The Evolution of Internet Services

The Evolution of Internet Services

Randy H. Katz, Eric A. Brewer, Steven McCanne
Electrical Engineering and Computer Science Department
University of California, Berkeley
Berkeley, CA 94720-1776

Abstract

The traditional focus of the Internet has been on the development of its protocols, primarily TCP/IP and HTTP. With the recent rise of the World Wide Web, and the explosion in the number of people with access to the Internet, the technical community has shifted towards new applications and network-based services to support them. It is our contention that while the protocol stack is well understood, the service architecture underlying these new applications is evolving rapidly and somewhat chaotically. We review the evolution of the Internet, with a particular emphasis on the rapid developments of the last five years, and propose a candidate service architecture based on the emerging structure of the Internet industry. We use this architecture to characterize the service offerings from several major Internet companies. We also speculate about the future path of Internet evolution.

1. Introduction

We divide the history of the Internet somewhat simplistically into two periods: before the rapid spread of the World Wide Web (WWW) and the emergence of graphical Web browsers, and afterwards. We set the pivotal year to be 1995, even though NCSA Mosaic appeared in 1993 and the WWW in 1991. The crucial event was the final privatization of NSFNet, enabling the formation of today's commercial Internet. This provided the critical flame that allowed commercial use of the Internet to ignite, brought on by the development of a real market: the simultaneous explosion in the number of individuals with access to the Internet (in part, due to lower cost PCs allowing higher home penetration rates) and the amount of increasingly useful content embedded in the Web.

As late as 1994, the correct answer to the question "what is the Internet?" was "anything that runs the TCP/IP protocol stack." Its essential feature was that applications could run on top of a small number of very well engineered protocols and APIs, thus hiding the complexities of the underlying diverse access networks. By this time, the Internet successfully spanned technology based on twisted pair, coax cable, telephone, fiber optic, packet radio, and satellite access. This conceptual architecture was so essential to the Internet's success that it was dubbed the "narrow-waist" model in Kleinrock's landmark National Academy study *Realizing the Information Future: The Internet and Beyond* [1] (See Figure 1).

By this time, a raging debate developed in the technical and policy communities on how best to converge data, television, and telephony networks and formats to achieve a

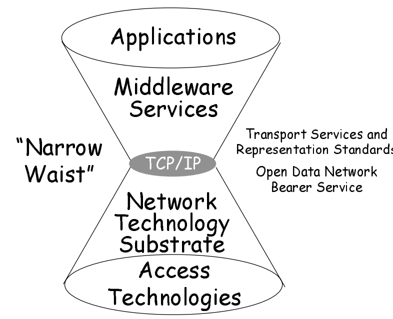

Figure 1. Kleinrock's "Narrow Waist" Model

TCP/IP provides to applications a small, well engineered interface to diverse access technologies. The Internet is any network running this set of protocols.

"National Information Infrastructure." Rather unexpectedly--at least from the view of the incumbent telecommunications and media giants--the killer application was not video-on-demand but information at your fingertips delivered via the WWW. The clear trend today is to move telephony and video onto the Internet as particular applications, and to enhance the underlying networking technology to better support these data types. Virtually all of the long distance telephone carriers have already announced that they are migrating their networks to become IP-based [Ref?].

However, the most rapid innovation has been in the proliferation of a new (and somewhat chaotic) service model on top of the Internet's narrow waist. On the one hand, this is being driven by the privatization of the Internet, leading to intense competition in access and backbone connectivity services (e.g., ISPs, NSPs, hosting, colocation, overlay networks, etc.), and on the other by the emergence of significant Web-based businesses: Portals, Content Delivery, E-Commerce, and in the near future, Entertainment.

While much of the current development is being driven by consumer access to Web content and business-to-consumer commerce, there is also significant focus on using Internet-based services to integrate enterprise business processes. The Internet is rapidly becoming the means that allows a corporation to outsource many elements of its information technology operations, not simply to present a Web presence to consumers or business partners.

Elements of this emerging service model include: network service providers (NSPs), Internet service providers (ISPs), Web hosting services, applications service providers (ASPs), applications infrastructure providers (AIPs), content delivery services, and so on. Some of these services are oriented towards corporate users (the *enterprise* market), others to ISPs, and still others to content developers.

This paper further elaborates on the service model we see emerging, within the context of the Internet as it exists at the end of 1999. We also makes predictions of how it may evolve from this point forward. Our focus is on the rich set of services above basic connectivity.

The rest of the paper is organized as follows. In the next section, we set the context for the new services model by giving a brief history of the Internet. Our particular focus is on the last five years, which has not been well documented in the technical literature. Section 3 presents our approach for describing the emerging Internet Service Architecture. Section 4 uses case studies of some existing Internet service companies to illustrate the architecture of Section 3. Section 5 contains our predictions of how this will evolve in the future. Our summary and conclusions are in Section 6. Section 7 is our references, and Section 8 is a glossary of terms.

2. History of the Internet

2.1 ARPANet to NSFNet, 1965-1995

The story of the ARPANet and the rise of the Internet has been previously told (e.g., [2], [8]), so we only cover the essential points here (see Figure 2). Based on independent conceptual work in packet-switching by Baran, Davies, and Kleinrock in the early-1960s, the Advanced Research Projects Agency's Information Processing Technology Office under Taylor and Roberts initiated a research program to develop a network based on this technology, called the ARPANet. The initial network, which became operational in 1969, consisted of nodes at UCLA, SRI, UCSB, and Utah.

By 1972, the packet switching concept had been applied to satellite (SATNet) and packet radio (PRNet) networks. Making these disparate networks interoperable to support such crucial applications as electronic mail and file transfer became a key technical challenge. This motivated the development of an open "internetworking architecture" based on the transport control protocol/internetworking protocol (TCP/IP) developed by Cerf and Kahn in 1975 [3]. The ARPANet officially switched over to the new protocols in 1983.

The Department of Defense adopted TCP/IP as the standard for its own operational networks. Many other research communities developed their own networks in this time frame, such as Department of Energy's ESNet and the NSF-backed CSNet. In 1985, the Department of Defense transferred the management of the research portion of the ARPANet to the National Science Foundation. Its evolution became known as NSFNet.

