John F. Kennedy School of Government
Harvard University
Faculty Research Working Papers Series

Universal Service in a Ubiquitous Digital Network

Jean Camp and Rose Tsang

February 2001

RWP01-006

Universal Service in a Ubiquitous Digital Network

L. Jean Camp, Jean_Camp@harvard.edu
Kennedy School of Government
L213
Harvard University
79 JFK St.
Cambridge MA, 02138

Rose P. Tsang, rtsang@ca.sandia.gov
Senior Member of Technical Staff
MS 9011
Sandia National Laboratories
PO Box 969
Livermore, California 94551

Abstract

Before there was the digital divide there was the analog divide -- and universal service was the attempt to close that analog divide. Universal service is becoming ever more complex in terms of regulatory design as it becomes the digital divide. In order to evaluate the promise of the next generation Internet with respect to the digital divide this work looks backwards as well as forwards in time. By evaluating why previous universal service mechanisms failed and succeeded this work identifies specific characteristics of communications systems -- in particular in billing and managing uncertainty -- and argues that these characteristics underlie success or failure in terms of technological ubiquity.

Developing a set of characteristics of services rather than a set of services is a fundamental break with the tradition of universal service. In fact, the implications of our proposal is that basic characteristics in the offering of the service rather than the absolute price are critical to close the digital divide: certainty of total charge, ability to avoid deposits or disconnection via best effort service, and payer-based control of all charges. While all of these principles sound obvious in fact none of these hold in the telephony network.

Universal service has evolved from common carriage (serve all with no discrimination) to a right to basic services (100% penetration). Universal service is now discussed as the digital divide, as the access to information as opposed to services becomes increasingly critical. However, we are discussing in this paper access to the bits and the network rather than access to the information (or intellectual property) once connected.

The provision of universal service is seen as technical problem only in that the technology costs money -- universal service debates have long been the domain of economists. Yet the design of protocols has been the domain of engineers, the building of systems the corporate domain, and the discussion of equity the interest of ethicists. The design of protocols can define the parameters of the corporate decision-makers, the variables of the economist, and the questions for the ethicist. The design decisions made at the fundamental levels can make communications equity more or less likely.

In this work I focus on the design of protocols for the next generation Internet, protocols which will fundamentally change the best-effort nature of Internet services. Building on the economic and ethnographic work of others I argue that the effects of protocols adoption on universal service can be predicted to some degree. By examination of past and current technologies I examine a set of technical mechanisms to determine how such mechanisms might harm or enhance universal service. I define each mechanism (e.g. denial of entry) and offer observations about each particular mechanism's implicit pricing assumptions. I close with a discussion of interest to ethicists and regulators on evaluating communications protocols with respect to universal access.

Protocols for developing multiple qualities of service for packet-switched networks have focused on economic efficiency (e.g. Mackie-Mason 1995; Choi, Stahl, & Winston, 1997; Shapiro & Varian, 1998), billing to encourage widespread adoption of network innovations (e.g. Xie & Sirbu 1985) and billing in a manner consistent with the underlying network (e.g. Clark, 1996). Here we examine a set of protocols which include varying quality of service mechanisms with respect to the compatibility of the protocols with universal access.

1. Introduction

The Internet and telephony were traditionally based upon very disparate engineering technologies (i.e., packet switching and circuit switching, respectively) as well as very different types of service offerings. The Internet has used best-effort data traffic for email and file transfers, and telephony has used guaranteed low bit-rate circuits for voice traffic. Best effort service means that all communications as treated exactly alike, and the failure or slowdown of communications is distributed through random statistical mechanisms. The convergence of these networks has resulted in the proposal of packet switched networks containing protocols which provide quality of service (QoS) mechanisms.

The economics of the resulting packet network will violate the fundamental assumptions on which universal service funding is built. Thus a new consideration of universal service is needed for next generation networks.

What is needed is a universal service model for a packet-switched network which supports both best effort traffic and traffic with guaranteed quality of service. It would not be ideal to have a new paradigm for universal service based on best-effort packet-switched networks developed just in time to be obsolete in the face of next generation networking.

In order to develop guidelines for such a universal access model, we examine the protocols proposed to create an Internet with quality of service guarantees. We begin by describing telephony's mechanisms for funding universal service through inter-customer, inter-company and intra-company payments. Historically, a constant quality of service has been in conflict with universal service (Mueller, 1993; Fischer, 1992). We argue that the future problems with universal service will not be the simple penetration rates problems of today, but will be more subtle issues of availability and quality of service. (The penetration rate is the percentage of households with POTS. Federal statistics are available with penetration rate broken down by income, race, age, geography, and family structure.) Reasons for loss of connectivity are primarily that toll pricing that is perceived as random, and an initial deposit is required. We discuss these and additional, secondary issues for loss of service and hypothesize how advanced services could help, or aggravate, the underlying problems. We argue that the critical components to support universal are end user feedback, entry controls, and support of both best-effort traffic as well as guaranteed service traffic. We show that the current regulatory mechanisms are inadequate for a competitive market where a continuous variable, quality of service, not a Boolean variable, existence of service, is key. We argue for a new conception of universal service.

