Internet growth: Myth and reality, use and abuse

Andrew Odlyzko
AT&T Labs - Research
amo@research.att.com
http://www.research.att.com/∼amo

Abstract

Actual Internet traffic growth rates of 100 percent per year are considerably less than the much-ballyhooed doubling every 3 or 4 months. But even these observed rates are still unprecedented and should be provoking new ways of planning.

1. Introduction

Almost all references to Internet growth claim astronomical rates of increase; the usual phrase is that "Internet traffic is doubling every three months." Even serious observers echo such claims. For example, former Federal Communications Commission Chairman Reed Hundt recently wrote, "In 1999, data traffic was doubling every 90 days. . . ." [Hundt]. Financial markets also appear to accept such estimates. During the meeting with financial analysts to discuss the report for the 3rd quarter of 2000, James Crowe, head of Level 3 Communications, cited the "doubling of Internet traffic every three or fours months" as proof that the supply of fiber in the United States could not possibly meet demand within the next few years. If this were true, the resulting imbalance in supply and demand would produce very pleasing revenue and profit prospects for carriers such as Level 3.

Amazingly enough for a claim that is so dramatic and quoted so widely, there have been no hard data to substantiate it. Indeed, careful scrutiny of existing evidence on traffic by a number of experts suggests that the truth is considerably more modest. Internet backbone traffic in the United States has been about doubling annually for the last four years, and currently appears to continue growing at about that rate. (Doubling is used here in a loose sense to cover growth rates between 70 percent and 150 percent per year.) Although this is extremely rapid growth, much faster than in any other communications service, it is not anywhere close to the 700 percent to 1,500 percent annual growth rates that a doubling of traffic each three or four months would imply.

Undoubtedly, annual growth rates of 100 percent per year require new ways of network budgeting and engineering. But the exaggerated myth of Internet growth rates on the order of 700 to 1,500 percent is leading to poor planning. Given the widespread belief in astronomical growth, what is the reality? And, finally, what does this mean?

2. The myth and the reality

An amazing phenomenon is that there has always been plenty of evidence of more modest growth rates, but the world at large continued to uncritically accept the “doubling every three months” story. As often happens, there is a grain of truth behind the claims of Internet traffic doubling every three or four months. Such growth rates did prevail for a short period during 1995 and 1996. This brief period of extraordinary growth seems to have colored popular perceptions.

The belief that Internet traffic could continue “doubling every three months” all this time shows an astonishing degree of innumeracy, the lack of simple quantitative reasoning. At this rate, traffic would be increasing by a factor of 16 per year. Hence, from the end of 1994 to the end of 2000, it would have grown by a factor of almost 17 million.

But what, indeed, happened? Until the end of 1994, the Internet backbone was funded by the National Science Foundation, and was well instrumented. Hence we know that it carried about 15 TB (terabytes) of traffic each month. Had that traffic grown by a factor of 16 million in the intervening 6 years, we would now have about 240 exa-bytes of traffic on Internet backbones each month. If we generously assume that there are 500 million Internet users in the world today, that volume of traffic would translate into about 1.5 million bits per second of backbone traffic for each user around the clock! This is enough for reasonably high quality video (if one uses appropriate compression). All this while most Internet users have access only to modems that transmit at best at 28 thousand bits per second. Moreover, those modems are in use typically for less than an hour per day, and on average transmit about 5 thousand bits per second while they are connected to the Internet. Even the vast majority of enterprises as well as some universities have links to the Internet that run at T1 speeds (i.e., maximal rates of 1.5 million bits per second). The bottom line is that current user behavior falls well short of projections based on late 1994 traffic and the extraordinarily high growth that occurred in 1995 and 1996.

Some claims have been made that the “doubling every three months” applies to network capacity, if not to network traffic. Yet a similar argument to that above shows that such a growth rate could not have applied for long. The Internet backbone at the end of 1994 was somewhat larger than two T3 (45 megabit per second) links crossing the continent. Growth by a factor of 17 million would have produced a network with over 600,000 OC48 (2.5 gigabit per second) links from coast to coast, far more than all the existing fiber strands could carry.

