

FORE
SYSTEMS



The logo for FORE SYSTEMS is a black square with the word "FORE" in large white capital letters above the word "SYSTEMS" in smaller white capital letters. A thin white horizontal line is positioned below the text, and a thicker red horizontal bar is at the bottom of the square.

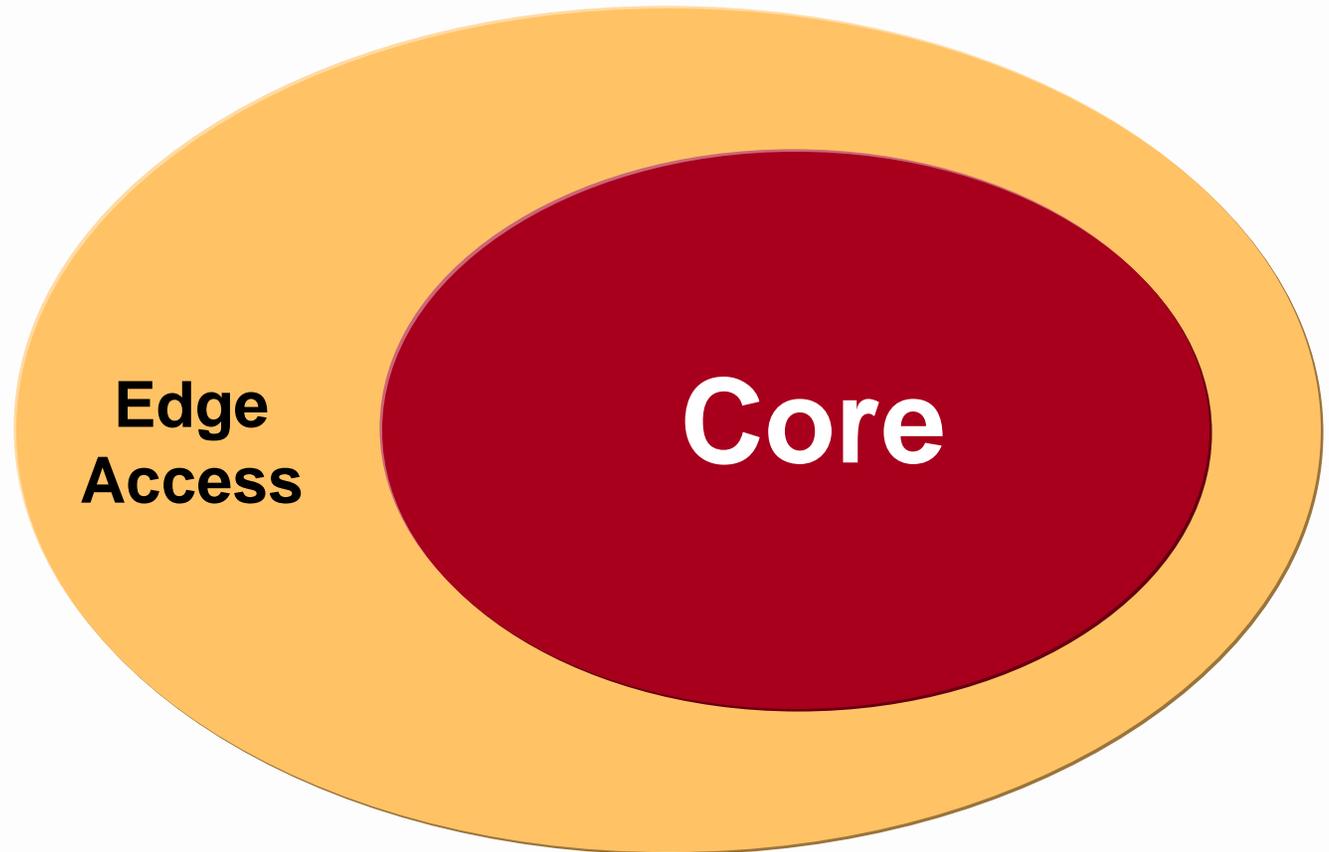
FORE
SYSTEMS

The Evolution of Routing

V1.0: Geoff Bennett

What is Routing?