2. Universal Service & Universal Access

Universal service as phrase was first used in 1907 by then-President of AT&T Theodore Vail (American Telephone & Telegraph, 1907). At the time universal service meant what is today meant by common carriage: interoperability and interconnection between competing networks. Universal service initially required AT&T to interconnect with the small independent telephone companies which AT&T had been purchasing after refusing interconnection to the independent subscribers. All other nations had or were in the process of nationalizing their telephone systems as an extension of the postal system, so-called postalization. In contrast, in the United States, despite a short interlude of government ownership during World War I AT&T became a regulated monopoly and remained one for decades (Gordon, 1995).

Universal service has been the regulatory and philosophical twin to natural monopoly, which dictated that there be only one dominant voice network provider. The meaning of the phrase universal has grown through the first three quarters of the last century to include ubiquitous access to voice service. By the nineteen fifties there was a system of settlements and separations, which by the seventies, was used to subsidize residential service using surcharges on commercial service (Brooks, 1975). Today, universal service is targeted at two groups: rural subscribers who may be unable to support the infrastructure necessary to provide service, and the urban poor. Wireless technologies have made the issue of wired infrastructure, and thus considerations of distance, increasingly unnecessary. As the politics of universal service are not at issue here, for the remainder of this work we will focus on the poor and not consider distance issues.

In fact, AT&T has never been the only phone company in the United States. Even after the creation of the FCC small rural phone companies existed. There is some argument over whether universal service was more effective pursued under a regulated monopoly or with competition. There is no debate that universal service with a single quality of service did not exist until after implemented by AT&T. Rural cooperatives had low qualities of service both in the times of provision and signal quality. In a farmer-owner cooperative the owners ran the switchboard in between other responsibilities, line repairs were made when possible, and phone lines were often wire fencing doubling as a transmission medium (Fischer, 1992). When independents were free to develop their own services, services with vastly different quality offerings existed as the demise of the fundamental patents of Bell allowed all to enter the market. Thus quality of service has historically been inversely related to the widespread adoption of services.

Today the universal service fund accepts payment from long distance service providers and selected residential services to subsidize connections for qualifying populations. Initially universal service funds went to all rural households, now households must qualify for universal service fund-supported rates. The universal service fund is actually a misnomer. Rather than a fund, universal service is supported by a series of exchanges and payments made between carriers. Long distance carriers pay local exchange carriers, with the assumption that long distance calls are dominated by the affluent subscribers who can afford the subsidy. Payments between companies are called separations. Payments within a company -- reimbursement of one division from the value-added services of another -- are called settlements. The payments subsidize basic services, qualifying individual households, and qualifying institutions.

2.1 An Overview of Universal Access Mechanisms

By considering universal service as one part of a analog divide, as opposed to the specific mechanism, encourages the consideration of other efforts to end various 'analog divides'. In this section we discuss the various divides created by communications technologies and how those Radio and television have not been subject to universal service in the United States; however, this is not true for all nations. For example, in the Soviet Union and China instead of having individual radios, loudspeaker systems piped news to all common areas. In West Germany everyone has the right to a radio for the public safety purpose of emergency planning and response. Yet in the United States universal service effectively continues to mean plain old telephone service with no subsidy for cable penetration, wireless penetration, or even availability of pay phones. However, a range of mechanisms dating from the Constitution have been implemented to increase access to communications technologies.

First, consider the analog divide created by postal service. Support for postal service, particularly post roads, was written into the Constitution (Article 1, Section 8, Clause 7), and Congress was given the responsibility to build post roads. Delivery of specific letters was paid for individually but the cost of offices and roads was paid for from general taxation and road tolls. Postage stamps and rates were well defined by 1799. In the intervening centuries postal rates have become increasingly simple and affordable. Subsidized book and magazine rates also provide for the extension of text media into homes across the country.

Of course, the adoption of mandatory education across the nation is a critical part of equal access to printed matter. Yet the politics, religious conflicts, and reasoning underlying widespread public education is far to complex and broad to be fairly classified as a universal service regime for literacy.

Second, consider widespread access to video programming. Cable companies are required to offer a minimal affordable service, and are bound by must-carry rules as defined in the 1992 Cable Act (Cable Television Consumer Protection and Competition Act of 1992, Pub. L. 102-385, 106 Stat. 1460). Must carry rules as mandated by the Cable Act (following extended court battles over a similar FCC ruling) require cable operators to carry channels of their competitors. Other common carrier type rules imposed on cable operators have included public access, leased channel rules, the fairness doctrine, and equal time rules.(Thorne, Huber, & Kellog, 1995)

Cable companies must comply with local contractual rules as well as federal law. The local authorities which grant cable franchises can require that public, educational and governmental ("PEG") channels be carried. In addition it is possible to change the dialogue from one about the number of channels to a dialogue based on the bandwidth reserved as cable systems become packetized.(Horwood, 1998) Thus there have been both universal service requirements in the form of common carrier requirements and limitations on pricing in the governance of cable systems.

The universal service fund has recently been expanded to include connecting schools and libraries to the Internet with the E-Rate program (NCES, 1998), which provides 20-90% discounts on telecommunications services.

