The Growth of Internet Overlay Networks: Implications for Architecture, Industry Structure and Policy

Dave Clark Bill Lehr Steve Bauer Peyman Faratin Rahul Sami

Massachusetts Institute of Technology

John Wroclawski USC Information Sciences Institute

** September 8, 2005 **

** Preliminary draft. Please do not cite without contacting authors¹**

Abstract

Over the past several years, we have seen the emergence of numerous types of socalled "overlay" networks in the Internet. There are many diverse examples of such overlay networks including the content-delivery-caching networks, implemented by companies like Akamai, the peer-to-peer file sharing networks associated with applications such as BitTorrent, the voice-over-IP services offered via Skype, and various testbed networks such as PlanetLab. These overlay networks enhance or modify the basic functioning of traffic handling within the Internet. Overlays exist in the blurry boundary between what we think of as "the Internet" (a globally interconnected network of IP networks) and the applications that exist on top of the Internet. Overlays also blur the boundaries between the network edges (what we think of as being associated with customer end-nodes) and the network core (what we think of as associated with the services that support the Internet). As such, overlays have important technological and policy implications for the evolution of next generation Internet architecture that historically has been based on the so-called "end-to-end" principle (SRC84], [BC01]) which relied on a relatively clear demarcation between applications and network services, and edge and core responsibilities.

Because of the Internet's growing role as basic infrastructure and increasingly central role in the communications industry, and hence, obvious focus for regulation, changes in Internet architecture have important policy and industry structure implications. For example, from a regulatory perspective, the debate over overlays in Internet-space is analogous to the on-going debate over "layered regulation" (see [Fre02], [Sic02], [Wer02]) and how one might distinguish between "basic telecom services" (which may be regulated under Title II as

¹ Bill Lehr is the corresponding author for this paper (<u>wlehr@mit.edu</u>).

common carriers) and "enhanced" or "information" services (which are unregulated). From an industry structure perspective, overlays are relevant to the tussle over what services are provided by ISPs as opposed to other third party service providers (so-called ESPs) or by customer-managed layers.

Since these overlays are important to the future industry structure of our communications industry and its regulation, including issues of VoIP, intercarrier compensation, universal service, and market-power regulation, we need a clearer understanding of what overlays are and might be. This paper provides a first attempt to help frame the issue. Herein we introduce a taxonomy for thinking about these overlays with some examples of their scale and growing importance in the Internet, and we suggest some preliminary thoughts on the implications of these overlays for industry structure and policy.

Table of Contents

I.	Introduction			
II.	Towa	rds a taxonomy of overlays7		
A	. Wł	nat is an overlay?7		
	1.]	Fitting Overlays into the architecture of the Internet		
	2.	Overlay Functionality		
	3.	Overlay Industry Structure		
E	8. Wł	ny do overlays emerge?11		
C	C. Th	e challenges Overlays pose for policy12		
	1.]	Implications for Internet architecture12		
	2.	Commercial Implications13		
	3.]	Implications for public policy14		
III.	Te	chnical, Commercial, and Policy Challenges for Different Types of Overlays 15		
A	A. Co	ntent Delivery Networks (CDNs)15		
	1.]	Introduction15		
	2.	Description16		
	3.	Technical implications of CDNs18		
	4. 0	Commercial implications of CDNs19		
	a)	Industrial Organization of Internet19		
	b)	Market Acceptance of Third Party CDNs		
	c)	Costs to Third Party CDNs		
	d)	Impact of Third Party CDNs: Tussle Situations		
	e)	Why Third-Party CDN?		
	f)	Growth of CDNs		
	5. I	Policy implications of CDNs		

В	. R	outing Overlay Networks	28
	1.	Example of Overlay Routing	30
	2.	Technical Implications of Routing Overlays	32
	3.	Implications of Routing Overlays on the Interests of ISPs	33
	4.	Future Growth of Routing Overlays	34
С	. S	ecurity and Privacy Overlays	34
	1.	Introduction	34
	2.	Description	35
	3.	Providing anonymity with overlays	36
	4.	Providing censorship resistance with overlays	37
	5.	Providing deniability with overlays	38
	6.	Regulatory impact	39
	7.	Commercial Impact:	40
	8.	Technical impact	42
IV.	C	Conclusions and Directions for Further Research	43
	1.	The economics of overlays	44
	2.	Policy issues	45
	3.	Social issues	46
V.	Bibliographic References		

I. Introduction

The Internet started out as a government-funded research network running on top of the Public Switched Telecommunications Network (PSTN). The Internet was a data application, mostly unregulated, that was supported on top of the public-utility regulated telephone networks. The Internet was an "overlay" that complemented the underlying basic infrastructure of the PSTN by adding new functionality (packet-switched data network) to support the special needs of the research community (peer-to-peer computer communications). Most of the incremental investment in routers, servers, and access devices (PCs) was undertaken by new types of providers (Internet Service Providers or ISPs) and by end-users (Customer Premise Equipment or CPE) to complement the PSTN basic infrastructure already in place.

With the commercialization of the Internet in the 1980s and its emergence as a mass market platform for broadband communications in the 1990s, the Internet has evolved into the principal platform for our global public communications infrastructure. Increasingly, IP packet transport is providing the basic transport medium for telephony and other multimedia applications (voice, video, and data). What was an "overlay" application has now become basic infrastructure. Over time, the traditional PSTN providers have come to play a larger role in managing the infrastructure and investment required to support the Internet. Deregulation, market growth, and innovation have resulted in a more complex and interdependent Internet infrastructure ecosystem.

The success of the Internet owes much to the interoperability and connectivity supported by ubiquitous adoption of the IP protocols and adherence to the "end-to-end" (e2e) design principles that have governed Internet architecture for so long. However, the Internet's success has also posed significant problems. With growth has come heterogeneous services (not everyone needs or wants the same capabilities); new needs and requirements (support for real-time services or enhanced security); and complexity and size issues (arising from the sheer magnitude of today's Internet measured in terms of traffic and connectivity).

To meet these challenges, the Internet needs to continue to evolve. In a process that looks at times like history repeating itself, the Internet is now spawning its own collection of "overlay" networks. There are many types and examples of overlays (see Table 1) that arise to meet a range of purposes and needs (see further discussion in Section II). The emergence of these overlays raises interesting questions for the future of Internet architecture and the role of the Internet as a common platform for global communications. For example, are these "overlays" precursors to the future architecture of the Internet? Or, are they nasty barnacles on the Internet that threaten the e2e connectivity and interoperability that have proven to be such a key aspect of the Internet's value? What are the implications of overlays for industry structure and for the regulation of our public communications infrastructure? Before it is possible to answer such questions intelligently, it is necessary to have a better understanding of what constitutes an overlay, the motivations for their deployment and use, and the potential conflicts and tensions that may arise among stakeholders. The goal of this paper is to frame such a discussion and provide further thought on the implications of overlays for Internet architecture, industry structure/business strategy, and public policy. As our analysis demonstrates, the policy questions raised by overlays are multifaceted and diverse.

The balance of this paper is divided into four sections. Section II offers a preliminary taxonomy for thinking about overlays that reflects the rationale for their existence/emergence and provides further elaboration of the sorts of technical, business/economic, and policy questions that overlays raise. Section III illustrates these questions in the context of three examples of overlays (content delivery, routing, and security). Section IV provides summary conclusions and suggestions for further research.

Туре	Function/Purpose	Example
Peer-to-peer (p2p)	File sharing (e.g., mp3s)	Napster, Gnutella
Content- delivery (CDN)	Content caching to reduce access delays and transport costs	Akamai, Digital Island
Routing	Reduce routing delays, resilient routing overlays	Resilient Overlay Network (RON), Akamai SureRoute
Security	Enhance end-user security, privacy	Virtual private networks (VPNs), onion routing (Tor, I2P), anonymous content storate (Freenet, Entropy), censorship resistant overlays (Publius, Infranet, Tangler)
Experimental	Facilitate innovation, implementation of new technologies, experimentation	General purpose (PlanetLab, I3)
Other	Various.	Email, VoIP (Skype), Multicast (MBone, 6Bone, TRIAD, IP-NL), Delay tolerant networks, etc.

 Table 1: Examples of Overlay Networks

II. Towards a taxonomy of overlays

In this section we provide a taxonomy for thinking about overlays that includes examining the different motivations for why they emerge. This proves relevant when thinking about the challenges overlays pose for the evolution of technology, industry structure, and communications policy. Before considering these motivations and the types of challenges, however, we need to define what constitutes an overlay.

A. What is an overlay?

The brief description in the introduction and list of example overlays in Table 1 includes a diverse array of networks and network technologies that appear to exist "on top" of the infrastructure that supports the general purpose Internet while also appearing distinct from pure end-user or "distributed applications." The services range from the email system, thought of by many as part of basic Internet infrastructure (and so not properly an overlay except in a technical sense) to experimental networks like PlanetLab that exists as a general platform for deploying multiple overlays.²

As a starting point, we offer this definition of an Overlay:

An Overlay is a set of servers deployed across the Internet that

a) provide some sort of infrastructure to one (or ideally several) applications,

b) in some way take responsibility for the forwarding and handling of application data in ways that are different from or in competition with what is part of the basic Internet,

c) are operated in an organized and coherent way by third parties (which may include collections of end-users) to provide a wellunderstood service that is infrastructure-like, but,

d) are not thought of as part of the basic Internet.

The rest of this section will elaborate the various dimensions of this definition. As we will see, the boundaries of the definition are necessarily fuzzy, and what is interesting is not the precision of the definition but the different dimensions and considerations that it implies.

² Planet Lab is a large set of servers, distributed all across the Internet, which host programs that support applications of one sort of another. By itself, it has some of the attributes of an overlay it is a set of servers deployed in an organized and coherent way to provide a well-understood service. But the service of Planet Lab itself is not to forward or otherwise manage application data; instead it is a platform on which one can host programs that do exactly that. In other words, Planet Lab is best thought of as a highly distributed service that makes it easy to deploy new overlays without having to deploy new physical hardware across the Internet (see, e.g., [PAST05]).

To understand overlays, one needs multiple perspectives (architectural, industry/commercial, and public policy). To elaborate, we consider how Overlays relate to the rest of the Internet in a loose architecture sense (what it means to be "on top" of the general purpose Internet), the functionality that may be provided by Overlays; and how Overlays relate to industry structure (who owns what).

1. Fitting Overlays into the architecture of the Internet

We offered the notion of an Overlay as existing "on top" of the basic Internet, while being "infrastructure-like" in that the Overlay is a component of or input to the applications/uses that use the Internet infrastructure. From this, it is tempting to resort to the computer science conception of protocol layers (e.g., 7 Layered OSI Reference Model) and to view overlays as a "middle layer" above the basic IP protocols but below the application layer. However, this model addresses only one aspect of the network architecture. To see how Overlays relate to the end-to-end design principles that have governed Internet architecture historically, it is necessary to consider how overlays relate to the evolving architecture of the Internet.

The Internet, in its most simple conception, has two sorts of components, *end-nodes*—the computers at the edge of the net that play the role of servers, user machines, and so on, and *routers*, which forward the packets between the end-nodes. In this simple view, one can think of the Internet as a cloud of connected routers with end-nodes connected around the edge of the cloud.

Applications (for example email, the web, games, and so on) do not run on the routers. Routers know nothing about application-level functions. They just forward packets. So in this simple model, an application is just a program on one end-node that talks to another similar program on another end-node.

As you might guess, this simple conception is actually too simple to be realistic. In particular, applications are often more complex than this. In the case of email, for example, when one user sends mail to another, it goes by way of intermediate servers that have names like "the smtp server" and "the pop server". In the case of the web, there are web caches and proxies. And so on. A network purist might make a definitional distinction and say that since these devices are not routers, they must be some form of end-node, and so the simple conception still applies. But these devices have some important characteristics. First, they are distributed around the Internet in a way that provides an infrastructure on which the application runs. From the perspective of the router, they may be an end-node but from the perspective of the application, they are infrastructure. Second, they tend to be provided and operated by third parties. If two users exchange mail, they depend on servers operated by others to forward that mail. In this context, the term "Overlay" signals the emergence, in an organized sense, of this new sort of capability.

In a later section, we return to consider what Overlays may imply for the future of the end-to-end design principles.

2. Overlay Functionality

A second way to understand overlays is to focus on the extra functionality offered by an overlay, beyond what is supported by the basic Internet. In the context of our discussion here, the basic Internet functionality is defined by the suite of core Internet protocols (IP, TCP, BGP, UDP, etc.) that comprise the minimal set of basic protocols that any network or node must support in order to be considered part of the Internet (collectively, we will refer to these as "the basic IP protocols" except when the context requires a more specific use). The ubiquitous deployment of these basic IP protocols helps account for the Internet's great success and growth. The ability for the basic IP protocols to be supported on many different physical infrastructures (ATM, Frame Relay, SONET, wireless) and to support many different types of end-user applications (data, voice, video) services helps promote connectivity and interoperability across heterogeneous infrastructure and applications and has helped the Internet evolve into a global communications platform. This view of the Internet is associated with the "hourglass" model [CSTB94] where the basic IP protocols are located at the waist.

