Chapter 2

Circuit and Packet Switching

2.1 Introduction

It is widely assumed that, for reasons of efficiency, the various communication networks (Internet, telephone, TV, radio, ...) will merge into one ubiquitous, packet-switched network that carries all forms of communications. This view of the future is particularly prevalent among the Internet community, where it is assumed that packet-switched IP is the layer over which everything else will be carried. In this chapter, I present evidence so as to argue that this will not happen. This stance is controversial, and is difficult to make concrete, as any attempt to compare the various candidates for the transport infrastructure[1] is fraught with lack of data and the difficulty of making apples-with-apples comparisons. Therefore, the evidence presented here is different from other chapters in this thesis. Observations, case studies, and anecdotal data (rather than controlled experiments, simulations and proofs) are used to take a stance and to predict how the network architecture will evolve.

Whatever the initial goals of the Internet, two main characteristics seem to account for its success: *reachability* and *heterogeneity*. IP, the packet-switching protocol that is the basis for the Internet, provides a simple, single, global address to reach every host, enables unfettered access between all hosts and adapts the topology to restore reachability when

[1] In this chapter, transport is used in the sense of the infrastructure over which many service networks run, not in the sense of the OSI protocol layer.

links and routers fail. IP hides heterogeneity in the sense that it provides a single, simple service abstraction that is largely independent of the physical links over which it runs. As a result, IP provides service to a huge variety of applications and operates over extremely diverse link technologies.

The growth and success of IP has given rise to some widely held assumptions amongst researchers, the networking industry and the public at large. One common assumption is that it is only a matter of time before IP becomes the sole global communication infrastructure, dwarfing, and eventually displacing, existing communication infrastructures such as telephone, cable and TV networks. IP is already universally used for data networking in wired networks (enterprise networks and the public Internet), and is being rapidly adopted for data communications in wireless and mobile networks. IP is also increasingly used for both local and long-distance voice communications, and it is technically feasible for packet-switched IP to replace SONET/SDH.

A related assumption is that IP routers (based on packet switching and datagram routing) will become the most important, or perhaps only, type of switching device inside the network. This is based on our collective belief that packet switching is inherently superior to circuit switching because of the efficiencies of statistical multiplexing and the ability of IP to route around failures. It is widely assumed that IP is simpler than circuit switching and should be more economical to deploy and manage. And with continued advances in the underlying technology, we will no doubt see faster and faster links and routers throughout the Internet infrastructure. It is also widely assumed that IP will become the common convergence layer for all communication infrastructures. All communication services will be built on top of IP technology. In addition to information retrieval, we will stream video and audio, place phone calls, hold video-conferences, teach classes, and perform surgery.

On the face of it, these assumptions are quite reasonable. Technically, IP is flexible enough to support all communication needs, from best-effort to real-time. With robust enough routers and routing protocols, and with extensions such as weighted fair queueing, it is possible to build a packet-switched, datagram network that can support any type of application, regardless of their requirements.

In spite of all the strengths of IP, this chapter will argue how it will be very hard for IP to displace existing networks. It will also conclude how many of the assumptions discussed above are not supported by reality, and do not stand up to close scrutiny.

The goal of this is to question the assumption that IP will be *the* network of the future. The conclusion is that if we started over - with a clean slate - it is not clear that we would argue for a universal, packet-switched IP network. In the future, more and more users and applications will demand predictability from the Internet, both in terms of the availability of service and the timely delivery of data. IP was not optimized to provide either, and so it seems unlikely to displace networks that already provide both. In this chapter, I take the position that while IP will be the network layer of choice for best-effort, non-mission critical and non-real-time data communications (such as information exchange and retrieval), it will live alongside other networks, such as circuit-switched networks, that are optimized for high revenue time-sensitive applications that demand timely delivery of data and guaranteed availability of service.

This is indeed a controversial position. Nevertheless, as researchers we need to be prepared to take a step back, to take a hard look at the pros and cons of IP, and its likely future. As a research and education community, we need to start thinking how IP will co-exist and co-operate with other networking technologies.

2.1.1 Organization of the chapter

Section 2.2 provides a more detailed description of circuit switching and packet switching than in Chapter 1. It also describes part of the earlier work on these two switching techniques. Section 2.3 dissects some of the claims about IP, especially when compared to circuit-switched networks. This section tries to demystify those claims that do not hold up to scrutiny. Section 2.4 discusses the implications for the network architecture. Section 2.5 concludes this chapter.

2.2 Background and previous work

Before starting our discussion about whether IP can be the basis of all communication networks, I will give some background about the two main switching techniques in use today: circuit switching and packet switching.

2.2.1 Circuit switching

Circuit switching was the first switching technique used in communication networks because it is simple enough to carry analog signals. This thesis will just focus on the digital version of circuit switching. Of course, the main example of its use is the phone system [72], but it is also used in the core of the Internet in the form of SONET/SDH and DWDM equipment [81, 126]. In circuit switching, the transmission medium is typically divided into channels using Frequency Division Multiplexing (FDM), [2] Time Division Multiplexing (TDM) or Code Division Multiplexing (CDM) [172]. A circuit is a string of concatenated channels from the source to the destination that carriers an information flow.[3]

To establish the circuits, a signaling mechanism is used. This signaling only carries control information, and it is considered an overhead. It is also the most complex part in circuit switching, as all decisions are taken by the signaling process. It is commonly assumed that the signaling and per-circuit state management make circuit switches hard to design, configure and operate.

In circuit switching the channel bandwidth is reserved for an information flow. To ensure timely delivery of the data, the capacity of the circuit has to be at least equal to the peak transmission rate of the flow. In this case, the circuit is said to be peak allocated, and then the network offers a connection-oriented service with a perfect quality of service (QoS) in terms of delay jitter and bandwidth guarantees, However, this occurs at the cost of wasting bandwidth when sources idle or simply slow down.

Contention only occurs when allocating channels to circuits during circuit/call establishment. If there are not enough channels for the request, the call establishment may be delayed, blocked or even dropped. In contrast, once the call is accepted, resources are not shared with other flows, eliminating any uncertainty and, thus, removing the need for buffering, processing or scheduling in the data path. When circuits are peak allocated, the only measure of Quality of Service (QoS) in circuit switching is the blocking probability of a call.

