Tussle in Cyberspace: Defining Tomorrow's Internet

David D. Clark
MIT Lab for Computer Science
ddc@lcs.mit.edu

John Wroclawski
MIT Lab for Computer Science
jtw@lcs.mit.edu

Karen R. Sollins
MIT Lab for Computer Science
sollins@lcs.mit.edu

Robert Braden
USC Information Sciences Institute
braden@isi.edu

Abstract

The architecture of the Internet is based on a number of principles, including the self-describing datagram packet, the end to end arguments, diversity in technology and global addressing. As the Internet has moved from a research curiosity to a recognized component of mainstream society, new requirements have emerged that suggest new design principles, and perhaps suggest that we revisit some old ones. This paper explores one important reality that surrounds the Internet today: different stakeholders that are part of the Internet milieu have interests that may be adverse to each other, and these parties each vie to favor their particular interests. We call this process "the tussle". Our position is that accommodating this tussle is crucial to the evolution of the network's technical architecture. We discuss some examples of tussle, and offer some technical design principles that take it into account.

Categories and Subject Descriptors

C.2.1 [**Computer Systems Organization**]: Computer Communications Networks—*Network Architecture and Design*; H.1 [**Information Systems**]: Models and Principles; K.4.1 [**Computing Milieux**]: Computers and Society—*Public Policy Issues*

General Terms

Design, Economics, Legal Aspects, Security, Standardization

Keywords

Tussle, Network Architecture, Trust, Economics, Design Principles, Competition

Work sponsored in part by the Defense Advanced Research Projects Agency (DARPA) and Air Force Research Laboratory, Air Force Materiel Command, USAF, under agreement number F30602-00-2-0553 at MIT, and agreement number F30602-00-1-0540 at ISI. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

SIGCOMM'02, August 19-23, 2002, Pittsburgh, Pennsylvania, USA.

1. INTRODUCTION

The Internet was created in simpler times. Its creators and early users shared a common goal—they wanted to build a network infrastructure to hook all the computers in the world together so that as yet unknown applications could be invented to run there. All the players, whether designers, users or operators, shared a consistent vision and a common sense of purpose.

Perhaps the most important consequence of the Internet's success is that the common purpose that launched and nurtured it no longer prevails. There are, and have been for some time, important and powerful players that make up the Internet milieu with interests directly at odds with each other.

Some examples are very current. Music lovers of a certain bent want to exchange recordings with each other, but the rights holders want to stop them. People want to talk in private, and the government wants to tap their conversations. Some examples are so obvious that they are almost overlooked. For the Internet to provide universal interconnection, ISPs must interconnect, but ISPs are sometimes fierce competitors. It is not at all clear what interests are being served, to whose advantage, to what degree, when ISPs negotiate terms of connection. It is not a single happy family of people dedicated to universal packet carriage.

We suggest that this development imposes new requirements on the Internet's technical architecture. These new requirements, in turn, motivate new design strategies to accommodate the growing tussle among and between different Internet players. The purpose of this paper is to explore what these requirements and strategies might be.

We begin by briefly discussing the Internet landscape - some fundamental differences between the mechanisms of engineering and society, and the players that populate our field. We then outline some proposed design principles intended to accommodate within the Internet mechanisms of society as well as those of engineering. We believe this accommodation is central to designing an Internet that is resilient to the challenges of society as well as those of technology. We conclude by discussing some tussle spaces, ways in which our principles might guide the technical response to these spaces, and specific technical research that may be of value in accommodating these tussles.

1.1 The natures of engineering and society

Engineers attempt to solve problems by designing mech-

anisms with predictable consequences. Successful engineering yields bridges that predictably don't fall down, planes that predictably don't fall out of the sky, and calculators that give the "right" answer. The essence of engineering is the development and codification of models, techniques and tools that deliver predictable, desirable behavior.

The technical development of the Internet has followed this path. As a community, we focus on design principles that deliver such virtues as robustness, scalability and manageability in the face of complexity, component failures, growth, and other challenges. However, as the Internet becomes mainstream it inevitably moves from being an engineering curiosity to being a mirror of the societies in which it operates. The Internet may have been designed by engineers, but its behavior (and its evolution) is by no means predictable today.

The operation of societies follows a different model. Historically, the essence of successful societies is the dynamic management of evolving and conflicting interests. Such societies are structured around 'controlled tussle' – regulated by mechanisms such as laws, judges, societal opinion, shared values, and the like. Today, this is the way the Internet is defined—by a series of ongoing tussles. Different parties adapt its mix of mechanisms to try to achieve their conflicting goals, and others respond by adapting the mechanisms to push back. Thus, conservative governments and corporations put their users behind firewalls, and the users route and tunnel around them. ISPs give their users a single IP address, and users attach a network of computers using address translation. There is no "final outcome" of these interactions, no stable point, and no acquiescence to a static architectural model.

The challenge facing Internet research and engineering is to recognize and *leverage* this reality – at minimum to accommodate it; if possible, to use it to *strengthen* the technical architecture. In other words, the technical architecture must accommodate the tussles of society, while continuing to achieve its traditional goals of scalability, reliability, and evolvability. This expansion of the Internet's architectural goals is a difficult, but central technical problem.

