PLURIS®

Avoiding Future Internet Backbone Bottlenecks

Through the Development of Inherently Scalable Terabit Network Routing Systems

Table of Contents

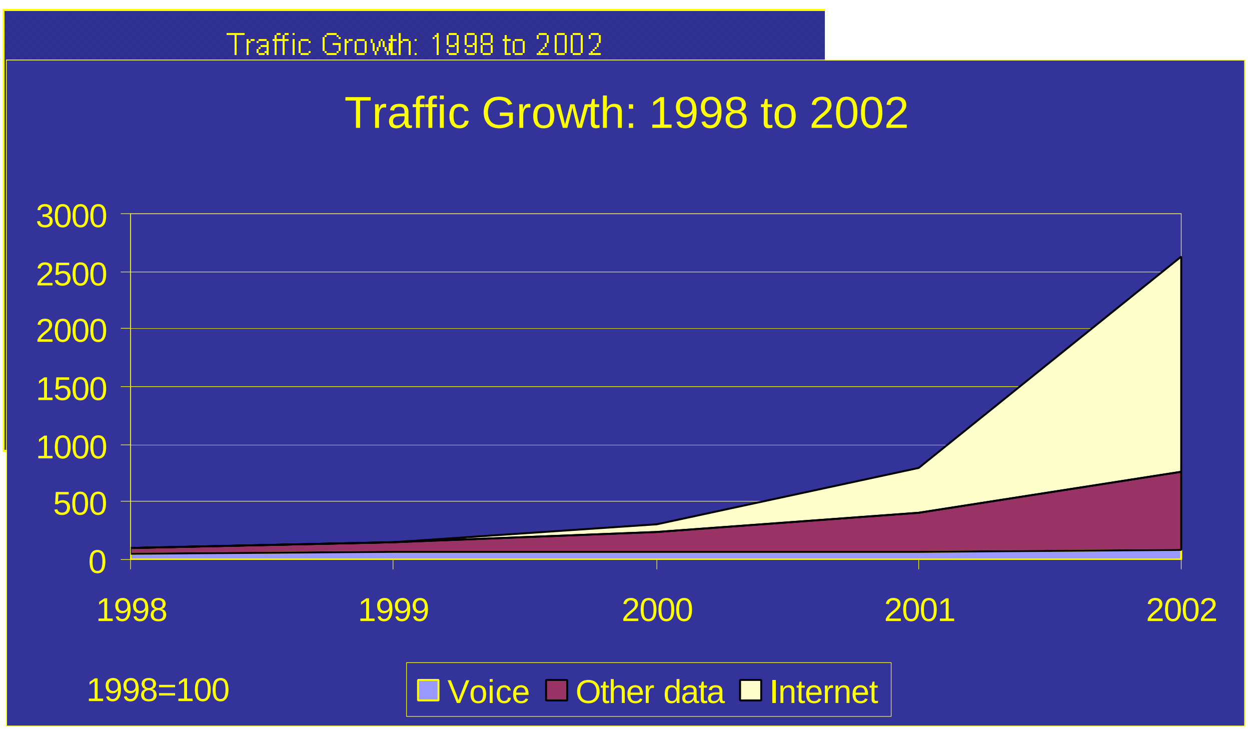
Overview

A number of market drivers and critical technology challenges will be central to the successful re-engineering of the Internet's core infrastructure to meet rapidly expanding demands for more capacity and throughput. Several architectural tenets have emerged for designing new fault-tolerant, highly scalable systems to meet tomorrow's multi-terabit Internet traffic requirements.

Escalating Demands on Backbone Networks

During the past two decades, the basic communications infrastructure throughout the world has experienced both unprecedented growth and sweeping change. This development has engendered a dramatic shift in the roles of many traditional communications service providers and given rise to whole new categories of service.

Throughout the Internet's existence, its traffic load has steadily increased at an average monthly rate of 10%, and it's currently projected to double every three to six months (see Figure 1).