However, the basic functionality of "best effort" service offered by the standard IP protocols is not always enough. The original Internet architecture was designed to support (unicast) communication between fixed locations where the source knew the address of the destination. Yet many applications have more general communication needs such as mobility, multicast, anycast, etcetera. These communication needs present new challenges that the current Internet architecture does not support, such as when the source does not know the destination address (as in multicast and anycast) or the location of the receiving host is not fixed (as in mobile communications).³ Table 1 hinted at a range of functional extensions, including mobility, customized routing, Quality of Service, novel addressing, enhanced security, multicast, and content distribution.

In order to address these needs and overcome the barriers inherent in the existing infrastructure, Overlays blur the clean Internet architecture distinction between packet forwarding and application processing. Overlays, as opposed to application-specific network solutions, are increasingly seen as the mechanism of choice for introducing functionality into the Internet.

3. Overlay Industry Structure

For many, the Internet technology is a "black box." The details of technical design are not regarded as relevant in themselves, so long as the capabilities continue to evolve to meet the demands of growing markets for enhanced electronic communication capabilities. From this perspective, it might be less interesting to focus on "what the

³ As [SAZS02] point out in their I3 proposal, "all attempts to implement these more general abstractions have relied on a layer of indirection that decouples the sending host from the receiving hosts; for example, senders send to a group address (multicast or anycast) or a home agent (mobility), and the IP layer of the network is responsible for delivering the packet to the appropriate location(s)". However, "implementing these more general abstractions at the IP layer poses difficult technical problems and major deployment barriers".

Internet does or should do," and to focus instead on the industry structure (who owns what) that supports the Internet.

As noted earlier, the Internet began as an overlay on the basic telephone network, which was itself subject to substantial government regulatory oversight. As part of this oversight, telephone companies were limited in the range of services they could offer and were required to provide non-discriminatory (common carriage) access to all users. Thus, public policy and industry dynamics (as already discussed) gave rise to an ISP industry. This consisted of new types of communication service providers, ISPs, who leased physical infrastructure from the underlying phone companies, and combined it with packet switching technology (routers and servers) to support the Internet. The ISP business is quite competitive and many of the services offered have been largely commoditized. In addition to providing the basic packet transport services supported by the basic IP protocols, ISPs also provide a host of other commoditized services. For many who take the industry perspective, basic Internet infrastructure should be defined as the services that are provided by the typical ISP since that is who historically has provided basic Internet services.

This suggests that Overlays might be identified with offerings from "third parties." That is, Overlay services are not provided typically by ISPs, but new types of service providers that operate in conjunction with the basic Internet services offered by the ISPs. Akamai, a provider of CDN services, is an obvious example. A less clear example is when the third party is comprised of a group of end-users as is the case with a peer-to-peer network. Many interesting and emerging overlays (peer-to-peer, routing) are first deployed by edge users in end-nodes and may not generally be thought of as "infrastructure providers."

Moreover, as traditional industry boundaries blur, the definition of what constitutes an ISP becomes less clear, rendering a definition of what constitutes an overlay less clear. As ISPs seek to differentiate themselves (to escape the pressures of competition in a commodity market), they add services, vertically-integrating to become so-called "enhanced service providers." These include ISPs that offer data storage and back-up services as well as overlays (e.g., CDN or routing services). Additionally, other types of information technology infrastructure providers (e.g., Oracle, Microsoft) and physical infrastructure providers (e.g., Verizon, Comcast) are integrating into the ISP space.

In this changing environment, the notion that overlays are offered by some "third party" needs to be thought of as relative to the changing role of ISPs and what constitutes the "basic Internet." Overlays exist between that which is provided by ISPs as part of our global communication infrastructure and the applications that ride on top of the infrastructure. The Overlays are separate from the applications in the sense that they provide a type of "infrastructure." That is, something that is not specialized to a single class of users or application).

B. Why do overlays emerge?

Overlays emerge for a variety of reasons. First, overlays may exist to support the special requirements of a particular class of application or user community. To the extent that the needs of a particular user community are different from those of the general Internet, it may make sense to address this group's needs through specialized functionality/capabilities that are separate from but work in conjunction with the basic infrastructure supported by the general Internet. As the Internet becomes more pervasive across all segments of the global economy and becomes the general platform supporting all types of electronic communications, we would expect the need to address specialized, heterogeneous interests to increase. Thus, the success of the Internet as an open standard leads to the need to satisfy heterogeneous requirements that, in itself, provides one justification for an overlay. It is worth noting, however, that even in this case, drawing the boundary between what functions are really general and hence should be implemented as "basic infrastructure" and which are specialized and so may be appropriate as "overlays" is neither static (i.e., yesterday's niche might be tomorrow's mass market need) nor clear (i.e., are mail servers basic Internet infrastructure or part of an overlay?).

Second, and related to the above, overlays may play a role in the dynamic evolution of Internet technology. One of the great advantages of the Internet's end-to-end architecture is the ability to incrementally deploy/adopt edge-based innovations. Applications can be deployed virally by a growing number of edge-nodes without requiring modifications to the basic Internet. The ubiquity with which TCP/IP are implemented provides a stable platform to support communications among and across heterogeneous edge-nodes. However, the very benefit of ubiquitous availability becomes a challenge when it comes to upgrading the Internet's own basic infrastructure. Coordinating the updating of all of the routers and servers that support the basic Internet represents a massive undertaking even if everyone agrees that an upgrade is needed and agrees on the nature (technology choice) for the upgrade.

Overlay networks can provide a way to first experiment with new routing and architecture designs (e.g., PlanetLab) and then as a way to incrementally deploy new solutions. Functionality that is missing in the current Internet may be deployed first in an overlay for those users/uses that most require (and are willing to pay for) enhancements that may not be available yet in the general Internet. This can include such things as enhanced quality of service (e.g., reduced delays from better routing) or security/privacy (e.g., onion routing to protect identity). Over time, successful innovations will become ubiquitously adopted and, as such, *de facto* components of basic Internet infrastructure. Whether the functionality offered by an overlay is viewed as an enduring specialized need (static) or simply the early version of functionality that is generally needed and will be deployed in basic infrastructure over time (dynamic) may be a question of perspective.

Third, overlays may arise because of conflicts in stakeholder interests, reflecting a tussle between and among customers, service providers, and policy-makers. For example, the missing functionality (*e.g.*, privacy, caching support for content delivery, or delay-minimizing routing) may be intentional. Routing overlays that seek to improve on the

basic Internet route selection process (*e.g.*, BGP), may be in conflict with policy-based routing implemented by peering ISPs in response to other, non-delay-related considerations (e.g., long-term interconnection agreements or regulatory-jurisdiction issues). Hence, an overlay that tries to select the "best" route based on global information about link delays may violate business agreements about traffic routing between ISPs that are seeking to manage traffic to minimize intercarrier payments. Or, overlays that implement privacy (obscure the source, content, or type of traffic) might be in conflict with public policy rules that seek to make traffic auditable by the police (e.g., CALEA) or to support carrier efforts to price discriminate differentiated services (e.g., price voice calls higher than data).

The forces that lead to each of these rationales (heterogeneous interests, dynamic evolution, and tussle) are fundamental and enduring and so we should expect that overlays will remain an important and growing feature of the Internet landscape.

C. The challenges Overlays pose for policy

Overlays pose an interesting challenge for policy in multiple dimensions. They raise important questions for the evolution of Internet technology (architecture), business/industry structure (how do overlays impact infrastructure costs and who owns/controls what investment?), and regulatory policy (how do overlays interact with open access, interconnection, or other basic infrastructure policies?). Before turning in Section III to a detailed discussion of these challenges in the context of specific types of overlay networks, we highlight some of the generic issues that overlays pose for technical, economic, and policy analysis.

1. Implications for Internet architecture

To understand how overlays pose a challenge for the evolution of Internet architecture, it is worth considering overlays in the context of the Internet's "end-to-end" principle.⁴ This principle has entered into policy debate of the desired nature of the Internet, so it is worth asking what Overlays imply for end-to-end. The end-to-end principle can be thought of as operating at two levels in the Internet, the packet level and the application level. At the packet level, the end-to-end principle leads to the original design of the Internet where the routers know nothing about applications, but just forward packets, and knowledge of the applications is confined to the end-nodes. This view is consistent with the simple model of the Internet we offered in our definition of overlays. But at the application level, the end-to-end principle could be interpreted as leading to an application architecture in which data is transferred directly and without intervention

⁴ The "end-to-end" design principle specifies that processing of the communications protocols ought to occur in the end-nodes. It suggests that application-specific functionality should not be built into the supporting network, but should be located in edge-nodes. This allows heterogeneous edge devices and applications running on them to share a common communications infrastructure that supports peer-to-peer communications. See [SRC84] or [BC01] for further discussion. See [LL00] or [Ise97] for discussion of policy rationales for preserving the principle in the Internet.

from the original source to the ultimate destination. By this definition, most real applications today are *not* consistent with end-to-end.

Consider the email example we discussed above. Normally, mail goes from sender to receiver via two servers, an smtp server and a pop server. There is no end-toend confirmation that the mail actually gets to the receiver—the sender and receiver depend on the servers to be reliable. The example of careful file transfer from the original end-to-end paper [SRC84], which involved confirmation from the destination to the source that the data was delivered correctly, is totally missing from email. From the application perspective, email is *not* consistent with end-to-end. Some applications do not involve any intermediate servers and services at all, and are consistent with a very simple version of end-to-end, some applications do involve intermediate services but do use end-to-end checks to confirm that the application is working properly, and many don't have this form at all.⁵

Using this framework, overlays that support one or a class of applications do not in themselves erode the end-to-end principle. They provide services in the Internet, but that is a consequence of the application design, not the overlay. Overlays that provide a new form of the basic Internet service, most specifically routing overlays, also do not erode the end to end principle. They signal that the user wants to purchase a service that the ISPs do not want to offer them, and thus signal a possible market failure, or alternatively, a conflict among market participants. Thus, the architectural implications of overlays vary depending on the type of overlay and the context in which one is examining the question.

Thus, even determining whether overlays enhance or undermine the end-to-end design principle is problematic and contentious.

2. Commercial Implications

Overlays have commercial implications because they may impact infrastructure costs. One of the benefits of ubiquitous adoption of a common technology are scale and scope economies. The global market for equipment and support services for basic Internet technology helps keep markets competitive which keeps costs low and expands consumer choice. The global network also realizes large positive network externalities. Overlay networks pose a threat of fragmenting the market if they lead to islands of incompatible and different overlay technologies and bundles of functionality. If this occurs, it may lead to reduced scale/scope economies and reduced network externalities. In the most extreme case, the proliferation of heterogeneous overlays may threaten Internet connectivity. Furthermore, overlays may result in redundant investments in overhead/functionality. For example, using an overlay to address routing deficiencies in the basic Internet may be more costly than upgrading the basic routing capabilities.

⁵ In contrast, Voice-over-IP is an example of a design with a very high-level end-to-end error check—if the message is garbled the human listener can say, "What?"

However, the need to upgrade implies adjustment costs and, perforce, invokes dynamic considerations.

As discussed, overlays can play a role in reducing dynamic adjustment costs of introducing new functionality to the Internet. Moreover, overlays which address specialized interests even at the expense of higher costs may still enhance total surplus (e.g., the overlay may support the price discrimination needed to recover the costs of meeting the enhanced service needs of a specialized group of customers).

In addition to its impact on general infrastructure economics, overlays are important to industry structure. As noted earlier, basic Internet infrastructure has historically been associated with ISPs and with customer-owned CPE. The growth of the Internet, deregulation, and industry convergence (telecoms v. cable, incumbents v. new entrants) have resulted in significant changes in the vertical and horizontal structure of service providers. For example, telephone and cable television companies now own and manage core Internet infrastructure like routers and servers that previously were owned and operated by independent ISPs. In the on-going competition among firms across the value-chain there is a tussle over what functionality is controlled where. For example, peer-to-peer overlays rely on edge-centric intelligence which is a substitute for networkcentric intelligence controlled by service providers. Also, service providers seeking to soften the impact of commodity competition try to differentiate their service offerings with enhanced features.

3. Implications for public policy

The tussle over Internet technology and industry structure leads directly to a tussle over public policy. In recent decades, industry convergence has been challenging the traditional silo-based model of communications regulation under which cable companies are treated differently from telephone companies, mobile providers from fixed line providers, and basic service providers from enhanced service providers. These regulatory distinctions are increasingly problematic in the emerging world of converged, broadband platform services. They impose asymmetric regulatory burdens and costs that distort prices and investment incentives.