1.2 The Internet landscape

Today, there are many parties that are part of the Internet milieu. These include:

- Users, who want to run applications and interact over the Internet.
- Commercial ISPs, who sell Internet service with the goal of profit.
- Private sector network providers who run a part of the Internet to facilitate their business or other undertaking.
- Governments, who enforce laws, protect consumers, regulate commerce, and so on.
- Intellectual property rights holders, who want to protect their materials on the Internet.
- Providers of content and higher level services, offered in search of profit or as a public service.

There is great diversity within each of these categories: there are "good users" and spammers, dominant ISPs and small players, private providers with more or less rigidity about usage, liberal and conservative governments, and so on. The resulting tussles span a broad scope: the rights of the individual vs. the state, the profit seeking of competitors, the resistance to those with malicious intent, those with secrets vs. those who would reveal them, and those who seek anonymity vs. those would identify them and hold them accountable. The list probably mirrors every aspect of society. For a detailed discussion of these various players and their impact on the Internet, see [1].

2. PRINCIPLES

The thesis of this paper is that the future of the Internet will increasingly be defined by tussles that arise among the various parties with divergent interests, and that the technical architecture of the Internet must respond to this observation. If this is so, are there principles to guide designers, and mechanisms that we should use in recognition of this fact?

In this paper we offer some design principles to deal with tussle. Our highest-level principle is:

- Design for variation in outcome, so that the outcome can be different in different places, and the tussle takes place within the design, not by distorting or violating it. Do not design so as to dictate the outcome. Rigid designs will be broken; designs that permit variation will flex under pressure and survive.

Within this guiding principle, we identify two more specific principles:

- Modularize the design along tussle boundaries, so that one tussle does not spill over and distort unrelated issues.
- Design for choice, to permit the different players to express their preferences.

2.1 Modularize along tussle boundaries

Systems designers know to break complex systems into modular parts. Modularity is typically used to manage complexity, allow for independent implementation and component reuse, or to meet other technical goals. But "tussle isolation" is perhaps a new principle.

- Functions that are within a tussle space should be logically separated from functions outside of that space, even if there is no compelling technical reason to do so. Doing this allows a tussle to be played out with minimal distortion of other aspects of the system's function.

The design of the DNS provides an example. The current design is entangled in debate because DNS names are used both to name machines and to express trademark. In retrospect, one might have predicted that fights over trademarks would be a tussle space, and made sure that the names that expressed trademarks were used for as little else as possible. In particular, one might imagine separate strategies to deal with the issues of trademark, naming mailbox services, and

providing names for machines that are independent of location (the original and minimal purpose of the DNS). One could then try to design these latter mechanisms to try to duck the issue of trademark.

- Solutions that are less efficient from a technical perspective may do a better job of isolating the collateral damage of tussle.

In contrast, the current trajectory of IP QoS design tries to isolate tussles. The use of explicit ToS bits to select QoS, rather than binding this decision to another property such as a well-known port number, disentangles what application is running from what service is desired. It can be anticipated that there will be tussles about what applications a user can run, and separately tussles about what service qualities can be provided. The designers felt that it was better to separate these ideas. This modularity allows tussles about QoS to be played out without distortions, such as demands that encryption be avoided simply to leave well-known port information visible or the encapsulation of applications inside other applications simply to receive better service.

2.2 Design for choice

Network protocols are designed so that different parties on the network can communicate with each other, consumers can make use of the resources of providers, and providers can interconnect with each other to provide service. It is important that protocols be designed in such a way that all the parties to an interaction have the ability to express preference about which other parties they interact with. Protocols must permit all the parties to express choice.

For example, the design of the mail system allows the user to select his SMTP server and his POP server. A user can pick among servers, perhaps to avoid an unreliable one or pick one with desirable features, such as spam filters. Users can select what news server they use, perhaps to prevent their children from encountering some of the more colorful news groups. This sort of choice drives innovation and product enhancement, and imposes discipline on the marketplace.

The form that the choice takes for the different parties may be different. A user of mail might choose her SMTP server by configuring a mail-sending program. An ISP might try to control what SMTP server a customer uses by redirecting packets based on the port number.[1]

Providing this sort of choice has a drawback—it adds to the complexity of configuring and using a service. For naïve users, choice may be a burden, not a blessing. To compensate for this complexity, we may see the emergence of third parties that rate services (the on-line analog of Consumers Reports) and parties that provide pre-configured software to relieve the user of dealing with the details of choice.

2.3 Implications

These principles, and the reality of tussle, have some further implications for design:

[1] An over-generalization of the tussle is that service providers exercise control over routing; end-users control selection of other end-points. End-users try to over-rule constrained routing with tunnels and overlay networks.

Choice often requires open interfaces. Open interfaces have played a critical role in the evolution of the Internet, by allowing for competition among algorithms, implementations, and vendors, and by enabling rapid technical progress through replacement of modular parts rather than entire systems. But open interfaces also allow choice, not just replacement. If a protocol allows a party to select among alternative providers of service, this usually implies that the interface to that service is well-defined, so that independent versions of the service can be constructed.

Tussles often happen across interfaces. Some tussles involve the use of different mechanisms by different parties. However, some tussles, such as the tussle among competitive ISPs, may involve technical interfaces between those parties. For example, BGP is used as the routing protocol among ISPs, who interconnect but are business competitors. If an interface occurs at a point of tussle, it will have different attributes from an interface that exists simply to foster interoperability, modularity, or competition among suppliers.