Source: Ryan Hankin Kent Telecom Industry Analysts

Figure 1: Traffic growth on the Internet

Much of this Internet data explosion is now being fueled by the combined effects of millions of new users, far richer content (e.g., multimedia), and the migration of corporate traffic onto the Internet. Overall data traffic is predicted to continue to increase rapidly throughout the next few years, while traditional voice-only traffic is predicted to grow at a much more modest rate.

By the year 2001, the volume of Internet data traffic is expected to overtake that of voice traffic on the existing Public Switched Telephone Network. Projected to continue unabated, Internet traffic will reach daily volumes of 1000 terabits by the end of 2005 (see Figure 2), a load more than 100 times the daily traffic on the telephone network.

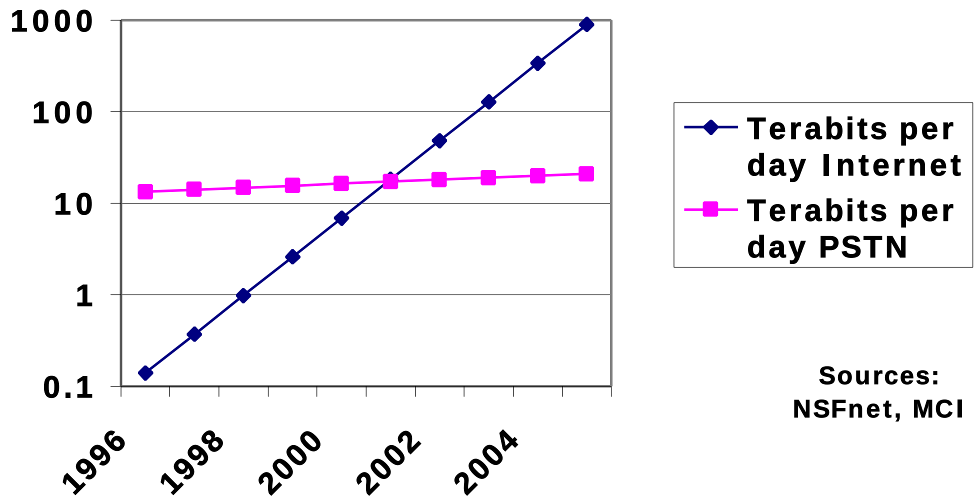


Figure 2: Total Internet data traffic

This fundamental shift toward data traffic, combined with the overwhelming rise in Internet IP-based communications, has led many traditional "telephone" companies to embrace the new opportunities and to re-invent themselves as IP-oriented datacom service companies. In addition, the Internet's popularity among both consumers and businesses has led to the evolution of a multi-tiered network of Internet Service Providers (ISPs), including Tier 3 retail providers, Tier 2 regional ISPs, and Tier 1 core ISPs, which provide high-bandwidth backbone connections.

The Complexity of Scaling Up Existing Internet Protocols

A core issue on the ever-shifting communications scene over the past decade has been the debate over the most effective protocols for carrying data traffic, both locally and across wide-area backbones. A three-way competition has emerged among the traditional isochronous circuit-switched international telephone system, the connectionless dynamically routed Internet Protocol (IP), and TDM-based architectures such as ATM, frame relay, and others.

Although IP has now clearly become the de facto standard for carrying data, fueled by the widespread growth of the Internet, the protocol has not been easy to scale up for use across high-speed backbones. As an alternative, many carriers have deployed ATM technology in the core of their networks to carry IP traffic within fixed-length TDM cells that can be transported in a much faster circuit-switched mode. Forwarding ATM cells is much simpler than routing IP packets because ATM uses a fixed size (53 bytes) and cells do not have to be independently routed. However, sending IP over ATM has inherent drawbacks that limit scalability when the backbone network becomes larger and more complex, as the Internet has grown over the past half-decade.

The extra overhead involved with encapsulating IP packets into ATM cells has become a major inhibitor to large-scale network efficiency, especially considering that many larger IP packets have to be segmented into multiple ATM cells and then reassembled at the end of the ATM switched circuit. In addition, the complexity of Layer 3 ATM control protocols required for maintaining point-to-point virtual switched connections in today's dynamic Internet environment has become unwieldy at best. While gigabit Ethernet has provided a compelling alternative to ATM for scaling enterprise business networks, it cannot be deployed in the wide area.