- **Hierarchy**
- **Scalability**
- **Security**



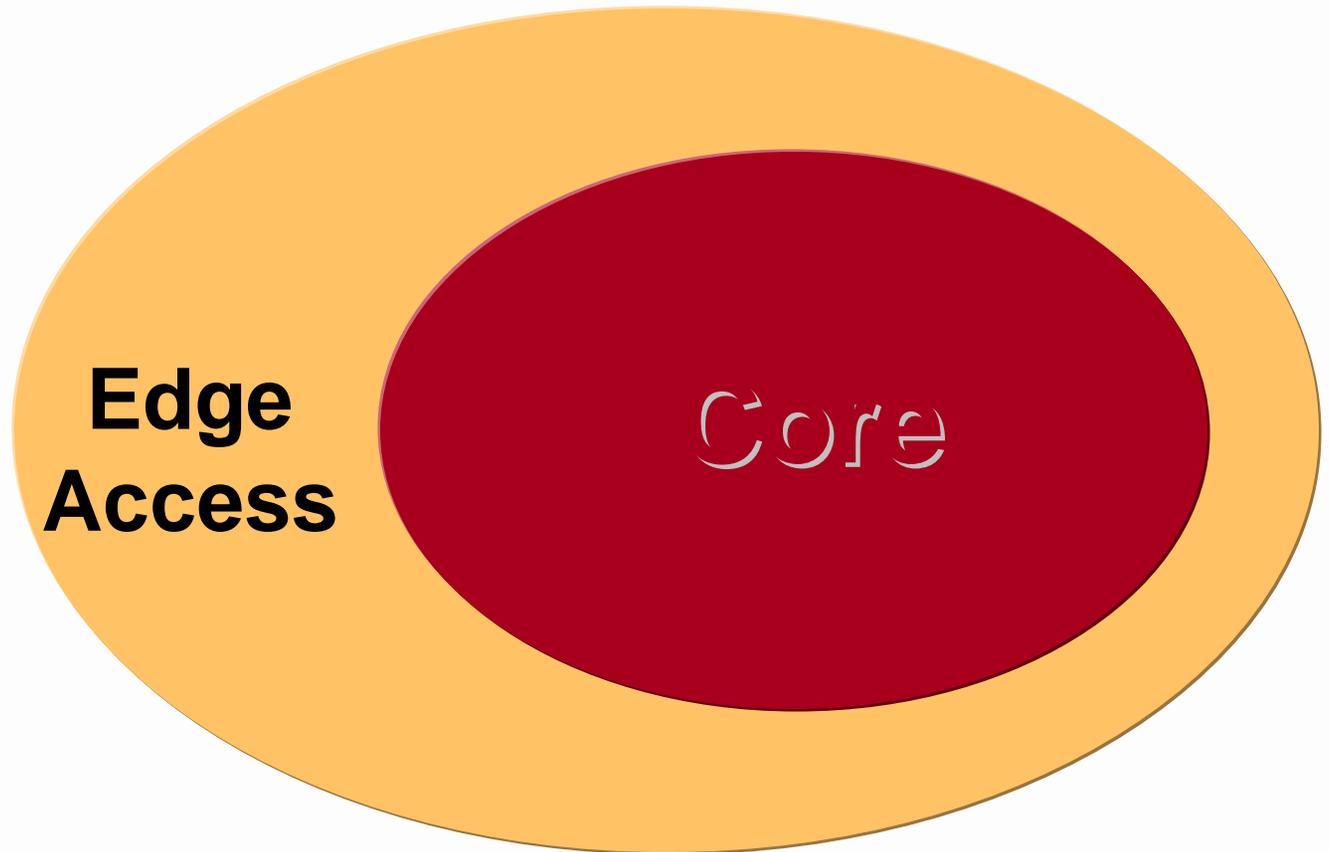
A very general definition of routing would be the interconnection of end systems on the basis of Network Layer addresses.

A further refinement of this definition would make it clear that Network Layer addressing schemes, like IP, include an address prefix, or Subnet ID, that uniquely identifies the network segment on which a given device is located.

Routers can use this identification to add three vital elements to our networks; Hierarchy, Scalability and Security.

What is Routing?