[2](Dense) Wavelength Division Multiplexing, (D)WDM, is a subclass of FDM that uses optical wavelengths as channels.

[3]Note that the source and the destination need not be edge nodes. They can be aggregation nodes in the middle of the network that combine several user flows into one big information flow.

To summarize, circuit switching provides traffic isolation and traffic engineering, but at the expense of using bandwidth inefficiently and signaling overhead. It is often said that these two drawbacks make circuit switching highly inflexible, especially in a highly dynamic environment such as the Internet. I will argue in this that these drawbacks are outweighed by the advantages of using more circuit switching in the core of the network.

2.2.2 Packet switching

Packet switching is the basis for the Internet Protocol (IP) [152, 172]. In packet switching, information flows are broken into variable-size packets (or fixed-size cells as in the case of ATM). These packets are sent, one by one, to the nearest router, which will look up the destination address, and then forward them to the corresponding next hop. This process is repeated until the packet reaches its destination. The routing of the information is thus done locally, hop-by-hop. Routing decisions are independent of other decisions in the past and in other routers; however, they are based on network state and topology information that is exchanged among routers using BGP, IS-IS or OSPF [148]. The network does not need to keep any state to operate, other than the routing tables.

The forwarding mechanism is called store-and-forward because IP packets are completely received, stored in the router while being processed, and then transmitted. Additionally, packets may need to be buffered locally to resolve contention for resources. [4] If the system runs out of buffers, packets are dropped.

With the most scheduling policies, such as FCFS and WFQ, packet switching remains work conserving; it keeps the link busy as long as there are packets waiting to be sent. This allows it to have a statistical multiplexing gain; that is, the capacity of an outgoing link can be much smaller than the sum of its tributaries and still have a packet delay or drop probability within certain statistical bounds. This gain is higher when traffic is more bursty. The buffering needs and the statistical multiplexing are the main characteristics of packet switching, and they will be crucial in its comparison with circuit switching.

In the Internet, the network service is connectionless and best effort; that is, it provides

[4] Resources have contention when they have more arrivals/requests that what they can process. Two examples are the outgoing links and the router interconnect.

no delivery guarantees. Reliability, flow control and connection-oriented services are provided by end-to-end mechanisms, such as with TCP [153]. Because the underlying service is best effort, there are no guarantees in terms of packet drops, maximum delay, delay jitter or bandwidth.

Much research was done in the early days of computer networking comparing circuit switching, packet switching and message switching (a variant of packet switching, in which the whole information flow is treated as a single switching unit) [96, 10, 164, 97, 175, 95]. Most of the work was done in the context of packet radio, satellite, and local area networks and shows how in these environments packet switching provided higher throughput for a given bound on the average delay. Packet switching not only made an effective use of the network bandwidth, but it also was robust and resilient to node and link failures.

Later work on different scheduling algorithms and signaling mechanisms, such as Weighted Fair Queueing (WFQ) [62], Generalized Processor Sharing (GPS) [141], Differentiated Services (DiffServ) [16], Integrated Services (IntServ) [20] and Deficit Round Robin (DRR) [113], showed how packet switching can also provide QoS guarantees if the admission of new flows to the network can be controlled.

2.3 IP Folklore

This section tries to identify some folkloric assumptions about IP and the Internet, and it examines each in turn. I will start with the most basic assumption, and the easiest to dispel: that the Internet *already* dominates global communications. This is not true by any reasonable metric: market size, number of users, or the amount of traffic. Of course, this is not to say that the Internet will not grow over time to dominate the global communications infrastructure; after all, the Internet is still in its infancy. It is possible — and widely believed — that packet-switched IP datagrams will become the *de-facto* mechanism for all communications in the future. And so one has to consider the assumptions behind this belief and verify whether packet-switched IP offers inherent and compelling advantages that will lead to its inevitable and unavoidable dominance. This requires the examination of some “sacred cows” of networking; for example, that packet switching is more efficient than circuit switching, that IP is simpler, it lowers the cost of ownership, and it is more robust when there are failures in the network.

2.3.1 IP already dominates global communications

It has been reported that the Internet already carries more traffic than the phone system [122, 162], and that the difference in traffic volume will become bigger and bigger over time because Internet traffic is growing at a rate of 100% per annum versus a rate of 5.6% per year for voice traffic [48].

Despite this phenomenal success of the Internet, it is currently only a small fraction of the global communication infrastructure, which consists of separate networks for telephones, broadcast TV, cable TV, satellite, radio, public and private data networks, and the Internet. In terms of revenue, the Internet is a relatively small business. The US business and consumer-oriented ISP markets have revenues of $13B each (2000) [28, 29], in contrast, the TV broadcast industry has revenues of $29.8B (1997), the cable distribution industry $35.0B (1997), the radio broadcast industry $10.6B (1997) [180], and the phone industry $268.5B (1999), of which $111.3B correspond to long distance and $48.5B to wireless [88]. The Internet reaches 59% of US households [133], compared to 94% for telephones and 98% for TV [127, 147]. Even though Internet traffic doubles every year, revenues only increase 17% annually (2001) [162], whereas long-distance phone revenues increase 6.7% per year (1994-97) [136]. If these growth rates were kept constant, IP revenues would not surpass those of the long-distance phone industry until 2017.[5]

If we restrict our focus to the data and telephony infrastructure, the core IP router market still represents a small fraction of the public infrastructure, contrary to what happens in the private enterprise data networks. As shown in Table 2.1, the expenditure on core routers worldwide was $1.7B in 2001, compared to $28.0B for transport circuit switches. So in terms of market size, revenue, number of users, and expenditure on infrastructure, it is safe to say that IP does not currently dominate the global communications infrastructure.

Figure 2.1 illustrates the devices currently used in the public Internet. The current communication infrastructure consists of a transport network — made of circuit-switched SONET/SDH and DWDM devices — on top of which run multiple service networks.

[5] It is interesting to note that for IP revenues to surpass those of long-distance telephony the Internet revenue per household would have to multiply by 358%.

Segment	Market size
Core routers	\$1.7B
Edge routers	\$2.4B
SONET/SDH/WDM	\$28.0B
Telecom MSS	\$4.5B

Table 2.1: World market breakup for the public telecommunications infrastructure in 2001 [161, 158, 159, 157].