Under Wolff's leadership at NSF, NSFNet was established on a scalable architecture of managed interexchange points with other Federal agencies, a high speed nation-spanning backbone, and regional networks to provide access to local universities and industrial research laboratories [4]. The backbone was strictly restricted from being used for any purpose other than research and education.

Once the architecture was firmly established, it became the goal to transition the backbone and the regional networks to a commercial service by 1995.

To encourage this shift, NSF lifted its usage restrictions in 1991. The backbone was upgraded to higher speed links and routers. The Network Access Point (NAP) architecture was established, with such facilities as MAE-East and MAE-West, allowing peering among independent regional networks (see Figure 3). The NSF-backed regional networks became early stage Internet Service Providers (ISPs), but the architecture allowed other ISPs to develop quickly. By April 1995, the NSF stopped funding the infrastructure of the NSFNet, and backbone interconnections were handled by companies like MCI and Sprint. The Internet was now completely in the hands of commercial entities.

Berners-Lee developed the concept of the World Wide Web at CERN, and released the first software to publish and access web documents in 1991. Its initial usage was by the high energy physics community. Andressen developed Mosaic at the National Center for Supercomputing Applica-

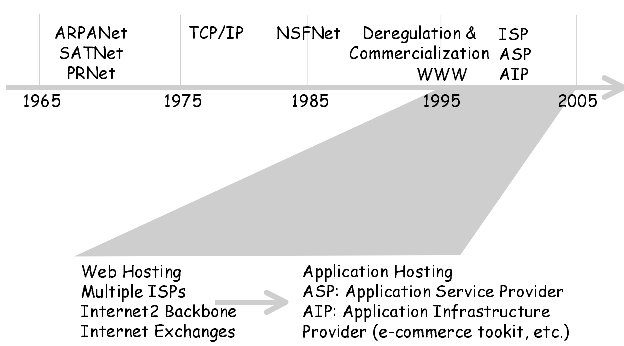

Figure 2. Evolution of the Internet

This time line shows the major stages in the evolution of the Internet. Between 1995 and today, the development has mainly focused on new services. New industries have arisen to provide these service.

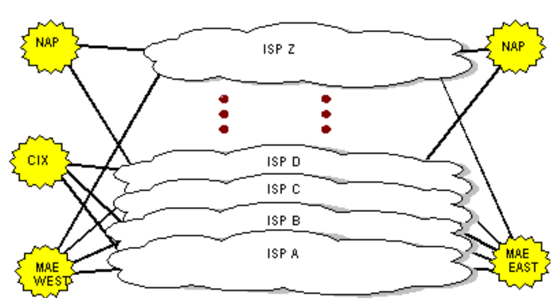

Figure 3. Network Access Point Architecture

NAPs provide a high speed, high bandwidth local area interconnection bridge among multiple ISPs. The inter-ISP routing is handled by boundary routers within the ISPs based on ISP-to-ISP negotiated peering agreements. MAE-East (DC), MAE-West (San Jose), and CIX (Palo Alto) were among the first NAPs to be deployed.

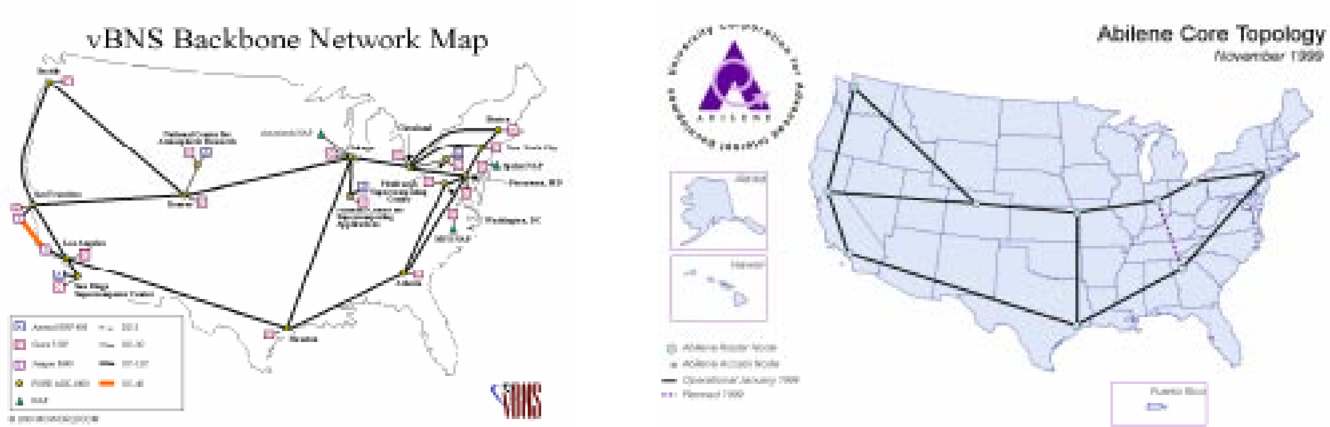

Figure 4. Backbone Maps for vBNS and Internet2’s Abilene Core

vBNS and Internet2 are two testbeds for advanced networking established in the wake of the NSFNet’s commercialization.The vBNS, primarily interconnecting supercomputing research centers, is being transitioned to Internet2.

tions in 1993, providing a graphical interface to the Web, which was essential to its wide-spread subsequent acceptance.

By 1995, the essential threads had come together to form the basis for the subsequent explosive growth of the Internet: a commercialized networking infrastructure, backed by major service and equipment vendors (e.g., MCI, Sprint, Cisco), with a compelling application in the form of Web access, and PCs becoming true appliances due to their ever declining price/performance.

2.2 Privatized Internet to Today

It should be noted that a strong research thread continues with the Very High-speed Backbone Network (vBNS) Project, and the more recent Internet2 Consortium with its initial Abilene project [5] (see Figure 4). The goal of these efforts is to develop advanced networking testbeds to support academic researchers’ experimentation with new networking technologies, protocols, and applications.

