

PSINet Architecture

This presentation discusses how PSINet is built and compares it to traditional network design.



PSINet Design Goals

- Highly Reliable**
 - Few single points of failure
 - Frame Relay
- Easily Maintainable & Expandable**
 - POP standardization
 - Remote management
 - Distributed service force
- Standards-based Technologies**
 - Compatibility
 - Economical

Copyright © 1997 PSINet Inc.

Confidential and Proprietary Information

T05.2

PSINet's primary goals when designing the network were reliability, ease of maintenance and expansion, and use of standards-based technologies.

Reliability is achieved by eliminating as many single points of failure as possible. The use of Frame Relay also increases the reliability of the network. Frame Relay is a telephony technology built for speed and reliability. It allows a logical network to be layered upon the physical network. Switching at the Frame Relay layer utilizes reliable, mature telephone technology.

POP design is consistent which allows our distributed service force to operate with efficiency whenever POP maintenance is necessary. PSINet uses remote management techniques to monitor and maintain network equipment. This reduces the number of on-site trips required, reduces costs and also reduces the amount of time needed to repair a problem. For ease of maintenance, PSINet has a distributed field service force making on-site maintenance and repairs quicker and easier.

PSINet has chosen to follow Internet standards; we do not support proprietary configurations when communicating with customers. This allows for the greatest variety when choosing equipment, for both PSINet and customers. As long as a device supports and implements the necessary standards, it should work with other equipment that implements the same standards.



PSINet *Traditional IP Design*

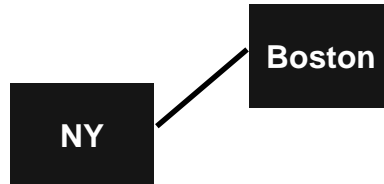
- Lease a phone circuit from Site A to B.
- Put routers on each end.
- Add new circuits and routers as needed.
- Leads to asymmetric network topologies.

In the traditional IP design a circuit between two sites would be leased from a telephone company and routers would be placed on each end. To add an additional site, another router and another circuit between one of the previous sites and the new one would be required. This leads to an asymmetric network which makes routing more difficult to manage, redundancy difficult, and provides many single points of failure.

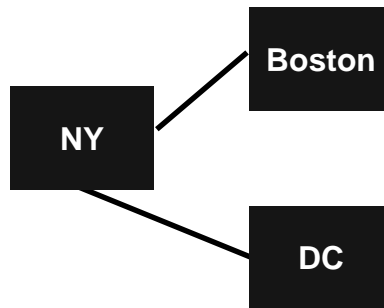
The network grows from need rather than from a well developed plan.



Traditional IP Design

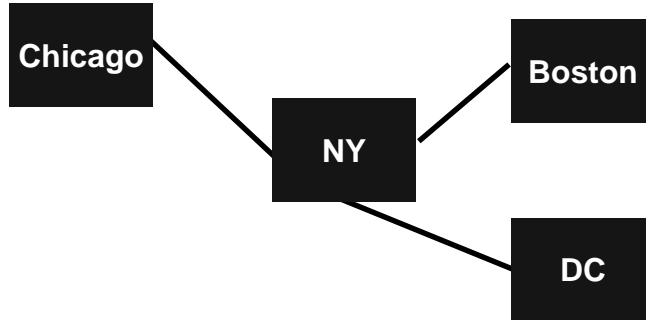


For example, assume an organization begins to develop a network with their primary offices - one in Boston and one in New York.



As the organization expands, the Washington DC office needs to be added to the network. Cost analysis is done and it is decided to lease a circuit between New York and DC.