The service networks include the voice network (circuit-switched), the IP network (datagram, packet-switched), and the ATM/Frame Relay networks (virtual-circuit-switched). Notice the distinction between the circuit-switched transport network, which is made of SONET/SDH and optical switches that switch coarse granularity ($n \times STS-1$, where an STS-1 channel is 51 Mbit/s), and the voice service circuit switches, which include Class 4 and Class 5 systems that switch 64Kbps voice circuits and handle various telephony-related functions. When considering whether IP has or will take over the world of communications, one needs to consider both the transport and service layers. In other words, for universal packet transport I am considering using a packet network to replace the transport infrastructure; and for voice-over-IP (VoIP) I am considering an application built on top of an IP network that replaces the traditional Class 4/5 TDM voice switches.

In order to examine the merits of a packet-switched IP network, one needs to compare it with an alternative. The obvious alternative is circuit switching. In one respect, this is not an apples-with-apples comparison; the packet-switched IP data network today already operates over a circuit-switched transport infrastructure. If we consider only the core of the network, we find essentially a central core of circuit switching surrounded by IP routers. It helps to think of the comparison as a question as to which one of two outcomes is more likely: Will the packet-switched IP network grow to dominate and displace the circuit-switched transport network, or will the (enhanced) circuit-switched TDM and optical switches continue to dominate the core transport network?

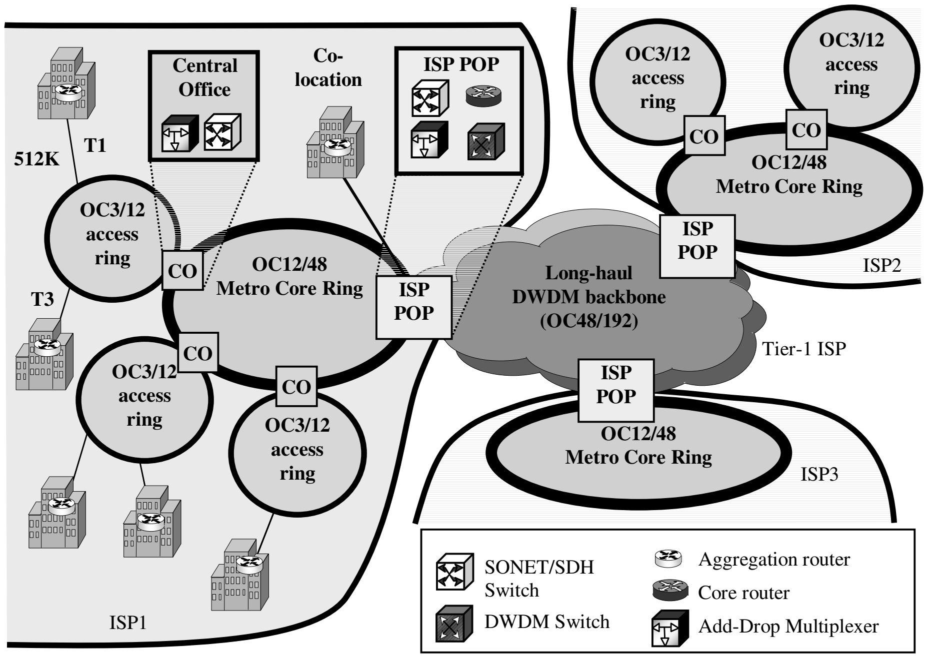

Figure 2.1: Architecture of the public Internet. There are also many large private voice and data networks that consist of IP routers, LAN switches and voice switches at customer premises.

2.3.2 IP is more efficient

> "Analysts say [packet-switched networks] can carry 6 to 10 times the traffic of traditional circuit-switched networks." — **Business Week**.

From the early days of computer networking, it has been well known that packet switching makes efficient use of scarce link bandwidth [10]. With packet switching, statistical multiplexing allows link bandwidth to be shared by all users, and work-conserving link sharing policies (such as FCFS and WFQ) ensure that a link is always busy when packets are queued-up waiting to use it. In contrast, with circuit switching, each flow is assigned its own channel, so a channel could go idle even if other flows are waiting. Packet switching (and thus IP) makes more efficient use of the bandwidth than circuit switching, which was particularly important in the early days of the Internet when long haul links were slow, congested and expensive.

It is worth asking: What is the current utilization of the Internet, and how much does efficiency matter today? Odlyzko and others [135, 47, 90, 23] report that the core of the Internet is heavily overprovisioned, and that the average link utilization in links in the core is between 3% and 20% (compared to 33% average link utilization in long-distance phone lines [135, 160]). The reasons that they give for low utilization are threefold: First, Internet traffic is extremely asymmetric and bursty, but links are symmetric and of fixed capacity; second, it is difficult to predict traffic growth in a link, so operators tend to add bandwidth aggressively; third, with falling prices for coarser bandwidth granularity as faster technology appears, it is more economical to add capacity in large increments.

There are other reasons to keep network utilization low. When congested, a packet-switched network performs badly, becomes unstable and can experience oscillations and synchronization. Many factors contribute to this. Complex and dynamic interaction of traffic means that congestion in one part of the network will spread to other parts. Further, the control packets (such as routing packets) are transmitted *in-band* in the Internet, and hence they are more likely to get lost and delayed when the data-path is congested. When routing protocol packets are lost or delayed due to network congestion or control processor overload, it causes an inconsistent routing state, and may result in traffic loops, black holes, and disconnected regions of the network, which further exacerbate congestion in the data path [107, 55]. Currently, the most effective way for network providers to address these problems is by preventing congestion and keeping network utilization low.

But perhaps the most significant reason that network providers overprovision their network is to give low packet delay. Users want predictable behavior, which means low queueing delay, even under abnormal conditions (such as the failure of several links and routers) [90, 77]. As users, we already demand (and are willing to pay for) huge overprovisioning of Ethernet networks (the average utilization of an Ethernet network today is about 1% [47]) simply so that we do not have to share the network with others, and so that our packets can pass through without queueing delay. We will demand the same behavior from the Internet as a whole. We will pay network providers to stop using statistical multiplexing and to instead overprovision their networks. The demand for lower delay will drive providers to decrease link utilization even more than it is today.