The Telecommunications Act of 1996 has played a critical role in opening the communications industry to wider competition [10]. The Act allowed cable operators to enter the phone business and phone companies to enter the cable business. The result has been intense merger activity among the traditional Interexchange Carriers (IXC), such as MCI/Worldcom and Sprint, and the Incumbent Local Exchange Carriers (ILEC), like SBC’s acquisition of PacBell and Ameritech. Qwest, a new generation network service provider deploying a nationwide fiber optic network based on IP technology, is in the process of acquiring US West.

AT&T’s acquisition of TCI, a cable operator, and @Home (which had earlier acquired the Web portal Excite), an ISP/NSP for the cable industry, is an interesting case in point. This brought cable modem local access and content capabilities to AT&T’s long-haul backbone business (see Figure 5). The strategy has raised the ire of America On-Line, which has sought to force the decoupling of local access from ISP and content provision in cable modem networks.

Significant beneficiaries have been the data-oriented Competitive Local Exchange Carriers (CLECs), such as Covad and Northpoint. These companies have capitalized on the deregulated access environment to provide an alternative to cable modems for broadband access based on digital subscriber loop (DSL) technology. Their current business model is to obtain access lines from an ILEC, pay to deploy access equipment in the ILEC’s central offices, and wholesale the broadband connectivity to ISPs.

With the shutdown of the NSFNet backbone in 1995, a remarkable number of NSP-deployed parallel backbones spanning North America have emerged. The structure of these backbones is very similar, with high bandwidth links spanning between San Francisco, Los Angeles, Seattle, concentrated in Denver or Chicago, with additional links to Washington, New York, and Boston. Additional backbone connectivity is provided to points of presence in the Southeast (e.g., Atlanta) and the Southwest (e.g., Dallas or Houston).

The aggregated North American backbone bandwidth has been growing rapidly. NSFNet started with 56 kbps links, upgraded to T1 (1.5 mbps) in 1989, and T3 (45 mbps) in 1991. Recall that this was the primary carrier of Internet traffic in this time frame. Commencing in 1995, the research-oriented vBNS consisted of OC-3 (45 mbps) and OC-12 (155 mbps) links. By the end of 1999, OC-48 (2 gbps) links had been deployed between San Francisco, Los Angeles, Denver, Chicago, Cleveland, New York, and

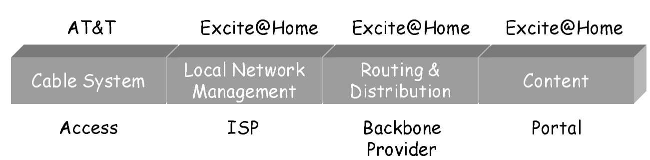

Figure 5. AT&T’s Internet Strategy

AT&T Cable, formerly TCI, provides local access via cable modems. AT&T’s Excite@Home unit provides ISP and backbone services to the cable systems. Excite also provides a Web Portal. The strategy spans access, backbone, and content. Competitors like AOL are concerned that their subscribers will be locked out of broadband access to the home.

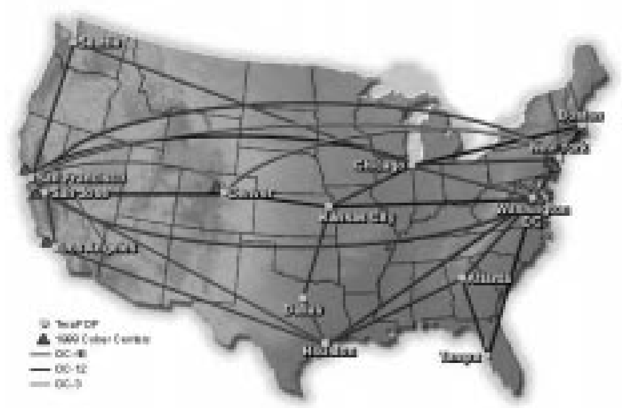

Figure 6. Qwest's IP Backbone (1999)

The bulk of Qwest's backbone is OC-48. OC-3's and OC-12's provide some regional and bypass bandwidth.

Washington. By 1997, it was common for NSPs to have OC-3 links in their backbones, replaced by OC-12 and -48 links by 1999 (see Figure 6). Cisco predicts widespread deployment of OC-192 (10 gbps) links in backbone networks by 2001. And of course, there are a very large number of national- and international-scale NSPs deploying such backbones today.

The distinction between NSPs and ISPs is somewhat fuzzy. In general, the relationship is many-to-many: an ISP will construct its network from long-haul capacity from more than one NSP, and an NSP will wholesale bandwidth to more than one ISP. While some NSPs only wholesale connectivity to ISPs, others will be their own ISPs for certain market sectors. While an NSP might sell national or international connectivity directly to major corporate clients, it might wholesale connectivity to ISPs that focus on the small/medium size businesses and consumers. To further complicate matters, it is not uncommon for an NSP to make an investment in an ISP, thereby gaining guaranteed demand for its backbone. An ISP may also make its own investment in an access network provider to obtain good access rates for its own customers.

3. The Emerging Internet Service Architecture

Beyond basic Internet connectivity, the first major service offering to emerge from ISPs was *Virtual Private Networks* (VPNs). This allowed a company to outsource its corporate interconnections to a third party. Rather than deploying and self-managing a private network, the customer shares connection resources, such as links and switches/routers, with other users. It negotiates a *service level agreement* (SLA) with the ISP to obtain loose guarantees on certain service attributes, such as worst case latencies, packet loss rates, and up times (typically the service provider pays a financial penalty if the SLA is not met). End-to-end encryption insures that company-confidential transmissions remain secure despite their transport in a "public" network.

The ISPs' second major offering was co-location and web hosting services. These allow a small or medium sized firm to gain a well-connected presence on the Web. This is accomplished by colocating the firm's own web server in the ISP's *Point of Presence* (PoP), a physical facility with high bandwidth connectivity to the Internet, usually via the ISP's own network service providers. Alternatively, the ISP may offer web servers in the PoP for lease by their clients. Some ISPs offer scalable servers based on processor clusters, to allow the web site to gracefully handle unexpectedly heavy access loads.