PSINet *Traditional IP Design*



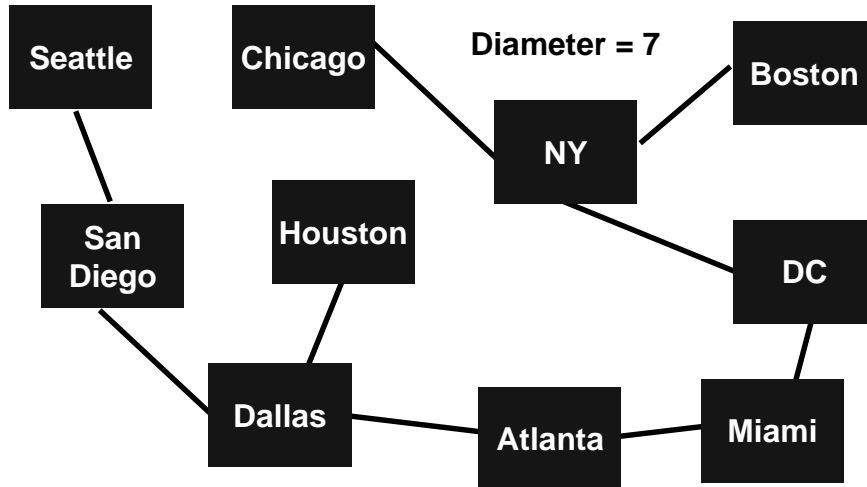
Copyright © 1997 PSINet Inc.

Confidential and Proprietary Information

T05.6

Chicago is the next office that needs to come on-line. Again, the network is examined and the decision is made to connect Chicago to New York.

PSINet *Traditional IP Design*

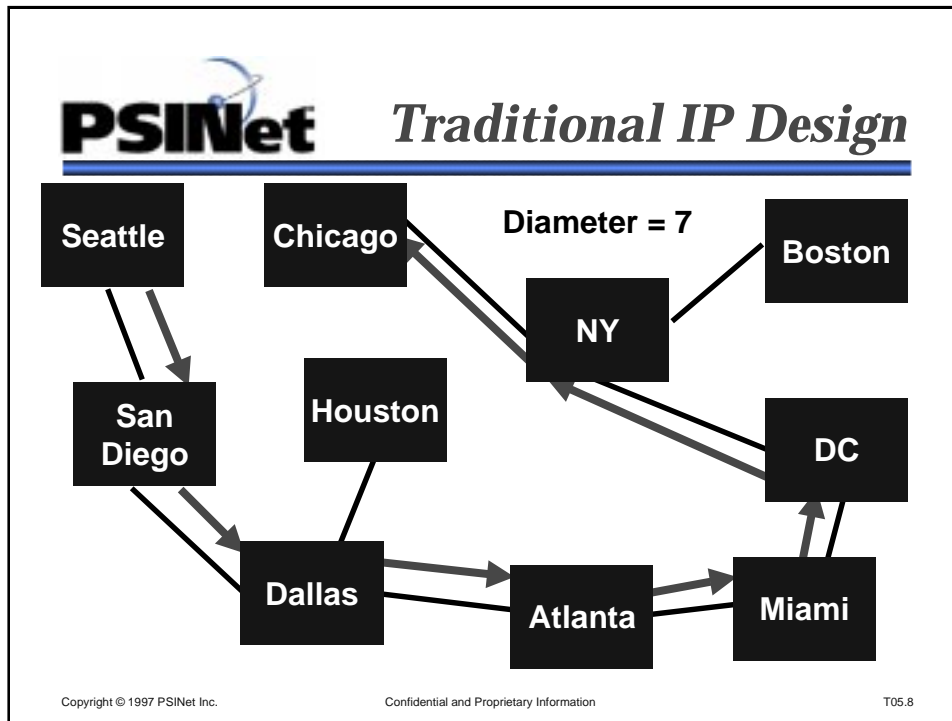


Copyright © 1997 PSINet Inc.

Confidential and Proprietary Information

T05.7

This process continues until the network is asymmetric, lacks redundancy and has a large diameter, where diameter is the maximum number of routers between any two sites on the network.



When a packet travels from Seattle to Chicago it goes through 7 different routers - Seattle - San Diego - Dallas - Atlanta - Miami - DC - New York - Chicago.

At each router the packet must be examined against the routing table and a decision must be made where to send the packet. As the size of the routing tables increases, each decision takes longer.