Therefore, even though in theory a statistical multiplexed *link* can potentially yield a higher network utilization and throughput, in practice, to maintain a consistent performance and reasonably stable *network*, network operators significantly overprovision their network, thus keeping the network utilization low.

But simply reducing the *average* link utilization will not be enough to make users happy. For a typical user to experience low utilization, the *variance* of the network utilization also needs to be low. There are two flavors of variance that affect the perceived utilization: variance in time (short-term increases in congestion during busy times of the day), and variance by location (while most links are idle, a small number are heavily congested). If we pick some users at random and consider the network utilization their traffic experiences, our sample is biased in favor of users who find the network to be heavily congested. This explains why, as users, we know the average utilization to be low, but find that we often experience long queueing delays.

Reducing variations in link utilization is hard. Without sound traffic management and traffic engineering, the performance, predictability and stability of large IP networks deteriorate rapidly as load increases. Today, we lack effective techniques to reduce the unpredictability of performance introduced by variations in link utilization. It might be argued that the problem will be solved by research efforts on traffic management and congestion control (to control and reduce variations in time), as well as work on traffic engineering and multipath routing (to load-balance traffic over a number of paths). But to date, despite these problems being understood for many years, effective measures are yet to be introduced.

We can expect that over time users will demand lower and lower queueing delays in the Internet. This means that as users, we collectively want network providers to stop using statistical multiplexing and to instead overprovision their networks *as if they were circuit switched* [115, 137, 77]. To date, network providers have responded to our demands by overprovisioning, by publishing delay measurements for their network, and by competing on the basis of these numbers. In the long term, the demand for lower delay will drive providers to make link utilization even lower than it is today, and network utilization will continue to decrease as the world economy becomes more dependent on the Internet.

One can take the demand for low delay one step further, and ask whether users experience the lowest response times in a packet-switched network. Intuition suggests that packet switching will lead to lower delay: A packet-switched network easily supports heterogeneous flow rates, and flows can always make forward progress because of processor sharing in the routers. In practice, it does not make much difference whether packet switching or circuit switching are used. This is studied in detail in Chapter 3, which (by analysis and simulation) studies the effect of replacing the core of the network with dynamic fine-granularity circuit switches, as described in Chapter 4. Let's define the user response time as the time it takes from when a user requests a file until this file finishes downloading. Web browsing and file sharing represent over 65% of Internet transferred bytes today [31], and so the request/response model is representative of typical user behavior. Now consider two types of network: one is the current packet-switched network in which packets share links and each flow makes constant, albeit slow, forward progress over congested links. The other network is a hypothetical comparison. Each new application flow triggers the creation of a low bandwidth circuit in the core of the network, similar to what happens in the phone network. If there are no circuits available, the flow is blocked until a channel is free. As we will see in Chapter 3, at the core of the network, where the rate of a single flow is limited by the data-rate of its access link, simulations and analysis suggest that the average user response time of both techniques is the same, independent of the flow length distribution.

In summary, even though packet switching can lead to more efficient link utilization, unpredictable queueing delays force network operators to operate their networks very inefficiently. One can conclude that while efficiency was once a critical factor, it is so outweighed by our need for predictability, stability, immediate access, and low delay that network operators will be forced to run their networks very inefficiently. Network operators have already concluded this; they know that their customers care more about predictability than efficiency, and we know from the dynamics of queueing networks, that in order to achieve predictable behavior, network operators must continue to utilize their links very lightly, forfeiting the benefits of statistical multiplexing. As a result, they are paying for the extra complexity of processing every packet in routers, without the benefits of increased efficiency. In other words, the original goal of "efficient usage of expensive and congested links" is no longer valid, and it would provide no benefit to users.

2.3.3 IP is robust

> "The Internet was born during the cold war 30 years ago. The US Department of Defence [decided] to explore the possibility of a communication network that could survive a nuclear attack." — **BBC**

The Internet was designed to withstand a catastrophic event in which a large number of links and routers were destroyed. This goal is in line with users and businesses who rely more and more on network connectivity for their activities and operations, and who want the network to be available at all times. Much has been claimed about the reliability of the current Internet, and it is widely believed to be inherently more robust and capable of withstanding failures of different network elements. Its robustness comes from using soft-state routing information; upon a link or router failure, it can quickly update the routing tables and direct packets around the failed element. In contrast, a circuit-switched network needs to reroute all affected active circuits, which can be a large task for a high-speed link carrying hundreds or thousands of circuits.

The reliability of the current Internet has been studied by Labovitz et al. [107]. They have studied different ISPs over several months, and report a median network availability equivalent to a downtime of 471 min/year. In contrast, Kuhn [102] found that the average downtime in phone networks is less than 5 min/year. As users, we have all experienced network downtime when our link is unavailable or some part of the network is unreachable. On occasions, connectivity is lost for long periods while routers reconfigure their tables and converge to a new topology. Labovitz et al. [106] also observed that the Internet recovers slowly, with a median BGP convergence time of 3 minutes, and frequently taking over 15 minutes. In contrast, SONET/SDH rings, through the use of pre-computed backup paths, are required to recover in less than 50 ms [51], a glitch that is barely noticeable to the user in a network connection or phone conversation.

While it may be argued that the instability and unreliability of the Internet can be attributed to its rapid growth and the ad-hoc and distributed way that it has grown, a more likely explanation is that it is fundamentally more difficult to achieve robustness and stability in packet networks than circuit networks. In particular, since routers/switches need to maintain a distributed routing state, there is always the possibility that the state may become disconnected. In packet networks, inconsistent routing state can generate traffic loops and black holes and disrupt the operation of the network. In addition, as discussed in Section 2.3.2, the likelihood of a network getting into a inconsistent routing state is much higher in IP networks because (a) the routing packets are transmitted in-band, and therefore are more likely to incur congestion due to high load of user traffic; (b) the routing computation in IP networks is very complex; it is, therefore, more likely for the control processor to be overloaded; (c) the probability of misconfiguring a router is high. And misconfiguration of even a single router may cause instability in a large portion of the network. It is surprising that we have continued to use routing protocols that allow one badly behaved router to make the whole network inoperable [105]. Conversely, high availability has always been a government-mandated requirement for the telephone network, and so steps have been taken to ensure that it is an extremely robust infrastructure. In circuit networks, control messages are usually transmitted over a separate channel or network. This has the added advantage of security for network control and management. In addition, the routing in circuit networks is much simpler.