One solution that has been proposed to this dilemma is to migrate toward a layered model of regulation ([Sic02], [Wer02], [Fre02]). While such a strategy may prove appealing, overlays challenge possible definitions of what constitute appropriate layers.⁶ Alternatively, overlays may be used to circumvent communication (e.g., access charges

⁶ For example, if "applications" are unregulated but "basic infrastructure" is, then policymaker needs to determine what constitutes an application as opposed to basic infrastructure. The recent Supreme Court "Brand X" decision confronted this issue in its determination that broadband access services offered by a cable television company are not a telecommunications service. And, even when the regulatory distinction can be made, enforcement is difficult when the technical architecture blurs boundaries. For example, there are many ways to deploy VoIP or DRM services that makes is challenging to craft a simple policy to enforce the phone tapping rules (CALEA) or copyright protection.

or CALEA)⁷ or other policies (e.g., bypass intellectual copyright protection or censorship rules).⁸ Finally, because overlays challenge industry structure, they impact competition policies (e.g., open access rules)⁹. For example, asymmetric access to the services of a content-delivery network may result in preferential access to some content at the expense of others. Whether this is a problem or not may depend on ones perspective (e.g., is it a specialized service to serve a specialized community that is paying for the incremental services provided, or is it a violation of common carriage non-discrimination rules).

III. Technical, Commercial, and Policy Challenges for Different Types of Overlays

In the following three subsections, we examine the technical, commercial, and policy challenges posed by three different types of overlays: Content Delivery Networks (CDNs), Routing Overlays, and Security Overlays. For each example, we provide a description of how the overlays operate and then identify issues raised by the growth of such overlays.

A. Content Delivery Networks (CDNs)

1. Introduction

The first class of network overlays we examine is Content Distribution Networks (CDNs). CDNs are overlay networks that dynamically cache content and services at distributed locations throughout the Internet. They are interesting overlays to examine because they represent a large share of the overlay traffic on the Internet today, and are associated with commercial offerings from Akamai and with popular peer-to-peer caching overlays such as BitTorrent.

CDNs are technologically fairly straightforward. When an application requests content or services hosted by a CDN, the CDN overlay services the request from one or more of the distributed servers throughout the Internet. Selection of the servers in the overlay to handle the request can depend upon multiple factors including the load on each server, which servers are topologically nearby, and the economic cost associated with servicing the request from each server. (We discuss in more detail below the technical mechanisms by which this is accomplished.) CDNs are overlays because the IP layer is

⁷ VoIP has been used to bypass regulatory-mandated intercarrier payments both domestically and internationally.

⁸ A key driver for the growth of peer-to-peer file-sharing programs like Gnutella or Napster was the opportunity to trade MP3 files in violation of copyright rules.

⁹ For example, a key policy for implementing open access in telephone services was structural separation of long distance and local telephone services. Local telephone providers were required to provide non-discriminatory interconnection to all long distance providers, thereby enabling the growth of competition in long distance. In the US, this was supported by restrictions that prohibited the largest local telephone companies from offering long distance services. Once local companies are allowed into long distance, they have an obvious incentive to discriminate against non-affiliated carriers.

responsible for delivering the packet to the appropriate destination but the decision about the source of packets is made at the application layer by the redirector, and not the original requestor.

While technologically straightforward, the impact of CDN overlays raises many interesting questions as CDNs evolve and grow. Fundamentally, CDNs change the patterns of traffic on the Internet, so they have a clear technical impact. Because traffic patterns also determine money flows between providers, CDNs influence the commercial relationships on the Internet, which in turn, gives rise to policy implications. For example, if CDNs provide superior access selectively to some content, would this give rise to a two-tiered Internet: one that is high quality for commercial content and one that is lower quality for non-commercial content? If so, would this raise concerns about equal, non-discriminatory access or free speech? Suppose an ISP sought to vertically integrate with a major CDN provider like Akamai. Would that raise antitrust concerns?

As we explain below CDNs arose because of an unmet need of end-hosts for lower latency access and delivery of content, as well as a desire to reduce the transport costs of content and Internet service providers. While CDNs have both technical and policy impacts, we focus on the commercial impact they give rise to because that is the one that we think is likely to be most salient in the near term.

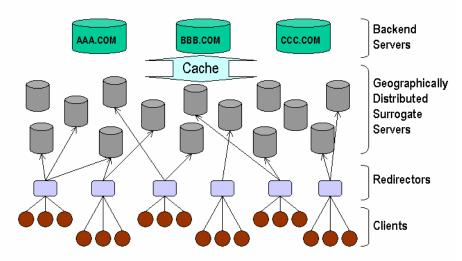
2. Description

CDN overlays address a fundamental challenge on the Internet – how to cost effectively distribute and acquire content while simultaneously lowering latencies experienced by end-hosts.

At a technical level CDNs consist of caches of content and services distributed across the Internet (see Figure 1). These caches contain copies of content and services retrieved either on-request or proactively from publishers and providers. The heart of a CDN is the method by which requests and content are routed and redirected in the overlays to accomplish the load balancing. Example ways of accomplishing this routing and redirection in distributed CDNs include DNS and URL rewriting and http redirection. In balancing the request and content load, CDNs optimize different criteria including technical measures such as response time and server loads and economic measures such as bandwidth costs.

Figure 1: Content Distribution Network (Source: Limin Wang presentation, from http://www.cs.princeton.edu/~lmwang/cdn/)

Content Distribution Networks



CDNs fall into three distinct categories 1) commercial 2) cooperative and 3) peerto-peer based overlays. Commercial CDNs distribute dynamic services such as on-line airline reservation applications and static content such as patches to Microsoft software products. Currently the largest commercial provider of CDN services is Akamai which claims to serve approximately 10-15% of web content and collocates with around 10,000 ISPs globally.¹⁰

Cooperative CDNs such as CoralCDN and OpenCDN seek to offer similar benefits to non-commercial users. Hardly surprising, the performance of these is often not as good since the cooperative CDNs rely on infrastructure that is contributed voluntarily. They do however offer the potential for content publishers to reach a larger audience and sustain service during larger flash crowds than would be possible from a resource limited server. Any web publisher for instance can "Coralize" their web URLs and cause their content to be cached in the CoralCDN by appending nyud.net:8090 to their URLs (e.g. rewriting http://www.x.com.into http://www.x.com.nyud.net:8090).

Finally many of the peer-to-peer overlays function as content distribution networks. Peer-to-peer content distribution networks differ from cooperative caches in terms of functionality (often including for instance search capabilities), content (larger percentage of files likely to raise issues of copyright infringement), and overlay structure (majority of the nodes are both servers as well as clients whereas in the cooperative caching networks many nodes are contributed purely altruistically and serve only as content caches.)

¹⁰ Estimate provided in private conversations with Akamai personnel.

Currently a significant amount of content is being served from the decentralized peer-to-peer caching overlays. BitTorrent, in particular, has been a popular distribution network in the past few years. Notably, BitTorrent first gained popularity as a distribution channel for Linux software distributions. BitTorrent remains vital to the distribution of content from a number of popular publishers that rely on the peer-to-peer CDN to lower their distribution costs. Some (much of?) this content might not be available in the absence of these cost savings. Thus, peer-to-peer content distribution may lower the costs of accessing diverse content (a public good!) while at the same time providing a platform for copyright infringement (a policy challenge!).

3. Technical implications of CDNs

Unlike routing overlays which do not alter the communication pair addresses (source/destination), a CDN dynamically changes the communication pair by redirecting communications to different destinations. Interestingly, CDNs -- unlike routing overlays -- respect the clean end-to-end architectural distinction between packet forwarding and application processing.¹¹ Thus, the CDN may be seen as an overlay infrastructure on top of the IP layer that supports multiple (web content and increasingly processing) applications. It may be argued, therefore, that CDNs enable scalable and architecturally sound methods for the distribution of content and services. In many ways they are essential to cost effectively dealing with traffic distributions that may be heavy tailed (i.e., have rare but occasional peak loads that are very large). CDNs enable service and content distribution costs to be shared among multiple providers, facilitating the distribution of the peak load associated with the heavy tailed traffic.

CDNs shift traffic patterns so that more content and services can be accessed locally. This benefits both ISPs on the receiving end and content publishers which do not have to pay the expensive transit traffic costs arising from redundant requests.¹² While efficient caching can help lower overall transport costs, how these savings are distributed can result in commercial tussles as we discuss further below.

The growth of CDNs also has implications for capacity planning and where infrastructure investment is likely to occur. To the extent CDNs result in a larger proportion of traffic being served locally, incentives to and the need for investments in long-haul end-to-end capacity will be reduced. Because such capacity, once in place, may be shared and may be largely sunk, this may shift the opportunity costs for accessing long-haul transport for services that are not part of the CDN. Additionally, because CDNs cost money to support, they will shift spending from transport to caching and access to

¹¹ That is, the IP layer is responsible for delivering the packet to the appropriate destination but the decision about the destination is made at the application layer by the redirector, and not the original requestor.

¹² For example, when many users in Australia wish to view a web page that may be stored on a CNN host computer in Atlanta, it is cheaper to download the relevant page once to a cache in Australia and then serve subsequent requests for the same information from the local cache instead of pulling the content again all the way from Atlanta.

such services may require additional payments (for example, as is currently the case with commercial caching services offered by Akamai). Those services or content that participate in a CDN may benefit from better performance (lower latency) and lower-cost access (because of the efficiency benefits of caching). Hence, in the future, CDNs may contribute to the creation of a two-class Internet, one that is high quality for commercial content, and one that is lower quality for non-commercial user-centric content.

Furthermore, continued growth of CDNs may raise questions regarding the value of application-specific (e.g., static content caching) vs. the basic Internet, resulting in reduced scale or scope economies for the Internet. Or, the proliferation of heterogeneous CDNs targeting different content may threaten the end-to-end connectivity which has contributed the growth of the Internet.

Finally, at a technical level, it is worth noting that public policies like police access to electronic communications (e.g., under CALEA) and intellectual property rights enforcement (e.g., Digital Rights Management) may impose design constraints on how CDNs are implemented. For example, implementing a wire-tap in a dynamically changing CDN could prove quite challenging, and especially difficult if it were a peer-to-peer or cooperative CDN (as opposed to one controlled by a single commercial provider like Akamai).

4. Commercial implications of CDNs

As mentioned above, CDNs may be implemented in a number of ways an the commercial implications are likely to vary depending on how the CDN is implemented (commercial, cooperative, or peer-to-peer). Due to space considerations and because there is already a significant literature on the implications of peer-to-peer systems (e.g., BitTorrent) we will focus on the implications of commercial CDNs like the one offered by Akamai.

a) Industrial Organization of Internet

We canonically represent the Internet by three hierarchical tiers of transport providers (see Figure 2): *Access providers*, who connect end-hosts to the global Internet, *regional providers*, who in turn connect access providers to the Internet backbone, and *backbone providers*, who have global reachability and connect transit providers to other regional transit and access providers. In this simplified example, the Internet is used to distribute content from a content provider, CNN, to an end-customer.

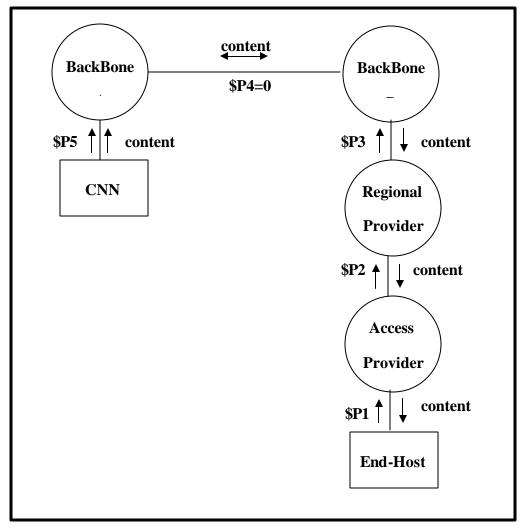


Figure 2: Industrial Organization of the Internet Transport

In this simplified example, there are two sources of revenue to the system: the content provider, for example CNN, who pays to have its advertiser-supported content served for free to consumers located around the globe (\$P5 in Figure 2); and the end-host, or consumer who pays a local access provider for Internet access (\$P1 in Figure 2) which is valued in part because it allows the user to access news media and other "free content" like CNN.¹³ Moreover, in this simple example, the traffic is very asymmetric. The end-user sends a request for CNN content which is then delivered downstream (requiring an order of magnitude more bits to deliver), hence we assume the traffic flows from CNN to

¹³ With paid-content, there may be additional revenue flows from end-user to the contentprovider. For example, the on-line Wall Street Journal provides charges a subscription fee directly to users that is billed via a credit card. Other publications like the New York Times provide a mix of free and paid content.

the end-user.¹⁴ The typical end-user pays access and usage fees to its local ISP. Although the usage fees might be metered with respect to volume of traffic (MB), peak or average traffic rates (Mbps), flat monthly "subscription" charges are common (e.g., \$40 per month for a DSL line with a maximum peak traffic rate of less than 1Mbps). The local ISP pays volume-based transport fees to its regional ISP (\$P2 in Figure 2), which in turn, pays volume-based transport fees to its backbone ISP (\$P3 in Figure 2).

The backbone providers, however, do not pay each other to exchange traffic. Under a "bill-and-keep" arrangement the assumes that traffic flows will be approximately symmetric, the price for exchanging traffic is set to zero (\$P4=0 in Figure 2). This is logical when the traffic volumes are roughly symmetric between them since it economizes on metering costs and the net payment flow would be zero, in any case, if traffic is appropriately balanced.¹⁵ This is most likely when the traffic flows are highly aggregated as is the case in the "best efforts" (single class of service) Internet. Finally, content providers can collocate at any level of the connectivity hierarchy but in general pay for their content to be hosted on the collocated ISP, e.g., CNN (P5 in Figure 2).