Broadcast companies have been subject to equal access requirements in the past, which are a form of common carrier regulation.

In short, regulation of communications technologies and rates for the purpose of transmitting information across the nation has existed from the authoring of the Constitution to the creation of local cable franchise authority (usually held by a city or town government). The most common controls mandate that the controllers of technical bottlenecks provide common carriage (e.g. interoperability) and rate controls. As communication becomes more affordable, consumers prefer simple pricing mechanisms. Thus the importance of absolute rates over the long terms are questionable as shown by the examples of postal service, telephony and telegraph services (Odlyzko, 1999).

2.2 Failures of Universal Service

The protocols selected for an information infrastructure may determine the availability of information technology. To develop hypothesis of why such systems would fail we examine the failures in previous universal service regimes. Why do Americans lose their phone service? What functions of protocols could support or undermine universal service? Why don't all Americans who are passed by cable television subscribe?

Universal service has traditionally been based upon two fundamental assumptions: that there exists a single service to which all are entitled and that providing this service is a function of geography. The single service to which all are entitled, plain old telephone service (POTS), is local voice access bundled with traditional toll service (i.e., long distance service). The first assumption is increasingly meaningless, as telephony and data networking merge. The second assumption is flawed in cases the infrastructure investment has been recovered, or is wireless.

The bundling assumption is an increasingly irrational selection of exactly two data services out of the many. Consider, for example, that text-based access to governmental data and job openings may be more valuable than voice toll service. Yet, despite its flaws, any universal access regime will be built upon the voice-based universal service regime. While the technological assumptions of universal service are dated, the importance of citizen access to information and communications are increasingly important. With the rise of xDSL and the continued popularity of telephone-based modems, telephony's universal service may be the backbone of the information age's universal access.

The most complete study of universal service losses to date (Mueller & Schement, 1996) showed that loss of phone service can be predicted by income, ownership of assets, and age of the head of household. In fact age of head of household predicts that young families are the most likely not to have phone service. Universal service has failed poor families users in urban areas; the young more than the old; and the native speaker less often than those who speak English as a second language.

The primary reason for loss of service is a high toll call bill. Toll call costs are unpredictable unless the caller understands the charging mechanism and can control all the calls made by all members of the household. Local call charges are billed periodically, not usage-based, and provide monthly feedback about costs. Tolls calls are increasing flat rate, usage-based and continue the historical pattern of providing monthly feedback about pricing. Historically toll call rates have been based on predicted congestion and have been priced by time of day. Toll calls were also historically priced according to cost of the infrastructure, and thus vary by distance. While time charges are not uncommon (e.g. 5 cent Sundays) telephony pricing mechanisms are becoming increasing based only on duration of the connection.

Other common reasons for loss of phone service are large calling card or collect call charges. In both cases the charges can be made by parties who are not responsible for the phone bill. In the case of collect calls and calling cards the cards may be for services provider by a different provider than the one selected by the party responsible for primary phone service. This results in a remarkable lack of price predictability. All these observations suggest that mechanisms which reduce uncertainty will reduce the incidence of loss of phone services. Mechanisms which reduce uncertainty can either create a constant bill for service, thereby reducing variability, or provide real time feedback about costs.

An unpredictable feature of loss of phone use is that the poor use more expensive telecommunications services per dollar of income than users in any other quintile. Hispanic and African American households spend more on cable TV, long distance, and advanced services (e.g. call waiting) than white households. Consumption of expensive telecommunications options is another not uncommon reason for disconnection. (Mueller & Schement, 1996)

Consider that homes without telephones often have cable television, even premium channels. Thus subscribers can manage their cable bills even though overwhelmed by the phone bill, as evident by disconnection of phone service. Cable television provides valuable services, including new programming. In neighborhoods where children are effectively under house arrest due to high levels of street violence and open drug trades, entertainment may provide a more important value than connectivity. Some proposals would link not only tools and local service, but also broadband entertainment access.

Examining the loss of telephony service and use of cable services suggests that to limit overall phone bills would serve universal service. Note that the same groups under-served by credit markets are conversely over-served by telecommunications markets, which function on the credit model. Other features which might characterize protocols compatible with universal service include unbundling entertainment from calling, concentrating control in the hands of the party responsible for the phone bill, and protocols with a low information threshold. These would support universal service regardless of the technical mechanism used.

Concentration of control in the hands of the party responsible for the bill will increase communications penetration. Calling cards, long distance calls and collect calls enable anyone with access to the physical device to charge calls; thus enabling adolescents, irresponsible guests and relatives to burden the phone owner with uncontrollable charges.

A low information threshold has not been previously identified as a driver of universal service. However, interviews with those who have lost phone service illustrate that there is the conception of uncertainty in billing.

A previous work on universal service characteristics focused upon the structure of the regulatory regime (Gillett, 1994). This work argued for a regulatory approach to universal service that is compatible with this set of technological specifications in that it argues for making explicit subsidies available directly to low-income users, not defining immature services as essential, ensuring competition, and, finally, favoring technologies that are digital, scalable, and extensible. In contrast we argue that digital, extensible and scalable technologies are not all created equal.