We are now in a period of rapid changes in the service architecture, supporting enterprise and web-based applications in the wide-area. As consumer users of the Internet, we are all aware of using the Internet for web access, content delivery, and electronic commerce. But another major trend has been the use of Internet technology for enterprise integration, within a firm as well as between the firm and its customers (other corporations as well as consumers) and suppliers. The result is a new model for information technology outsourcing: outsourced networking (ISPs/VPNs), outsourced web servers (Hosting services), outsourced facilities (Internet exchanges/NAPs and specially constructed buildings), outsourced data centers, and outsourced applications for enterprises, ISPs (and content providers) do not build their own networks, but rather overlay them on top of ANPs and NSPs (and even other ISPs!). ISPs obtain caching services from third parties, while web sites subscribe to still other caching services from third parties.

In the remainder of this section, we explore the structure and implications of this rich and complex service environment.

3.1 Infrastructure Technology Providers

Today's Internet services demand a range of technologies well beyond VPN and web hosting. The main new thrust is in the general area of content delivery services with the view of achieving high performance for latency/loss sensitive data streams: streaming media support, access device-specific transformations, and cache management.

There is an increasing use of real-time media on the Web. A number of radio and television stations are now "webcasting." Yahoo's Broadcast.com Portal provides a comprehensive set of links to publicly available sources of streaming media. Companies and educational institutions are also using these techniques for "desk-top" training and education. The essential enabling technology is streaming *media servers* from companies like RealNetworks (www.realnetworks.com).

The kinds of devices attaching to the Internet, from televisions to advanced cellular telephones and PDAs, create a demand for *services that transform web content* into a form for presentation on these devices. ProxiNet (www.proxinet.com) is an example of a technology provider in this arena. They provide PDA client software, called ProxiWeb, which supports Web access via their intermediate proxy servers, running their ProxiWare. These are scalable cluster servers inside the Internet that transform on demand Web content for presentation on the particular kind of end device.

While others provided their own proprietary Web search portals, Inktomi pioneered in developing the concept of a search service that would be sold as a component to enable rather than compete with portal sites. Inktomi's core technology is *scalable servers and resource management*, which enable specific portal and network-level services like search engines, cache management, and content transformation/filtering/transformation.

Numerous companies have developed *services for content delivery*. These include Sandpiper Networks, Akamai, BackWeb, and Marimba. Sandpiper distributes web content to servers deployed around the world. Akamai offers a similar overlay network of managed caches throughout the Internet to improve access performance to Web content. BackWeb collects content from multiple sources and "pushes" it to targeted subscribers in a time critical fashion. It provides service components for carefully managing the bandwidth allocated to distributing information, ensuring intended recipients become aware of new information directed to them, and tracking services to insure that the content provider can know that the needed content has been received and viewed by the receiver. Marimba offers push technology for keeping software applications up-to-date across the Internet.

Multicast Network Overlays? (e.g., FFNetworks)

3.2 Architectural Models

Figure 7 is a high level illustration of our architectural model for Internet services. The model consists of *Applications* constructed on top of *Application Infrastructure Services*. In turn, these services are delivered over *Application-specific Overlay Networks*. At the bottom are the *Internetworking* services, which provide the basic connectivity.

3.2.1 Applications

At the top are the end-user applications. These include *Portals*, *E-commerce* (electronic commerce), *E-tainment* (electronic entertainment), and *Media*. Note that there is not a clear separation among these--it is not uncommon to find elements of all of these applications classes under the umbrella of a given portal site.

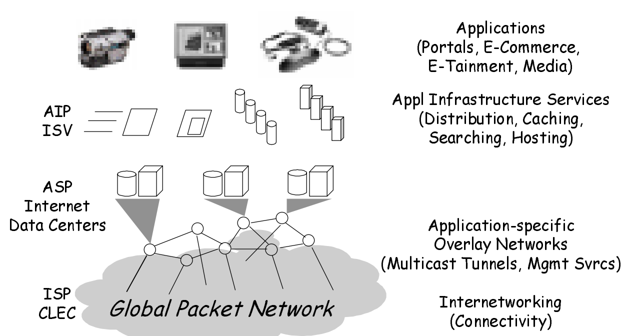

Figure 7. Architectural Model for Internet Services

Portals are web sites providing a gateway to a broad collection of information services for end users. These include news, stock quotations, and weather, as well as chat groups, electronic mail, and real-time messaging. These sites require infrastructure services for searching, content dissemination, content transformation, real-time messaging, and mail delivery.

E-commerce are sites for business-to-consumer (e.g., Amazon, Priceline), business-to-business (e.g., industrial supplies, UPS), and consumer-to-consumer (e.g., EBay) buying and selling. These applications require services for customer support (e.g., product presentation, comparison shopping, "shopping basket" support, order support) while integrating enterprise applications like inventory management and order tracking with the Web. A set of electronic business services such as credit card verification and clearing houses are also needed.

E-tainment are applications for interactive entertainment. While they have not as yet proven themselves, on-line game playing sites fall into this category. These place the heaviest demands on the underlying network to be "class of service" aware.

By "Media," we are referring to prestored streaming audio and video delivered over the Internet. These consist of television and radio clips, as well as real-time broadcast streams. The required services are related to efficient media dissemination and content delivery, and place demands on the underlying network technology to support classes of service.

3.2.2 Applications Infrastructure Services

The challenge faced by network-centered applications, such as the ones described above, is that they must achieve high and predictable performance in a highly heterogeneous environment spanning multiple administrative domains which cannot be completely controlled by the developer. The Internetwork-environment also presents an opportunity in the form of leveraged services that can be accessed over the network. This has led to the emergence of diverse third-party services for constructing applications. These services provide rich capabilities for information *distribution*, *caching*, *transformation*, *searching*, and *hosting*.

Distribution services are those capabilities for routing information through the connectivity provided by the Internet. One motivation is to achieve fault tolerance for content publishers, by distributing copies from the publisher to multiple servers distributed throughout the Internet. This can also achieve improved access performance for subscribers, by enabling regional access centers with mirrored content. Distribution also supports delivery to end subscribers, for such streaming media applications as Internet-based television and radio. Distribution services may also be used to "push" content to specific users, based on need-to-know or subscriptions. Stock quotations or news items are examples. Since the performance of distribution depends to some extent on the nature of the underlying connectivity between publishers and subscribers, a distribution-specific network may be overlain on top of the Internet. See the discussion in the next subsection.