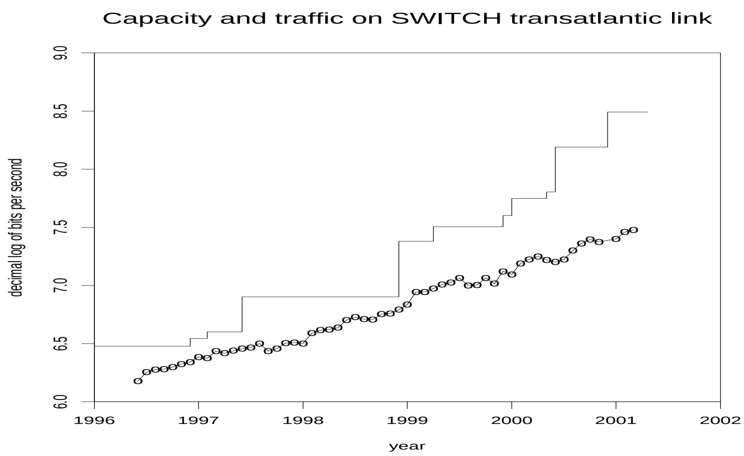
The “doubling every three months” over the last six years story is simply not consistent with reality. Such growth rates can appear for only brief periods. In fact, before 1995-1996, Internet backbone traffic in the U.S. in the early 1990s was doubling each year, as the publicly available statistics for the backbone of that period show. Since early 1997, the growth rate in traffic has reverted to 100 percent a year. This was apparently first pointed out by my colleague Kerry Coffman and myself [CoffmanO1] (which is based on data through the end of 1997). A recent update, based on information through mid-2000, is available in a second paper we wrote [CoffmanO2]. Admittedly, since most ISPs do not release their traffic statistics, the evidence for this estimate is often circumstantial. The details are presented in those two papers.

New studies have provided additional evidence of relatively slower growth rates than the mythological doubling every three or four months. RHK, Inc., a telecommunications consulting firm, now estimates that traffic is growing about 200 percent per year, which while greater than 100 percent per year is still far short of the vaunted 700 to 1,500 percent. A recent joint study by J. P. Morgan and McKinsey, based partially on the work we did [CoffmanO1, CoffmanO2], concludes that the growth rate is close to 100 percent per year. There were even earlier reports from other consulting firms, such as Probe Research or Ovum, which also estimated growth rates closer to a doubling each year.

The detailed description of the data and arguments that show that Internet backbone traffic is about doubling each year are presented in the previously cited studies done with Coffman [CoffmanO1, CoffmanO2]. It appears that not only is overall traffic growing at about 100 percent per year, but that even traffic at individual institutions that are moderately large and have a well-developed electronic communications infrastructure is also growing at about that rate in a novel form of “Moore’s Law” [CoffmanO2].

Here I will just illustrate this phenomenon, and several other relevant ones, with the example of SWITCH, the Swiss academic and research network. (For a more detailed discussion of SWITCH, as well as of numerous other networks, see the previously cited work with Coffman [CoffmanO1, CoffmanO2].) In addition to extensive connections within Switzerland, SWITCH has several international links, with current traffic statistics available at their Website [SWITCH]. The link to the U.S. has traditionally been the most expensive, and at least on some occasions cost more than the rest of the network. Thus, it is natural that this link tends to be the most congested in the SWITCH network.

Although the SWITCH transatlantic link has been expensive, its capacity and the traffic it carries have been approximately doubling each year for the past decade. Users have apparently found it important enough to pressure the network administrators to continue increasing the bandwidth of the connection. The graph below shows the capacity of the link since the beginning of 1996, and traffic from the U.S. to Switzerland on that link from May 1996 to February 2001. (The detailed data were supplied by SWITCH, through courtesy of Willi Huber. The traffic in the opposite direction, from Switzerland to the U.S., has traditionally been considerably lower, by factors of two or three, and is not shown.)



Capacity of link between the Swiss SWITCH network and the U.S., and traffic on it towards Switzerland.

The graph shows various minor perturbations in traffic volumes, some associated with university vacations, others with random factors. Overall, though, the growth rate of traffic over this period of almost five years has been remarkably steady at 88 percent per year. This growth rate was not affected much even by large changes in the capacity of the link. At the end of 1998, this link was extremely congested (which resulted in very poor service, with high packet loss rates, etc.). When capacity was tripled (from 8 Mb/s in November 1998 to 24 Mb/s in December), there was a brief spurt in traffic growth, but it was not large. More recently, in the summer of 2000, the capacity of the link increased from 64 Mb/s to 155 Mb/s. At the end of November 2000, it was further doubled to 310 Mb/s. Yet these capacity increases did not result in any major new traffic growth beyond the customary doubling each year.