In this example, we assume that the prices paid for access by end-users, contentdistribution by content providers like CNN, and for transport by regional and access ISPs are competitive. It is worth noting that the transport contracts between successive levels of the hierarchy are asymmetric. Access providers pay transit charges to the upstream regional providers who do not pay termination charges to the access providers. Regional providers in turn pay backbone providers who do not pay regional providers for termination. This makes sense as long as the traffic flow is asymmetric, but this depends on where content providers and their subscribers are actually located. As long as most

¹⁴ Although because end-users might be located on many different access networks, including those directly connected to back-bone providers, the traffic flow between any two interconnecting ISPs in the hierarchy may be symmetric over time.

¹⁵ Under the Telecommunications Act of 1996, incumbent carriers were required to negotiate reciprocal compensation agreements with interconnecting competitive local carriers. Because these rates were initially set substantially above economic costs, they gave rise to a significant incentive to arbitrage the interconnection payments by creating situations with asymmetric traffic patterns. Thus, a number of competitive exchange carriers signed up ISPs as customers in order to benefit from reciprocal compensation payments from the incumbents. These arise when dial-up ISP customers called their ISP, making a local call that originated on the local incumbent carrier's network but terminated to the ISP which was a customer of the competitive local exchange carrier. Because Internet calls are typically of much longer duration than regular voice calls, this resulted in a significant imbalance of excess terminating minutes going to the competitive carrier. The incumbent carriers renegotiated much lower interconnection charges with the competitive exchange carriers and sought to deny reciprocal compensation payments owed on the grounds that these were not the sorts of calls that the Telecommunications Act of 1996 anticipated in its interconnection rules.

content providers are located on backbone providers and most of the traffic is associated with this content, this is not an unreasonable assumption.¹⁶

b) Market Acceptance of Third Party CDNs

In this section we will demonstrate the rationale for third party CDNs, given the canonical connectivity pattern shown in Figure 2. First, observe that the content provider, access provider and the end-users all have an incentive to reduce the latency and increase throughput of content delivery through caching. We distinguish between two types of content providers: static and dynamic content providers. Static content providers (such as Microsoft patch updates) have an incentive to optimize their costs in large volume global distributions. Dynamic content providers (such as CNN or Hilton Hotel) on the other hand are interested in increasing the probability of a successful revenue generating transaction -- lower latency leads to a better user experience. For this reason, dynamic content providers are willing to pay a surplus for lower-latency content delivery. Access providers also have an incentive to cache because of reduced transit costs to regional ISPs (P2 in figure 2). End-hosts benefit from access provider caching because they experience lower latencies and higher throughput in content transfers.

However, given the current industrial organization of the ISPs, realizing the potential value gain from caching was not possible without a CDN. The combination of asymmetric pricing of transport (payment flows upwards, and not downwards, in the connectivity hierarchy) and "revenue neutral" bill-and-keep interconnection contracts between backbone ISPs, there was no mechanism for transferring the willingness-to-pay for improved content distribution from content-providers to the downstream ISPs on the other side of the peering point. For example, in Figure 2, suppose that CNN stands to gain significantly from having its pages cached at access providers (e.g., lower payments to Backbone provider A for volume-based content delivery charges and improved performance – lower latency – realized by CNN subscribers). Access provider D or C both have incentives to cache content to lower transit charges they would otherwise owe to upstream providers but they do not have a ready mechanism to allocate caching responsibilities between them.¹⁷ Moreover, since the transit charges relate to the volume of traffic but do not distinguish among content-sources, there is no way for CNN to pay more to have its content cached relative to the content of other potentially less-valuable content sources. Although CNN could negotiate with each access provider individually to provide caching services, this would likely entail significant transaction costs.

In addition to resolving the quandary arising from the asymmetric interconnection agreements in the Internet, CDNs can also benefit content providers in several other

¹⁶ However, if more of the content end-users want is served by end-users as part of various peerto-peer user communities, this may no longer be a valid presumption.

¹⁷ The problem becomes even more difficult if transport charges are not competitive (e.g., backbone, regional, or access providers have market power) or in the face of significant sunk/fixed costs (i.e., incremental termination costs are significantly below long-run incremental termination costs).

ways: 1) decreased server load, since the content server is not responding to each request individually; 2) decreased capacity planning risks: related to the above point, since source requests are served through CDN servers, rather than original destination, then the content provider does not need to make costly capacity planning decisions and investments, especially in the face of uncertain demand for content; and, 3) increased security: CDNs, given the economy of their scale (with servers distributed throughout the globe), can offer valuable security services to content providers against attacks such as Distributed Denial of Service (DDoS) attacks. Increased reliability also increases end users' satisfaction because DoS attacks don't bring down the content they are trying to access. These added benefits provide another source of potential willingness-to-pay that might be captured.

The potential to capture the surplus associated with improved content delivery creates an economic opportunity and provides the motivation for a third-party to enter to solve the "money routing" problem posed by the current Internet's industry structure.

c) Costs to Third Party CDNs

Although there are benefits to be realized, a third-party CDN also must incur capital and operating costs. Major factor input costs to deliver CDN services include capital, management and transport costs of content. CDNs currently sign collocation contracts with all access, regional and backbone ISPs. A CDN pays collocation fees with larger scaled ISPs. These, collocation prices are volume-based, vary from one to three years, and are independent of the services provided by the CDN. In addition, a CDN typically pays these ISPs based on their traffic volume; in other words, they are largely treated as any other customer of that ISP.

d) Impact of Third Party CDNs: Tussle Situations

CDNs can result in a large efficiency gain, economically as well as in performance. Their introduction can provide widespread benefits, but can also potentially have a negative impact on certain players in the content distribution game. In general, because of the efficient caching at access ISPs, regional and backbone ISPs will carry less traffic and face lower revenues. Peering backbone ISPs may also face lowered aggregate traffic volume across their peering points. Importantly, whether traffic reductions at the peering point are symmetric -- thereby maintaining the peering contract condition on which "bill and keep" is based -- is conditional on the presence of CDNs on *both* sides of the peering point. Thus, the CDN (assuming it carries a significant fraction of Internet traffic) can potentially unravel the peering relationship *or* help correct existing imbalances.

In general, whether ISPs benefit from third-party caching is dependent on a number of conditions that include transfers made from the CDN provider to the ISPs, the pricing regime, and the cost of caching (which in turn depends on the distribution and scale of the CDN, as well as the content itself¹⁸). For example, if the average cost of

¹⁸ For example, how dynamic is the content and how much and widely is it accessed?

caching content is much smaller than the cost of transporting the content from the original server, or if there is tremendous value gained in serving the content at a low latency, then there is a clear benefit to caching, and it should be possible to distribute the resulting surplus among all the parties involved. At the other extreme, if caching is costly and does not add value, then we would expect that content not to be cached. However, this is not necessarily true always; one interesting scenario that could arise is when the cost of caching is higher than the cost of hosting and transport, but is smaller than the total *price* that regional or backbone providers charge the access provider (for transit) and the content provider (for hosting). In this case, the access provider will be glad to have the CDN in its network, and the content provider will be happy to use the CDN service to cache its content. The content will be cached even though it is not economically efficient. The CDN itself can extract a share of the content owners' surplus.

This last scenario highlights another role that the CDN can assume in the current industry structure: that of a broker who *arbitrages* prices across the core of the Internet. The existence of a CDN thus prevents the prices for hosting and transit charged by the core ISPs from getting too high. The overall effect of the CDN entry is to reduce the pricing power of the core ISPs, and increase the power (and surplus share) of the access providers and the content owners.

e) Why Third-Party CDN?

Since offering CDN services has the potential to meet a real customer demand (for content providers and their customers seeking lower cost, higher quality distribution services), it is worthwhile asking why ISPs do not assume this role directly? It turns out there are a number of possible reasons why ISPs do not provide CDN service themselves. For example, negotiating inter-provider quality-of-service (OoS) agreements has proven challenging for providers. While ISPs have been providing differentiated QoS Service Level Agreements (SLAs) for on-net customers for a number of years, it remains difficult to provide such services across multiple provider networks. The basic technologies to support these exist (e.g., MPLS, IntServ, DiffServ, etc.), but appropriate commercial arrangements to implement these have yet to be developed.¹⁹ This is especially true for delay-inelastic applications such as streaming video that require more than a "best-effort" transport service. Implementing an end-to-end QoS-differentiated SLA requires not only standardization of technical interfaces and algorithms but more importantly mechanisms for revenue and cost sharing. CDNs, which reduce the need to transfer traffic end-to-end, help reduce the impediments to low-latency traffic caused by the absence of appropriate industry SLAs.

The coordination costs are further exacerbated by the need to maintain current transit and peering contracts. If one ISP caches, this could upset the symmetry of traffic on which current peering arrangements are based. Caching by one creates a symmetric need for caching by the other. Because of the costs associated with implementing a CDN

¹⁹ Work to develop such standards is currently underway (see, for example, <u>http://cfp.mit.edu/qos</u>).

and the risk it may pose to existing peering arrangements, both ISPs may find it preferable to not implement caching.

Furthermore, and particular to certain access ISPs, there is the concern that common carrier/open access regulations might make it difficult for such ISPs to offer CDN services except on an open access basis. For example, if CDN services were deemed a basic telecommunications service then they might be subject to the wholesale open access rules implemented under Title II of the Communications Act. With the recent Supreme Court "Brand X" decision (which determined that cable modem service was not a basic telecommunications service) and with the current trend in the FCC's enforcement of the Telecommunications Act of 1996 (e.g., its recent decision to classify DSL as an information service), regulatory barriers to access ISPs providing more CDN or other enhanced Internet-related services have been reduced.

Finally, the lack of caching may induce content-providers and end-users to invest in higher capacity access links that may be lower cost to support and may produce spillover benefits for other network services.²⁰

f) Growth of CDNs

As noted earlier, CDNs are the type of overlay associated with the highest volume of traffic in today's Internet. The largest commercial CDN (Akamai) is alleged to account for up to 15% of Web traffic by itself, and peer-to-peer CDNs²¹ collectively also account for a large (but difficult to measure) volume of traffic. The clear commercial justification for CDNs may help explain why these overlays developed relatively early and have grown relatively large. Consequently, CDNs may provide insight into the lifecycle patterns of other overlays.

The history of CDNs mirrors the history of other overlays which are generally much younger in their lifecycle (e.g., routing or experimental overlays). Content and services were initially hosted on end-user machines. Users tended to operate as both content authors as well as system administrators publishing their own content. ISPs eventually began providing web hosting as a service and content hosting became a significant commercial activity. For users with popular sites and many downloads though, the cost of hosting content can be significant. Payments to ISPs for hosting popular sites can be in the tens of thousands of dollars a month. This created the economic incentive for content providers to look for ways of reducing their costs. CDNs arose as a response to this need.

²⁰ Note, under the assumption that ISP rates are competitive, and hence, reflect economic costs, the only benefit from inducing customers to shift to higher capacity links is to shift costs from network providers to end-customers.

²¹ Growth of the peer-to-peer CDNs like BitTorrent is fueled in part (perhaps, mostly by) the desire by end-users to access copyright-protected media content without having to pay the rights holders. This provides another source of "value" that CDNs can capture.

This growth is in part because there is a universal demand by both end-hosts and content providers for better than best-effort service offered by the IP network. Furthermore, adoption of such networks does not involve end-hosts technologically or financially. Users' web requests are transparently redirected to closer CDN servers and the costs are born by ISPs and content providers and not the users. In contrast, general demand for routing and other overlays is currently much smaller. Furthermore, as discussed elsewhere in this paper, the tussle between routing overlays and ISPs is much more pronounced, thereby increasing the complexity of the legal and technical responses required of ISPs. Therefore CDNs are relatively more likely to enjoy faster growth and positive network externalities than other types of overlays.

Another consequence of this growth of content networks is the incremental introduction of latency competition between content providers. Currently there is no real competition to deliver lower latency content. However, CDNs could be seen as a mechanism that induces content providers to compete not only on content but also on latency because of new opportunities for product differentiation. Users may switch between different content providers because of the perceived latencies. For example, users may choose among competing media sources (e.g., CNN or the New York Times) based on their perceived latency if one is served via a CDN and the other is not. The perceived latency becomes another dimension in the user's preference function to consider when choosing which content to access.²²

In the future, it is possible that third-party CDNs will continue to grow. Whether the market for such services will be competitive or not remains to be seen. Currently, as already noted, Akamai is the largest provider by far. There are scale and scope economies and network externality benefits associated with operating a global CDN. It remains to be seen whether Akamai will continue to strengthen its lead or will face stronger competition in the future. Additionally, ISPs may become more directly involved in providing CDN services. For example, the challenges to offering inter-provider QoS SLAs are being worked out. Additionally, consolidation among ISPs and the emergence of mega-access providers like Comcast, Verizon, or SBC may make it feasible for their access ISP subsidiaries to contract directly with large content providers (e.g., Disney). By combining the CDN functionality with other access operations these providers may realize additional scale and scope economies. Even if vertical integration does not make sense from a cost perspective, an access provider might seek to vertically integrate to enhance its ability to differentiate its offerings, or if the CDN market is imperfectly competitive, to leverage the market power of a CDN into local access services.