Frame Relay

- Layers 1 & 2 telco technology
- PVC - Permanent Virtual Circuit
- Bandwidth utilization
- Logical network
 - Simplified IP routing
- Backup connections

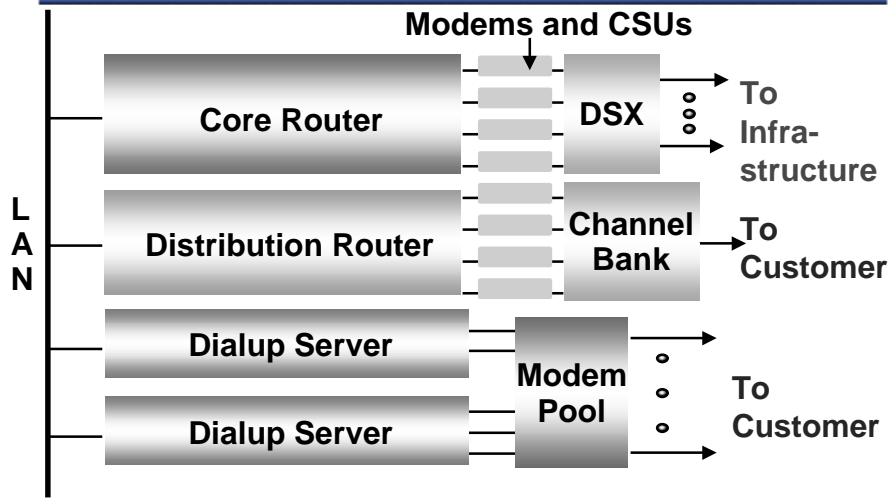
Frame Relay is a telephone technology used within PSINet. Frame Relay operates in layers 1 and 2 of the OSI model. When traffic reaches a switch and needs to be routed, telephone switching technology is used; this is more efficient since the information does not have to be examined against all the information in an IP routing table.

There are many advantages to using Frame Relay in PSINet's infrastructure:

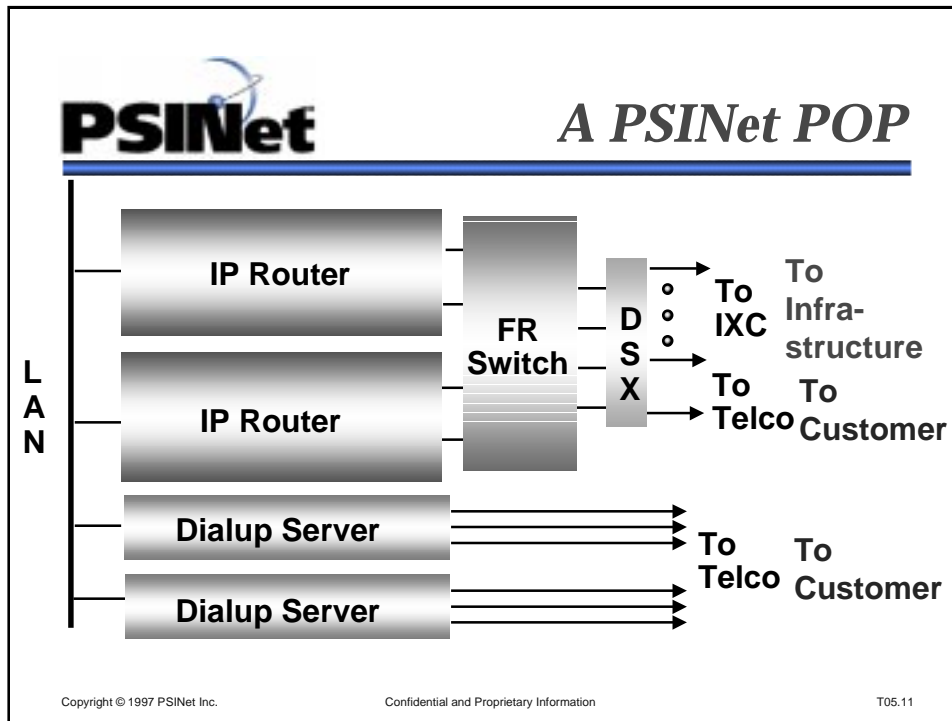
Permanent Virtual Circuits can be created which allow management of bandwidth utilization and specification of a logical network.

IP routing can be simplified with the creation of a symmetric, organized logical network.

Backup PVCs can be implemented which allow traffic to automatically route around network problems.



In a traditional POP there are many CSUs and modems for incoming circuits. Each interface is a potential point of failure. For a T1 there would be 24 CSUs and 48 wires. Each CSU has 2 interfaces that connect to another interface on the router which makes for $24 \times 2 + 48 + 24 = 113$ points of potential points of failure.