Open interfaces at tussle points may benefit from the following sorts of properties, which are not always important in other cases.

- Visible exchange of value.
- Exposure of cost of choice.
- Visibility (or not) of choices made.
- Tools to resolve and isolate faults and failures.

It matters if the consequence of choice is visible. Choices made in public are sometimes different than those made in secret. In some cases, there is no way to hide a choice. Often, the choice can be secret, even if its consequences are visible. The routing arrangements among ISPs are generally not public, even though everyone can see the consequences at the BGP level. A link-state routing protocol requires that everyone export his link costs, while a distance vector protocol makes it harder to see what the internal choices are.

Tussles have different flavors. In some cases, the interests of the players are simply adverse, and there is no win-win way to balance them. But in many cases, players' interests are not adverse, but simply different. A user wants to send data; a provider wants to be compensated for carrying it. While this implies a natural tussle over pricing, in the end both parties realize that they must meet the other's needs.

To support this class of tussle, recognize that there is often an exchange of value for service. Value need not be "money" but often will be. Napster is a non-monetary example that illustrates the "mutual aid" aspect of peer-to-peer networking. Whatever the compensation, recognize that it must flow, just as much as data must flow. Sometimes this happens outside the system, sometimes within a protocol. If this "value flow" requires a protocol, design it. (There is an interesting case study in the rise and fall of micro-payments, the success of the traditional credit card companies for Internet payments, and the emergence of PayPal and similar schemes.)

Tussles evolve over time. A traditional engineering design produces a result that is constant until the mechanism is redesigned. But tussle is ongoing and evolutionary.

Each side finds new ways to gain an advantage, and then the other side responds. This implies, first, that any thinking about tussle must view it as a multi-round process, and second, that as mechanism is drawn into an ongoing tussle, it may be used in unexpected ways, and require redesign to survive in this new role.

There is no such thing as value-neutral design. What choices designers include or exclude, what interfaces are defined or not, what protocols are open or proprietary, can have a profound influence on the shape of the Internet, the motivations of the players, and the potential for distortion of the architecture.

Don't assume that you design the answer. You are designing a playing field, not the outcome.

3. TUSSLE SPACES

In this section we discuss some specific aspects of the Internet in which different players with competing interests come together. In each case, our goal is to examine the nature of the tussle and to illustrate how our principles can be applied in specific cases. We suggest some specific research areas that would benefit from application of our principles.

3.1 Economics

One of the tussles that define the current Internet is the tussle of economics. The providers of the Internet are not in the business of giving service away. For most, it is a business, run to make a profit. This means they are competitors, and look at the user, and each other, as a customer and a source of revenue. Providers tussle as they compete, and consumers tussle with providers to get the service they want at a low price.[2]

How can we, as engineers, shape the economic tussle? In fact, we have great power to shape this tussle, but first we have to understand the rules that define it. A standard business saying is that the drivers of investment are fear and greed. Greed is easy to understand—it drove hundreds of billions of dollars worth of investment in telecommunications over the last decade, much of which now sits at risk of bankruptcy. But fear is more subtle. The vector of fear is competition, which results when the consumer has choice. The tussle among providers and consumers in a competitive landscape is the most basic attribute of a marketplace. Most economists of a "western" bent would argue that competition is good: it drives innovation, disciplines the market, insures efficiency, and removes the need for intervention and regulation of a market. To make competition viable, the consumer in a market must have the ability to choose. So our principle that one should design choice into mechanism is the building block of competition.

Here are some specific examples, with implications for research and network design:

3.1.1 Provider lock-in from IP addressing

ISPs would like to find ways to lock in their customers; customers want to preserve the ability to change among providers. This illustrates the basic consumer-producer tussle in a competitive world. For hosts that use static addresses, renumbering is a complex task. Because renumbering hosts can be hard, there is a very explicit tension today between the desire to have addresses reflect topology to support efficient routing and the desire of the customer to change providers easily. Either a customer is locked into his provider by the provider-based addresses, or they obtain a separate block of addresses that are not topologically significant and therefore add to the size of the forwarding tables in the core of the network. Many ISP's refuse to route small address blocks, nominally protecting the routing tables but also locking the customer to their address range. The responses by the consumer include dynamic host numbering (DHCP) and dynamic update of DNS entries when the host is renumbered.

- A desire for vigorous competition would suggest that the consumer should have the choice to move from ISP to ISP. Given that, the Internet design should incorporate mechanisms that make it easy for a host to change addresses and to have and use multiple addresses. Addresses should reflect connectivity, not identity, to modularize tussle. This would relieve problems with end-node mobility, improve choice in multi-homed machines, and improve the ease of changing providers.

3.1.2 Value pricing

One of the standard ways to improve revenues is to find ways to divide customers into classes based on their willingness to pay, and charge them accordingly—what economists call value pricing. An example from another sector is the "Saturday night stay" criterion for airline travel. It costs the airline no more to carry a passenger if she does not stay over Saturday night, but this restriction tends to separate the business and pleasure traveler, which is useful because the business traveler seems to have a greater willingness to pay. Airlines impose Saturday night stay restrictions, and consumers respond by buying multiple tickets, and using only some of the segments of the flight. Airlines respond by declaring this behavior unacceptable. And thus the tussle evolves.