The major problem in scaling the Internet backbone to higher levels lies in the Layer 3 software complexity required to maintain routing protocols. Backbone routers must constantly exchange IP accessibility information through routing updates to build and maintain accurate routing tables. In addition, exterior gateways to the Internet are multiplying daily, requiring a complex mesh of IP subnets controlled through Layer 3 protocols, such as the Border Gateway Protocol (BGP). BGP is used to build TCP sessions between neighboring routers and therefore must exchange IP prefixes for subnets throughout the Internet, which currently number more than 50,000 and are growing rapidly.

The challenge of scaling up inter-router and ATM circuit-control protocols is becoming especially problematic as the global backbone topology of the Internet gets more complex. The Internet is composed of many parallel backbone networks interconnected via IntereXchange Points (IXPs), also called Network Access Points (NAPs). These backbones consist of many different Points of Presence (POPs) serving particular regions. The high-speed connections between major subsections of the Internet, such as POPs and IXPs, represent the concentration points for transactions from millions of users and therefore constitute the most critical potential bottlenecks as overall Internet traffic increases.

As shown in Figure 3, the myriad combination of physical internal and external connections plus the complexity of Layer 3 software protocols required to implement

Internet POPs present an extremely intricate challenge. Current routers are interconnected via line cards, an approach that not only is expensive but also limits the overall transit throughput of the POP to the speed of an individual router port. No matter how big the individual router backplanes become, the hundreds and thousands of gigabits of traffic transiting through the POP must funnel through the thin straws of OC-12 and OC-48 links between the separate routers and ATM switches. Because today's POPs must be pieced together out of such disparate pieces, the results are generally sub-optimal throughput, degraded Quality of Service, reduced fault-tolerance, and a complete lack of scalability.