In datagram networks, inconsistent routing state may cause black holes or traffic loops so that the service to existing user traffic is disrupted – i.e., inconsistent routing is *service impacting*. In circuit networks, inconsistent routing state may result in unnecessary rejection of request for new circuits, but none of the established circuits is affected. In summary, currently with IP, not only are failures more common, but also they take longer to be repaired and their impact on users is deeper.

On the face of it, then, it seems that packet-switched IP networks experience more failures and take longer to re-establish connectivity. However, it is not clear that reliability and fault tolerance are a direct consequence of our choice of packet switching or circuit switching. One can attribute much of the growth of the Internet to the ad-hoc and distributed way that it has grown; so it should not be surprising that there are frequent misconfigurations of routers and poorly maintained equipment [114]. Table 2.2 shows that router operations are the most common source of network failures.

The key point here is that there is nothing inherently unreliable about circuit switching, and there is an existence proof that it is both possible and economically viable to build a robust circuit-switched infrastructure, that is able to quickly reconfigure around failures.

Type of failure	Frequency of occurrence	description
Router Operations	36.8 %	Maintenance, power failures, congestion
Link Failure	34.1 %	Fiber cuts, unreachable, interface down
Router Failures	18.9 %	Hardware and software problems, routing problems, malicious attacks
Undefined	10.5%	Miscellaneous and unknown

Table 2.2: Frequency of occurrence of recorded network failures in a regional ISP in a one-year period [107].

There is no evidence yet that we can define and implement the dynamic routing protocols to make the packet-switched Internet as robust. Perhaps the problems with BGP will be fixed over time and the Internet will become more reliable. But it is a mistake to believe that packet switching is inherently more robust. In fact, the opposite may be true.

2.3.4 IP is simpler

> "IP-only networks are much easier and simpler to manage, leading to improved economics." — **Business Communications Review**

It is an oft-stated principle of the Internet that the complexity belongs at the end-points, so as to keep the routers simple and streamlined. While the general abstraction and protocol specification are simple, implementing a high performance router and operating an IP network are extremely challenging tasks.

In terms of router complexity, while the general belief in the academic community is that it takes 10's of instructions to process an IP packet, the reality is that the complexities of a high performance router has as much to do with the forwarding engine as with the routing protocols (BGP, IS-IS, OSPF etc), where all the intelligence of the IP layer resides, as well as the interactions between the routing protocols and forwarding engine. A high performance router is extremely complex, particularly as the line rates increase. One subjective measure of the complexity is the failure rate of the start-ups in this space. Because of the perceived high growth of the market, a large number of well-financed start-ups with very capable talents and strong backing from carriers have attempted to build high performance routers. Almost all have failed or are in the process of failing— putting aside the business/market-related issues, none have succeeded technically and delivered a product-quality core router. The core router market is still dominated by two vendors, and many of the architects of one came from the other. The bottom line is that building a core router is far from simple, mastered by only a very small group of people.

If we are looking for simplicity, then we would do well to look at how circuit-switched transport switches are built. First, the software is simpler. The software running in a typical transport switch is based on about three million lines of source code [154], whereas Cisco's Internet Operating System (IOS) is based on eight million [66], over twice as many. Routers have a reputation for being unreliable, crashing frequently and taking a long time to restart, so much so that router vendors frequently compete on the reliability of their software, pointing out the unreliability of their competitor's software as a marketing tactic. Even a 5ESS service telephone switch from Lucent, with its myriad of features for call establishment and billing, has only about twice the number of lines of code as a core router [179, 67].

The hardware in the forwarding path of a circuit switch is also simpler than that of a router, as shown in Figure 1.1 and Figure 1.2. At the very least, the line card of a router must unframe/frame the packet, process its header, find the longest-matching prefix that matches the destination address, generate ICMP error messages for expired TTLs, process optional headers, and then buffer the packet (a buffer typically holds 250ms of packet data). If multiple service levels are added (for example, differentiated services), then multiple queues must be maintained, as well as an output link scheduling mechanism. In a router that performs access control, packets must be classified to determine whether or not they should be forwarded. Further, in a router that supports virtual private networks, there are different forwarding tables for each customer. A router carrying out all these operations typically performs the equivalent of 500 CPU serial instructions per packet (and we thought that all the complexity was in the end system!).

On the other hand, the linecard of an electronic transport switch typically contains a SONET framer to interface to the external line, a chip to map ingress time slots to egress time slots, and an interface to a switch fabric. Essentially, one can build a transport linecard (Figure 1.2) by starting with a router linecard (Figure 1.1) and then removing most of the functionality.

One measure of this complexity is the number of logic gates implemented in the linecard of a router. An OC192c POS linecard today contains about 30 million gates in ASICs, plus at least one CPU, 300 Mbytes of packet buffers, 2 Mbytes of forwarding table, and 10 Mbytes of other state memory. The trend in routers has been to put more and more functionality on the forwarding path: initially, support for multicast (which is rarely used), and now support for quality of service, access control, security and VPNs.[6] In contrast, the linecard of a typical transport switch contains a quarter of the number of gates, no CPU, no packet buffer, no forwarding table, and an on-chip state memory (included in the gate count).

In terms of power consumption, a high-end router dissipates 75% of the power in the linecards, half of which comes from inter-chip I/O communication. IP linecards require many chips, and thus they consume much power. The use of Ternary Content Addressable Memories (TCAMs) for parallel route lookups further exacerbates this power consumption. In contrast, electronic circuit switches consume less power because they use simpler hardware, allowing more linecards (and thus more capacity) to be placed in a single rack.

It should come as no surprise that the highest capacity commercial transport switches have two to twelve times the capacity of an IP router, and sell for about half to one twelfth the price per gigabit per second, as shown in Table 1.1. So, even if packet switching might be simpler for low data rates, it becomes more complex for high data rates. IP's "simplicity" does not scale.