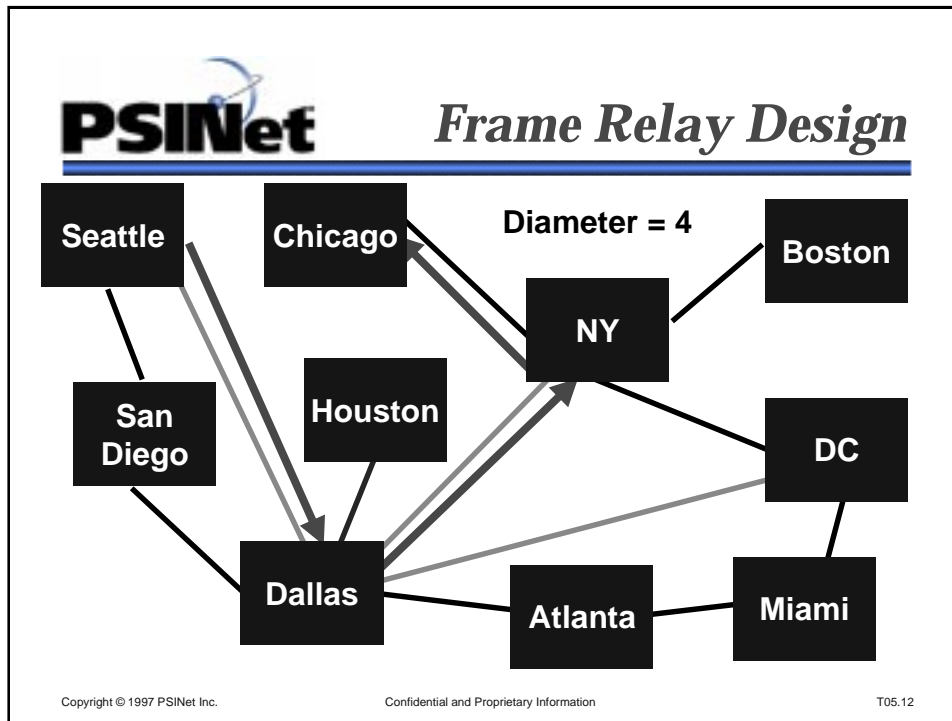


Notice the reduction of points of potential failure. There are no CSUs, just a DSX board and a FR switch. The one port on the FR switch replaces the 113 points of potential failure in the typical IP POP design.

The IP routers are rated at thousands of packets per second, support up to 24 direct connections and are capable of hundreds of virtual (FR) connections.

Dialup servers serve both ISDN and analog lines and serve up to 96 lines total.

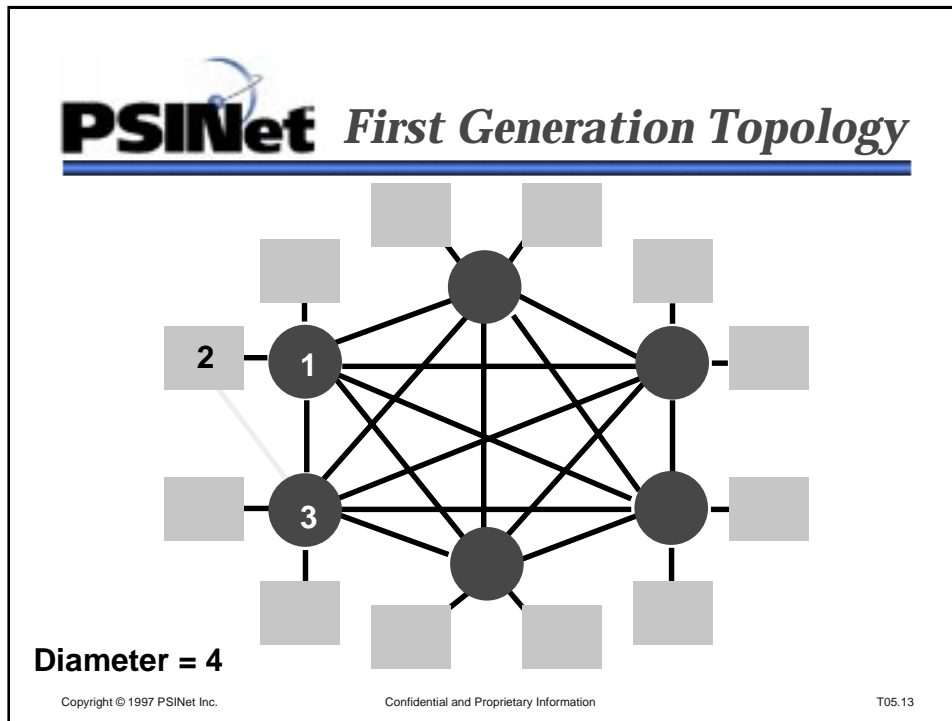
POP standardization makes it easy to swap components. The Frame Relay switches allow us to deliver circuits to any IP router and eliminate the need for discrete DSUs and channel banks for substrate circuits.