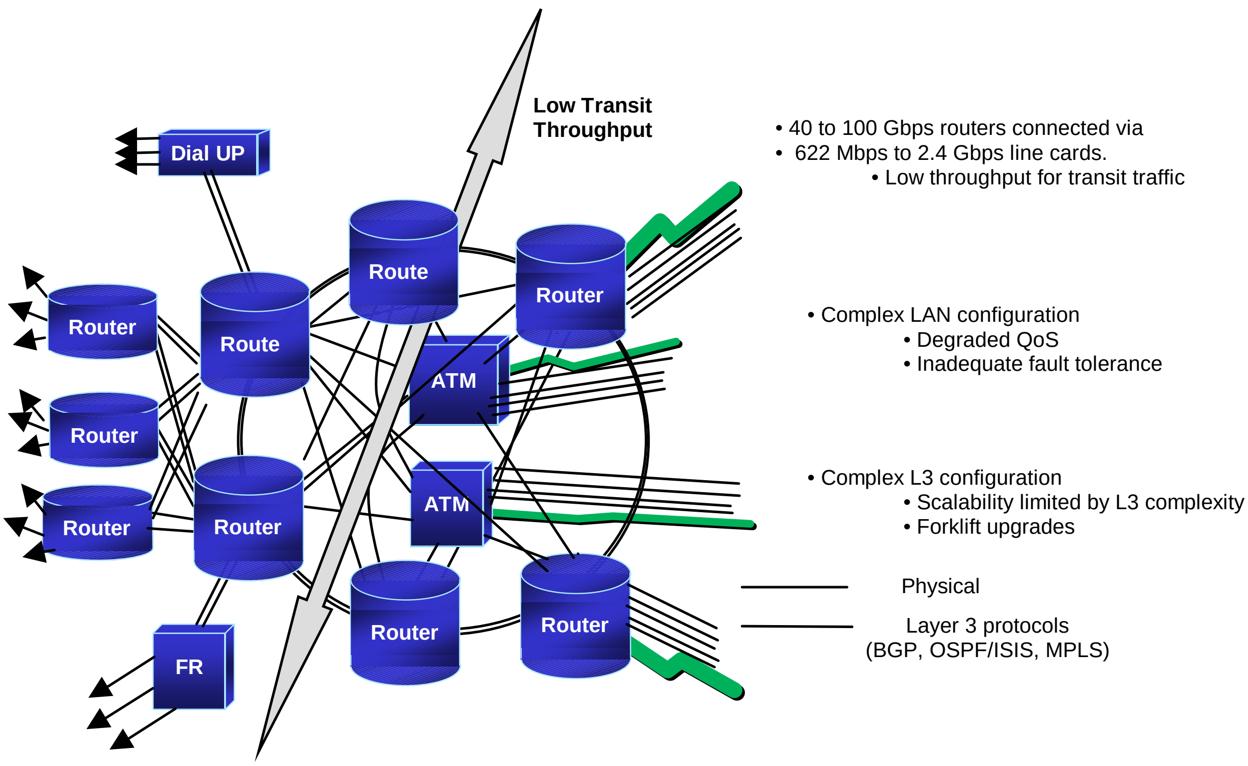


Figure 3: Example of today's non-scalable POP designs

Routing IP Over Optical Links with Wave Division Multiplexing

As backbone architectures have moved toward wider deployment of fiber-optic connections between POPs, IXPs, and telecom COs, new opportunities have also opened up for simplifying the backbone routing of native IP packets. Instead of carrying IP long distances over a series of ATM-based DS-3 transmissions, these new configurations are terminating the DS-3 links in the telecom CO and then carrying the IP data packets directly over long-haul optical links, such as OC-48 and OC-192. This evolution to next-generation transmission technology using high-speed switched IP has the potential to eliminate the overhead associated with encapsulating IP in a Layer 2 TDM carrier , and to reduce the complexity of Layer 3 software structures required to manage the network.

One of the key breakthroughs in fiber-optic technology that has opened the door for routing IP directly over the backbone is the advent of Wavelength Division Multiplexing (WDM). WDM has made it practical to transmit multiple data streams on individual wavelengths within the same fiber link (see Figure 4). With WDM, the carrying capacity of a single fiber can now be boosted to as much as 400 Gbps, with the future addition of more channels potentially able to drive raw transmission rates into the terabit range.