In fact we propose that there exist specific characteristics for quality of service mechanisms which can undermine universal service, or conversely can increase the connectivity of the poor. Furthermore, Mueller & Schement study showed that the services targeted specifically for the poor are not reaching poor young families. Thus the systematic causes of denial of service in a QoS network need to be addressed at a fundamental level in a network which provides both best effort and QoS transmission.

3. Hypotheses of Technical Characteristics and Universal Service

The previous discussion provides suggestions about what mechanisms may or may not serve to enhance universal service. Here we first provide simple descriptions of theses technical mechanisms and then argue that such mechanisms support universal service.

Depending on the underlying economics as built in the network, the network may be more or less accessible to families with either consistently low or uncertain means. Companies can provide highly affordable best effort service at a consistent price. This would level network usage in a real time manner and remove consumer uncertainty. Today companies use rough measures -- providing price incentives in time periods other than peak. This creates uncertainty for less informed consumers. Companies build for peak to ensure that highly profitable corporate data does not encounter congestion. Best effort service could provide affordable, constant rate, toll access off-peak. There would be short term denial of service as opposed to long term loss of service. Publicly available AT&T data suggests that this denial of service would occur only between 9-10 am weekday mornings. Currently companies can, and some do, offer a maximum monthly charge yet the customer can neither negotiate this maximum with any degree of granularity nor negotiate it periodically. The ability to control one's own bill would remove uncertainty and provide families the ability to control their monthly expenditure on communications services. All of these services and changes are possible with any set of protocols which provide end user feedback, entry controls, and support of best-effort traffic.

Alternatively companies could focus on expensive guaranteed delivery services, and high-end products. Billing may be built in to the system in such a manner as appear uncontrollable to the end user. Large deposits may be required to protect the companies, thus locking many off the network and into an information blackout.

3.1 Best Effort Service

A fundamental difference between the Internet and traditional media types is the allocation of resources. With print, telephony , and broadcast there is a given resource (pages, bandwidth, and spectrum) and access to this resource is provided in a predetermined manner. In telephony the access is provided based on order of arrival -- that is first come, first served. In print and broadcast there are far tighter constraints and thus systems which far exceed the concept of mere bureaucracy to control the final decisions.

On the Internet the constraint is bandwidth. However, the bandwidth is used with orders of magnitude more efficiency than in other media because of statistical sharing. Return tot he phone example. (Of course these days phone calls are mostly packet-switched as is the Internet but still the telephone system is the canonical circuit-switched system.) A system which is circuit switched gives every connection or calls its own channel. And the connection or call holds that channel even when it is not in use. In contrast, packet switched networks use statistical sharing meaning that no single call 'owns' a channel.

Thus when all the channels of the phone company are in use (likely to happen only on Mother's Day in the United States) The next people who try are denied service, that is they cannot make calls because all the channels are taken. In contrast when a packet-switched network is filled (called congestion in both cases) and more people enter the network service slows down for everyone. That is, instead of denying an entire connection every previous connection is given a lower bandwidth. Notice the marginal cost of a packet delivered on a packetized network is zero until the network is congested. When the network is congested there is a cost, of other's time, of packets dropped and networks slowed. But transmitting n+1 packets cost the same as transmitting n packets when the network is not congested.

Thus there is the same amount of bandwidth in each case what differs is how the bandwidth is shared. Now quality of service mechanisms are being developed which will allow both for reserved bandwidth -- like a circuit or telephone call -- and best effort service -- like the original packet-switched design.

Once the network has been built the cost of sending one additional packet on a network which is not congested is zero. One could easily imagine a system whereby all the capacity was reserved in circuit-like segments. However, these would not be true circuits, just reserved bandwidth. Thus others could use the bandwidth at no cost as long as the individual who was paying to have the bandwidth could call upon the total reserved with no notice. Alternatively the bandwidth could go unused when the individuals who had demanded bandwidth reservation had no demand.

3.2 Real-time Feedback

Consider billing as economic feedback. Phone calls are billed monthly, thus providing periodic pricing feedback. The decision to accept a connection or make specific call is not directly linked to the payment for that call.

Telephone service providers use explicit, delayed feedback to influence user behavior. Through pricing, providers attempt to offer only as much as their networks will bear. Telephone companies can provide real-time feedback to users about service availability; for example, through denial of service as soon as a specific call reaches a given cost threshold, or at some user-selected monthly total charge

Traditionally telephone services have attempted to provide real time feedback about pricing by having differential pricing based on distance called, time of day, and call duration. Yet these charges are invisible to the user until the bill arrives. Once a bill arrives it may be too late for the economically disadvantaged user. Through pricing differentials companies attempt to level network use, and succeed in increasing pricing uncertainty.

Real time feedback can consist of real time pricing feedback. Often feedback means notifying the user that a particular connection is congested (by setting a flag on the transmitted packet or dropping a packet). However useful this may be to the transport mechanism this is not particularly useful to the consumer. Real time feedback can provide pricing information during a connection or purchase, or provide a running tally of all charges made to an account.

3.3 Differentiation of Applications

The previous media types were based on differentiation of application: video, voice, and print. Convergence has created not only new applications (on-line shopping, chat) but also placed all media on the same wires.