Now we return to our asymmetric network example and add Frame Relay technology. This allows a logical network to be layered upon the underlying physical structure by creating PVCs.

For example, using straight IP technology to get from Seattle to Dallas requires an IP routing decision be made in San Diego. Frame Relay technology allows the creation of a logical IP link between Seattle and Dallas. Thus the IP routing of a packet between Seattle and Dallas is direct; the IP packet goes from Seattle to Dallas without ever entering the San Diego IP router. The traffic will have to enter the San Diego Frame Relay switch but this is a different type of routing based upon switching.

Note that the addition of three PVCs (Seattle - Dallas, Dallas - New York, and Dallas - DC) reduces the IP diameter of the network to 4.



This diagram represents the logical structure of PSINet's infrastructure between 1992 (when Frame Relay was first implemented) and 1996 (when the architecture was redesigned).

Core routers (the circles) are connected in a full mesh; that is, every core router is one IP-hop away from every other core router. Thus, if any one core router goes down, all other core routers can continue to communicate with each other.

Core routers serve to pass transit traffic between customers and enforce policy routing with other providers.

Leaf routers (the squares) serve to transmit routing information from customers to the core and manage routing within the leaf area of the network. Leaf routers pass transit traffic for directly connected customers only.

Each leaf router has a primary PVC to one of the core routers. The leaf also has one or more backup PVCs to other core routers for redundancy.



PSINet *First Generation Topology*

- Advantages
 - Logical network
 - Redundancy
- Disadvantages
 - Full mesh difficult to maintain (20+ cores)

This topology creates a very symmetric, redundant network. Thus the diameter of the network is small and the network is robust with respect to outages.

However, maintaining the full mesh of the core is difficult and requires a large number of PVCs. As the number of core routers increases, the number of PVCs that must be created, maintained and monitored becomes unmanageable.

For these reasons, PSINet determined it was necessary to redesign the logical layout of our infrastructure. Keep in mind, without Frame Relay there is no way to redesign network topology without changing the underlying physical layout.

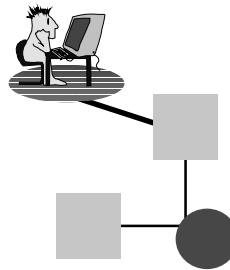
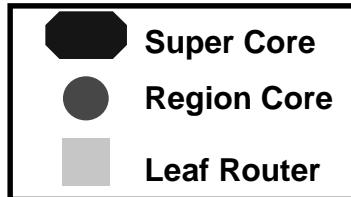


Design Goals

- Regionalized routing
- Redundancy & backup routing
- Extensible

When creating the newest topology, PSINet took many things into consideration and decided to regionalize the network. A clear, easily managed backup plan was also necessary. Support for large scale expansion was also a goal of the new architecture.

- Routing Protocol(s)
 - RIP - customers
 - IGRP - region
- Traffic
 - Within the leaf
 - Customer <-> Region
- Connections
 - Customers

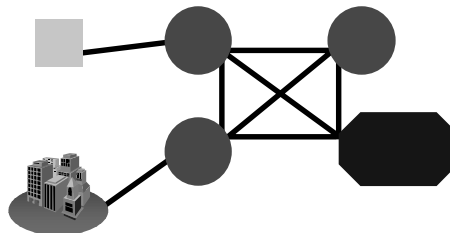
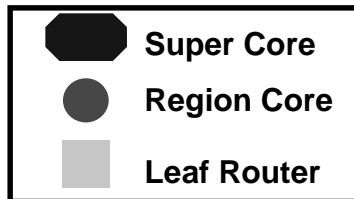


Taking the new topology one piece at a time, let's begin with the leaf.