²² With equivalent distribution technologies, the choice between media depends more on editorial content (e.g., New York Times v. Boston Globe); however, with asymmetric distribution media the attributes of distribution become important factors also (e.g., USA Today on-line v. ABC on-line with streaming media).

5. Policy implications of CDNs

CDNs reduce the network load and improve efficiency, and benefit participating content providers, users, and ISPs. In contrast to the case of routing overlays, they are not in direct competition with ISPs, because they provide a service that ISPs generally do not offer. However, if CDNs grow to the point that a large part of Internet traffic is cached and served by CDNs, they raise important questions regarding the value of application-specific *vs.* general networks. As mentioned earlier, such an outcome poses a threat of fragmenting the market which may lead to educed scope and scale economies, and network externalities, factors that fueled the explosive growth of the Internet.

As discussed earlier, the moneyless peering agreements between backbone ISPs support and re-enforce classless best-effort service in the Internet. Traffic from a content owner to a user (such as the CNN.com front page) received the same quality of service as user-to-user communications (such as email traffic). To the extent capacity investments are justified by the need to support commercial traffic and then this capacity is shared with non-commercial traffic, the commercial traffic may be seen as subsidizing user-to-user traffic. This picture could change if a larger percentage of commercially valuable traffic is eventually cached and served by CDNs. The commercial websites are then likely to be less sensitive to latency and congestion in the backbone, because it will not affect the user experience much. Other uncached (or uncachable) traffic will still be adversely affected by backbone latency and congestion, but the parties involved might not be willing to pay more to get better service for this traffic. This might be a coordination problem or a cream-skimming problem.²³ In the worst case, this could lead to a reduced level of investment in backbone capacity and maintenance, leading to deteriorated performance for traffic other than the CDN-cached traffic.

Interestingly, we note that unpaid peer-to-peer content distribution networks like BitTorrent can raise similar policy issues, for different reasons. These networks rely on the fact that users currently pay a flat fee for Internet connectivity, as opposed to a volume-based charge, and hence do not mind caching and serving data for other users. However, the huge growth of peer-to-peer traffic volumes, and the significantly higher usage of users who cache peer-to-peer content as compared to other users, is leading to a call for volume-based pricing to be introduced for end users. If ISPs introduce volumebased prices for end users, it could have a dramatic impact on user behavior: Users will

²³ A coordination problem would exist if the willingness-to-pay is higher than the cost of increasing capacity but the transaction costs of collecting this willingness-to-pay under current market mechanisms exceed the net benefits to be realized. Alternatively, a cream-skimming problem arises if the willingness-to-pay for capacity investments of the uncachable content exceeds the incremental costs of such capacity, but only if the high-value content is also contributing to the shared fixed costs. If the high-value content is diverted to a less expensive network for them, it might no longer be feasible to support the uncachable content. Similar concerns regarding the sustainability of what was perceived to be a natural monopoly in telephone infrastructure motivated regulatory restrictions against competitive bypass for many years until changes in the technology made it plausible that efficient competition (not cream skimming) might exist.

be less likely to initiate traffic that has no direct value (including for instance contributing to an online encyclopedia, or posting to a blog), because the marginal price of traffic will be nonzero. A tiered pricing scheme might mitigate this to some extent, but it would still be very different from having a flat rate.

Both these potential scenarios – reduced investment in backbone capacity and usage-based pricing – could be construed as being positive or neutral developments, not causes for concern. Indeed, in a static snapshot of the world, they can both be viewed as price differentiation leading to greater economic efficiency. One could argue that if users are not willing to pay more to support an application, or get better performance from a communication-based application, then it is not generating any (or enough) value, and so it is not worth improving the network capacity for such traffic. However, network applications have very marked network effects: The value a user derives from participating in a particular application or community depends to a large extent on the number of other users participating in the same network. For example, a user today might be willing to pay a few cents per email if she really had to, but she may not have done so ten years ago when email was still nascent. Thus, volume-based pricing can make it harder to introduce and test new applications; similarly, applications that need lowlatency user-to-user communication might never be launched. In fact, it has been suggested that the rapid expansion of the Internet itself is due to the prevalence of flatrate pricing [Got03].

Finally, CDNs growth could introduce new policy issues as they become basic and essential infrastructure, much like the current underlying IP network. In such circumstances, the question of which content gets higher quality access may become a "free speech," or First Amendment issue that might provide a basis for imposing open access restrictions. And, if the CDN market proves to provide a nexus for market power, it might attract regulatory oversight in its own right. For example, it is unclear how antitrust authorities might view an acquisition of Akamai by Verizon or SBC. Certainly, the other ISPs would likely be strongly opposed.

B. Routing Overlay Networks

A routing overlay is an overlay that exists for the purpose of controlling or modifying the path of data through the network. In a routing overlay the endpoints of the information exchange are unchanged from what they would have been in the absence of the overlay, but the route through the network that the packets traverse between these endpoints may be different.²⁴

The routing overlay is unique among classes of overlays we discuss in this paper because the overlay network performs a function that is already implemented by the existing Internet infrastructure. In contrast with other classes of overlays, which exist to provide new functionality, routing overlays in their purest form exist to change the way

²⁴ For example, this contrasts with a CDN, wherein the source/destination addresses of the communicating pairs of nodes may be changed by the CDN.

an existing function is performed. It is this overlap, between routing as a base Internet function and routing as an overlay network function, that leads to the most interesting properties of routing overlays.

Before considering routing overlays further, we briefly discuss the routing function of the existing Internet.

Routing -- the determination of a path between the source and destination of transmitted data -- is a basic function of all computer networks. In the simplest case, this determination is trivial. If there is only one path between source and destination, data packets must follow it, or the communication will fail.

In any reasonably large network, the situation will be more complex, because there will be several possible paths between any given source and destination. In this case, routing becomes an *optimization* problem. When more than one path is available, the object of routing is to choose the "best" path. As we will see, this choice is not necessarily obvious even in a single network.

In our situation, however, the problem is even further complicated because there is no single, omniscient observer to choose a globally optimal path. Instead, the path taken by data packets is a result of individual decisions taken by each of the many ISPs that combine to form the Internet. These decisions are driven by a number of factors. Chief among these are the internal structure of the ISP, which determines the cost to carry a packet from when it arrives at the ISP to when it leaves, and the business arrangements between each ISP and its peers in the network, which determine the cost to hand off traffic to the next ISP, and/or receive traffic from a previous ISP in the path.

These individual decisions are coordinated by a network protocol known as the Border Gateway Protocol (BGP). Broadly speaking, BGP allows each ISP to express its policies for accepting, forwarding, and passing off packets using a variety of control knobs. BGP then performs a distributed computation to determine the "best" path along which packets from each source to each destination should be forwarded.

This formulation raises two difficulties, one fundamental and one pragmatic. The first of these is that the notion of "best" is in fact insufficient to fully express the routing task. "Best" is a single dimensional concept, but routing is a multi-dimensional problem. Individual ISPs, in making their routing decisions, may choose to optimize a wide variety of properties. Among these might be

• The cost of passing on a packet²⁵

²⁵ The problem is further complicated when the costs of passing a packet reflect other costs such as those reflected in carrier interconnection agreements which may only imperfectly reflect network costs. See further discussion at footnote 27 below.

- The distribution of traffic among different physical links within their infrastructure to maximize utilization and minimize congestion -- so-called traffic engineering
- Performance in some dimension, such as bandwidth available to the traffic or transmission delay across the ISP.

Further, because the management of each ISP chooses its own objectives, different ISPs may choose to optimize *different* quantities, leading to an overall path that captures no simple notion of "best", and rarely if ever is best for the user.

A second, pragmatic problem with the current Internet routing infrastructure is that it has evolved over time from one in which simple technical objectives dominated to one in which ISPs often wish to express complex policy requirements.²⁶ For this reason the knobs -- the methods available within BGP to control routing choices -- have also evolved over time, and are presently somewhat haphazard and baroque. This compounds the fundamental problem outlined above by making it harder for ISPs to express precisely the policies they desire, even after those policies are known.

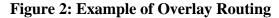
In summary, we observe that two broad statements may be made about the present IP routing system. First, the route used for data is determined entirely by the ISPs, without input or control from the end user or application. Second, what is optimized by the routing system is an imprecisely defined mix of cost and ISP operational efficiency, rather than any metric directly related to application performance.

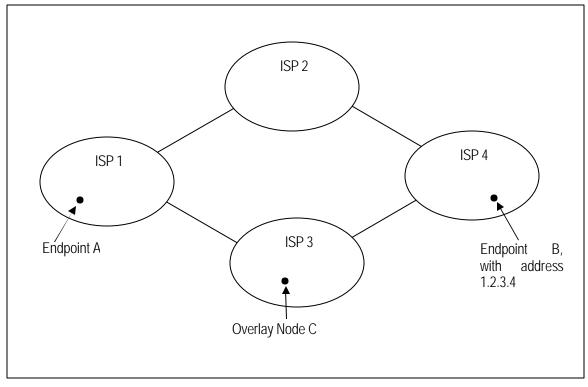
We turn now to the concept of the routing overlay itself. The objective of a routing overlay is to override, in some fashion, the base Internet routing process described above. We illustrate with an example.

1. Example of Overlay Routing

The simple network in Figure 2 has sufficient structure to illustrate the concepts of interest. Endpoint A in ISP 1 wishes to communicate with endpoint B in ISP 4 and potentially could use a BGP-based route offered by ISP 1 or a route offered by an overlay network. These alternative routes are determined as follows.

²⁶ See note 25 *supra*.





For the BGP-based path, assume that ISP 4 uses the BGP protocol to advertise a range of addresses starting at 1.2.0.0 (which contains the address of endpoint B, 1.2.3.4), to both ISP 2 and ISP 3. Each in turn advertises this address range to ISP 1. These advertisements indicate that address 1.2.3.4 can be reached through the advertising ISP. ISP 1 thus knows of two alternative routes for packets whose destination address is within this given address range. ISP 1 could have a variety of considerations in choosing between the two routes. For whatever reason ISP 1 chooses the route through ISP 2 and for this example does not vary its choice. Thus, the BGP-based route available for endpoint A is via ISP 2.

Now suppose that endpoints A and B, situated within ISPs 1 and 4 respectively, participate in an overlay routing network. (Each endpoint might simply be a PC with software installed associated with the overlay, or, with some abuse of the term, might be a corporate, government or university location where a server runs software associated with the overlay network on behalf of users within the location.) Suppose the overlay network has a node C in ISP 3, and, though not relevant for the present example, might also have nodes in other ISPs as well. Through some measurements, endpoint A is aware of the current performance parameters, such as loss and latency, of the BGP-based route via ISP 2 and the overlay-network route via the overlay node C in ISP 3. Based on this information, endpoint A may choose either of the routes, depending on the achieved performance. Further, it might switch back and forth at will, using each route for a period of time when it seems preferable.

An analogous scenario can be described for the reverse direction.

2. Technical Implications of Routing Overlays

Early research on routing overlays focused on their ability to improve application performance by selecting a higher-quality path through the network than that selected by the BGP routing. Experimental results reported by a number of researchers demonstrate this capability in practice. With overlay routing, end-users can potentially attain lower latency, lower loss, higher throughput and/or increased availability.

These results are obtained for a number of reasons. Among these are:

- The overlay can select a route that is intrinsically tuned to the specific needs of the application, rather than relying on the generic route chosen by BGP. This effect is particularly important because, as we have seen, the generic route is rarely optimized for application performance.
- Overlays that use active characterization of network paths can choose non-default paths with low load, and thus minimal congestion delay.
- Overlays can compensate for intermittent failures in the network or the network routing protocols by choosing alternate, functioning paths.
- Overlays can "work around" the effects of ISP load management and traffic engineering, as detailed in the discussion of CDNs.

At first glance this capability appears to heavily favor the use of routing overlays. However, the situation is not so clear cut. To date, the deployment of routing overlays has been minimal and primarily experimental. Should overlays become widely deployed, a corresponding set of negative technical effects is likely to become apparent. These effects are due to the uncoordinated control of routing by many different entities acting independently.

- Severe performance degradation will occur if several overlays simultaneously shift traffic from a highly loaded path to a path with lower load. Since the activities of the different overlays are not coordinated, one possible effect would be to over-shift traffic, leading to extreme congestion on the newly chosen path.
- A more serious negative effect may occur due to interaction between routing at the IP layer and routing within an overlay when both are simultaneously responding to a disruption. In this case, research suggests that sustained oscillations can occur if the traffic volume on the overlay is appreciable. Similar oscillations may be expected if the simultaneous reactions are by multiple overlays. Such oscillations are known to cause sustained, ongoing performance loss due to rapidly fluctuating traffic loads and changing paths.

These negative effects may occur whenever simultaneous routing decisions made by different entities are *uncoordinated*. Should routing overlays become widespread, it may

be possible to mitigate these effects with additional technical mechanism, to loosely coordinate the actions of multiple overlapping overlays.

