple enough to enable a concrete scaling study. An important question for this architecture is how it would support competition in the access and in the backbone and the economics of that competition.

7.11 Lessons for Internet

The Internet architecture is much more complicated than that of 100 × 100. Many researchers have come to the conclusion that the only viable approach to introduce new services is via a combination of application-layer overlay and middle boxes. We want to model the economic aspects
of such overlay and middle box services. Intuition suggests that they introduce a competition for resources that, if not well managed, may result in poor network behavior. As in the case of service differentiation and increased reliability, the poor behavior does not result from poorly designed protocols. Rather, it results from the unexpected interactions of strategic users that all attempt to improve their own utility. That is, by providing more control to the users, these overlay and middle box services enable them to be more strategic and, consequently, to move to a Nash equilibrium instead of social optimum. Unfortunately, in realistic models of congestion, the price of anarchy can be very large. We believe that these questions deserve some careful study.

8 Results of Prior NSF Support

8.1 Pravin Varaiya

Award No: 0099824; Amount $180,000; Date: 08/15/2002; Duration: 36 months

Title: Hierarchical Systems

Summary: For the past three years Kurzhanski and Varaiya have developed techniques for approximating the reach sets of dynamical systems. The basic idea is this. Consider the differential equation

\[ \dot{x}(t) = f(x(t), u(t), v(t)), x(0) \in X_0, \]

where \( x(t) \in \mathbb{R}^n \) is the state, \( u(t) \in U \) is the control, \( v(t) \in V \) is the disturbance, and \( X_0 \) is the set of possible initial states. Let \( X(t, X_0) \) be the set of states that can be reached at time \( t \), by choosing an appropriate control, independent of the disturbance. The problem is to calculate (an approximation) of the reach set \( X(t, X_0) \). This calculation is needed in a variety of applications.

Our approach to this problem is to convert this problem into an optimization problem, show that the value function of this optimization problem—call it \( V(x, t) \)—satisfies a forward partial differential equation, similar in form to the Hamilton-Jacobi-Bellman (HJB) equation, and is such that \( X(t, X_0) = \{ x \mid V(x, t) \leq 0 \} \), i.e. it is a level set of the value function.

It is computationally very difficult to solve the HJB equation directly. We have focused on the special case of linear differential equations (the function \( f \) in (1) is linear), and sought to approximate the reach set externally and internally by ellipsoids. The linear case makes it possible to use powerful convex analysis methods, combined with an “ellipsoidal calculus” that we have invented, to produce computationally tractable algorithms. Paper [3] describes how to handle uncertainty, [1] describes the optimization approach, [2] shows how we obtain internal approximations, and [4] gives the full story. Conference papers [5,6,9] work out the the difficult cases of state space constraints and hybrid or switched systems.

8.2 Jean Walrand

Program Title: WebTP : A User-Centered Web Transfer Protocol


This project demonstrates a working implementation of a new protocol that is aware of user preferences. The protocol is receiver driven and uses application level framing. The activity of this project can be found at http://webtp.eecs.berkeley.edu. Work done includes research on the design of user-centric optimization of protocols, possible improvements of TCP, and architecture for web-optimized transmission protocols. The students supported by this grant were Linhai He, Jeonghoon Mo, and Wilson So.


Online Statistics for ATM Networks

NCR-9628818; July 1, 1996 to June 30, 1999

The objective of this research is to clarify the possibilities and methods of control of high-speed networks. These networks include not only the asynchronous transfer mode networks (ATM) but also high-speed networks based on the TCP/IP protocols. The high-speed creates new problems related to the large rate-delay product that tends to make traditional window-based congestion control mechanisms ineffective.

The control of high-speed networks poses new challenges. The large rate-delay product of the connections makes the usual end-to-end window-based congestion control mechanisms less
effective than in lower speed networks. Intuitively, the large rate-delay product requires a large window size for a retransmission protocol to be efficient. However, the source cannot control where the large number of packets in its window are stored in the network. It may happen that most packets pile up in one router which is then congested. This intuitive argument suggests that a more effective control mechanism might be to pace the transmissions of packets with rate control mechanisms such as leaky buckets, as is recommended for ATM networks. In addition to congestion control, we study call admission control, which is needed if quality of service is guaranteed instead of being provided on a best-effort basis.

The following students were supported in part by this award: I. Hsu, S. Papadakis, G. de Veciana, N. McKeown.