Caching services distribute web content to regional caches to improve the access performance experienced by end subscribers. In addition to distribution, these offer services for managing the caches themselves. Because of the high degree of dynamic content on Web pages, cache services have evolved to separate the advertising content and other dynamic components of the web page from the static portions. Regional banner advertisements may be inserted into cached web pages. Cache residency may be reserved for particular publishers, or it may operate more traditionally, holding the most frequently accessed pages by a given user community. The former can be sold as a service to content providers seeking better performance for their end users, the latter as a service for ISPs to obtain both better performance and reduced access charges to the rest of the Internet.

Transformation services adapt web content to the capabilities of access devices, which may be limited in their screen, input, or processing capabilities. These may include personal digital assistants, smart phones, set-top boxes, televisions, or even automobiles. In general, a full-fledged web browser cannot be implemented within these devices at this point in the state-of-the-art. Even if sufficient processing were available in the end device, there are still advantages in terms of scalability and availability by executing the transformation capabilities on servers inside the network. Ad insertion into web pages based on regionalized or user-specific information is another kind of transformation service.

Searching is a common activity on the Web, and search services are widely available to end users via general and community-specific portals. Effective searching depends on having a comprehensive collection of information to be searched, as well as easy-to-use tools for specifying user queries and presenting results. Search services are among the first to be outsourced over the network. In this model, a particular service provider specializes in building the most comprehensive collection of Web information, utilizing its own web crawling and indexing techniques. It then resales search services to individual sites requiring the search capability. The portal site leverages the large collection maintained by the third party, and need not dedicate any resources to managing the every growing collections itself.

The Web search service provides an example of outsourced resources: a third party service provider is responsible for collecting and managing the collection, making sure it can process large numbers of simultaneous searches while maintaining high availability. This is an indirect example of hosting services, in this case, for storage, scalability, and availability. Hosting services provide external computational and storage resources colocated with high Internet connectivity for running web servers and other applications.

3.2.3 Application-Specific Overlay Networks

The Internet has long been described as a “network of networks.” By 1995, the Internet already encompassed over 50,000 different networks. The vast number of these were edge networks interconnected by a small number of backbone networks such as NSFNet. Since 1995, the number of regional and international networks, deployed by NSPs and ISPs, has grown very rapidly. The result is a rich and complex mesh of connectivity.

Yet it is remains difficult for applications to manage--or even to adapt to--the performance of the underlying connectivity upon which they are built. The research community has focused on developing technologies for network service guarantees (*integrated services*) or “promises” (*differentiated services*). Unfortunately, developers cannot wait for these capabilities to be standardized, incorporated in products, and deployed in the network. They have taken the expeditious approach of constructing their own *application-specific overlay networks*, with application-specific servers strategically placed within the resulting logical network.

The application-specific servers are specialized for supporting such services as caching, advertisement insertion, streaming media, transformation, network measurement, and so on. The servers may be standalone generalized or specialized computers, scalable clusters for high performance and fault tolerance, or may be “virtual” machines allocated within a hosting service.

The application-specific overlay networks are constructed from general connectivity and specialized tunnels among the servers. The tunnels may be used to form a logical backbone for content distribution or for other purposes that are specific to the application. The tunnels may be constructed so as to support global multicast, even though the underlying network backbones do not support multicast. To achieve the needed performance goals of the application, it is important that extensive performance measurements be maintained. This information is exploited to monitor and configure the overlays and placement of services within the network to assure the required level of performance.

3.2.4 Internetworking

The overlay networks are constructed on top of the basic connectivity mesh provided by the collection of NSPs and ISPs. Within a given provider’s network cloud, it is possible to obtain service level agreements, quantifying the worst case latencies and packet loss rates. These are valuable for configuring intra-network provider tunnels, but offer little assistance in understanding the performance of tunnels that span multiple providers.

Most ISPs also provide colocation services, allowing their customers to place services close to the ISP’s high bandwidth connectivity to the rest of the Internet. These offer ideal locales to place the servers of the overlain content delivery network.

3.3 Enterprise Evolution

Against this backdrop of service differentiation and outsourcing of information technology capabilities, we look at the way the Internet has led to a gradual outsourcing of enterprise information technology functions.

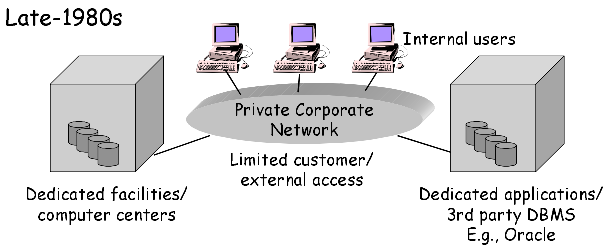

Figure 8. Dedicated Network/Applications/Facilities

This was the common enterprise situation in the late 1980s.

The state-of-the-art in late 1980s is illustrated in Figure 8. The typical enterprise had dedicated "everything," with the possible exception of some major portions of enterprise software, such as database management systems or transaction processing systems obtained from third party developers like Oracle. That is, the information technology of the enterprise consisted of dedicated computing facilities interconnected by a private corporate network assembled from lines leased from a telephone company. The applications were accessible primarily by users from within the organization, with very limited access by external customers or business partners. In specific business domains, third party managed "value-added networks" emerged, linking suppliers and customers. Think in terms of the old days of calling a phone number to check on flight status: there was no direct access by consumers to the flight status databases maintained by the airlines.