As an example of similar behavior in the Internet, some acceptable use policies for residential broadband access prohibit the operation of a server in the home. To run a server, the customer is required to pay a higher "business" rate. Customers who wish to sidestep this restriction can respond by shifting to another provider, if there is one, or by tunneling to disguise the port numbers being used. The probable outcome of this tussle depends strongly on whether one perceives competition as currently healthy in the Internet, or eroding to dangerous levels.

- This discussion illustrates the observation that there may be no such thing as value-neutral design. The design and deployment of tunnels (or other mechanisms to mask what services are being used by a consumer) shifts the balance of power from the producer to the consumer. Given that value pricing is not a moral wrong, should the consumers be aided in their quest to bypass the controls of the producers? Those who

[2]There is now considerable interest in the economics community in the nature of the Internet. Some of the seminal papers are published in [9]. For an overview of the current literature on Internet economics, see the Web site maintained by Mackie-Mason at `http://china.si.umich.edu/telecom/net-economics.html`.

see the consumer as "the little guy" being abused by the "big providers" will design such mechanisms, and this is part of the tussle, not something that happens outside the tussle. What mechanisms get designed, and what standards get approved, are all part of the tussle.

3.1.3 Residential broadband access

There is concern today that the advent of broadband residential access will be accompanied by a great reduction in competition. Today there are almost 6000 dialup Internet service providers. A pessimistic outcome five years in the future is that the average residential customer will have two choices—his telephone company and his cable company, because they control the wires. This loss of choice and competition is viewed with great alarm by many, who fear that it may lead to higher prices and restrictions on what the user may do, and there are many forces aligning to fight this loss of competition. Some are regulatory, calling for laws to mandate "open access", to force the owners of the wires to allow multiple ISPs to use them. Economists and regulators hope that multiple providers will install their own cables, to increase competition.[3] However, in a tussle of competition, one cannot compel a potential provider to invest and enter a market.

Using the principles of this paper, one should speculate on what sorts of investments are actually likely to be made, and to think about what choice, and what tussle modularity, would improve the outcome of such an investment. One investment option that is gaining momentum now is municipal deployment of fiber, because fiber installed by a neutral party such as a municipality can be a platform for competitors to provide higher level services (e.g. phone, Internet or television). This requires that the equipment lighting the fiber support multiple service providers. Most of the equipment made today is not "naturally open" in this way, having been designed without consideration of this particular modularity boundary (or indeed with the specific goal of confounding it).

- An important R&D project is to design and demonstrate a fiber-based residential access facility that supports competition in higher-level services. Technical questions include whether sharing should be in the time domain (packets) or color domain, how the fairness of sharing can be enforced and verified, an approach to fault isolation and other operational issues, and how incremental upgrades can be done. This project is motivated both by the principle of "design for choice", and as well by recognition of new tussle boundaries.

Most of today's "open access" proposals fail to balance the interests of concerned parties because they are not modularized along tussle space boundaries. For example, the capital costs and deployment pragmatics of broadband infrastructure differ greatly from those of operating mail and web servers. This creates a natural boundary between the two tussle spaces of broadband facilities provision and ISP services. Proposals that implement open access at this modularity boundary are more likely to benefit the Internet as a whole, because they allow each tussle to play out independently.

[3]For an analysis of issues in residential broadband access, see [3].

3.1.4 Competitive wide area access

Today, the Internet system does not let the individual customer select his "long distance provider" the way the telephone system does. This is an example of designers failing to appreciate a competitive tussle space.

At the time that equal access was being introduced into the telephone system, there was a call for Internet routing to support the same capability. The Internet designers deemed this not necessary. They reasoned that there would be sufficient competition in the market because there were going to be many ISP's directly competing to serve the customer. Letting the local provider enter into a wholesale arrangement to obtain wide area service seemed adequate, because if one local provider made an unsatisfactory choice in wide area provider, the customer could just switch to a new local provider.

But this decision may be having undesirable consequences today. It is possible that customers today would be much more likely to see more service diversity, e.g. quality of service support for applications, if there were more competition.

- The Internet should support a mechanism for choice of source routing that would permit a customer to control the path of his packets at the level of providers. A design for such a system must include where these user-selected routes come from or how they are constructed, how failures are managed, and how the user knows that the traffic actually took the desired route. The capability must also be approachable by a broad class of users of varying sophistication. This is a very complex design challenge,[4] but could have a great influence.

This example illustrates another important point about competition. One should be prepared to pay for what one uses, or there is little incentive for a provider to offer it. Today, service providers do not like loose source, because ISPs don't receive any benefit when they carry traffic directed by a source route. ISPs enter into business arrangement that determine which traffic they agree to carry across which interfaces, and a source route has the effect of overriding these arrangements. Why should they be enthusiastic about this? Since source routes don't work effectively today, researchers propose even more indirect ways of getting around provider-selected routing, such as exploiting hosts as intermediate forwarding agents. (This kind of overlay network is a tool in the tussle, certainly.) Another, perhaps simpler, approach is to compensate the provider for carrying the packets. But this idea tends to upset designers as well as customers, because they fear they will end up in an onerous "pay by the byte" situation, which does not seem to have much market appeal.

- The design for provider-level source routing must incorporate a recognition of the need for payment. There

[4]In particular, today's loose source routes, even if widely implemented, would provide only a small portion of what is needed.

must be enough generality in the payment schemes that the market can select an outcome that works for all parties. (Remember, we are not designing the outcome, only the playing field for the tussle.)