The positive aspect of the unification of media type is that users cannot be denied one type of information. For example, because the Internet provides both streaming audio and text one need not choose between the morning paper and a radio. Conversely, if a newspaper subscription is canceled for lack of payment, radio reception is no affected. The threat inherent in the promise of convergence is that loss of a single point of connectivity can place a family in an information blackout, where job listings are not available because too much was spent on pay-per-view.

One way to address this problem is to never end connectivity -- only to allow the degradation of service to best effort but never allowing absolute disconnection. A second way to address this is to attempt to unbundle the applications so that each is a separate 'account'. Of course, the weakness of this solution is that one bill would then become many, conversely creating the uncertainty which has been identified as the core problem.

However, consider that bundling is the largest current source of disconnection. In effect, local phone companies serve as collection agents for long distance companies. The bundling cause very different types of services: local phone calls for jobs, services and delivery; toll-free calls for information government services, and goods; and toll calls are bundled together. Thus the inability to pay for toll calls removes local service, despite the fact that it is quite easy to separate the two in the network.

Similar bundling of distinct services could become far worse as the technical convergence is associated with a series of mergers of companies in previously distinct business lines. An ability to differentiate between applications or qualities of service would enable connection of one type even after another type of connection has been unpaid.

The differentiation by application is a discernible element of a protocol. If applications can be differentiated then such applications could be disabled singularly, rather than disabling all services and applications from one unpaid bill. Current mergers and the linkage between local service providers and ISP increases the potential for bundling of services so that all connectivity may be lost from one service bill unpaid.

3.4 Denial of Entry

Denial of entry is a technical term meaning that a specific request for transmission is denied. In current networks denial of entry usually means disconnection - telephone or cable service is terminated and the entire process of reconnection must be repeated communications to occur. By requiring deposits, they limit their exposure to customer abuse.

For example, the need for a deposit illustrates how telephone companies' needs to limit risk and the consumers' need for service have been in conflict. Quality of service and bandwidth reservation technologies combined with best effort service can resolve some of these conflicts. The potential for denial of service at the entry point would remove the need for a deposit. A customer could quantify an acceptable bill, and for each billing period the appropriate service would be discontinued as soon as the amount was reached. In today's networks, for example, all toll calls would be disabled as soon as a monthly maximum were met.

3.5 Information Overload

Information overload is the least quantifiable measure. Information overload refers to the idea that too much data or data in an inappropriate format (for example, if this paper were in Aramaic) are as useless as no data. Information overload simply means that the flow of data are too great to be absorbed. Information overload occurs at different thresholds for different people, in different contexts, and on different subjects. In this case I refer to the existence of too much real time feedback, or feedback that is simply at a level of detail or abstraction so that it is not useful.

3.5 Billed Party Authentication

Authentication is determination that a particular request is allowable. Authentication for billing purposes means verification that the person attempting to create a debt has a right to do so.

Currently phones are device-authenticated. This means that access to the device (the phone) allows one to make charges to the party who authorized the device. At an extreme example, a criminal could break into another's house and make phone calls. The device would verify the criminal as authorized.

In contrast Web sites use both device and user-based mechanisms for authentication. Users supply pass phrases and individual email accounts to authorize charges. The use of cookies to store passwords changes this personal authentication to device authentication.

Without authentication at the personal level, as opposed to the machine level, (the current use of cookies is an obvious example) heads of household have higher levels of financial exposure.

4. Providing Quality of Service

In the previous section, we identified mechanisms which would facilitate universal service and thus push the digital divide towards closure. . Such mechanisms include: network support for best-effort services as well as toll services, realtime feedback about price, entry controls, authentication of the responsible billing party before charges are accepted, unbundling of applications, and low information thresholds for the user.

It is accepted in many research circles that phone calls over the Internet, or any packet-based networks, require a higher level of service than email. This is because phone calls occur in the real time, rather than happening at any time. For example, if two parts of an email or a few seconds apart the email still arrives whole. However, if two parts of a sentence are separated or arrive out of synch in a voice conversation the result is garbled. The protocols that would provide better certainty in terms of having packets arrive on time and in order are called quality of service (QoS) mechanisms.

In this section we discuss three QoS network protocols in light of their potential effect on universal service. The first two schemes perform dynamic bandwidth allocation at the packet level. The third scheme performs dynamic bandwidth allocation at the flow level, i.e., it provides connection-oriented resource reservation rather than packet-oriented.

4.1 Expected Capacity Service

The Expected Capacity Service is a two level priority scheme proposed in (Clark, 1996). It provides a guaranteed minimum capacity service with a guaranteed burst size allowance provided by a token bucket. Clark considers the scheme in terms of a simple sender-pays model or a more complex receiver-pays model. The sender-pays model is as follows. Each user selects a profile. The profile consists of two measures: minimum rate and token bucket depth (comparable to TCP's window size.) In response to the user's selected profile, the service provider offers a price for the time period (e.g. a month) in which the profile will be enforced. Thus users are provided a fixed and predictable price for their selected profile.