Now consider the time frame of 1995, just before the Web's explosive growth. The situation is captured in Figure 9. The period is characterized by the rise of major third party applications developers, such as PeopleSoft and BAAN. These are called *Enterprise Resource Planning (ERP) applications*, providing automation and integration for certain common business processes such as procurement, human resource management, accounting, treasury management, project management, sales and logistics, materials management, supply chain planning, and revenue management. Previously, an organization might build some of these subsystems for its own use using in-house or consulting staff. The difference brought about by the ERP developers was the sheer scale of applications components that could be assembled from off-the-shelf, dramatically decreasing

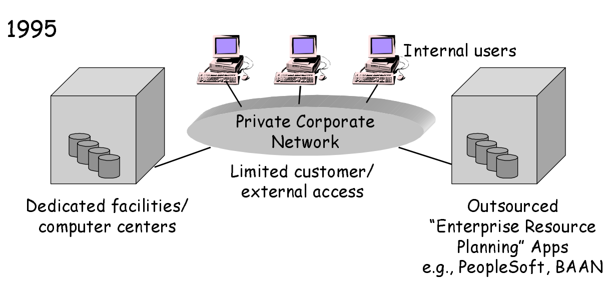

Figure 9. Rise of ERP Applications

Enterprise Resource Planning applications developed by third party software firms found wide-spread acceptance in the mid-1990s.

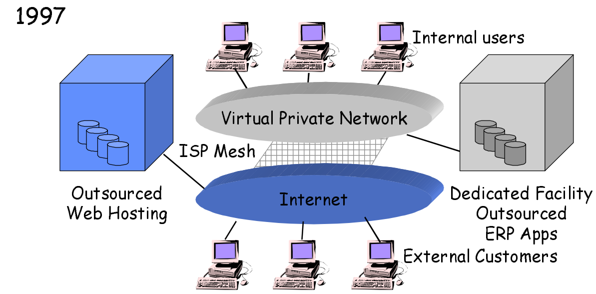

Figure 10. Outsourced Connectivity and Web Hosting

Driven to provide connectivity to external customers and business partners, corporations move to web hosting and VPNs provided by ISPs.

the cost of outsourcing applications development (at least in theory--the easy fit of the standard components with a given organization's business practices is more promise than reality). A collection of components is often specifically assembled and customized for broad industry segments such as Communications, Government, Financial Services, Health Care, Higher Education, Manufacturing, Retail, Service Industry, and Transportation.

With the establishment of Internet Service Providers after the 1991 commercialization of the Internet and the 1996 Telecommunication Act, an organization could outsource its wide-area network as a VPN. For the time frame of 1997, this is shown in Figure 10. In effect, encryption enables a secure tunnel through the public Internet to interconnect the remote private local area networks of an organization. In the same time period, outsourced Web Hosting became a valuable service. This allowed the enterprise to provide Web content to external users without opening their private corporate networks to direct access by outsiders. Third party Web hosting was particularly useful for small and medium sized businesses, who could not afford to establish their own professionally staffed computer centers with high bandwidth connectivity to the Internet. Note the continued separation between external and internal users. The latter still access enterprise applications inside dedicated computer center facilities.

Around this time, portals had emerged as a major new kind of Internet application. These are implemented on top of services located other places and accessed over the Internet, such as search, caching, ad insertion and management, e-commerce. This is shown in Figure 11.

By 1999, these outsourcing trends have come together. Third-party enterprise applications are now provided by Applications Service Providers, and run external to the enterprise, with connectivity provided by VPNs. ISPs no longer construct their own physical buildings for holding the web hosting data centers. These are now provided by third-party facility management companies. Some web hosting companies no longer offer access to their managed backbone, but rather construct the connectivity between their facilities from that provided by multiple ISPs. Such overlay networks are now widespread, for cache management and content delivery. This is shown in Figure 12.

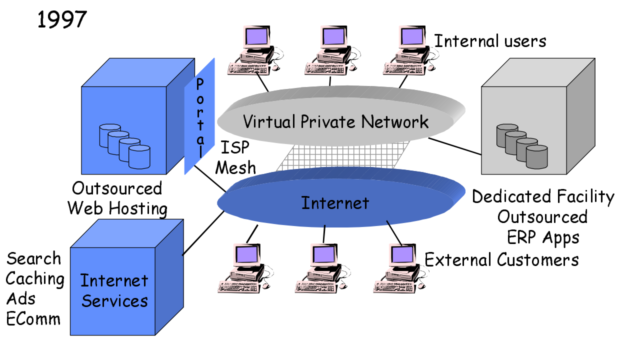

Figure 11. Portals and Enabling Services

Component services accessible over the network are now used to construct the portals themselves.

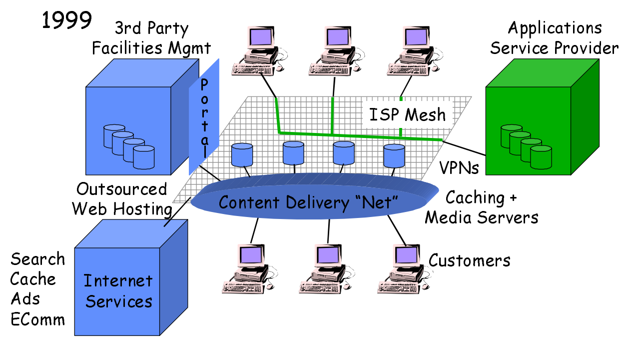

Figure 12. Outsourced Applications

The latest thrust in outsourcing consists of applications that can be run on machines managed by external service providers.

3.4 Emerging Internet Business Layers

Bringing this discussion together, Figure 13 offers a layered architecture for describing the way in which Internet services are built on top of more primitive services. Table 1 summarizes the Service Levels, the functions they provide, and offers examples of companies operating at the indicated level.

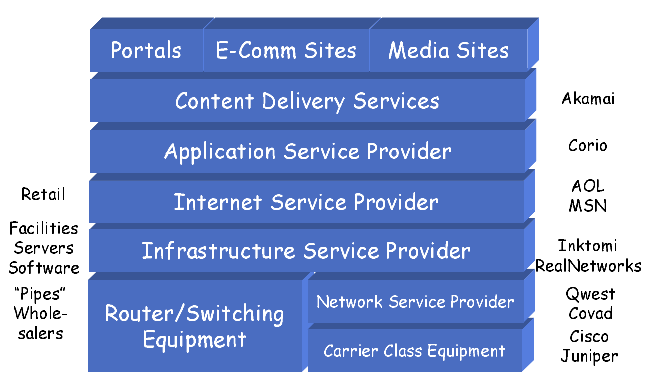

Figure 13. Layered Internet Service Architecture

Applications are built on top of content delivery services. These sit on top of Application Service Providers, which in turn are on top of Internet Service Providers. Supporting the latter are the Infrastructure Service Providers.