The new form of “Moore’s Law” we have proposed [CoffmanO2] predicts that the natural growth rate of data traffic is about 100 percent per year. This growth rate results from a complex interaction of technology, economics and sociology. It has even less of a firm technical foundation than traditional "Moore's laws" that appear to hold for semiconductors, magnetic memories, photonic transmission and other areas. Yet it appears to apply to a wide variety of institutions over extended periods of time, as the data from SWITCH illustrate.

One final remark about the graph of SWITCH traffic: It shows a general decrease in utilization. While traffic grew at an annual rate of 88 percent between May 1996 and February 2001, capacity grew at the rate of 165 percent in the same period. This is, again, a general phenomenon: As prices of transmission capacity decrease (and they have been decreasing rapidly recently on transatlantic links) and data networks become more important, utilizations also decrease, producing better service. Thus, supply (i.e., capacity) has been out-pacing demand (as measured by data traffic), resulting in improved service.

3. Abuses of Internet growth myth

The myth of Internet traffic that doubles every three or four months is dangerous. It leads to bad decisions. It surely helped inflate the current bubble in optical networking stocks. After all, if demand is outpacing supply of transport capacity, then money making opportunities are virtually limitless.

This Internet growth myth may also have helped inflate the general tech stock bubble. By creating an impression of extremely rapid development of the Internet, it may have propelled the rush to invest into every dot-com that came along. After all, if the Internet grows by a factor of 16 each year, then the first mover may have an unbeatable advantage, and so no price is too high to be the first to stake a claim in another "California gold rush."

Yet the reality is quite different, since diffusion of new technologies on the Internet is not notably faster than in the old world. With the notable exception of the Web and browsers, it still takes about a decade for new products and services to achieve high penetration in society [Odlyzko1]. This is even true of basic Internet technologies, such as IPv6.

The myth of astronomical growth rates has also been damaging on a technical level. It has retarded development of an understanding of what users want from the Internet, and what they are likely to get. If traffic demand were really to outpace supply, then it would be necessary to ration access through pricing and prioritization schemes (commonly referred to as QoS, or "Quality of Service"). Traffic might then indeed fill the links, producing (at least at some high level aggregation) smooth flows, which would again fit in well with QoS, or even with the ATM (asynchronous transfer mode) packet switching technology. Under excess demand conditions, in short, engineering efficiencies take priority.

However, in an environment where supply and demand are in approximate balance, user preferences are likely to dominate, as they have in utilization of computers and local area networks. What people want is quick response times (also known as "low transaction latency") [Odlyzko2]. That is why they buy 800 MHz Pentium III computers and leave them idle over 99 percent of the time. That is also why they install ever-faster Ethernet networks, and again utilize them at ever-decreasing fractions of their capacity. Evidence from long distance networks, especially private line ones that corporations lease from carriers, shows that this same dynamic operates in long distance transport. As prices decrease, users opt for high bandwidth links with low utilizations in order to get low transaction latency and faster responses. In that environment, neither QoS nor ATM is appropriate.

This argument does not say that we will have "unlimited bandwidth," any more than that we will have "unlimited computing power." Links will tend to fill up if they are not upgraded. There will be a continuous process of network capacity growth stimulating traffic growth, through increased usage and new applications, and increased traffic growth leading to pressure for upgrades of the network. In a similar process, computer hardware and software have been growing in parallel. What the argument predicts, though, is that in this dynamic process, user preferences for low transaction latency will dominate.

4. Positive effects of Internet growth myth

Myths are not always harmful. Technological forecasting has a poor track record, and many successes have been built on false assumptions. Even the development of the Internet (or, more precisely, its predecessor, the Arpanet) was motivated by the expectations that packet switching would provide (1) resilience in the presence of faults and (2) more efficient use of transport capacity through statistical multiplexing. However, neither expectation has been met so far. The traditional phone network continues to be more reliable than the Internet. Further, utilization rates of data networks, including the Internet, are considerably lower than of the voice network, producing surprisingly high costs [Odlyzko2].

The importance of e-mail is yet another example of poor prediction that caused no lasting damage. E-mail, which was not chosen as a necessary feature of the Arpanet [Abbate], proved to be the "killer app" for the first two decades of the Arpanet/Internet. By some counts, it is still the most important service of the Internet.

The reason the Internet has seemingly trumped the traditional telephone network is that it has provided an unparalleled platform for innovation, with its open standards allowing rapid development of new services. The ability to achieve low response times through leasing of high capacity links that are run at low utilizations was a key factor in this revolution.