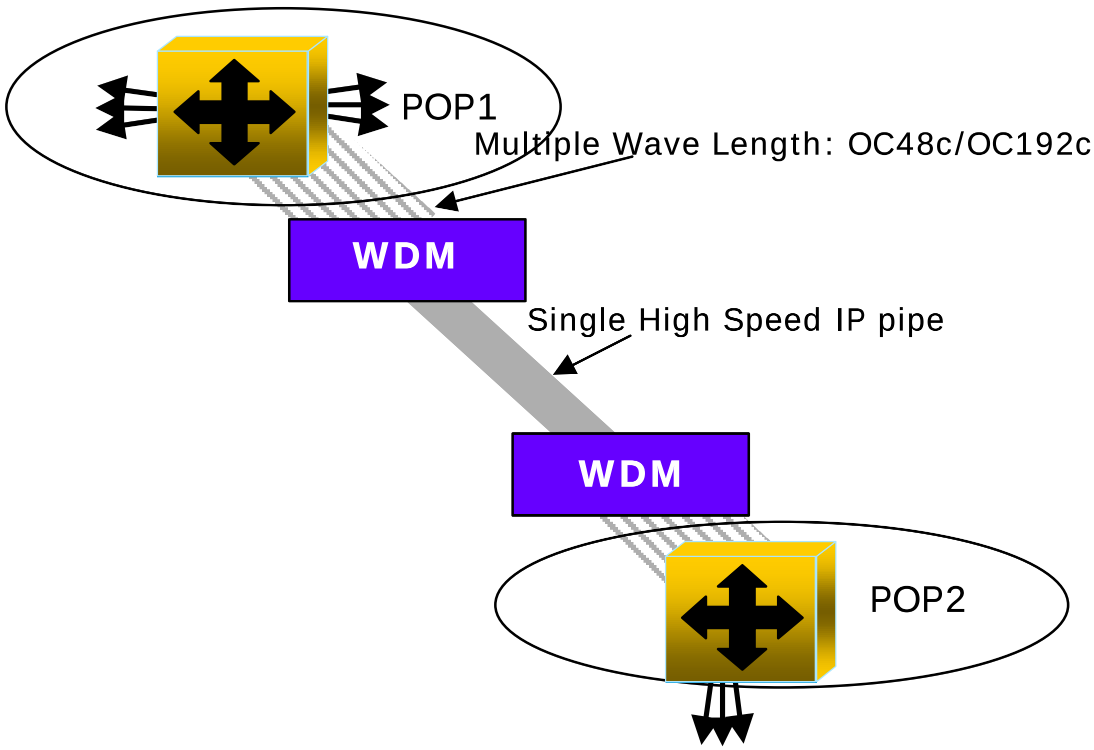


Figure 4: WDM-based high-speed backbone pipes

IP over WDM can provide significant benefits by eliminating the overhead associated with re-packing IP into ATM and reducing the complexity of maintaining Layer 3 routing protocols. In addition, the ability to carry more channels on a single fiber link has the potential to reduce the cost of SONET equipment, reducing the number of ADMs (Add Drop Multiplexers) needed, for example.

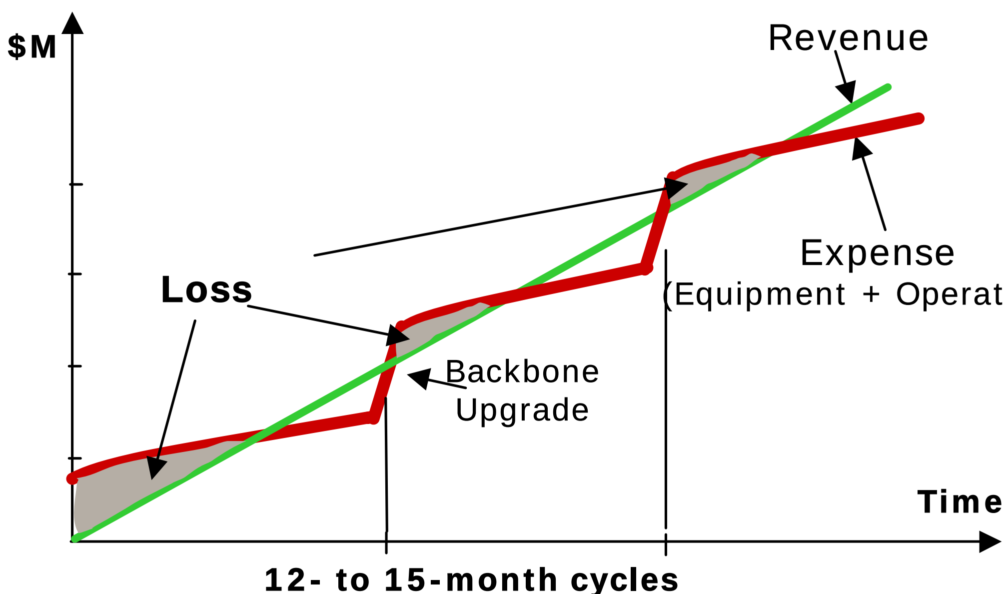
Pure WDM networks, however, currently lack the network management and restoration capabilities that exist in SONET infrastructures. Next-generation WDM infrastructures will need to deliver equivalent functionality by providing more intelligent interaction between the data layer and the optical layer. In addition, direct IP routing across the backbone using WDM will require new highly scalable terabit routing systems that can seamlessly handle both today's and tomorrow's raw IP traffic levels. From the perspective of backbone service providers and Tier 1 ISPs, this need for terabit-level, highly scalable, carrier-class IP routing systems is perhaps the most critical challenge to staying competitive in tomorrow's Internet marketplace.