- Overlay architectures should be evaluated for their ability to isolate tussles and provide choice. A comparison is warranted between overlay architectures and integrated global schemes to understand how each balances the relative control that providers and consumers have, and whether economic distortion is greater in one or the other.

3.2 Trust

One of the most profound and irreversible changes in the Internet is that by and large, many of the users don't trust each other. The users of the Internet no longer represent a single community with common motivation and shared trust. There are parties with adverse interests, and some genuine "bad guys" out there. This implies that mechanisms that regulate interaction on the basis of mutual trust should be a fundamental part of the Internet of tomorrow.[5]

Most users would prefer to have nothing to do with the bad guys. They would like protection from system penetration attacks, DoS attacks, and so on. This is a profound tussle, between people who want to be left alone, and people who want to bother them. Since host security today is of variable and mostly poor quality, this desire for protection leads to firewalls. Firewalls change the Internet from a system with transparent packet carriage between all points (what goes in comes out), to a "that which is not permitted is forbidden" network. This is a total reversal of the Internet philosophy, but pure transparency is not what most users long for. For over ten years, Internet purists have been bemoaning the fact that firewalls inhibit innovation and the introduction of new applications (fifteen years ago they were called "mail gateways"), but firewalls have not gone away.

The principle of "design for choice" would imply that users should be able to choose with whom they interact, and users should be able to choose the level of transparency they offer to other users. The principle of "tussle isolation" suggests that these mechanisms should not be overloaded on to any other mechanism, but should be separated. Further, one should consider if within the broad topic of trust, there are separable issues.

The first topic is control over which parties are willing to exchange packets with each other.

- In the abstract, there is a technical question as to whether each end-node can implement sufficient trust-related controls within itself, or whether delegation of this control to a remote point inside the network is required—a "trust-aware firewall". As a practical matter, the market calls for firewalls. Firewalls of the future must be designed so that they apply constraints based on who is communicating, as well as what protocols are being run and where in the network the parties are. Such a device would imply the design of new protocols and interfaces, to allow the end node and the control point to communicate about the desired controls.[6] Issues of choice arise: who gets to pick which firewall a user uses?

- To prevent DoS attacks, protocols could be changed so that end-nodes do not have to establish state or otherwise invest effort until they have verified that they want to talk to the party initiating communication. This concept is challenging, first because of a difficult balance between cost and function, and second, because the idea of "first packet trust verification" is at odds with the layered model of protocols, in which transport establishes a connection before any higher level information is exchanged.

Another tussle about firewalls is worth noting. Who gets to set the policy in the firewall? The end user may certainly have opinions, but a network administrator may as well. Who is "in charge"? There is no single answer, and we better not think we are going to design it. All we can design is the space for the tussle. But this illustrates the point about visibility of decision-making. If a system administrator has installed control rules in a firewall that affect an end user, should that end user be able to download and examine these rules? One way to help preserve the end to end character of the Internet is to require that devices reveal if they impose limitations on it. However, there is no obvious way to enforce this requirement, so it becomes a courtesy, not a real requirement.

Another dimension of trust is the fact that most users don't trust many of the parties they actually want to talk to. We connect to web sites but are suspicious that they are gathering information on us, stealing our credit cards, not going to deliver what they promised, and so on. In this case, the solution is more complex; we depend on third parties to mediate and enhance the assurance that things are going to go right. Credit card companies limit our liability to $50, or sometimes nothing, in case of dispute. Public key certificate agents provide us with certificates that assure us we are talking to the party we think we are. Web sites assess and report the reputation of other sites. The fact of these third parties contrasts with our simple model of two-party end-to-end communication among trusting parties. Each individual interaction may be two-party end-to-end, but the application design is not.

- An important engineering principle for future applications is that there should be explicit ability to select what third parties are used to mediate an interaction, and to act as an agent for the end-user in improving his trust in the operation. The parties must be able to choose, so they can select third parties that they trust.[7]

Another space in which trust is eroding is that users less and less trust the software they have to run. They suspect

[5] A thoughtful analysis of trust that has shaped our thinking is provided by [10].

[6] The IETF is working on such standards, e.g. the MIDCOM working group.

[7] An interesting debate relevant to this topic centers on the IETF proposal to charter the Open Pluggable Edge Services (OPES) working group, and the IAB deliberation on policy concerns. The IAB has focused on issues of whether one end or both have to concur with the insertion of an intermediate node in the communication, and what tools the user should have to detect and recover from a faulty node.

their operating system and browser of gathering information on them and passing it on without their knowledge, or turning them in for software license violations. There are web sites that claim to look at the outgoing data stream from the user's machine and detect and remove any information that is leaking out.

- This problem may best be dealt with using non-technical means—regulation, public opinion and so on. Just because a problem manifests in a technical space, it does not mean it has to be solved there. But it is an interesting exercise to consider whether there are technical means to protect a user from software running on their own machine. The history of mandatory security controls and security kernels suggests that this problem is thorny.

3.2.1 The role of identity

One obvious point about trust is that if communication is to be mediated based on trust, then as a preliminary step, parties must be able to know to whom they are talking. Otherwise, one has little basis for judging how much to trust others.