3. Implications of Routing Overlays on the Interests of ISPs

Implicit within the notion of application routing overlays is that control of the route selection is, at least to some extent, wrested away from the network operator and shifted to the end user. This loss of control over a basic function of network operations has strong implications for the interests of the ISP.

We explore these implications with an example. In Figure 2, for the given destination, ISP 1 has selected a route that passes through ISP 2. However, the end user (A), using the overlay routing network, has effectively overridden the ISP's decision and is instead sending traffic via ISP 3. For appreciable traffic volumes, as might occur when multiple end users and destinations are aggregated, this could be detrimental to ISP 1 for a number of reasons.

- ISP 1 could have economic reasons for its selection. The cost it pays for transit services through ISP 2 could be much less than for transit through ISP 3. In this case the effect of the overlay would be greater cost to ISP 1, reduced revenue to ISP 2 and increased revenue to ISP 3, the high-cost provider.²⁷
- ISP 1 could have engineering reasons for its selection. To balance load on its links (to achieve maximum efficiency and minimum congestion), ISP 1 may apportion traffic between ISPs 2 and 3 based on historical traffic volumes. In this case, the benefit to some users of the overlay will be compensated for by a degradation of service to other users, and a loss of overall efficiency for the ISP. This move away from the optimum will continue until the ISP rebalances its traffic, at which point the cycle is likely to begin again.

From these examples we see that many of the most interesting questions raised by routing overlays are not technical; instead they are related to the changing relationship between ISPs and their customers. As long as ISPs retain complete control over routing decisions within the network, there is little call for the technical routing mechanisms to resolve the "tussle" between the choices of the ISPs and those of end users.

Routing overlays change this equation by giving the users an input into the routing decision. To date, however, they do not provide a coordinated way to resolve conflicting objectives between the various parties. Instead, they simply allow end users to override the ISP in certain situations. It is this lack of coordination that leads to many of

²⁷ ISPs may have long-lived, legacy interconnection agreements that reflect contracts negotiated at different times and in different market conditions that give rise to pricing that does not reflect current opportunity costs. These agreements are often negotiated bilaterally and are private, making it difficult for outsiders to identify private ISP route preferences while embedding in those preferences cost and revenue considerations that are not obviously related to what BGP focuses on.

the negative effects of overlays. For many researchers, the logical next step in the development of routing overlays and related choice mechanisms is to add technical and economic mechanisms that better coordinate the now multi-party routing decision.

4. Future Growth of Routing Overlays

At present, routing overlays are primarily a topic of experiment and research, although some preliminary signs of commercial interest have appeared. Whether the implications of routing overlays on network technology and economics become important depends on whether these overlays grow to handle significant amounts of traffic.

Two broad forces may lead to widespread deployment of routing overlays. First, it is possible that routing overlays will become an intrinsic part of the Internet's technical design, or architecture. While it is beyond the scope of this paper to discuss the rationale for this in detail, there is some reason to believe this may occur due to *scalability* considerations – the ability of purpose built overlays, containing only a subset of the Internet's routing nodes, to provide richer functionality and respond more quickly to disruptions than the current single-layer routing architecture. In this case the adoption of routing overlays would be based on purely technical criteria, and it is likely that such overlays would be operated at least in large part by existing ISPs.

The second scenario is that overlays gain widespread significance as a method for end users and third parties to affect routing decisions. In this circumstance, the broader implications of deployment become fully apparent, because the routing overlay becomes as a vehicle for contention over the routing decision. We have seen above that this use of overlays may create substantial negative effects for ISPs, and potentially for the overall stability of the Internet. At minimum, additional research and development is required before routing overlays can safely fulfill this role.

C. Security and Privacy Overlays

1. Introduction

The final class of overlay networks we discuss are ones that we broadly characterize as "security overlays." These overlay networks provide different forms of communication protection [Her99], user or server anonymity [DMS04], [CSW00], censorship resistance for online content [WM01], [FBH02], or deniability of the knowledge of traffic [CSW00] or content [WRC00]. This is a particularly interesting class of overlays because even if the volume of traffic on these overlays is not large, the policy and social implications can be significant.

In many ways these overlay networks mirror the content and routing overlays discussed in the previous two sections. Security overlays change the routing and caching behavior of communications and content on the Internet. The difference is that instead of changing the behavior to optimize performance or money flows, these overlays enhance some aspect of end-user security. Some provide for secret communications or anonymity for end users; others make content robust against attempts of powerful adversaries to remove it from the Internet and enable users to establish legal deniability of traffic or content ownership.

This class of overlays tends to make the Internet opaque to regulation, easily frustrating policy makers' objectives. Through clever use of cryptographic techniques and system engineering these networks provide provable properties about how hard they are to break or the legal deniability afforded to network participants. In many ways, this class of overlay networks re-raises questions from the encryption debates in the 1990's [HAHH94]. While encryption hides the content of communications in a network, some of these overlays hide the entire network.

However, the beneficial uses of this type of overlay network are significant. As we discuss below, one of the organizations funding research in this area is the United States government. Notably we demonstrate that the government's interest extends beyond the academically fascinating challenges these networks raise to an actual operational interest in the properties these overlays provide. This suggests that the policy and legal issues raised by these networks are multifaceted. Security overlays have unexpected benefits that challenge common, preconceived notions that these networks further illegal activities more often than they further civil liberties.

2. Description

The most widely recognized type of overlay network providing a security property is a virtual private network (VPN). VPNs provide encrypted tunnels between points on the network, extending, for instance, a corporate network across multiple geographic regions or between the home office and a traveling executive's laptop computer. VPNs do not raise any new policy issues that have not been discussed elsewhere [MBSK95] or earlier in this paper, so we do not mention them further.

The security overlays we do cover in this section are more limited in current deployment, but are significant for the policy questions that they raise. In particular we discuss overlays that provide anonymity, censorship resistance, and deniability. We first describe what the networks do and how they function. We then discuss the implications in terms of their regulatory, commercial, and technical impact.

Note that we do not provide an exhaustive survey of this class of overlay networks (see [Din05] and [AS04].) Instead we examine prominent overlay networks within this class. Table 2 below provides a descriptive summary and examples of each type of overlay network.

Onion routing overlays	Onion routing networks, or mix nets, are overlay networks that enable pseudo-anonymous communication over the Internet. Current examples include the Tor [DMS04] and I2P [I2PNET] networks.
Anonymous content storage and retrieval overlays	These overlays protect the identity of authors, publishers, and content providers when they store, query, and download content from the Internet. Current examples include the Freenet [CSW00] and Entropy [Entropynet] networks.
Censorship resistance overlays	These overlay networks attempt to make it very difficult for powerful adversaries to remove content or pollute the overlay network with distracting materials. Current examples include Publius [WRC00], Infranet [FBH02] and Tangler [WM01] networks.

 Table 2: Description and examples of prominent security overlays

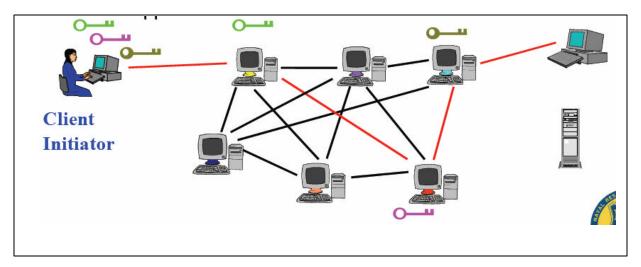
Many of the security overlays have overlapping goals, so our classification of example networks is not strict. Rather we classify the networks by functionality they are most commonly associated with. The rest of the analysis in this section will be structured according to the properties provided by the overlays.

3. Providing anonymity with overlays

One of the main focuses of security overlays is to provide anonymity – hiding identity or authorship or making a network participant unrecognizable in a larger crowd. The desire for this anonymity stems from a variety of motivations. One or both participants in a communication may desire to conceal their actual network location (either from each other or from observers elsewhere in the network.) In other cases the author, content server, or user downloading content may wish to remain anonymous.

At a technical level anonymity is accomplished through the use of encryption, tunnels through the Internet, and proxies that rewrite or re-label packets. Figure 3 below depicts an onion-routed overlay network, Tor, which enables a client to encrypt his or her communication and tunnel traffic through multiple nodes in the overlay until finally communication is established with a host on the regular Internet. Servers can also be hosted anonymously if a rendezvous point is established to tunnel packets back to the anonymous server. In general onion-routing networks are generic transport or network layer overlays capable of providing anonymity to any application. Care must still be taken as applications may leak the identity of an anonymous host in other ways than are protected by the overlay network. Ensuring anonymity is therefore still a non-trivial task for most users of overlays.

Figure 3: Depiction of encrypted tunnels through an onion routed overlay network protecting the identity of the client initiator in communications with the end server. (Source: online presentation of the Tor network available at http://??)



An anonymous content and retrieval overlay network operates in a slightly different way. In the case of the Freenet network, when content is inserted, searched for, or retrieved, the original source of the messages is obfuscated. A node can claim itself or another arbitrarily chosen node as the data source. Any replies are simply forwarded back to the neighboring host that initially sent the message so no information about the original source of the message is visible past the first hop into the network.

Anonymity can be potentially compromised in any of these networks by an observer capable of monitoring many links and performing a traffic timing analysis. By correlating sending and receiving activity an adversary can establish with high probability that two nodes are communicating. While the cost of this traffic analysis is likely prohibitive for arbitrary nodes on the overlay, it would be feasible to monitor known pairs of nodes to determine when they are communicating. Again, anonymity is not absolute.

4. Providing censorship resistance with overlays

Censorship resistance is the commonly used term to describe overlay networks that are designed to resist the attempts of powerful adversaries to remove, or make inaccessible, certain types of content on the Internet. A variety of ways of architecting an overlay network for censorship resistance exist. The anonymity mechanisms described in the previous section are a general mechanism that makes it difficult to locate the actual providers or users downloading content. Anonymity thus serves as a technical and legal shield against any entity trying to remove content from the Internet.

If an attempt is made to insert content into the Freenet network that would overwrite an existing file, the original file will be spread back along the insertion path to the malicious content publisher. This means that any attempt at a technical level to knock content out of the Freenet network will result in the content being cached at more network locations than before the attempted removal.

Another general strategy to avoid censorship is to automatically cache content at many locations (preferably in many different legal jurisdictions) on the Internet. The Chord File System [DKKM01] for instance caches file contents at distributed locations. This potentially frustrates attempts on a legal level to remove content by making the organization interested in removing content pursue the action in multiple jurisdictions for each file they want removed.

Censorship resistance at a technical level can be accomplished in a number of other ways as well. In AChord [Achord] the system in designed in such a way that makes it difficult for any one node participating in the system to assume responsibility for a chosen document. This makes it difficult for nodes to pretend to host content while not actually serving it to the rest of the Internet.

5. Providing deniability with overlays

Deniability is the final property provided by some overlay networks. Deniability is the ability to disclaim connection with or responsibility for either stored content or communications. It becomes much more difficult to establish the responsible party in such an overlay. This property while being independently useful in certain circumstances also contributes to an overlay's censorship resistance by providing a defense that no intent existed to host illegal content.

Deniability of stored content is often accomplished by having nodes in a distributed overlay store encrypted files but not the decryption keys. Each node therefore can plausibly assert that they do not know the content of the files on their system. This approach is taken in the Publius [WRC00] and PAST [RD01] systems. This property also lends itself to the censorship resistance of a network in that individual nodes cannot choose the content they host.

A slightly different approach is taken in the distributed stenographic storage system of Mnemosyne [Mnemosyne]. In that system the existence of files cannot be determined from local knowledge alone; the local storage device appears to be random bits that do not contain any apparent structure. Even if the participant wanted to remove content they would not know which pieces of data to remove from their storage device. The Mnemosyne system also automatically contacts multiple nodes and downloads random files to obfuscate the actual desired file being downloaded.

The anonymous connection overlays and storage and retrieval overlays also afford a degree of deniability. The true origin of a request or traffic cannot be definitively determined allowing a participant in such a network to claim that they were not the one responsible. This is similar to the "Trojan-horse defense" employed in recent legal cases [Reu03] where individuals claimed that content or traffic was originated on their computers by requests from outside individuals.

6. Regulatory impact

While the amount of traffic carried by security overlays providing anonymity, censorship resistance, and deniability is likely to be small, the regulatory impact maybe significant. In some ways these networks make the job of law enforcement or national security more difficult (in other, non-obvious, ways they make it easier). Unquestionably, for better or worse, they complicate notions of identity and responsibility on the Internet.

Notions of identity and responsibility on the Internet have been debated extensively in the past [cite, cite, cite]. But security overlays further challenge what can be assumed about network activity. These networks provide technically justifiable excuses for most network traffic or digital content on a computer. They fundamentally change the notion of identity. While the binding between an IP address and an end-user was never absolute, these networks completely break the correspondence.

From many people's perspective these networks serve a beneficial purpose – namely safeguarding civil liberties. This is the most often stated motivation of developers and operators of security overlay networks. They assert that anonymity is essential to healthy societies [CSW00], [DMS04] providing protection to individuals concerned about the reactions of repressive governments or other powerful adversaries.

Inevitably this leads to a tension with the interests of law enforcement and national security. In cases where a criminal act was committed using these networks, law enforcement is left with few ways of determining the culpable parties. If all the evidence is digital, and all the digital evidence is anonymized, crimes becomes much more difficult to solve.