The Business Imperatives of Infrastructure Planning

Companies primarily in the business of providing high-bandwidth backbone services include Tier 1 ISPs, local central office (CO) access providers, and long-haul telecom/datacom networking companies. Most leading service providers have already committed multi-billions of dollars toward the construction of fiber backbones and high-capacity network infrastructures. Similarly, Tier 1 ISPs, such as UUNet, have invested heavily in creating POPs and SuperPOPs designed to reliably aggregate traffic from millions of regional Internet users, while delivering high-speed access to and from the Internet backbone. In the central office model, much of the current investment focus is on converting existing switching systems from DS-3 interfaces to OC-48 and OC-192 connections that will carry local IP traffic directly onto the long-haul fiber backbone.

The basic value proposition under which these companies must constantly compete consists of bandwidth, bandwidth, and more bandwidth. The only way to keep up with the market's escalating appetite for bandwidth is to ensure that every facet of network investment combines both today's fastest performance and the built-in scalability to meet tomorrow's growth demands. To reliably meet customer expectations under conditions of exponentially increasing market growth, all of these backbone communications providers

must be able to deploy scalable IP switching infrastructures that can maximize their overall investments in the long run, without the requirement for periodic "forklift" upgrades. Figure 5 shows that without a scalable solution, the 12-15 month upgrade cycle will severely erode service provider profit opportunities.



PLURIS

In addition to containing their capital investments in new hardware and software, service providers must grapple with the significantly larger operational cost impacts of continuous network upgrades and unnecessary network complexity. "Touch-administration" staff expenses are the primary cost component in establishing, maintaining, and administering modern sophisticated networks. Simply increasing the number of entities in the network will steadily drain providers' budgets as the increased network complexity will demand commensurate increases in staff hours dedicated to network management.

Therefore, as network backbones grow toward terabit levels, service providers must have new solutions that offer scalability and long-term capacity growth, without requiring exponential increases in operational costs. To the extent that terabit-level networking solutions can provide integrated capabilities in unified, fault-tolerant, scalable systems, service providers will be better able to contain their overhead and operational costs and will be able to maximize profits from their rapidly expanding markets.

The Inherent Limitations of Either "More Boxes" or "Bigger Boxes"

Figure 5: Service provider profit and loss model (showing how "forklift" backbone upgrades create periodic losses)

The cr
will re
generally have consisted of either adding more boxes or trying to build bigger single-backplane boxes. The traditional operational approach of simply adding more limited-scale conventional routers carries with it an exponential increase in overall network complexity and Layer 3 software overhead that eventually bogs down the network in its own control logic. To reach the required capacity at major Internet concentration points, such as SuperPOPs and IXPs, carriers have had to build complex networks within networks, while experiencing a downward curve of ever-diminishing efficiency from each incremental addition.

The maximum speed of conventional routing systems is inherently limited by the speed of the semiconductor technologies used to implement them and the aggregate bandwidth of their internal backplane architectures. From a semiconductor standpoint, even the continued march toward finer process geometries and the advent of high-speed processes, such as Galium-Arsenide (GaAs) and Silicon-Germanium (SiGe), are keeping chip-level speeds up to the growth pace predicted by Moore's Law. But even if semiconductors do continue to double in speed every 18 months, that means chip-level performance is falling further behind every day as the Internet continues to multiply at a sustained rate 3 to 6 times faster than semiconductors.

Backplane bandwidth constraints have traditionally posed the major barrier to simply creating faster versions of conventional router architectures. Initial router designs were based on single-bus architectures, which provided the most straightforward and robust implementations for limited-scale systems. Although the evolution of dual-bus and crossbar structures have done much to improve backplane throughput, the critical issue is now shifting away from the raw backplane speed of single-box systems and toward the need for scaling to handle many more ports than any single box can support.