One might argue that the reason the circuit switches cost less is that they solve a simpler problem. Instead of being aware of individual application flows, they deal with large trunk lines in multiples of 51 Mbit/s. So for the sake of comparison, it is worth considering the cost and complexity of building a core transport switch that could establish a new circuit

[6] Interestingly, these features are added to provide traffic isolation and engineering, features that are intrinsic to circuit switching.

for each (TCP) application flow. Let's assume that each user connects to the network via a 56 Kbit/s modem; this will define the granularity of the switch. While such a small circuit might not be the best way to incorporate circuit switching into the Internet, using such small flow granularity provides an upper bound on the complexity of doing so. A 10 Gbit/s linecard needs to manage at most 200,000 circuits of 56 Kbit/s. The state required to maintain the circuits, and the algorithms needed to quickly establish and remove circuits, would occupy only a fraction of one ASIC. This suggests that the hardware complexity of a circuit switch will always be lower than the complexity of the corresponding router.

It is interesting to explore how optical technology will affect the performance of routers and circuit switches. In recent years, there has been a good deal of discussion about all-optical Internet routers. As was mentioned in Chapter 1, there are two reasons why this is not feasible. First, a router is a packet switch and so inherently requires large buffers to hold packets during times of congestion, and currently no economically feasible ways exist to buffer large numbers of packets optically. The buffers need to be large because TCP's congestion control algorithms currently require at least one bandwidth-delay product of buffering to perform well. For a 40 Gbit/s link and a round-trip time of 250 ms, this corresponds to 1.3 GBytes of storage, which is a large amount of electronic buffering and (currently) an unthinkable amount of optical buffering. The second reason that all-optical routers do not make sense is that an Internet router must perform an address lookup for each arriving packet. Neither the size of the routing table, nor the nature of the lookup, lends itself to implementation using optics. For example, a router at the core of the Internet today must hold over 100,000 entries, and must search the table to find the longest matching prefix — a non-trivial operation. There are currently no known ways to do this optically.

Optical switching technology is much better suited to circuit switches. Devices such as tunable lasers, MEMS switches, fiber amplifiers and DWDM multiplexers provide the technology to build extremely high capacity, low power circuit switches that are well beyond the capacities possible in electronic routers [15].

In summary, packet switches and IP linecards have to perform more operations on the incoming data. This requires more chips, both for logic functions and buffering; in addition, these chips are more complex. In contrast, circuit switches are simpler, which allows them to have higher capacities and to be implemented in optics.

2.3.5 Cost of ownership of IP is small

> "Packet technology is just inherently much less expensive and more flexible than circuit switches." — CTO of Sonus.

IP networks are usually marketed as having a lower cost of ownership than the corresponding circuit-switched network, and so they should displace circuit switching from the parts of the network that it still dominates; however, this has not (yet) happened. For example, Voice over IP (VoIP) promises lower communication costs because of the statistical multiplexing gain of packet switching and the sharing of the physical infrastructure between data and voice traffic. Despite these potential long-term cost savings, less than 6% of all international traffic used VoIP in 2001 [38, 98]. VoIP has become less attractive because fierce competition among phone companies has dramatically driven down the prices of long-distance calls [26]. In addition, the cost savings of a single infrastructure can only be realized in new buildings.

One of the most important factors in determining a network architecture is the total cost of ownership. Given two options with equivalent technical capabilities, the least expensive option is the one that gets deployed in the long term. So, in order to see whether IP will conquer the world of communications, one needs to answer this question: Is there something inherent in packet switching that makes packet-switched networks less expensive to build and operate? Here, the metric to study is the total cost per bit/s of capacity.

As we saw in Section 2.3.1, the market for core routers is much smaller than that of circuit switches. One could argue that the market difference is because routers are far less expensive than circuit switches and that carriers are stuck into supporting expensive legacy circuit-switched equipment; however, IP, SONET/SDH and DWDM reached maturity almost at the same time,[7] so a historical advantage does not seem to be a valid explanation for the market sizes. A more likely explanation is that there are simply more circuit switches than routers in the core because routers are not ready to take over the transport infrastructure, and thus the market size cannot be used as a good indication of the equipment cost.

[7] In April 1995, commercial Internet was born after the decommissioning of the NSFnet. In March 1994, Sprint first announced its deployment of directional SONET rings. The first deployments of WDM were from June 1996.

To analyze the total cost of packet and circuit switching, I will start breaking down the cost structure of an ISP. Table 2.3 shows the capital expenditure (capex), operation expenses (opex) and transport costs (interconnection fees) of an Internet carrier [184]. Similar numbers are found in [119].

Routing/switching equipment (capex)	20%
Network management and staff (opex)	45%
Transport/transmission	35%

Table 2.3: Cost structure for an Internet carrier averaged over ten tier-1 and tier-2 ISPs in the US and Europe [184].

Capital expenditure is the cost to build a network. Because there is little difference in the links and link terminations in routers and circuit switches, the difference in capital expenditure lays in the cost of the boxes. Production and design costs are related to the complexity of the system. Figures 1.1 and 1.2 show how routers need more components, and these are more complex, and thus routers are more expensive to design and produce. It should not be surprising that an OC192c packet-over-SONET (POS) linecard for a router costs \$30-40K, whereas the equivalent SONET TDM linecard costs only \$10-20K. If we consider that linecards are the most expensive part of a full router/switch, it is fair to say that it is more expensive to build a router than a circuit switch of the same capacity.

Anyhow, capital expenditure is the smaller part of the pie, and operating expenses represent the biggest cost factor for an ISP. To grasp the importance of the latter, let me point out to a study by McKinsey and Goldman-Sachs [118] that shows that unless per-bit operating expenses are reduced 25%-30% per year through 2005, no reasonable amount of per-bit capital expenditure reduction will allow carriers to achieve sustainable Return on Invested Capital (ROIC). However, this reduction in operating cost is not easy to achieve, as operating expenses are difficult to quantify, and their reduction may have a direct impact on the service quality.