Users make connections and use network services without pricing each individual connection. Pricing feedback should be directly available as the user is selecting the profile. At the edge of the network, the service provider tags each packet as either being "in" or "out" of the user's profile. At a congested network switch, packets tagged as "out" will be dropped. If no congestion occurs, packets tagged as "out" may be delivered thus providing the user with more effective bandwidth than their selected profile specifies. At non-peak hours, this effectively allows users to consume resources, in a best-effort manner, irrespective of their selected profile without penalizing the performance of other users. This provides a potential mechanism for the distribution of network usage amenable to universal service.

Notice having the receiver pay in all conditions means the sender must know of the receiver's usage profile to properly label packets. For example, if the initial sender requests a web page, the responding server would be required to know the usage profile of the requester to know how to tag the requested item for the receiver-pays model to serve the client's request. To overcome this Clark suggests that the request be sent with the usage profile, so that packets may be tagged appropriately. To send all complexities of possible profile requires active or executable usage profiles. As the author note, this complicates the scheme remarkably.

This model for pricing the Internet provides the possibility for real time user feedback. As packets get dropped or slowed the users can be notified. If the user is not notified, he or she may perceive the slowed service and may understand this as indirect notification. The addition of explicit user notification that a connection has changed priority and the reason for the change (total bandwidth out of profile, use at this time not in profile, etc.) would add one mechanism (real time feedback) hypothesized as being supportive of universal service. In contrast, this would add complexity and introduce the possibility of information overload.

Authentication could allow only the party responsible for the bill to alter the profile, thus other parties cannot increase the bill. As the profile is altered periodically and then set for all device users for a given period, this protocol is amenable to user, as opposed to device, authentication.

4.2 Dynamic Packet Auctions

In (Mackie-Mason and Varian, 1995) a Smart Market for responsive pricing in the Internet is proposed. In this model the user attaches to each packet a bid or amount that the user is willing to pay to get a specific packet through the network. When packets are dropped those packets with the lowest bids are dropped first. Each packet with a higher bid is charged the amount of the lowest bid submitted into the network; i.e. for each packet, the user is charged a threshold that is less than or equal to the amount bid. In the Smart Market approach, pricing is explicitly identified as a form of economic, as opposed to network, feedback. In this, we certainly agree with the authors.

The authors include in the discussion of the Smart Market the proposal that some users can choose to pay nothing and receive only best effort service only. The authors further note that they expect the smart market to be implemented with software so that users do not have to bid on every packet - the authors suggest a possibility that the users halt sending when the price goes above, for example, $0.0001 per packet.

By implementing the packet auctioning as an active response (not controlled by the user at the packet level) which observes user selections and adapts within user controls, this protocol could serve as effectively as the previously described Expected Capacity Service. In this case an agent would search for optimal routes and bids, and the total monthly bill would be some predictable amount. This would also remove the high user information overhead.

Conversely, when implemented with an active user those people bidding must understand the probability of congestion occurrences, and the bandwidth demands their applications make on the network. Users would not know how much money was spent to complete a connection until after the connection is complete. Two transactions which seem identical to a user could have very different costs. Issues of receiver-pays mechanisms are not addressed. Again when the case is that a sender pays only for the request, while the party contacted pays to transmit an entire Web page, this mechanism may prove flawed.

This protocol is subject to information overload. Note that the need for the head of household to be able to control charges would require authentication for every packet initiated. If the total bill could be periodically set, as with a profile, then occasional user authentication would be adequate. This could be implemented with session authentication, where each user can spend a certain amount. Session authentication and user-specific pricing would increase information overload.

4.3 RSVP

RSVP (Zhang et al, 1993), Resource reSerVation Protocol, is an protocol designed for IP-based networks for setting up end-to-end QoS across a heterogeneous network. This means that the design of RSVP is ideal for telephone calls and viewing real time broadcast (e.g. television).

RSVP assumes that in addition to toll RSVP traffic there exists continuous best-effort traffic. Thus the core mechanism (use of best effort) hypothesized to support universal service is assumed in RSVP.

In RSVP resource reservation is based upon flows, i.e., streams of data from a source to a particular destination. Each flow (e.g. a single voice conversation would have two flows, one to and one from) has an associated specification (flowspec) which contains information about the QoS. The variance of these specifications would determine not only the specific resources reserved for the flow but also may determine in some configuration the price. When a user requests a flow setup (e.g. makes a call, orders a movie), the flowspec is propagated to the appropriate branch in the multicast tree where, if each intermediate router reserves the requested resources for the flow's QoS, the request is accepted by the tree. Once a flow is setup it corresponds to a logical dedicated circuit; the flow's QoS are guaranteed until the circuit is explicitly torn down, regardless of potential congestion. This means that RSVP can make the Internet act like the telephone network.

The initial motivation for RSVP was to support more efficient multicast operations. Thus it is based upon a receiver-initiated request model. In more traditional (non receiver-initiated) multicast protocols, a sender would have to know the QoS requirements of all the receivers in their group (e.g., receiver A requests 128 kbps, receiver B requests 2 Mbps, etc.). Clearly, this would require an exchange of a large amount of information between the sender and all receivers, as well as the maintenance of a large amount of dynamic state information. Thus RSVP, being receiver-initiated, is naturally amenable to the receiver-pays model. In the previous configurations there was the problem of having a receiver pay for requested resources. Now consider the reverse of the situation: the user sends a request and pays for the transmission of a Web page over which he or she has no control. To support the sender-pays model, an extension to the RSVP protocol must used where the QoS requests are propagated further up the multicast tree to the source node.