- **Hierarchy**
- Scalability
- Security

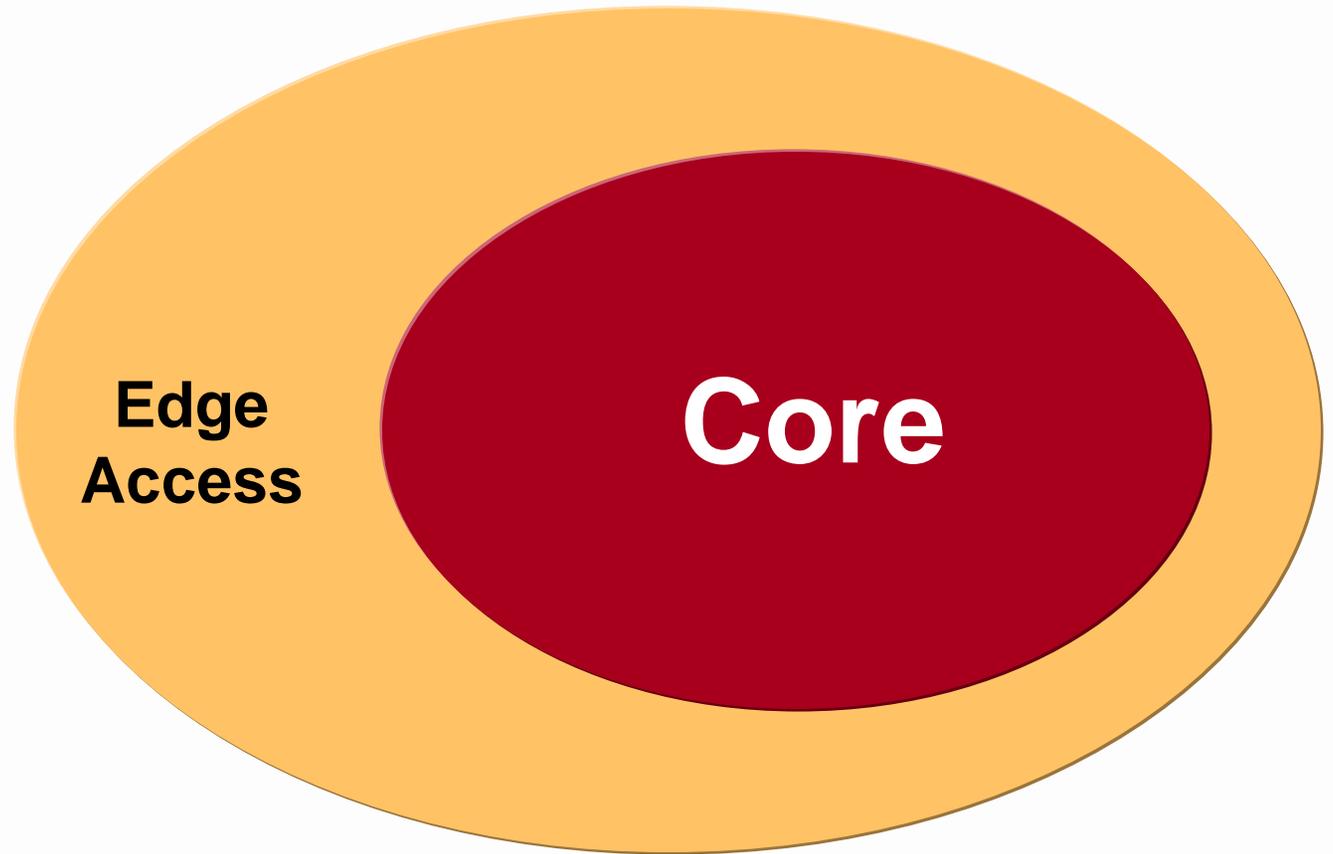


Hierarchy is really visible at the edge of the network. Specifically we can create LAN segment workgroups using Network and Subnet IDs.



What is Routing?