Operating these overlay networks has not been directly addressed by any court, nor do any laws specifically address this type of overlay. They represent unexplored space from a legal and regulatory standpoint [TorFAQ]. Copyright and intellectual properties protection are the most obviously impacted regulatory space in which a challenge to these overlays may arise. To quote from the Freenet website:

"[Y]ou can't allow those in power to impose "good" censorship, without also enabling them to impose "bad" censorship." "You cannot guarantee freedom of speech and enforce copyright law. It is for this reason that Freenet, a system designed to protect Freedom of Speech, must prevent enforcement of copyright.

This is a conflict of ideologies that is difficult if not impossible to resolve. The only policy point we would add is to consider the research, or experimental nature of many of these security overlays. The research domain is often given more leeway for exploration even when the research interests conflict with other communities needs.

All this may make it seem like governments would regard overlays that provide anonymity as a nuisance to be tolerated at best. Seemingly paradoxically, the U.S. government is one of the primary sources of research funding in this area. The latest generation project on onion routing was sponsored and hosted by the U.S. Naval Research labs.

This is intriguing and surprising to many. Why would the government be interested in anonymous overlays? On one level they are supporting academic research. But they are also interested in anonymity for operational reasons as well. As indicated in the presentation by Syverson [Syv04] of the Naval Research Laboratory, the government conducts open source intelligence gathering on the Internet, and, at times, is interested in hiding the fact that queries come from the government. The government or law enforcement agency may for instance want to covertly monitor a web-based bulletin boards used by adversaries of the United States or other criminal organizations. According to the Tor project, "A branch of the U.S. Navy uses Tor for open source intelligence gathering, and one of its teams used Tor while deployed in the Middle East recently."

Purchasing multiple point of presence on the Internet would be another way of distributing the source of queries, but certainly individuals, and law enforcement agencies, even at the state level, lack the budget resources. Cooperating in an anonymous network would enable many organizations to highly distribute monitoring efforts, effectively hiding numerous surveillance activities.

However, this raises a policy question of whether the government should be able to look at U.S. websites anonymously. The recent monitoring of websites by the FBI's counterterrorism task force before the presidential conventions included generating over a thousand pages of documentation on the ACLU and Greenpeace [cite]. One way the groups could have potentially discovered the monitoring would have been by looking at their web logs and seeing that the queries originated from the governments networks. If the government had instead been monitoring these sites anonymously, detection of this activity, even if it is ultimately determined to be permissible, may not have been possible. Perhaps the government should not be able to search domestic websites anonymously without a court order.

7. Commercial Impact:

The commercial impact of security overlays is likely to be limited in the near future.²⁸ ISPs are not likely to offer anonymity, censorship resistance, or deniability enhancing services in the future for reasons we discuss below. A commercial enterprise operating an overlay network is somewhat more likely, but still the commercial impact is

²⁸ However, if such overlays do grow to handle a significant amount of traffic, they could have adverse impacts on capacity investment and planning costs. That is, because security overlays can hide the source/destination and purpose of traffic, they would make it difficult for ISPs or other infrastructure providers to condition their investments based on the value of the traffic, and hence, may fail to optimally allocate investments. Additionally, the technologies used to implement such overlays are often at cross purposes to the technologies and motivations underlying other overlays such as CDNs or RONs. In contrast, security overlays use network resources inefficiently to implement other policy goals (privacy instead of latency or transport cost reduction).

still likely to be limited in the near term. For now, security overlays are likely to remain edged-based phenomena, operated freely by like minded communities of users.

In many cases security overlays have not passed beyond the research test bed phase (e.g. [FBH02], [WM01], [WRC00]) or are no longer being actively developed or promoted (e.g. [Entropynet]). One Canadian company, Zero Knowledge Systems did operate a general purpose "Freedom Network" that was an onion-router like overlay network service providing user anonymity, but the company has since discontinued the overlay network and no similar general-purpose commercial offering has reemerged.

The reason for this is likely the fact that the commercial demand for these security overlays is fairly limited in most cases. Anonymity, censorship resistance, and deniability are not services that are generally required by most of the consumer population. ISPs may also be unable to offer such services given regulatory requirements already in place such as CALEA. Also these overlays deliberately trade performance for additional security, but most users are more interested in the former rather than the later. Couple this with the unsettled legal liabilities of a company offering these services, and the commercial prospects are likely not significant at this time.²⁹

A major barrier to adoption for end users of a commercial offering is how to purchase service while maintaining anonymity. When the Freedom Network was a commercial offering, the company had an elaborate protocol for how users paid for the service, but users still had to trust that the company was not logging IP information or recording information that would allow them to correlate traffic with payment records. Relying on the company as a trusted third party conflicted with the anonymity aims of customers. (Were anonymous electronic-cash deployed this problem would be alleviated.³⁰)

A commercial opportunity may eventually develop as interest from corporations or government organizations increases. Governments and corporations have expressed initial interest in these networks as a way of enabling anonymous tips or suggestions, experimenting with new services without exposing the corporate identity, protecting against denial of service attacks, gathering information covertly from adversaries or competitors, and enabling anonymous elections and voting systems. The Electronic Frontier Foundation actively promotes these networks for use by political dissidents and whistleblowers [TorFAQ].

Edged-based, user run, and free security overlay networks however are steadily gaining in popularity. In 2005, the Tor network had 200 nodes in Europe and the U.S., each node routing between one and ninety gigabytes of traffic in a day [Din05]. One of

²⁹ Additionally, barriers to deploying these may be lower for criminals who may have a greater incentive to make use of such services (analogous to the argument that if guns are outlawed, only criminals will have guns).

³⁰ Indeed, an important reason why cash money is so valuable is because it supports anonymous trade.

the reasons anonymity is growing in popularity is the needs of peer-to-peer users. In Japan, the most popular peer-to-peer filesharing clients, Winny [cite] and its successor Share [cite] form an overlay network which provides some anonymity for clients. In the United States, the Freenet network is in operation and used by a small population of clients wanting to share content. Finally, a popular Bittorrent client, Azureus, is capable of utilizing the I2P and Tor networks for anonymous communications.

For now it is likely that users interested in such security properties are technically sophisticated enough to download, install, and run one of the cooperative, free network tools available on all the major operating systems. Technically the tools are not particularly sophisticated, requiring for instance around 30,000 lines of C code in the case of Tor [TorFAQ]. These tools have the added benefit, comforting to many users, that their source code is available for inspection.

8. Technical impact

Security overlays change some of the most basic underlying assumptions about networks and user activity on the Internet. They fundamentally change common conceptions of identity on the Internet. They represent a significant challenge to the field of computer forensics, already struggling to attribute network activity or content found on a server to an individual. They also challenge the service model of providing geographic dependent services. A technical benefit of these overlays though may be to mitigate some types of denial of service attacks, a prevalent and vexing problem on today's Internet.

The most significant technical impact of security overlays is the challenge they represent to notions of identity on the Internet.³¹ The correspondence between IP address and user identity has never been absolute --multi-user machines, and network address translators (NATs), are two of the existing ways in which an IP address potentially represent multiple individuals. But in both these cases the multiple individuals represented were at least likely from the same organization or group. With these security overlays, a request or traffic from an IP address may represent a proxy request for another computer anywhere on the Internet.

On reason that this change is significant is that authorization on the Internet is often based upon IP address. Access to certain network resources is often restricted to limited IP address ranges. Universities for instance negotiate access to digital repositories from outside companies that implement access control by only serving queries from the IP addresses assigned to the university. If an overlay node is inside the universities network it may be making queries on behalf of a computer anywhere on the Internet.

Similarly prevalent is the use of IP address filters in firewalls. Administrators of firewalls often configure their systems to extend greater trust to limited parts of the IP

³¹ And, in the future, we may expect individuals to want multiple identities that may be context dependent (work/play, on-the-road/home).

address space for certain activities. Security overlays again represent a hard to detect challenge to the assumption of trusted IP address ranges.

Security overlays also represent a significant technical challenge to the field of computer forensics. Establishing evidence in a court of law is difficult if a user can claim and prove technical deniability. With a security overlay running on their computers, users have plausible excuses that the overlay was responsible for storing the content or generating the network activity. A defense lawyer could definitively demonstrate how the overlays could make the requests on behalf of others without the computer owner being involved in any way.

These security overlays also represent an interesting challenge to notions of the geographic origin of traffic, which poses an obvious challenge for enforcement of national sovereignty. Many services on the Internet take the apparent geographic origin into account when severing content, for instance displaying results in the language most likely appropriate for the query origin. Results from some search engines are also optimized for the geographic location of the query node. However a security overlay may be again making a query for a person half-way around the world in a completely different actual linguistic or cultural context.

A technical benefit of security overlays may be as a defense against the ever problematic denial of service attacks. Rendezvous points on the Internet could forward legitimate traffic onto the secret, anonymous servers, protecting the actual resources from denial of service attacks.

IV. Conclusions and Directions for Further Research

The Internet emerged as an overlay on the telephone system, and triggered a massive shift in the structure of the telecommunications industry, with economic, policy and social implications. We believe that overlay systems on top of the Internet may signal yet another shift, and while overlays may not be as dramatic or a far-reaching as the Internet itself, they again have important economic, policy and social implications.

Overlays exist for several reasons, which this paper has tried to sort out. One reason for overlays is that specialized groups of users have specialized niche requirements. If these requirements imply some sort of function that is distributed across the network, (so that these users cannot satisfy their requirements with code that runs only on their own end-nodes), then the structure of the solution will probably resemble an overlay. To the extent that overlays allow the general functions of the Internet to be specialized for smaller sets of users, this is a benign situation that signals the expected richness of a maturing product. Another motivation is that overlays can allow the early deployment of new and unproven next generation applications. Systems such as Planet Lab, which are essentially a highly distributed platform for deploying new distributed applications, can be viewed as an overlay in their own right or as a tool to make overlays on demand. In either case, they are a valuable part of an ongoing research process. In this sense, overlays may play a useful and important role in the dynamic evolution of the Internet and in the deployment of new infrastructure. A third reason for overlays to exist

is that the Internet Service Providers are failing to sell to the customer what the customer wants to buy, and this creates an opportunity for a third party to enter the market. Routing overlays seem to fit into this category, and signal some sort of misalignment between customer and provider. A final reason for overlays is that they capture an intrinsic tension between the interests of different parties, and their structure evolves as part of the tus sle among these parties. The use of overlays to allow anonymous communication is a good example.

An overlay may serve several of these roles simultaneously, and may evolve over time. For example, today's niche or new functionality may evolve into basic infrastructure over time. Or, as in the case of content delivery networks (cdns), an overlay may serve to reduce transport costs in a way that might benefit many yet still provide a nexus for tussle as providers and customers compete for the allocation of surplus that the added functionality can deliver.

1. The economics of overlays

In economic terms, there are two sorts of overlays today. One is a commercial offering such as Akamai, where we can already see the business case and the economic structure. The other is the academic experiment, some of which (like Planet Lab) are very large scale and very successful as a platform for experimentation. In the latter case, however, it is not clear what the business model is—who will run this service when it grows up, and who will pay for it. Who is the provider, and who is the customer?

Some of the current academic experiments seem to be an exercise is suspending disbelief in economic realities. In the case of Planet Lab, for example, the servers are nothing but PCs, and machines of that capacity are easy to capitalize. The network connections are in most cases being donated by the researchers. The operational costs are being paid out of research grants. Since the platform is thus "free," the services are also free—the goal is not to make money but to learn about and prove the value of the services. There is nothing at all wrong with this, but it provides little evidence about the eventual economics.

If the traffic loads on an overlay build up, the servers will cease to be simple PCs, and will become more complex and expensive servers, and the loads on the net will be such that the costs will have to be accounted for somehow. At this point, the platform will not be free. Who will provide them? One answer is a third-party provider such as Akamai. Another answer is a coalition of ISPs.

Technically, the ISPs themselves are in an excellent position to offer services such as these at low cost, since they own or have access to the infrastructure and locations to host the service. But for them to deploy an overlay with global reach, it might require a degree of negotiation and cooperation that would be hard for competitor ISPs to achieve. The cooperation might raise issues of antitrust, as well as difficulties of sharing revenues and operational data. But if these difficulties raise barriers to the provision and operation of overlay services by ISPs, this leaves the door open to third-party overlay service providers, who could reap the commercial benefits of the service and leave to the ISPs only the commodity business of raw Internet packet carriage. This outcome might not be a financially healthy one for the ISPs. If most commercial overlay services are operated by third party providers, there is a risk that the incentives of the ISPs and the overlay providers may not be well aligned. This can lead to under-investment and a stagnation in innovation and upgrades. So it is important to observe whether these new overlay services become a revenue opportunity for ISPs, or just yet another source of traffic.

There is a third possibility—a middle ground. Overlays, even if operated by third parties, may help solve some of the "flow of revenues" problems created in the existing Internet by the current bill-and-keep interconnection contracts. Overlay providers may collect revenues in one part of the Internet, and expend them in other parts of the Internet, which represents a flow of revenue from the source of the value, even across the bill-and-keep interfaces. So a third party overlay might deprive the ISPs of some overlay service revenues, but also bypass some difficulties in routing money across the Internet.