By enabling receiver-initiated resource reservation, each receiver can reserve the resources it requires and at the same time be assigned a corresponding price for the specific QoS request. Each receiver needs to know only about its own QoS requirements. This means pricing may be provided as a fixed and predictable function of the user's requested QoS. The problem with this approach is that network resources become inherently more valuable as utilization, i.e., potential congestion, increases. Thus the price charged would not necessary reflect the immediate cost of the resources to the network services provider.

RSVP is also suited to realtime pricing. In this case a RSVP flow would be as a function of both the requested QoS flow and the network load. However, the problem with this approach is unpredictable pricing for users. The user may not know when they were using more "expensive" network resources, i.e., during peak hours, until after the resources were consumed.

Since in RSVP each application's flow requires a separate flowspec, the unbundling of services follows naturally. However since users must be able to specify the QoS requirements for each application's flowspec, information overload may be associated with RSVP. A possible solution is to provide direct pricing feedback as the user is selecting the parameters for the flowspec, as described in the first pricing alternative, and combining that with a threshold mechanism to prevent overspending.

Authentication allowing only the person who is responsible for network charges to change the flowspec would provide certainty of pricing on specific connections, but not of the entire service if the number of connections were left uncontrolled. Thus the threshold would be implemented with session authentication or by periodically setting a total bill and letting the software determine the requirements, as with the smart market.

5. Protocols for Universal Service

The next generation Internet is being built. The first generation Internet, by being accessible to most only through telephone connections, exacerbated the failures of universal service and created a digital divide. The next generation Internet will be implemented with technologies with specific characteristics. We hypothesized that these characteristics can push to close the digital divide, or force it to open ever wider. Recall in the earlier section we presented a set of mechanisms and an argument as to why those mechanisms would support universal access. The following table illustrates which proposals for guaranteed quality of service use these mechanisms. To test these hypotheses we examined three mechanisms for providing quality of service in Internet services. Concluding this work we explain the impact of the features of the previously discussed QoS protocol mechanisms on the viability of universal service.

We predicted the following mechanisms will enhance universal service and help close the digital divide by creating predictability in price, low information overhead, and unbundling of services. In order to test this we first examined the QoS mechanisms to see which had the characteristics. If pricing characteristics are associated with particular mechanisms design for universal service is simplified, from a concept to a set of mechanisms. It is, after all, easier to evaluate a design with respect to the inclusion of specific mechanisms than with conceptual characteristics. Thus we now consider the technical mechanisms which support the economic characteristics favorable to universal service.

	Expected Capacity	Dynamic Auction	RSVP
Best Effort Service	Yes	Yes	No
Realtime User Feedback	Possible (dropped packets)	Possible (dropped packets)	Possible (receiver pays)
Entry Controls	Yes	No	Yes
Differentiation of Applications	Possible	No	In flowspec

Table 1: Mechanisms Predicted to Support Universal Service

The previous table lists the characteristics of the QoS mechanisms examined in this work. The above table implies that expected capacity would be most supportive of universal service and dynamic auction the least. The following table weakly supports that contention.

	Expected Capacity	Dynamic Auction	RSVP
Price Predictability	Yes	Within a Range	Per Connection
Information Overhead	Limited	High	Medium
Unbundling	Through Profile	No	Through flowspec
Authentication needed for payer control	To alter profile	Session/User	Session/User

Table 2: Characteristics of Protocols with Respect to Universal Service

The first table examined characteristics which we argued in Section 2 affect universal service at least in terms of telephony penetration. The second table listed technical mechanisms which could reasonably be expected to support the economic characteristics favorable to universal service. Do these mechanisms in fact support universal service?

It appears that, in this case, the intuitively obvious is also true: that the availability of best-effort service is critical to supporting universal service. Surprisingly, if it possible to set a maximum in the system and use best effort for continued connectivity it appears that real time user feedback is not critical. In fact, it is possible that real time user feedback can lead to information overload. The ability to unbundle applications is also clearly important. Notice that the technical ability to unbundle services is necessary but not sufficient to unbundling billing, as shown by the case of long distance and local phones bills.

The Internet, or IP-based services, have promise in that there are technical characteristics which provide the potential to expand universal service. Specifically the ability to offer flat rate best effort service, the potential for authentication before adding a usage-based charge, the potential for unbundling at a detailed level, the ability to offer pricing without requiring an understanding of the underlying billing mechanisms, and finally, the potential for customized service level with specific monthly charges selected in advance. All these features offer promise towards the goal of true universal service. End user feedback may be of use in providing universal service if real time cost information enables users to control their bills, thus removing billing uncertainty. Yet the existence of real time feedback may not be critical. Real time user feedback requires the users to track all of their phone bills. If the user is not in fact the party responsible for billing, real time feedback is meaningless.