Certainly there seems no reason to believe that IP networks are simpler to operate and maintain. Indeed, a report by Merrill Lynch [121] shows that the normalized operating expenditure for data networking is typically significantly larger than for voice networks. If we look at the number of network administrators present in most companies, usually there are far more operators for the IP network than for the phone network.[8]

Operating expenses are tied to the reliability, manageability and complexity of the network, and IP does not seem to win in any of these three fronts: First, as argued in Section 2.3.3, IP has not demonstrated to be as reliable as SONET/SDH, and thus requires more attention. Second, Internet management platforms are rudimentary and lack integration and interoperability, and tools for capacity planning, traffic engineering and monitoring are almost non-existent in IP [184, 118]. Finally, as mentioned in Chapter 1 and Section 2.3.4 routers do not scale as well as circuit switches in terms of switching capacity. Consequently, one needs more routers than circuit switches to carry the same traffic. This creates a more complex network that is more expensive to build, harder to control and with more network elements demanding attention from operators.

However, there is one area in which IP can potentially reduce costs. IP networks require less network capacity to carry the same information (especially when traffic is bursty) because of the statistical multiplexing gain of packet switching. However, as we saw in Section 2.3.2, carriers do not take advantage of this characteristic of IP, and they prefer to operate their networks at very low utilization, as to ensure the reliability of their network.

To summarize, packet-switched networks seem to be more expensive to build and operate than circuit-switched networks. While some of the causes for the high costs of IP may be addressed in the future (better router software and software tools), others will remain (more complex boxes, less scalable routers). Nevertheless, IP is more flexible than circuit switching, and so there is a tradeoff between cost and flexibility. It is up to the carriers to decide when the need for flexibility justifies the extra cost of packet switching.

2.3.6 Support of telephony and other real-time applications over IP networks

> "All critical elements now exist for implementing a QoS-enabled IP network."
> **— IEEE Communications Magazine**

There is a widely-held assumption that IP networks can support telephony and other

[8] Stanford University (with a population of about 15,000 people) employs 80 full-time telephone engineers, 25 full-time IP network engineers, and 350 part-time local IP network administrators.

real-time applications that require minimum guaranteed bandwidth, bounded delay jitter and limited loss. If one looks more closely, one finds that the reasons for such an optimistic assumption are quite diverse. One school holds the view that IP is ready today. There are two reasons for such a belief. First, IP networks are and will continue to be heavily overprovisioned, and the average packet delay in the network will be low enough to satisfy the real-time requirements of these applications. Second, most interesting real-time applications, including telephony, are *soft* real-time in the sense that they can tolerate occasional packet delay/loss and *adapt* to these network variabilities. While today's IP networks are heavily overprovisioned, it is doubtful whether a new solution (far from complete yet) that provides a worse performance can displace the reliable and high quality service provided by today's TDM-based infrastructure (which is already paid-for).

Another school believes that for IP to succeed, it is critical for IP to provide Quality of Service (QoS) with the same guarantees as TDM but with more flexibility. In addition, the belief is that there is no fundamental technical barrier to build a connection-oriented service (Tenet [75] and IntServ [20]) and to provide guaranteed services in the Internet. The technical ingredients for a complete solution include efficient packet classification and scheduling algorithms. Unfortunately, after more than ten years of extensive research and efforts in the standards bodies, the prospect of end-to-end per-flow QoS in the Internet is nowhere in sight. The difficulty seems to be the fact that there is huge culture gap between the connection and datagram design communities. By blaming the failure on "connections", a third school holds the view that a simpler QoS mechanism such as DiffServ is the right way to go. Again, we are several years into the process, and it is not at all clear that the "fuzzy" QoS provided by DiffServ (with no route pinning support and no per flow QoS scheduling) will be good enough for customers who are used to the simple QoS provided by the existing circuit-switched transport networks.

The truth is that many of these QoS mechanisms, such as DiffServ and IntServ, are implemented in most routers deployed in the Internet; however, few service providers enable them and use them. The reasons are that these mechanisms are difficult to understand and configure and that they require an active cooperation among ISPs for them to provide end-to-end QoS.

Finally, no matter what technology we intend to use to carry voice over the Internet, there are few financial incentives to do so. As Mike O'Dell[9] recently said [134]: "[to have a Voice-over-IP (VoIP) service network one has to] create the most expensive data service to run an application for which people are willing to pay less money everyday [...] and for which telephony already provides a better solution with a marginal cost of almost zero." The result is that despite the promised cost reductions of Voice over IP, in 2001 less than 6% of all international voice traffic out of the US used VoIP.

On the other hand, because circuits are peak-allocated, circuit switching provides simple (and somewhat degenerate) QoS, and thus there is no delay jitter. The user (or server) can inform the network of a flow's duration, and specify: its desired rate and blocking probability (or a bound on the time that a flow can be blocked). These measures of service quality are certainly simpler for users to understand and for operators to work with, than those envisaged for packet-switched networks.

2.4 Discussion

Up until this point, I have considered some of the folklore surrounding the packet-switched Internet. The overall goal is to provoke discussion and research on fundamental issues that need to be addressed so that IP can continue to revolutionize the world of communications. As a research community, we need to think beyond the daily challenges of maintaining and optimizing the expanding Internet, and move on to consider the enormous challenges that lie ahead.

It seems that there are two main limitations to the widespread adoption of IP: dependability and the right way for IP to co-exist with circuits. In what follows, I will discuss each in turn.

2.4.1 Dependability of IP networks

High dependability, in the broadest sense, is a must if IP is to become a successful transport technology (to compete or displace circuit-based transport networks), and if the Internet is

[9] Former Senior Vice President of UUNET, responsible for technical strategic direction and architecture of the network.

to become the universal infrastructure for high value applications. For example, voice services are a high-revenue, and very profitable business. Trusting them to today's unreliable, and unpredictable IP networks would be an unnecessary risk, which is why — despite predictions to the contrary — telephone carriers have not done so.

High dependability means several things: robustness and stability, traffic isolation, traffic engineering, fault isolation, manageability, and last but not least, the ability to provide predictable performance in terms of bounded delay and guaranteed bandwidth (QoS). In its current form, the Internet excels in none of these areas. Although it is clearly a challenge to achieve each of these goals, they must all be solved for IP to become dependable enough for use as a transport mechanism.

2.4.2 Interaction of IP and circuits

The current Internet is based on packet-switched routers in the edges, interconnected by a circuit-switched transport network. Given the benefits of circuit switching, it would seem perverse for the packet-switched network to grow to subsume the transport network. It is inconceivable that the network providers would remove the existing, robust, reliable, predictable and largely paid-for transport network, and replace it with a technology that seems more complex, less reliable, more expensive and not yet installed.