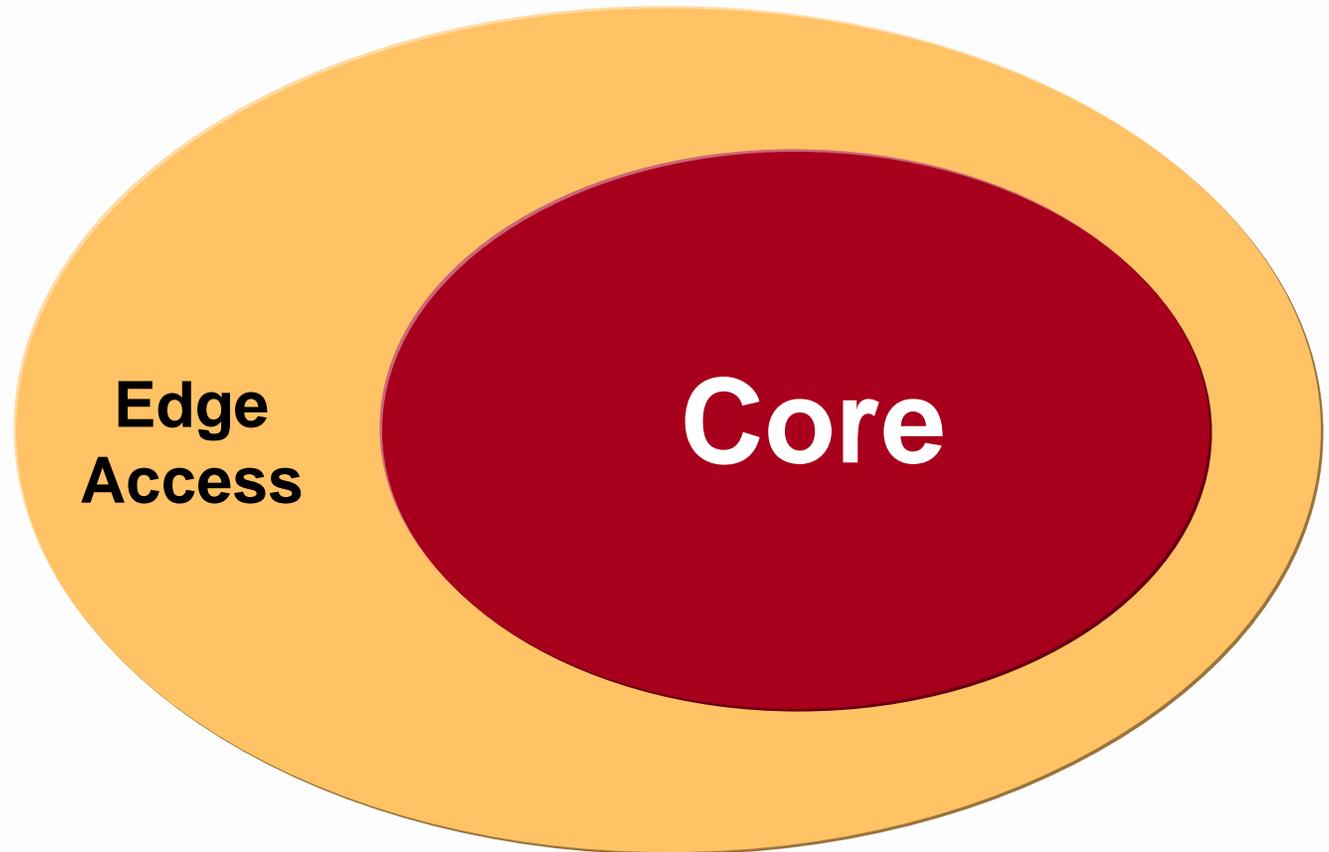
- Hierarchy
- **Scalability**
- Security



Routing allows us to scale up to large network populations because the router is able to preserve inter-subnet bandwidth.

What is Routing?

- Hierarchy
- Scalability
- **Security**



Routing is a *component* in a security system. Routers are able to act as Network Layer addressing filters, which is the first step in the creation of a firewall.

However, firewalls are much more than just packet filters, and so a conventional router alone cannot guarantee security in our networks.



Routing Evolution

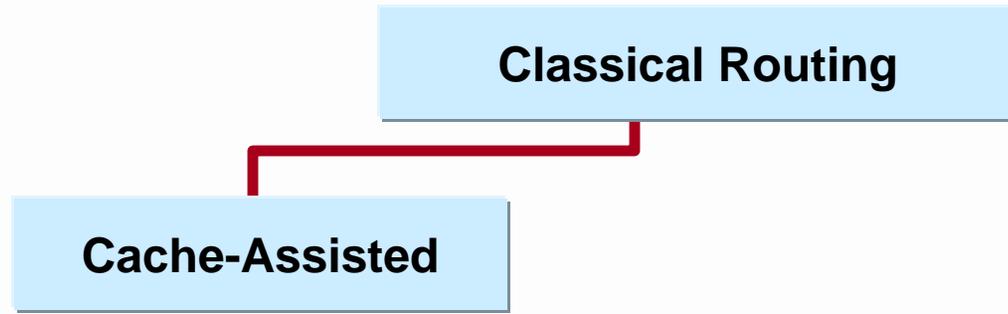
Classical Routing

All routing originates from Classical Routing.

In simple terms, Classical Routing states that end systems with different Network Identifiers may only communicate through a router.

In this context, Classical Routing is entirely CPU-dependant. A simple rule of thumb is that an internetwork transfer of a typical IP packet requires around 1,000 CPU instructions.