One could take this as a call for the imposition of a global namespace of Internet users, with attached trust assessments. We believe this is a bad idea. It is hard to imagine a global system that is really trustworthy. More importantly, there are lots of ways that parties choose to identify themselves to each other, many of which will be private to the parties, based on role rather than individual name, etc. What is needed is a framework that translates these diverse ways into lower level network actions that control access. This implies a framework for talking about identity, not a single identity scheme. We suggest that such a framework could usefully share and arbitrate information across many layers of the protocol stack.

The need to know to whom we are talking will challenge a current precept of the Internet, which is that it is permissible to be anonymous on the Internet. There is a fundamental tussle between the ideas of anonymous action, and the idea that in a society where "that which is not forbidden is permitted", one can be held accountable for ones actions. A possible outcome of this tension is that while it will be possible to act anonymously, many people will choose not to communicate with you if you do, or will attempt to limit what you do.[8] A compromise outcome of this tussle might be that if you are trying to act in an anonymous way, it should be hard to disguise this fact. This illustrates the observation that one must think about whether the consequences of choice are visible, or can be hidden.

3.3 The tussles of openness

One of the most profound fears for the Internet today is that it will lose its "open" qualities: the openness to innovation that permits a new application to be deployed, the openness of access that allows a user to point their Web browser at any content they please, the openness that allows a user to select the servers and services that best meet their needs.

[8]An analog is the current situation with Caller ID, where a sender can block the caller's information, but the receiver can refuse to accept calls from a sender that does.

The openness to innovation—to new applications and new uses—has perhaps been the most critical success factor for the Internet. But openness is not an unalloyed virtue for service providers. Openness often equates to competition, which creates the fear factor that demands costly investment and drives profits to a minimum. Many telephone company executives remember the good old monopoly days, with a comfortable regulated rate of return and no fear. And many current ISPs may long for a return to those less open, high margin days, if they could only figure out how to get there. The keys are closed or proprietary interfaces and vertical integration.

Motivations concerning open vs. proprietary systems have much to do with economics. Economists have studied the motivation of providers with various degrees of market power to choose open or proprietary interfaces; see [6]. Industry understands that interfaces, or lack thereof, can shape a market.[9] There is probably a whole paper on the tussles surrounding open vs. closed systems. However, as a starting point, the first exercise should be to speculate about whether these various openness tussles can be modularized and disentangled, and what this means for mechanism design.

Vertical integration—the bundling together of infrastructure and higher-level services—requires the removal of certain forms of openness. The user may be constrained to use only certain providers of content, or to pay to run certain protocols, and so on. However, vertical integration has nothing to do with a desire to block innovation. Even in a market with a high degree of vertical integration, innovation that brings new value to the customer is likely to benefit all parties. So it would be wise to separate the tussle of vertical integration, about which many feel great passion, from the desire to sustain innovation.

The technical characteristic of the network that has fostered innovation is transparent packet carriage—the ability to deploy a new protocol without having to modify the inside of the network. But transparency is not the same thing as openness, though they are related. With this brief motivation, we consider some old design principles of the Internet, including the principle that is usually equated with transparency, the end to end arguments.

4. REVISITING OLD PRINCIPLES

4.1 The future of the end to end arguments

One of the most respected and cited of the Internet design principles is the end to end arguments, which state that mechanism should not be placed in the network if it can be placed at the end node, and that the core of the network should provide a general service, not one that is tailored to a specific application [11]. There are two general dimensions to the arguments: innovation and reliability.

[9]While technical network designers may not think about open interfaces as a tool to drive market structure, industrial players understand this fully. When then Senator Gore announced his vision for a National Information Infrastructure (NII) in the early 1990s, at least two organizations produced requirement documents for the "critical interfaces" that would permit the NII to have a suitable structure [4, 5, 2].

Innovation: If the core of the network has been tailored to one specific application, this may inhibit the deployment of other applications. If the core of the network must be modified to deploy a new application, this puts a very high hurdle in front of any unproven idea, and almost by definition, a new idea is unproven.

Reliability and robustness: If bits of applications are "in the network", this increases the number of points of failure that can disable the application. The more simple the core of the network, the more reliable it is likely to be.

The simplest application of the end to end arguments produces a network that is *transparent*: packets go in, and they come out, and that is all that happens in the network. This simple idea was very powerful in the early days of the Internet, but there is much fear that it seems to be eroding, for many of the reasons discussed above:

- The lost of trust calls for less transparency, not more, and we get firewalls.
- The desire for control by the ISP calls for less transparency, and we get application filtering, connection redirection, and so on.
- The desire of third parties to observe a data flow (e.g. wiretap) calls for data capture sites in the network.
- The desire to improve important applications (e.g. the Web), leads to the deployment of caches, mirror sites, kludges to the DNS and so on.

This is a lot of mechanism, a large potential loss of transparency, and an increasing focus on improving existing applications at the expense of new ones. So what is the future of the end to end arguments? We argue that the end to end arguments are still valid and powerful, but need a more complex articulation in today's world. The discussion to this point gives us some guidance.

Evolution and "enhancement" of existing, mature applications is inevitable. As applications become popular, lots of players, including application providers and the ISPs, will want to get involved in them, whether as a move toward vertical integration, enhancement of performance or reliability, or some other reason. This will almost certainly lead to increased complexity, perhaps decreased reliability or predictability, and perhaps an evolution of the overall application away from the original vision. We should not imagine that anyone can do much about this. If we design applications so that the user can control what features "in the network" are invoked, we may have done as much as we can.