Service Layer	Function	Example Companies
Network Service Provider	Backbones + POPs	Qwest, Level3, Covad
Internet Service Provider	Managed connectivity, VPN service	Concentric, Verio, PSINet, Earthlink, AOL
Hosting	Scalable hosting platform	Exodus, AboveNet
Facilities Provider	Secure, protected buildings; Internet Exchanges	Equinix, Concentric Super-POP
ERP Apps Provider	Specific Enterprise Applications	PeopleSoft, BAAN
App Support Provider	App Integration and Management	Corio
App Infrastructure Support Provider	Support for App Development	BEA Systems, Vitria, NEON, Mercator
Content Delivery Network Provider	Distribution/Caching	RealNetworks, Marimba, Sandpiper, Akamai

Table 1. Internet Service Layers

4. Case Studies

4.1 Next Generation Network Equipment Provider: Redback/Siara

Siara's focus is on building equipment for Class of Service (CoS)-aware networks. These are enabling for content delivery networks, and incorporate certain features for ISP's who wish to charge for differentiated services.

Such equipment supports mechanisms for flow policing, shaping, CoS marking, policy routing, and per-flow statistics gathering. Packets are inspected and classified based on rules. They are filtered and selectively forwarded at the network edge. Packets are sorted and queued in the network core.

An essential feature is ability to collect per packet service statistics. This enables SLAs and billing services for network providers. Multiple service levels are also provided, and are customizable by the user or the application. Given that all network equipment is programmable, new features can be configured and made available very rapidly. Bandwidth allocations to flows can be also be adjusted dynamically.

A goal is direct support for dynamic service delivery. The programmability of the network equipment allows accelerated service development and enhancement, and rapid service provisioning/configuration. Also supported is

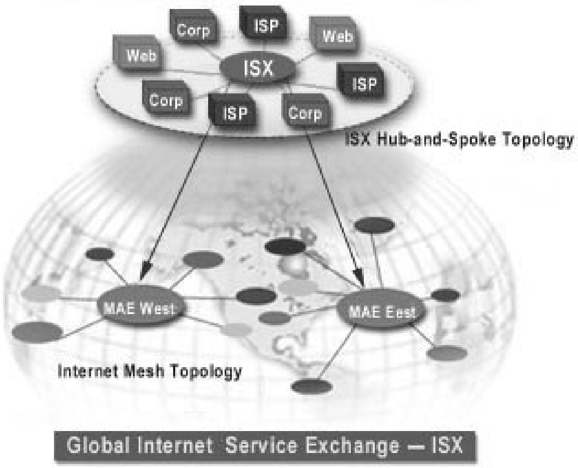

Figure 14. AboveNet's Colocation Concept

The ISX is a distributed facility that combines hosting, Interne exchange, and a high speed global backbone.

intelligent service routing. The equipment provides capabilities for making the network CoS-aware, thereby supporting network self-adjustment via signaling, routing, and switching.

4.2 An Internet Business Exchange Provider: Equinix

Equinix builds and maintains secure and sophisticated utility buildings. Facilities include back-up power generation and fire suppression systems. Such buildings are the ideal locals for ISPs, hosting service providers, and content providers to locate their equipment for connection to the Internet.

4.3 A Hosting Service Provider: AboveNet

AboveNet offers a service they call the "Global Internet Service Exchange (ISX)." They provide data centers that allow ISPs and content providers to be colocated. Their facilities offer one hop connectivity to their own high speed, globe spanning backbone. Figure 14 captures their architectural model.

4.4 An ASP: Corio

Corio defined the concept of the Applications Service Provider (ASP). They provide for managed outsourced applications by creating a centralized location for managing a customer's applications, providing service management, application administration, systems administration, and network administration. Their architectural model is captured in Figure 15.

Corio's data centers are configured within the physical hosting facilities provided by ISPs further down the food chain. In Corio's case, the ISPs are Concentric and Exodus.

Corio is responsible for ensuring that the connectivity between the customer's facilities and the Corio data center is secure. It is essential that an ASP provide a secure net-

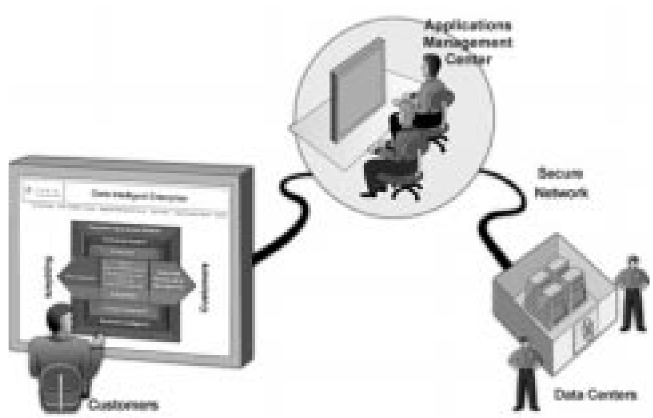

Figure 15. Corio's ASP Service Model

Enterprise applications are run in Corio managed data centers over a secure network. The management of the applications, the network, and the data center are outsourced to Corio.

work, implemented with the necessary firewall, intrusion detection, and application level security technologies.

4.5 Content Delivery Service Provider: Sandpiper

Sandpiper is an example of a Content Delivery Service Provider (Sandpiper recently merged with Digital Island, an ISP). Sandpiper defined Content Delivery Networks (CDNs) as a dedicated network of servers, deployed throughout the Internet, to which web publishers distribute their content based on subscriptions.

Sandpiper's distribution service is based on a distributed infrastructure, replication technology to distribute content to multiple servers, rendezvous technology to route users to the nearest server, multiple content support, publication tools, extensive logging capabilities and performance monitoring tools, security services, and service provisioning capabilities for bandwidth, storage, and processing.