What seems more likely is that packet switching will continue to exist at the edge of the network, aggregating and multiplexing traffic from heterogeneous sources for applications that have no delay or quality requirements. In other words, packet-switched IP will continue to provide a simple service abstraction for a variety of applications. However, this does not preclude the existence of highly specialized service networks living alongside IP and using other switching techniques. In fact, it is unlikely that the phone or TV cable service networks will be completely replaced by an IP network any time soon as it would require a huge amount of capital to build a new network.

At the core of the network, one can expect the circuit-switched transport network to remain as a means to interconnect the packet-switched routers and as a means to provide high reliability and performance guarantees. Over time, more and more optical technology will be introduced into the transport network, leading to capacities that (necessarily) electronic routers cannot achieve.

One remaining question is whether or not the circuit-switched network will be controlled by IP. In other words, will the IP network decide dynamically when to create new circuits between routers? For example, a router could monitor the occupancy of its queues or the number of active flows and periodically add or remove circuits to other routers based on current demand [7, 181]. Such a system has the benefit of enabling IP to gain the benefits of fast optical circuit switches in the core, yet maintain the simple service model for heterogeneous sources at the edge.[10]

However, while a complete control by IP of the circuit-switched backbone seems appealing to IP, one needs to remember that the majority of the revenue for the circuit switches will still be from other applications, such as voice. Since the packet-switched network is unlikely to provide the predictability needed for voice traffic, it will continue to operate over its own, separate circuit-switched edge network and to be carried over the shared transport network at the core. In this environment, it is unlikely that the routers will be allowed to control the entire capacity of the transport switches, unless the revenue for the Internet exceeds that of telephony. At the current growth rates, it will take over 15 years for data traffic to surpass telephony as the main source of revenue in telecommunications. In the future, it is more likely that the routers will be allocated a fraction of the circuit-switched transport infrastructure, which they can control and adapt to best serve their needs.

With the dynamic control of circuit-networks (possibly by an IP-based control plane), it is also conceivable that the IP routers at the edge can signal to the transport network to dynamically create new circuits or change the bandwidth of existing circuits.

2.4.3 What if we started with a clean slate?

In the preceding discussion, an outcome was depicted based on historical conditions, in the context of a pre-existing circuit-switched transport network. So if we started again, with the benefit of hindsight, would we build a network with circuit switching at the core, and packet switching at the edge? I believe that we would, and that it would look something like this:

[10] Chapter 4 and Chapter 5 describe two ways of integrating IP and circuit switching in the core.

- **Addressing scheme.** A simple, unique and universal addressing scheme (like IP's) would allow us to communicate with any sort of device or application anywhere in the world. This addressing scheme defines the routing algorithms in the intermediate network nodes, but it is completely independent of the forwarding or switching mechanisms that they use.

- **Switching in the edges of the network.** Packet switching would be used in the edges of the network as well as in those links where bandwidth is scarce (such as some satellite and wireless links, and underwater cables). The reasons for this are threefold. First, packet switching makes a very efficient use of the bandwidth in these cases. Second, as will be emphasized in Chapter 3, it can greatly improve the end-user response time by borrowing all available link bandwidth when other users are not active. Finally, packet switches can be cost effective for lower link rates. The packet-switched network should ideally gather traffic from disparate sources, and multiplex it together in preparation for carriage over a very high capacity, central, circuit-switched core. In this environment, local switching at the edge of the network is an optimization that may or may not be necessary. Without it, the packet-switched network is simply a hierarchy of statistical multiplexers, with little or no forwarding decisions. All traffic can be multiplexed towards the core, then demultiplexed again towards the edge. While less efficient, it provides a simplified environment in which to deploy the delay guarantees needed by telephony. And so it might be feasible to carry the traffic from access voice switches to the core over the statistically multiplexed edge network.

- **Switching in the core of the network.** At the core of the network, there seem a number of compelling reasons to use circuit switching. First, circuit switching has already demonstrated its robustness and its ability to quickly recover from failures. Circuit switching is inherently simpler than packet switching, requiring less work to forward data, and consequently will cost less as a result, will consume less power, and will take up less space. Last, but not least, circuit switching provides an easy way to adopt the huge potential of high capacity optical switches. Without electronics on the forwarding path, one can expect optical switches to provide abundant capacity at low cost.

- **Integration of both switching mechanisms.** Rather than working independently, both these mechanisms would be tightly integrated, in such a way that an action in one provokes an appropriate reaction in the other. For example, packet switching would have to export the QoS and connection-oriented nature of the circuit-switched core to the applications that require it. On the other hand, circuit switching has to respond to the increases in activity of packet switching, by adapting its capacity among core/edge gateways accordingly. Additionally, we will find more hybrid switches that can do both circuit and packet switching, serving as gateways between the two worlds. Chapter 4 and Chapter 5 describe two ways of bridging packet switching and circuit switching. Finally, the idea of using circuit switching to interconnect distant routers can also be extended to using a circuit-switched crossconnect to interconnect the packet-switched linecards of a router.

2.5 Conclusions and summary of contributions

While it is technically pleasing to believe that IP will dominate all forms of communication, our delight in its elegance is making us overlook its shortcomings. IP is an excellent means to exchange data, which explains its success. This chapter has demystified some of the proclaimed advantages of IP, such as the claims that IP is simpler, more robust, more efficient, that it dominates world communications, and that it can support QoS-aware applications. I have reserved the rebuttal of what is probably the most important claim for next chapter; namely, that IP can achieve better response time for the end user.

IP remains ill suited as a means to provide many other types of service, and is too crude to form the transport infrastructure in its own right. To allow the continued success of IP, we must be open-minded to it living alongside, and co-operating with, other techniques (such as circuit switching) and protocols that are optimized to different needs.

The conclusion is that while packet-switched IP will continue to dominate best-effort data services at the edge of the network, the core of the network will use circuit switching as a transport platform for multiple services. Circuit switching allows the construction of networks with very high capacity, scalability, flexibility, self-healing, reliability and auto-adaptation to current network traffic conditions; thus, IP will have a hard time replacing the circuit switching that already exists in the core. We should instead start thinking of how to integrate the two technologies: circuit switching in the core and packet switching in the edges.