The myth of Internet traffic doubling every three months probably has had some positive effects. It has made it clear to many that we are in a new environment, that “this is not your father’s network.” However, even a doubling of traffic each year leads to the same conclusion, although it does not sound anywhere near as dramatic.

From a comparative perspective, traditional voice networks have been growing at about 10 percent per year. Cellular services have been growing faster, on the order of 30-40 percent per year recently. Data traffic apparently used to grow at about 20-30 percent per year in the 1980s, and 30-40 percent in the 1990s [CoffmanO1, CoffmanO2, Galbi]. That was while those data networks were dominated by private line networks, run largely by corporations to provide internal communications. Now that the public Internet is becoming the dominant network, its 100 percent annual growth rate is beginning to set the pace of the entire telecommunications infrastructure.

In an environment of change that is this rapid, old methods of planning no longer apply. For example, capacity on undersea cables continues to be sold in the traditional form of 25-year contracts. Yet if you lease an OC3 link today, within three years of annual traffic doubling you will need almost 10 times as much capacity. Under those conditions, keeping the OC3 link in addition to another one that is about 10 times as large will probably be more bother than help. Thus a 25-year contract is really good only for about 3 years, and should be amortized accordingly.

Further, in an environment of traffic doubling each year, careful network capacity planning is impossible. The premium is on simplicity. Simplicity makes the upgrades less expensive, and, in particular, minimizes the costly labor that is required. That appears to be a major factor leading the industry to simplify the network hierarchy (moving away from the old IP over ATM over SONET over WDM systems towards IP over WDM) and also to move towards Gigabit Ethernet and later 10-Gigabit Ethernet. In such an environment, overprovisioning a link can often be much more economical than careful engineering or QoS, since it corresponds to what David P. Reed calls “a constant phase shift in an exponential capacity curve,” putting in new capacity just a few months before it would be needed anyway. (See the paper [FishburnO] for economic models of such processes.)

5. Conclusions

The story of Internet traffic doubling every three months is a fable that seems to have arisen from a rather brief spurt of traffic growth in 1995-1996. The astronomical growth rates of the popular fable can be dangerously misleading in leading to poor choices in technology and unnecessary costs. When seen over the decade of the 1990s, traffic appears to be doubling about once each year. That is already unprecedented growth, growth that requires new modes of capacity and feature planning from carriers and equipment suppliers.

References

[Abbate] J. Abbate, *Inventing the Internet,* MIT Press, 1999.

[CoffmanO1] K. G. Coffman and A. M. Odlyzko, The size and growth rate of the Internet. *First Monday,* Oct. 1998, ⟨http://firstmonday.org/⟩. Also available at ⟨http://www.research.att.com/∼amo/doc/complete.html⟩.

[CoffmanO2] K. G. Coffman and A. M. Odlyzko, Internet growth: Is there a "Moore's Law" for data traffic?, *Handbook of Massive Data Sets,* J. Abello, P. M. Pardalos, and M. G. C. Resende, eds., Kluwer, 2001, to appear. Available at ⟨http://www.research.att.com/∼amo/doc/complete.html⟩.

[FishburnO] P. C. Fishburn and A. M. Odlyzko, Dynamic behavior of differential pricing and Quality of Service options for the Internet, *Proc. First Intern. Conf. on Information and Computation Economies (ICE-98)*, ACM Press, 1998, pp. 128-139. Extended version to appear in *Decision Support Systems* (2000). Available at ⟨http://www.research.att.com/∼amo/doc/complete.html⟩.

[Galbi] D. Galbi, Bandwidth use and pricing trends in the U.S., *Telecommunications Policy*, vol. 24, no. 11 (Dec. 2000). Available at ⟨http://www.galbithink.org⟩.

[Hundt] R. Hundt, *You Say You Want a Revolution*. Yale Univ. Press, 2000.

[Odlyzko1] A. M. Odlyzko, The slow evolution of electronic publishing, pp. 4–18 in *Electronic Publishing - New Models and Opportunities,* A. J. Meadows and F. Rowland, eds., ICCC Press, 1997. Available at ⟨http://www.research.att.com/∼amo/doc/complete.html⟩.

[Odlyzko2] A. M. Odlyzko, The current state and likely evolution of the Internet, pp. 1869–1875 in *Proc. Globecom'99*, IEEE, 1999. Available at ⟨http://www.research.att.com/∼amo/doc/complete.html⟩.

[SWITCH] SWITCH traffic statistics, available at ⟨www.switch.ch/lan/stat/⟩.