4.6 Content Delivery Service Provider: SkyCache

SkyCache provides a satellite-based broadcast overlay network to improve the distribution of Internet information. The motivation is to reduce router-to-router hops and packet losses within the Internet, eliminate ISP Internet clogs during peak traffic or one-time event spikes, achieve high bandwidth even beyond the current reach of fiber deployment, and enable high quality streamed content and high volume data transmission.

4.7 Content Delivery Service Provider: Akamai

Akamai is yet another content delivery service provider, focusing on distributed cache management. Akamai's service is based on a four step process: (1) the user enters a standard URL, (2) the Web server returns HTML with embedded URLs pointing to servers within Akamai's own

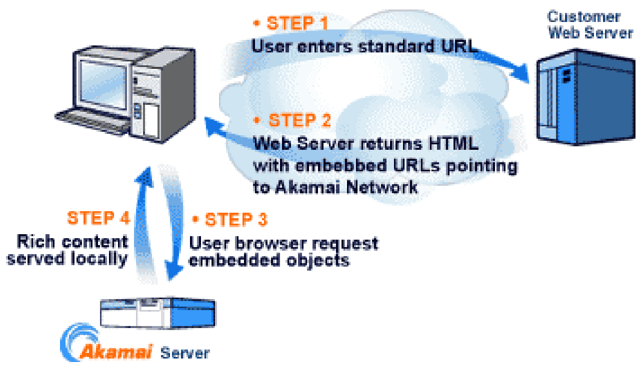

Figure 16. Akamai Rendezvous Process to the Nearest Cache Server

Akamai's overlay network provides capabilities to remap generic URLs into a URL for a copy of the web object that is closest to the requesting node.

overlay network, (3) the user's browser requests the embedded objects, and (4) these are served from the closest server. The process is indicated in Figure 16. Akamai currently has over 1700 servers deployed across 100 ISPs spanning 30 countries.

5. The Future of Internet Services

Similar to the complex distinction between an ANP, an ISP, and a NSP, service providers with expertise in one layer are branching out into other layers. Will the horizontal service model take off? Or are we stuck with competing vertical stovepipes, until there are just a handful of service companies leading to defacto standard architectures? Is there a standard API for search services for example?

Other than searching and caching, can you make a profitable business out of being a narrow service provider?

What is the charging/billing infrastructure needed to make horizontal service provision take off? Monitoring/measuring infrastructure and services? Consumer Reports services, referral services.

Quality/class of service aware event networks to support message oriented applications (e.g., TIBCO, Vitria). Event classes include: unsolicited broadcast/notification, broadcasting, anonymous pub/sub, sophisticated filters and rules engines, XACTional semantics. How are these built on top of multicast networks, making use of smart agents? Transactional messages, perishable messages, selection criteria, times, push or pull, peeking or browsing.

Will real-time streaming media on the Internet be the next wave of killer applications?

Intelligent collaborating agents in the network, "adaptive, collaborating, self-organizing," leading to increasing automation? Cogents=collaborative agents

6. Summary and Conclusions

7. References

[1] L. Kleinrock, et al., "Realizing the Information Future: The Internet and Beyond," National Research Council, National Academy Press, Washington, DC, 1994. See http://www.nap.edu/readingroom/books/rtif.

[2] B. Leiner, et al., "The Past and Future History of the Internet," Communications ACM, V 40, N 2, (February 1997), pp. 102-108. Version also at http://www.isoc.org/internet/history/brief.html.

[3] V. Cerf, R. Kahn, "A Protocol for Packet Network Interconnection," *IEEE Trans. Communications*, V. COM-22, N. 5, (May 1974), pp. 627-641.

[4] Merit Networks, "The NSF Backbone Project, 1987-1995," http://www.merit.net/merit/archive/nsfnet/.

[5] University Corporation for Advanced Internet Development, http://www.internet2.org.

[6] B. Chinoy, T. Salo, "Internet Exchanges: Policy Driven Evolution," in B. Kahin, J. Keller, eds., *Coordinating the Internet*, The MIT Press, Cambridge, MA, 1997.

[7] D. Mills, H. Braun, "The NSFNet Backbone Network," Proceedings ACM SIGCOMM Conference, 1987.

[8] K. Hafner, M. Lyon, *Where Wizards Stay Up Late: The Origins of the Internet*, Simon and Schuster, New York, 1996.

[9] R. Haynal, "ISP Page," http://navigators.com/isp.html.

[10] Telecommunication Act of 1996, http://www.fcc.gov/telecom.html.

8. Glossary

Network Service Provider (NSP): Network backbone providers, usually of a national or global scale, incorporating very high speed links within their network design (e.g., OC-12/48 migrating towards OC-128). NSPs peer at Internet Exchange Points (IXCs) also known as Network Access Points (NAPs) in earlier terminology. Note that many NSPs may have a portion of their business that is ISP-like while wholesaling connectivity to other ISPs. The distinction may be market driven, by region or size of the customer. Excite@Home has created its own wide-area routing and distribution network for the cable industry.

Access Network Provider (ANP): Generally operate in a wholesale mode, providing access lines to an ISP, who in turn retails connectivity to a consumer or a business customer. Examples include Covad and NorthPoint. AT&T Cable is a cable modem access provider.

Internet Service Provider (ISP): More retail-oriented than an NSP. Will build own access and backbone network from pieces provided by various NSPs. AOL is an example, as is Concentric and Verio. Excite@Home also provides local network management, making it an ISP as well as an NSP.

Portal: A web site providing a broad index to end user-based information services, like news, weather, and stock quotations.

Application Service Provider (ASP): Provides secure access to enterprise applications on its own servers colocated with ISPs to insure excellent network connectivity. Corio is an example of an ASP.

Application Infrastructure Provider (AIP): tools for building web enabled applications, including e-commerce building blocks for credit card verification. Inktomi’s search engine is an example of an infrastructure component that supports the development of portals.

Internet Infrastructure Service Provider (IISP): web searching and caching services sold to portals and Internet Service Providers.

Content Delivery Networks: intelligent routing + replicated content + service from closest/least congested location)

Web Hosting Services: Servers colocated with high bandwidth Internet connectivity, allowing small to medium size business to maintain a high quality presence on the Web. The hosting platform may provide scalable resource allocation, allowing the hosted site to grow gracefully in response to increased user access demand.