Entry controls can assist in universal service; Entry controls enable networks to either refuse entry after reaching some trigger point, or refuse services for which there would be further charge. These features would remove all uncertainty and help to control bills. The ideal use of best-effort traffic with respect to universal service is to allow for low cost service without disconnection when subscribers have overspent for premium services. Entry controls limit periodic charges without requiring other controls or bookkeeping by the responsible user.

IP-based services can allow access to multimedia, as well as e-commerce opportunities that could allow connection charges to spiral with little or no advance warning. The ability to offer complex payment plans increase the potential for information overload, so that subscribers may see unexpected increases in price. The ability to offer customized service can allow consumers to get into debt more quickly by ordering more customized services than their budgets can support.

An unexpected but in retrospect obvious conclusion is that price predictability is inversely related to predictability in terms of available bandwidth. From an examination of the above protocols, it appears that there is a tradeoff in predictability in quality and predictability in price. Predictability in quality requires reserving network resources regardless of the presence of congestion, which means regardless of the unpredictable actions of others. Predictability in price requires cost-controlling response to scarcity, which leads to variable QoS delivery.

	Expected Capacity	Dynamic Auction	RSVP
Quality Predictability	Limited	Limited	High
Price Predictability	Yes	Within a Range	Per Connection

Table 3: Certainty in Service v Certainty in Price

Predictability in price requires either entry controls or that a simple subscription model will recover costs. Predictability in quality means that certain resources are certain to be available. The cost, and therefore presumably the price, of the bandwidth reserved in advance cannot be exactly known because the price of the bandwidth is a function of the traffic of others across the network while the bandwidth is reserved. Future work would include proving the situations under which this statement holds.

By looking at the digital divide from the far side of its analog counterpoints we offer an innovative approach to the problems of disconnection. Protocols can be viewed through a lens tinted with an understanding of why families are disconnected. Such a view suggests that the systematic construction of the network may have a far larger long term impact on the digital divide than the size of any universal service fund.

Before the deployment of any significant network mechanism such as the previously discussed QoS mechanisms, it is crucial to understand the mechanism's impact on the viability of the services for the poor. As seen from experiences with telephony networks, as well as the Internet, once a specific networking technology is deployed its impact on subsequent service offerings is indeed lasting. Thus it is crucial that universal access be considered in the design of protocols for the next generation Internet. This paper provides the information for protocols designers and early adapters to develop and select protocols which can serve universal access, and help to close the digital divide.

References

American Telephone & Telegraph, 1907, Annual Report, American Telephone & Telegraph Company, New York, NY.
Anania & Solomon, 1988, User arbitrage and ISDN, Intermedia, London, January.
Brooks, 1975, Telephone: The First Hundred Years, Harper & Row, NY, NY.

Clark, 1996, Explicit Allocation for Best Effort Packet Delivery Service, The Telecommunications Policy Research Conference, Solomons Islands MD.
Choi, Stahl, & Winston, 1997; The Economics of Electronic Commerce, Macmillan Technical Publishing, Indianapolis, IN.
Fischer, C., 1992, America Calling: A Social History of the Telephone to 1940, University of CA Press, Berkeley, CA.
Gillett,S.E., 1994, Technological Change, Market Structure, and Universal Service , Telecommunications Policy Research Conference, Solomons Island, October.
Gordon, 1995, "Postalization", American Heritage, vol. 46 (6), pp. 16-18.
Horwood, 1998, "Cable Franchise Renewal and Local Right-of-way Management ", Alliance for Community Media 1998 Western Regional Conference: Harness the Power, October 22-24.
McKnight, L.W. & Bailey, J.P., 1997, Internet Economics, MIT Press, Cambridge, MA.
Mackie-Mason, & Varian, H., 1995, "Pricing the Internet", Public Access to the Internet, ed. Kahin &Keller, Englewood Cliffs, NJ: Prentice Hall.
National Center for Education Statistics (NCES), 1998, Public libraries in the United States:, National Center for Education Statistics, US ED, Washington, D.C.
Mueller, 1993, "Universal service in telephone history", Telecommunications Policy, Vol. 6, pp. 352-370.
Mueller & Schement , 1996, Universal Service from the Bottom Up: A Study of Telephone Penetration in Camden, New Jersey", The Information Society, Vol. 12, pp. 273-292.
National Telecommunications and Information Administration, 1998, Falling Through the Net: Defining the Digital Divide, United States Government Printing Office; Gaithersburg, MA.
Odlyzko, 1999, " Trends in communication pricing and their implications for the Internet ", Workshop On Internet Service Quality Economics Massachusetts Institute of Technology Research Program on Communications Policy (MIT RPCP), Cambridge, MA, USA December 2-3.
Shapiro &Varian, 1998, Information Rules, Harvard Business School Press, Cambridge MA.
Thorne, Huber, & Kellog, 1995, Federal Broadband Law, Little, Brown & Company, Boston, MA.
Xie. J & Sirbu M., 1985, "Price Competition and Compatibility in the Presence of Positive Demand Externalities ", Management Science, June 1995.
Zhang, Deering, Estrin, Shenker & Zappala, 1993, "RSVP: A New Resource ReSerVation Protocol", IEEE Network Magazine, September 1993.