The most we can do to protect maturing applications is to bias the tussle. If application designers want to preserve choice and end user empowerment, they should be given advice about how to design applications to achieve this goal. This observation suggests that we should generate "application design guidelines" that would help designers avoid pitfalls, and deal with the tussles of success.

Keeping the net open and transparent for new applications is the most important goal. Innovation and the launch of new applications is the engine that has driven the growth of the Internet and the generation of new value. So barriers to new applications are much more destructive than network-based support of proven applications. Since new applications must, almost of necessity, launch incrementally, they most benefit from the transparent simplicity that the end to end arguments fostered. By the principle of isolation of tussle, any barriers that are put into the network as a result of the desire to control mature applications or issues of trust should not prevent parties that want transparency from getting it.

Failures of transparency will occur—design what happens then. Today, when an IP address is unreachable, there is little in the way of helpful information about why. A sophisticated user can run traceroute, but today's normal user just gets frustrated. Tools for fault isolation and error reporting would help – the hard challenge is not so much to find the fault but to report the problem to the right person in the right language. That person may be someone who can fix the problem, or someone who can decide to choose a different path or provider – fault reporting is as much a tool of tussle management as it is a tool of technical repair. Of course, some devices that impair transparency may intentionally give no error information or even reveal their presence, and that must be taken into account in design of diagnostic tools.[10]

Peeking is irresistible. If there is information visible in the packet, there is no way to keep an intermediate node from looking at it. So the ultimate defense of the end to end mode is end to end encryption. In most discussions of the need for encryption, the putative threat is someone who wants to steal your information or modify it. The ISP as a "threat" is not the normal assumption. But if the ISP is trying to control or modify what you are doing, then the ISP is the issue.

Of course, encrypting the data stream has drawbacks. One is that the actions of the ISP might actually be making things better. They might be offering performance improvements or other benefits that the end user actually wants. But this situation is not an issue; if the user has control over whether the data is encrypted or not, the user can decide if the ISP actions are a benefit or a hindrance. The other drawback is that encrypting the stream might just be the first step in an escalating tussle between the end user and the network provider, in which the response of the provider is to refuse to carry encrypted data. It is probably not the case that a commercial ISP would escalate to this level. In the U.S., competition would probably discipline a provider that tried to do this. But a conservative government with a state-run monopoly ISP might. And in that case, policy will probably trump technology in any case. Then the advantage of having the encrypted mode is that it would force the government to be explicit about what their policy was. Forcing the choice to be public and visible is about all that technology can do to moderate this situation.[11]

Note that in a multi-way application, where third parties are involved to insure the validity of the transaction, the meaning of "end to end" gets more complex, and so does the proper use of encryption.

[10]See the footnote above on the deliberations by the IAB on the charter for the OPES working group.

[11]The next step in this sort of escalation is steganography—the hiding of information inside some other form of data. It is a signal of a coming tussle that this topic is receiving attention right now.

4.2 Separation of policy and mechanism

Another design principle of great age and uncertain origin[12] is that technologists should design policy-free mechanism, and allow those who use the system (whether literal "users", administrators, etc) to adjust the mechanisms to match their specific needs. This paper challenges this principle as perhaps being too simplistic. True value-neutral design is, at best, extremely difficult. Mechanism defines the range of "policies" that can be invoked, which is another way of saying that mechanism bounds the range of choice. So in principle there is no pure separation of policy from mechanism.

However, this does not negate the principle. The chief advantage of attempting to separate mechanism and policy is to isolate some regions of the system from tussle. Even if the attempt is not completely successful, these isolation regions can serve to separate different tussles from each other, and can serve as technological 'fixed points' that allow different tussles to play out at different speeds.

- Perhaps the most challenging intellectual puzzle in this design space is to discover parts of mechanism that really can be divorced from policy — which, in other words, actually *are* value-neutral.

One value (or bias) that is shared by many people is user empowerment. This is the preference that the user, rather than the service provider or the software provider, be able to pick what applications to run, what servers and services to use, and so on. User empowerment, to many, is a basic Internet principle, but for this paper, it is the manifestation of the right to choose—to drive competition, and thus drive change.

One could argue that user empowerment is a bias, of the "David and Goliath" sort—a bias imposed on the tussle between the little guy and the provider, who is seen as "big and bad". This view would suggest that to the extent one tries to be value-neutral in the design of mechanism, one should not favor user empowerment. One could also argue that the fundamental design goal of the Internet is to hook computers together, and since computers are used for unpredictable and evolving purposes, making sure that the users are not constrained in what they can do is doing nothing more that preserving the core design tenet of the Internet. In this context, user empowerment is a basic building block, and should be embedded into all mechanism whenever possible. This paper suggests that the latter view is the defensible one, because choice is a basic tool to deal with tussle.

The recognition of tussle as a fundamental behavior does give one further hint at how to try to separate mechanism from policy. If one can find spaces where tussle are unlikely, then (as noted above) the interfaces and mechanisms can be simpler. If one can truly separate tussles, then one can do a better job of matching mechanism to problem. So the instruction to "separate mechanism from policy" is not incorrect, but just requires careful thought to carry out as best one can.

[12]An early articulation of the principle can be found in [8].