Routing Evolution

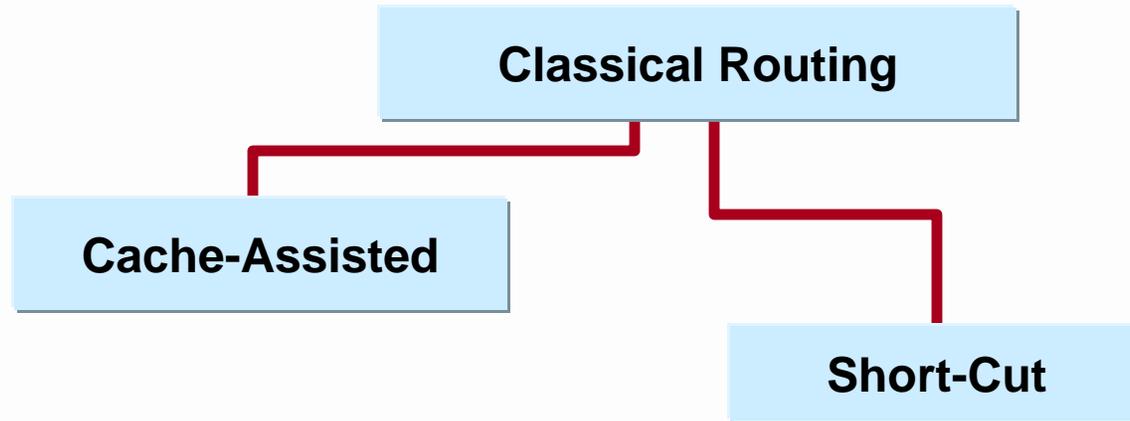


Strictly speaking, a connectionless communication system such as IP requires the router to process each packet as though it had no prior information. So, even though a router might be sending thousands of packets between two end points, each one *should* go through the same 1,000 CPU instructions.

In a real network, routers are able to take advantage of recognising flows in the packet stream and abbreviating the code path dramatically.



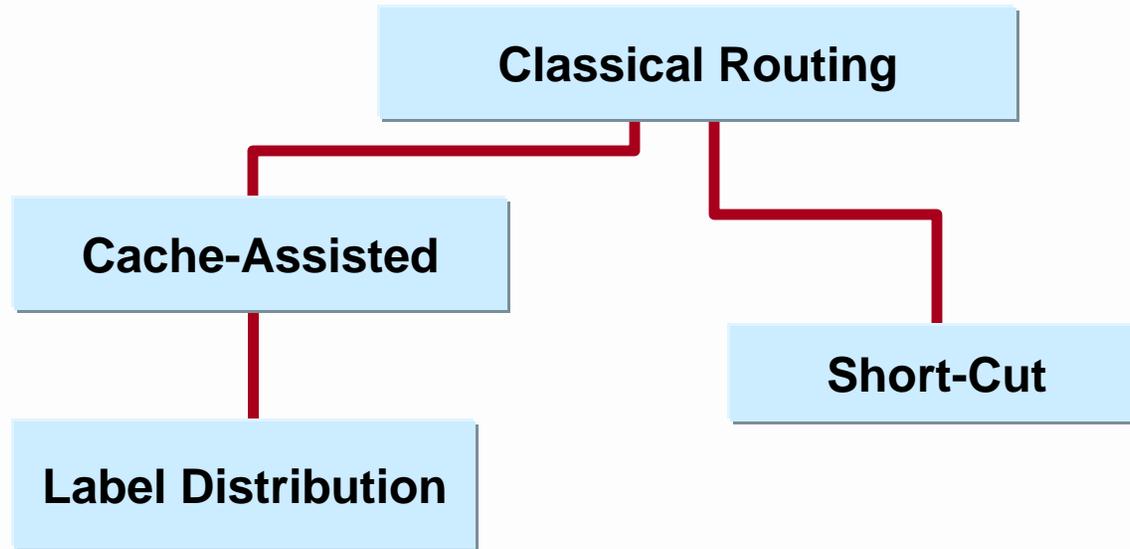
Routing Evolution



An alternative approach is to make use of a connection-oriented underlying network (such as ISDN, Frame Relay SVCs or ATM). The connections can act as short-cuts within the routing procedure.