Each leaf router is connected to customers and also to a region core router. Thus the leaf router passes traffic between the region core and customers. In addition, the leaf router is responsible for managing traffic of connected customers. For example, if two customers that are connected to the same leaf are passing traffic, that traffic should remain in the leaf and not pass to the region core.

Customers connected to the leaf broadcast their network routes to PSINet using RIP. This information is converted to IGRP and passed from the leaf router to the region core router. The routing table in the leaf contains specific information about all customers connected to that leaf and a default route to the region core; all traffic not destined for a customer in the leaf is sent to the region core for further examination and routing.

- Routing Protocol(s)
 - IGRP - leaf, core
 - OSPF - IMAN, core
- Traffic
 - Within the region
 - Region <-> Core
 - Leaf <-> Leaf
- Connections
 - InterMAN area



Now we will examine a region core. Each region core router is connected to one or more leaf routers and also to a super core router. The region core router transits traffic between the leaf area and the super core. In addition to passing transit traffic, the region core is responsible for routing traffic within the region. Thus, if traffic is being passed between two leaf routers in the same region, the region core router is responsible for managing this traffic; the traffic does not need to enter the super core to reach its destination.

The region core router communicates with leaf routers using IGRP. This protocol is also used when region cores communicate with each other.

Another function provided by the region core is related to PSINet's InterMAN service. InterMAN customers are connected to PSINet via a region core router. Since InterMAN routes are propagated using OSPF, region core routers communicate using this protocol in addition to IGRP.

A routing table in a region core router contains information specific to that region. Thus, each region core router has information about all leaf routers in the region and all customers connected to those leaf routers in addition to any InterMAN routes in the region. Any routing information that is non-region specific is handled with a default route to the super core. If a packet is destined for some location outside of the given region, it is passed to the super core for further examination and routing.

Routing Protocol(s)

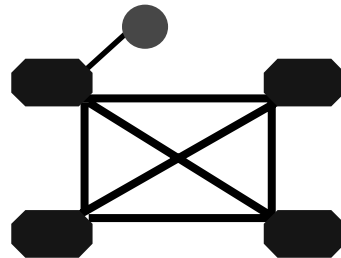
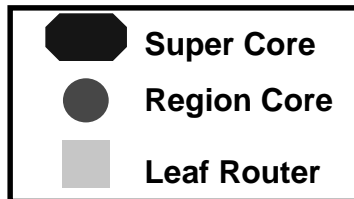
- IBGP - core
- BGP4 - external peers
- OSPF - IMAN routes

Traffic

- Transit between regions
- To/from external peers

External Connections

- CIX, MAE-East, MAE-West
- International



Finally, we will discuss the super core. The super core contains one router for each region so the super core passes transit traffic between different regions of PSINet. In order to provide this function, each super core router must have complete knowledge of PSINet routing information; the super core routers are unable to use a default route, they must have complete routing tables.

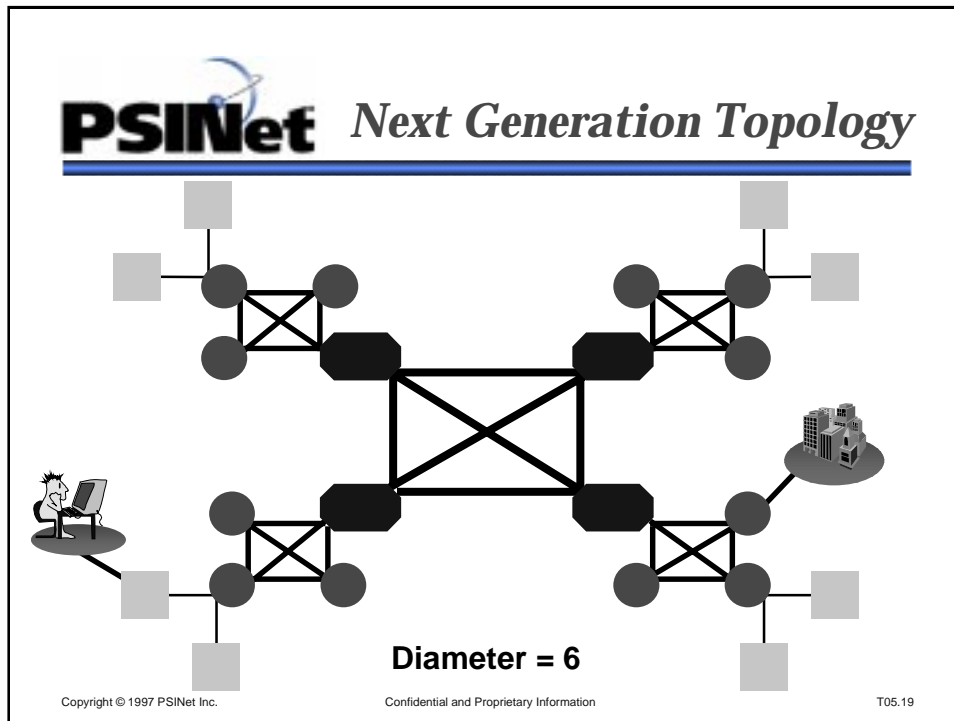
Traffic is exchanged between PSINet, Inc. (in the United States) and our international partners and subsidiaries through the super core.