The challenge of effectively terminating thousands of WDM channels from hundreds of OC-48 and OC-192 connections goes well beyond the capacity of any single monolithic system. For instance, as the 200,000+ voice lines between core central office installations transition instead toward thousands of DSL lines, high-speed cable modems, and wireless connections, the new infrastructure will require hundreds of OC-48 and OC-192 optical ports to carry the traffic over WDM channels (see Figure 6).

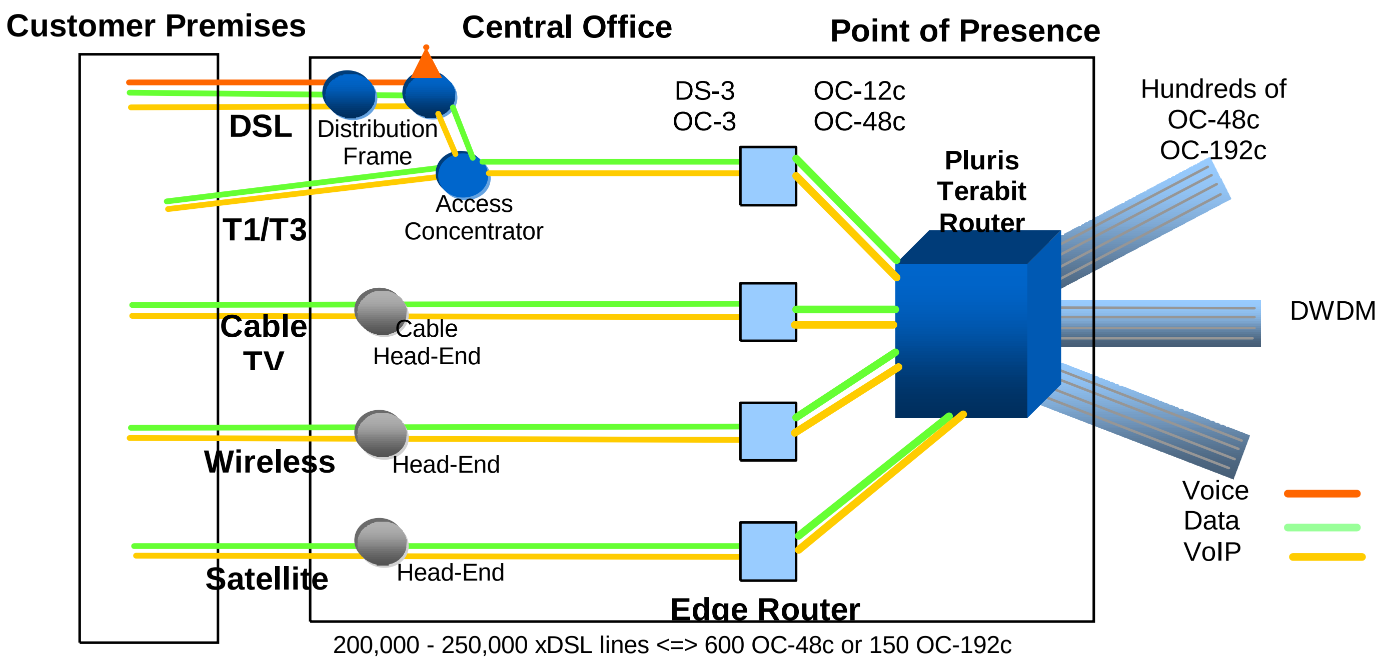


Figure 6: Terabit routers – an aggregation model

No matter how big or fast the backplane, a single-box system will always be fundamentally constrained by the amount of physical real estate in the box for providing port connections. This port-density limitation cannot be overcome simply by attempting to aggregate a number of monolithic individual systems, because the sheer complexity of load-sharing port traffic between the separate boxes quickly becomes prohibitive.

The emphasis for next-generation systems must be on creating inherent capacity to cost-effectively terminate the maximum number of ports within each box and then being able to gracefully and cost-effectively expand the overall system to encompass additional capacity. In essence, the basic architecture for tomorrow's backbone-class, terabit-level devices will need to transition from the concept of a "single box" to that of a "single system" in which all system components can be mutually optimized and smoothly scaled.