5. LESSONS FOR DESIGNERS

This more complex interpretation of old design principles, and the introduction of new principles, needs to be seen in terms of system synthesis. How can we, as designers, build systems with desired characteristics and improve the chances that they come out the way we want? If we try to design a system that is open, for example, which means we will encounter the tussles surrounding vertical integration and capture of value in exchange for investment, how can we proceed?

One can learn from the past. To some of us in the research community, a real frustration of the last few years is the failure of explicit QoS to emerge as an open end-to-end service. This follows on the failure of multicast to emerge as an open end-to-end service. It is instructive to do a postmortem on these failures.[13] Here is one hypothesis. For the ISPs to deploy QoS, they would have to spend money to upgrade routers and for management and operations. So there is a real cost. There is no guarantee of increased revenues. Why risk investment in this case? If the consumer could exercise effective competitive pressure in ISP selection, fear and greed might have driven ISPs to invest, but the competitive pressures were not sufficient. On the other hand, if ISPs use the new QoS mechanisms in a closed way, rather than an open way, they greatly enhance revenue opportunities. Thus, for example, if they deploy QoS mechanisms but only turn them on for applications that *they* sell, they reduce the open nature of the Internet and create opportunities for vertical integration. If Internet Telephony requires QoS to work, and they only turn on QoS for *their* version of Internet Telephony, then they can price it at monopoly prices.

One can thus see the failure of QoS deployment as a failure first to design any value-transfer mechanism to give the providers the possibility of being rewarded for making the investment (greed), and second, a failure to couple the design to a mechanism whereby the user can exercise choice to select the provider who offered the service (competitive fear). The argument about choice here is actually subtle. The user had the power to choose the level of QoS needed—that could be expressed in the ToS bits. What was missing was routing, to allow the user to favor one ISP over another if that ISP honored the bits.

- Anyone who designs a new enhancement for the Internet should analyze the tussles that it will trigger, and the tussles in the surrounding context, and consider how they can be managed to ensure that the enhancement succeeds. As noted above, a powerful force is the tussle of competition. Protocol design, by creating opportunities for competition, can impose a direction on evolution.

6. CONCLUSION

As the Internet evolves to become a full component of society, the person most likely to be dismayed is the fabled cypherpunk. [7] summarizes the cypherpunk view of privacy as follows: "[T]he cypherpunk's credo can be roughly paraphrased as 'privacy through technology, not through leg-

[13]The case study of the failure to deploy multicast is left as an exercise for the reader.

islation.' If we can guarantee privacy protection through the laws of mathematics rather than the laws of men and whims of bureaucrats, then we will have made an important contribution to society. It is this vision which guides and motivates our approach to Internet privacy." Our position is that the laws of men and the so-called whims of bureaucrats are part of the fabric of society, like it or not. They are some of the building blocks of tussle, and must be accepted as such. We, as technical designers, should not try to deny the reality of the tussle, but instead recognize our power to shape it. Once we do so, we acquire a new set of hard, technical problems to solve, and this is a challenge we should step up to willingly.

7. ACKNOWLEDGMENTS

The authors gratefully acknowledge essential and ongoing discussions with members of the NewArch project,[14] particularly Mark Handley, Noel Chiappa, Ted Faber, and Aaron Falk. Sharon Gillette, Jean Camp and the Sigcomm reviewers provided welcome comments and feedback, shaping our future work as well as this paper. Sally Floyd provided invaluable encouragement at a well chosen moment. Our sincere thanks to all.

8. REFERENCES

[1] Blumenthal, M. S., and Clark, D. D. Rethinking the design of the Internet: The end to end arguments vs. the brave new world. *ACM Transactions on Internet Technology 1*, 1 (August 2001). Version appeared in Communications Policy in Transition: The Internet and Beyond, B. Compaine and S. Greenstein, eds. MIT Press, Sept. 2001.

[2] Computer Science and Telecommunications Board, National Research Council. Realizing the information future: The Internet and beyond, June 1994.

[3] Computer Science and Telecommunications Board, National Research Council. Broadband: Bringing home the bits, January 2002.

[4] Computer Systems Policy Project. Perspectives on the national information infrastructure: Ensuring interoperability, 1994.

[5] Cross-Industry Working Team. An architectural framework for the national information infrastructure, 1994.

[6] Economides, N. The economics of networks. *International Journal of Industrial Organization 14*, 6 (1996), 670–699.

[7] Goldberg, I., Wagner, D., and Brewer, E. Privacy-enhancing technologies for the internet. In *Proceedings of IEEE COMPCON 97* (1997), pp. 103–109.

[8] Levin, R., Cohen, E. S., Corwin, W. M., Pollack, F. J., and Wulf, W. A. Policy/mechanism separation in HYDRA. In *Symposium on Operating Systems Principles* (1975), pp. 132–140.

[9] McKnight, L., and Bailey, J., Eds. *Internet Economics*. MIT Press, 1997.

[10] Nissenbaum, H. Securing trust online: Wisdom or oxymoron. *Boston University Law Review* (2001). Available as `http://www.princeton.edu/~helen/BU-final-trust.pdf`.

[11] Saltzer, J., Reed, D., and Clark, D. D. End-to-end arguments in system design. *ACM Transactions on Computer Systems 2*, 4 (Nov. 1984).

[14] `http://isi.edu/newarch`