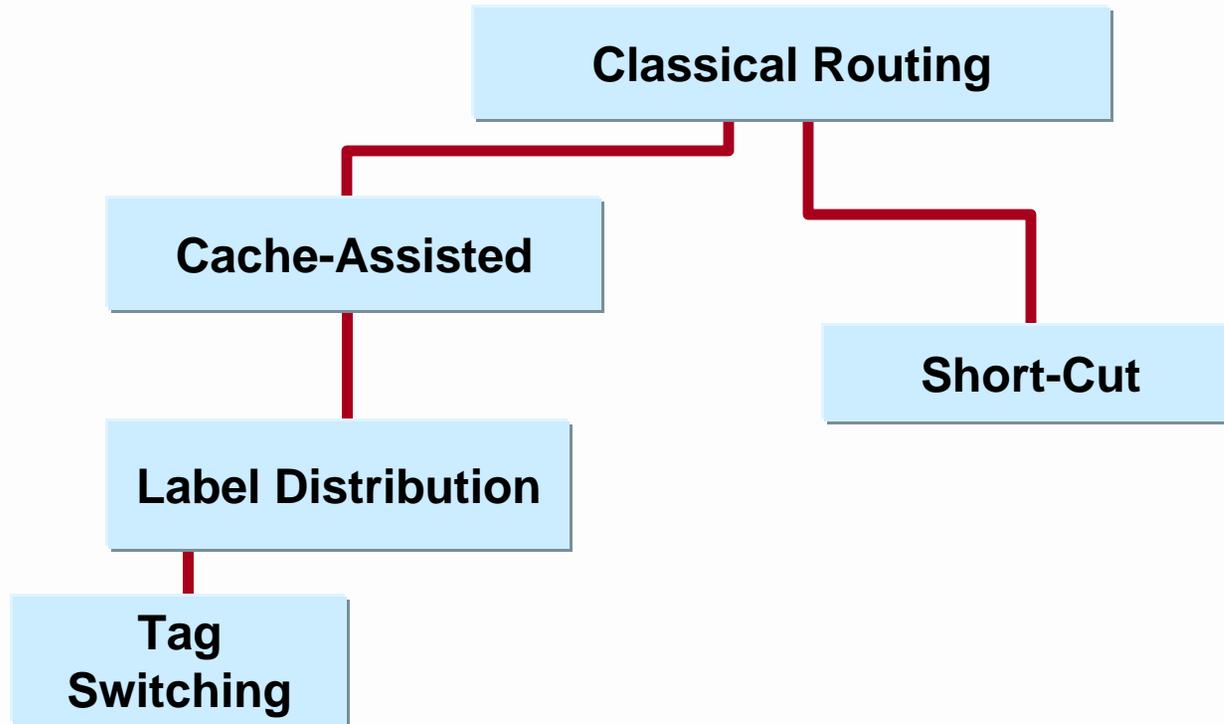
Routing Evolution



Chronologically the next development came by extending route caches between routers. This is done by assigning labels to cache entries.