Similarly, PSINet must also communicate with the rest of the world. This function is also handled by the super core. For example, PSINet has a router at MAE-East, one of the main inter-exchange points. This router is part of the super core.

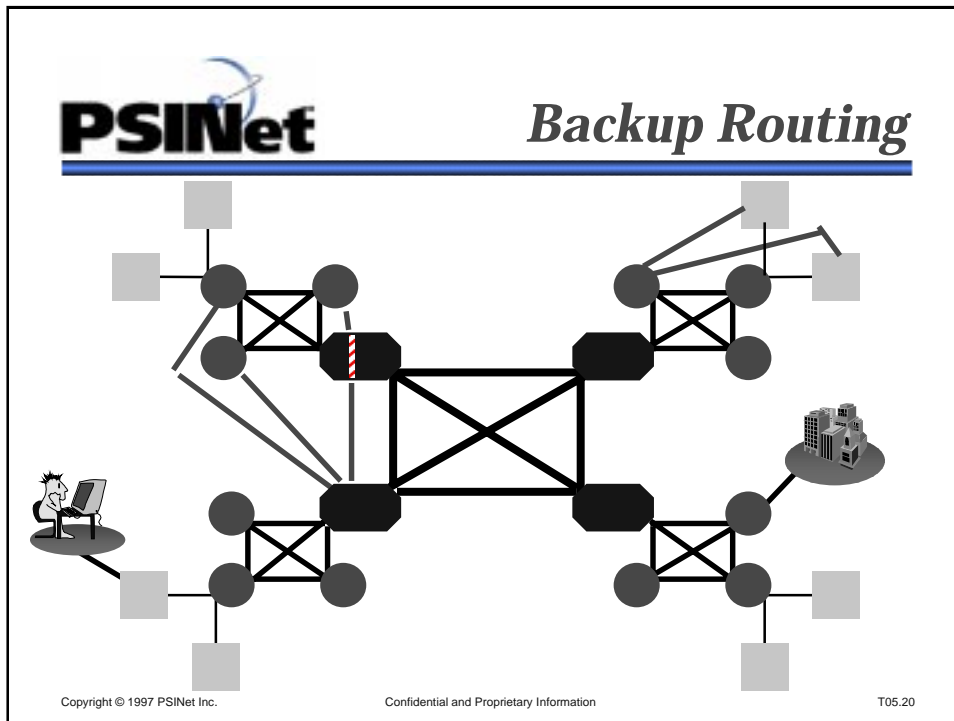
PSINet routing information, that is, routes belonging to PSINet's infrastructure and customers, is communicated using IBGP. IBGP is a variation of BGP used within a single Autonomous System.

InterMAN routes are propagated throughout PSINet using OSPF. The super core routers are the backbone area in an OSPF network.

When communicating with external peers, PSINet uses BGP4, the accepted protocol for exchanging routing information between different Internet service providers.



Putting all the pieces together, this is the conceptual picture of how PSINet's architecture is organized.



PSINet included a backup procedure when designing this architecture.

Each leaf router connected to a given region core also has a backup connection to a different region core within the same regional area. Each region core router has a backup connection to a second core router. This configuration results in a dynamic network that is easy to maintain and expand.