Of course, there is a spectrum of complexity required in the coordination to provide different overlay services. Some services may require little negotiation if the right standards exist—ISPs today run the servers that support email, and little ongoing negotiation is required there (perhaps because email is free). But the difficulty agreeing on an approach to control of spam illustrates the structural problems there.

2. Policy issues

If ISPs have problems negotiating to set up a global overlay, another business option would be for an ISP or group of ISPs to purchase an existing overlay provider. Would this create any regulatory issues about competition and market power? Would the market accept a global service owned and offered by a single ISP, or would there be concerns about fair and equitable treatment? Would this be an issue the marketplace could sort out on its own? These are future possibilities we might consider and debate.

The third parties that we have described as providers of overlay services have a different regulatory tradition and structure than Internet Service Providers. ISPs, especially if they own their own facilities, usually come from a sector with a history of governmental regulation, for example telecoms providers or cable providers. They tend to have a physical locus to their operations, based on where their facilities are situated. While there are certainly global ISPs, even these large firms have a sense of being located in a physical space, and that is where their primary regulation arises. Overlay service providers, in contrast, sit at a much higher level in the "service layers": they are information providers, with no tradition of regulation, at least in the United States. They may own servers and other computing equipment, but their hardware base is not tied to communications facilities, and they can "go global" much more quickly. Indeed, they have to, if their overlay service is going to be available everywhere. They will much more resemble multi-national firms, with a similar presence in many countries, regulated nowhere and everywhere.

These firms may carry out functions that are of great interest to society. They may store and forward inter-personal communication, and disseminate important content. These functions may raise questions of lawful intercept and wiretap, control of illegal content dissemination, free speech and so on. But the different form of the industry may confound traditional approaches to regulation. It may not be clear what agencies have the jurisdiction to regulate, if regulation is necessary, or what laws apply to these firms, if they are not "communications" firms. And this will have to be sorted out in every country that has concerns.

Another important regulatory issue will have to do with openness and equal access. The Internet was founded on a tradition of open access—open to any application provider, any user, and any new entrant into the ISP market. The telecoms industry is shaped by the non-discrimination requirement for common carriage. But there is no reason to think that overlay services will be similarly open. They lack either a historical expectation or a regulatory expectation. Commercial overlays will probably be open based on willingness to pay, but they may cater to classes of customers (for example large volume customers) and in doing so preclude small or non-commercial users from having access to the same services. It is possible that there might be a response—a different set of overlay service providers that cater to smaller or non-commercial users, but this is just a possibility. The other possibility is that overlays will transform the Internet into a two-tier network, with overlays providing service enhancements to larger customers, and a lower-function "basic" Internet the only option for the smaller users.

3. Social issues

The Internet has the feature that anyone can talk to anyone. This feature is also a drawback when one of the communicants is malicious or disruptive. Overlays may be a means to build "gated communities in cyberspace", where like-minded participants agree to talk only among themselves, and others are closed out. Whether this happens, and what it might mean to the future of the Internet, should be a topic of observation and discussion.

To some extent, the Internet has matured, which implies that the landscape of functions and applications is somewhat set, and disruptive innovation is harder to pull off. Overlays represent a way to innovate at a higher level, and create a new order, with a new set of players, a new economic landscape, and a new set of rules. The computer science research community has seen this, and has flocked to study overlays as a way to achieve change and have an impact. It is just possible that overlay services are the Next Big Thing.

V. Bibliographic References

[Achord] Hazel, Steven and Brandon Wiley (2002), "Achord: A Variant of the Chord Lookup Service for Use in Censorship Resistant Peer-to-Peer Publishing Systems," presented at IPTPS-02 (available at: http://thalassocracy.org/achord/achord-iptps.html).

[Alt00] Altmann, Jorn (2000), "A Reference Model of Internet Service Provider Businesses," ICTEC2000, 3rd International Conference on Telecommunication and Electronic Commerce, Dallas, Texas, USA, November 2000 (available at: <u>http://www-networking.eecs.berkeley.edu/~altmann/publications/altmann_fv1_internetmodel_ictec20</u> 00.pdf).

[APMS04] Akella, Aditya, Jeffrey Pang, Bruce Maggs, Srinivasan Seshan, and Anees Shaikh, "A Comparison of Overlay Routing and Multihoming Route Control," SIGCOMM 2004, Portland, Oregon, August 30-September 3, 2004.

[AS04] Androutsellis-Theotokis, S. and Spinellis, D. 2004. "A survey of peer-to-peer content distribution technologies." *ACM Comput. Surv.* 36, 4, December 2004.

[BC01] Blumenthal, Marjory, and David Clark, "Rethinking the design of the Internet: the end-to-end arguments vs. the Brave New World," *ACM Transactions on Internet Technology (TOIT)*, vol. 1, issue 1 (August 2001) pages 70-109.

[BSG00] Boucher, P, A. Shostack, and Ian Goldberg. "Freedom Systems 2.0 Architecture" Zero Knowledge Systems, Inc. White Paper, December 2000.

[CC05] Chuang, John and Nicolas Christin (2005), "A cost-based analysis of overlay routing geometries," in *Proceedings of IEEE INFOCOM'05*, Miami, Florida, March 13-17, 2005.

[CCWB04] Chun, Byung-Gon, Kamalika Chaudhuri, Hoeteck Wee, Marco Barreno, Christos Papadimitriou, and John Kubiatowicz, "Selfish Caching in Distributed Systems: A Game-Theoretic Analysis," mimeo, Computer Science Division, University of California, Berkeley, presented at PODC'04, St. Johns, Newfound land, Canada, July 25-28, 2004.

[Cha81] Chaum, D., "Untraceable electronic mail, return addresses, and digital pseudonyms" In Communications of the ACM 4(2), February 1981.

[Cla05] Clark, David (2005), "Architecting a Networked Application," mimeo, MIT, January 12, 2005.

[CSTB94] CSTB (1994), Computer Science and Telecommunications Board, National Research Council, Realizing the Information Future: The Internet and Beyond, National Academy Press, Washington DC.

[CSW00] Clarke, I., O. Sandberg, B. Wiley, and T. Hong, "Freenet: A Distributed Anonymous Information Storage and Retrieval System" In the Proceedings of Designing Privacy Enhancing Technologies: Workshop on Design Issues in Anonymity and Unobservability, July 2000, pages 46-66.

[Din05] Dingledine, R. (2005), "Anonymity bibliography" Available at http://www.freehaven.net/anonbib

[Din05b] Dingledine, Roger (2005), "Tor: Anonymous Communications for the United States Department of Defense...and you." Presented June 3, 2005.

[DKKM01] Dabek, Frank, M. Frans Kaashoek, David Karger, Robert Morris, and Ion Stoica, *Wide-area cooperative storage with CFS*, ACM SOSP 2001, Banff, October 2001.

[DMS04] Dingledine, Roger, Nick Mathewson, and Paul Syverson, "Tor: The Second-Generation Onion Router" In the Proceedings of the 13th USENIX Security Symposium, August 2004.

[Entropynet] "Entropy (Emerging Network To Reduce Orwellian Potency Yield)" Project website available at <u>http://entropy.stop1984.com/</u>

[FBH02] Feamster, N., M. Balazinska, G. Harfst, H. Balakrishnan, and D. Karger, "Infranet: Circumventing Web Censorship and Surveillance" In the Proceedings of the 11th USENIX Security Symposium, August 2002.

[Fre02] Freiden, Robert (2002), "Adjusting the Horizontal and Vertical in Telecommunications Regulation: A Comparison of the Traditional and a New Layered Approach," mimeo, Pennsylvania State University, 2002.

[Got03] Gottinger, Hans-Werner (2002), <u>Economies of Network Industries</u>, London and New York: Routledge, 2003.

[HAHH94] Hoffman, L. J., Ali, F. A., Heckler, S. L., and Huybrechts, A. 1994. "Cryptography policy." *Commun. ACM* 37, 9 September1994.

[Her99] Herscovitz (1999) "Secure virtual private networks: the future of data communications." *Int. J. Netw. Manag.* 9, 4 August 1999.

[HKSC04] Hosanagar, Kartik, Ramayya Krishnan, Michael Smith, and John Chuang, "Optimal Pricing of Content Delivery Network (CDN) Services," *Proceedings of the 37th Hawaii International Conference on System Sciences*, 2004.

[I2PNET] "I2P Anonymous Network": Project website available at http://www.i2p.net/

[Ise97] Isenberg, David (1997), "The Rise of the Stupid Network," AT&T Research Labs, June 4, 1997 (see <u>http://www.isen.com/stupid.html</u>).

[KTCI04] Ram Keralapura, Nina Taft, Chen-Nee Chuah, and Gianluca Iannaccone (2004), "Can ISPs take the Heat from Overlay Networks," mimeo, available at: http://networks.cs.ucdavis.edu/publications/2004_rubinet_2005-01-29_03_16_46.pdf

[KW02] H. Kung and C. Wu, "Content Networks: Taxonomy and New Approaches," in *The Internet as a Large-Scale Complex System*, edited by Kihong Park and Walter Willinger, Oxford University Press, 2002.

[LL00] Lemley, Mark A. and Lessig, Lawrence (2000), "The End of End-to-End: Preserving the Architecture of the Internet in the Broadband Era" (October 2000) UC Berkeley Law & Econ Research Paper No. 2000-19.

[LM05] Lanzi, Diego, and Massimiliano Marzo, "Content Delivery and Vertical Integration in On-line Content Markets," *Review of Network Economics*, vol. 4, issue 1 (March 2005) pages 63-74.

[LMRT01] Laffont, Jean-Jacques, Scott Marcus, Patrick Rey, and Jean Tirole, "Internet Peering," *Papers and Proceedings of the American Economic Association, American Economic Review*, vol. 91, No. 2, (May 2001) pages 287-291.

[LMRT03] Laffont, Jean-Jacques, Scott Marcus, Patrick Rey, and Jean Tirole, "Internet Interconnection and the off-net-cost pricing principle," *Rand Journal of Economics*, vol. 34, no. 2 (Summer 2003) pages 370-390.

[MBSK95] Milberg, S. J., Burke, S. J., Smith, H. J., and Kallman, E. A. 1995. "Values, personal information privacy, and regulatory approaches." *Commun. ACM* 38, 12 Dec. 1995.

[Mnemosyne] Hand, Steven and Timothy Roscoe (2002), "Mnemosyne: Peer-to-Peer Steganographic Storage," Proceedings of the 1st International Workshop on Peer-to-Peer Systems (IPTPS'02), Boston MA 2002 (available at: <u>http://www.cs.rice.edu/Conferences/IPTPS02/</u>)

[MWA04] Mahajan, Rahul, David Wetherall, and Thomas Anderson, "Negotiation-based Routing Between Neighboring ISPs," mimeo, University of Washington, 2004.

[PAST05] Peterson, Larry, Thomas Anderson, Scott Shenker and Jonathan Turner, "Overcoming the Internet Impasse Through Virtualization," *IEEE Computer Magazine*, April 2005, pages 62-69.

[RD01] Rowstron, Antony, and Peter Druschel, "Storage management and caching in PAST, a large-scale, persistent peer-to-peer storage utility," mimeo, 2001.

[Reu03] Reuters, "Hackers Defense: the computer did it.", Available at http://www.cnn.com/2003/TECH/internet/10/28/hacker.defense.reut/index.html.

[SRC84] Saltzer, Jerome, David Reed, and David Clark (1984), "End-to-end arguments in system design," *ACM Transactions on Computer Systems (TOCS)*, vol 2, no. 4 (1984) pages 195-206.

[SAZS02] Ion Stoica, Daniel Adkins, Shelley Zhuang, Scott Shenker, Sonesh Surana, "Internet IndirectionInfrastructure," *Proceedings of ACM SIGCOMM*, August, 2002.

[Sic02] Sicker, Douglas (2002), "Further Defining a Layered Model for Telecommunications Policy," draft mimeo, University of Colorado, Boulder, September 1, 2002.

[Syv04] Syverson, P. (2004), "Making Anonymous Communication", Presented at the National Science Foundation, June 8, 2004.

[TorFAQ] Tor: Legal FAQ. Available at <u>http://tor.eff.org/eff/tor-legal-faq.html</u>

[Wer02] Werbach, Kevin (2002), "A layered model for Internet policy," *Journal of Telecommunications and High Technology Law*, vol 1 (2002) pages 58-64.

[Wer04] Werbach, Kevin (2004), "Breaking the Ice: Rethinking Telecommunications Law for the Digital Age," mimeo, Wharton School, December 2004.

[WM01] Waldman, M. and D. Mazières, "Tangler: a censorship-resistant publishing system based on document entanglements". In the Proceedings of the 8th ACM Conference on Computer and Communications Security (CCS 2001), November 2001, pages 126-135.

[WRC00] Waldman, M., A. Rubin, and Lorrie Cranor, "Publius: A robust, tamperevident, censorship-resistant and source-anonymous web publishing system" In the Proceedings of the 9th USENIX Security Symposium, August 2000, pages 59-72.