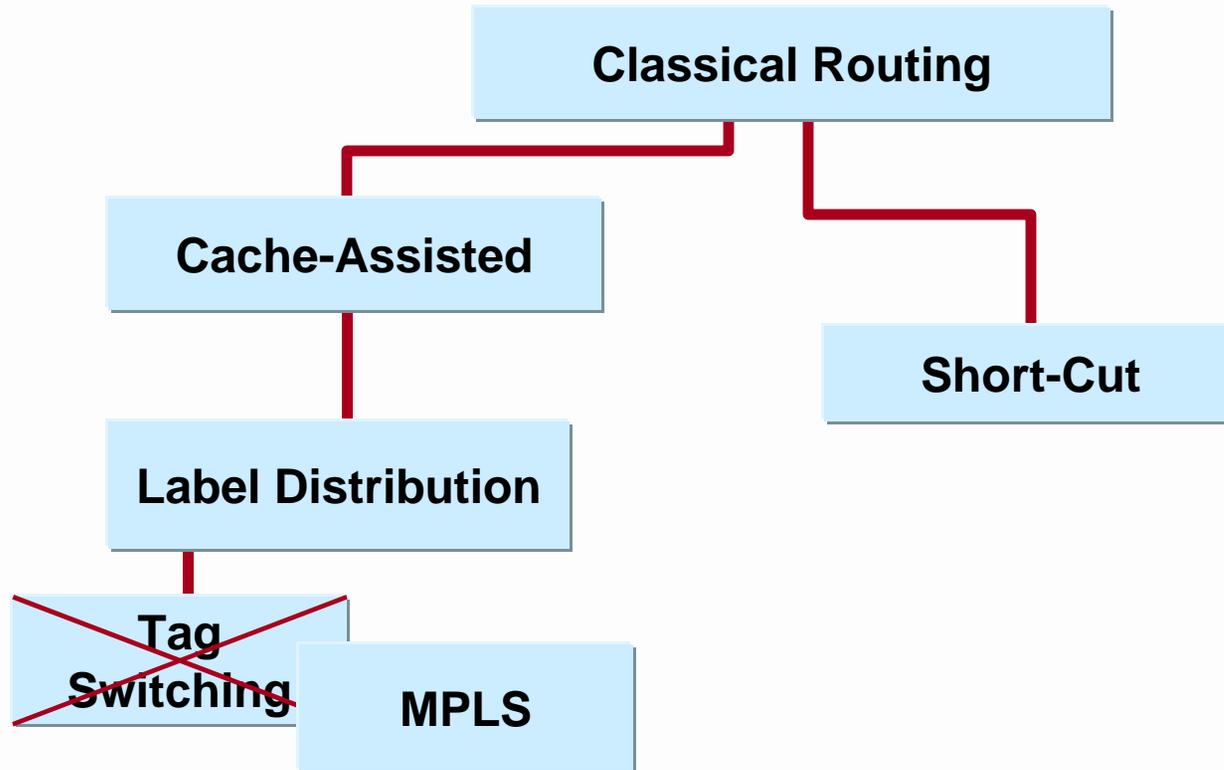
Routing Evolution



The first experimental implementation of label distribution came from Cisco in the form of Tag Switching.

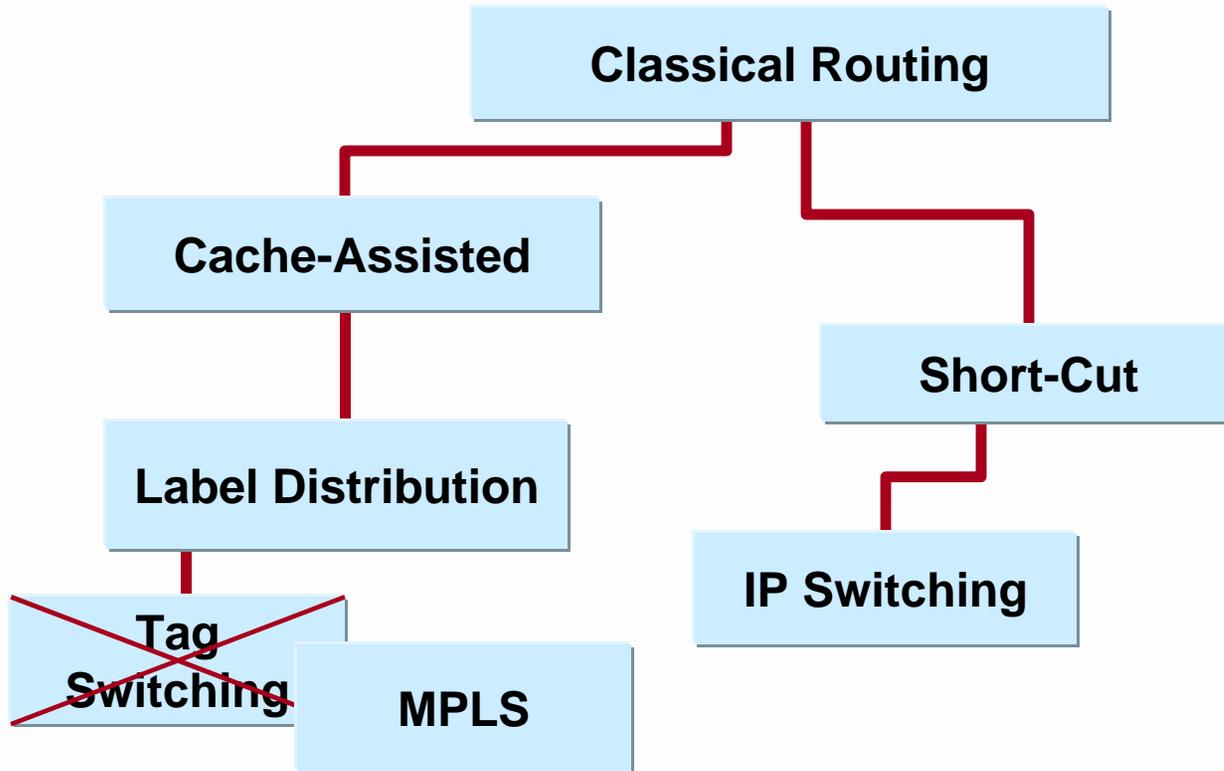


Routing Evolution



But Tag Switching is not the accepted direction of development for cache distribution in the IETF. Multiprotocol Label Switching (MPLS) is the standards-based routing acceleration technique.

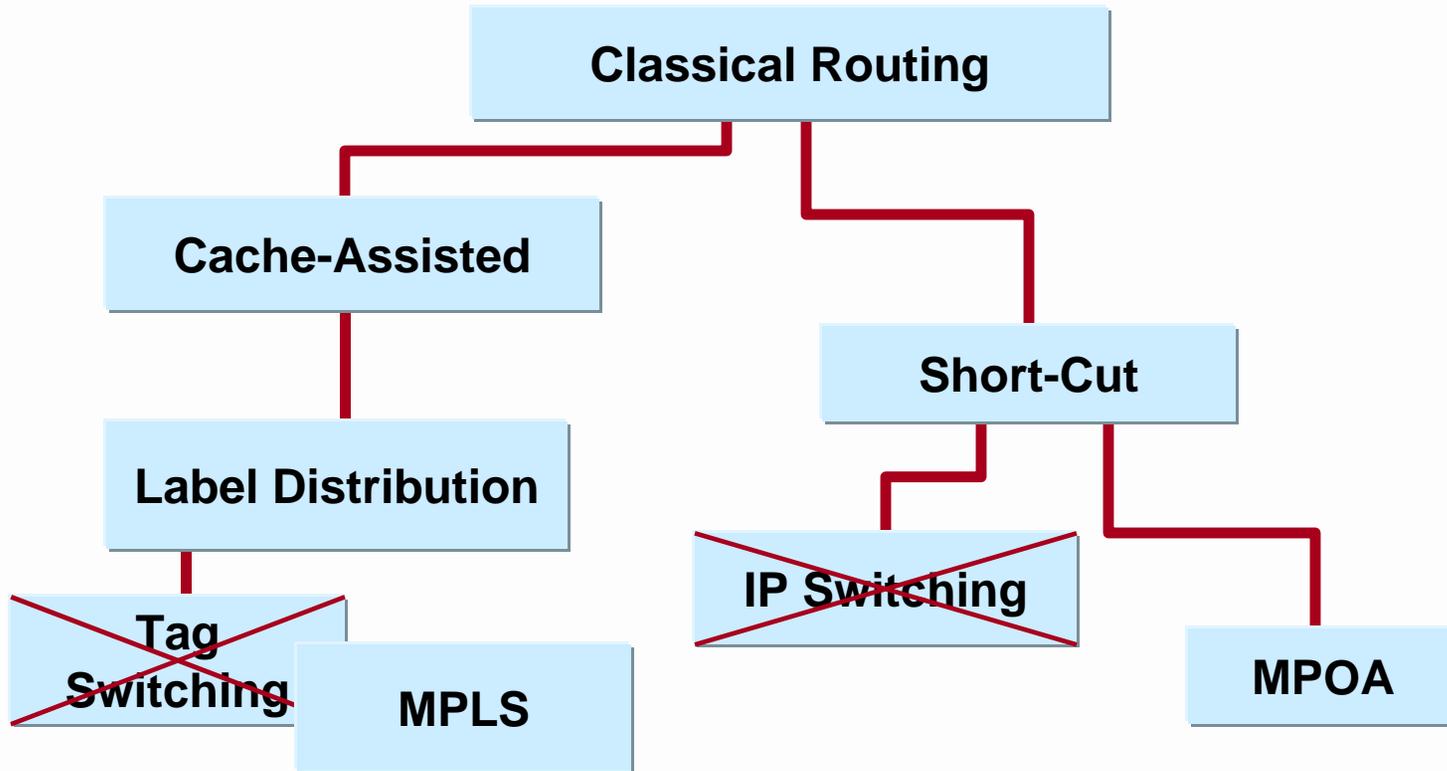
Routing Evolution



Short-cut routing took a similar proprietary side road for a time. Ipsilon proposed a simplified short-cut technique known as IP Switching in response to the ATM Forum's rather sluggish development of MPOA. However, IP switching removes one of the basic benefits of an ATM backbone, its ability to integrate data with voice and video traffic (IP Switching removes ATM Forum signalling from the ATM switches).



Routing Evolution

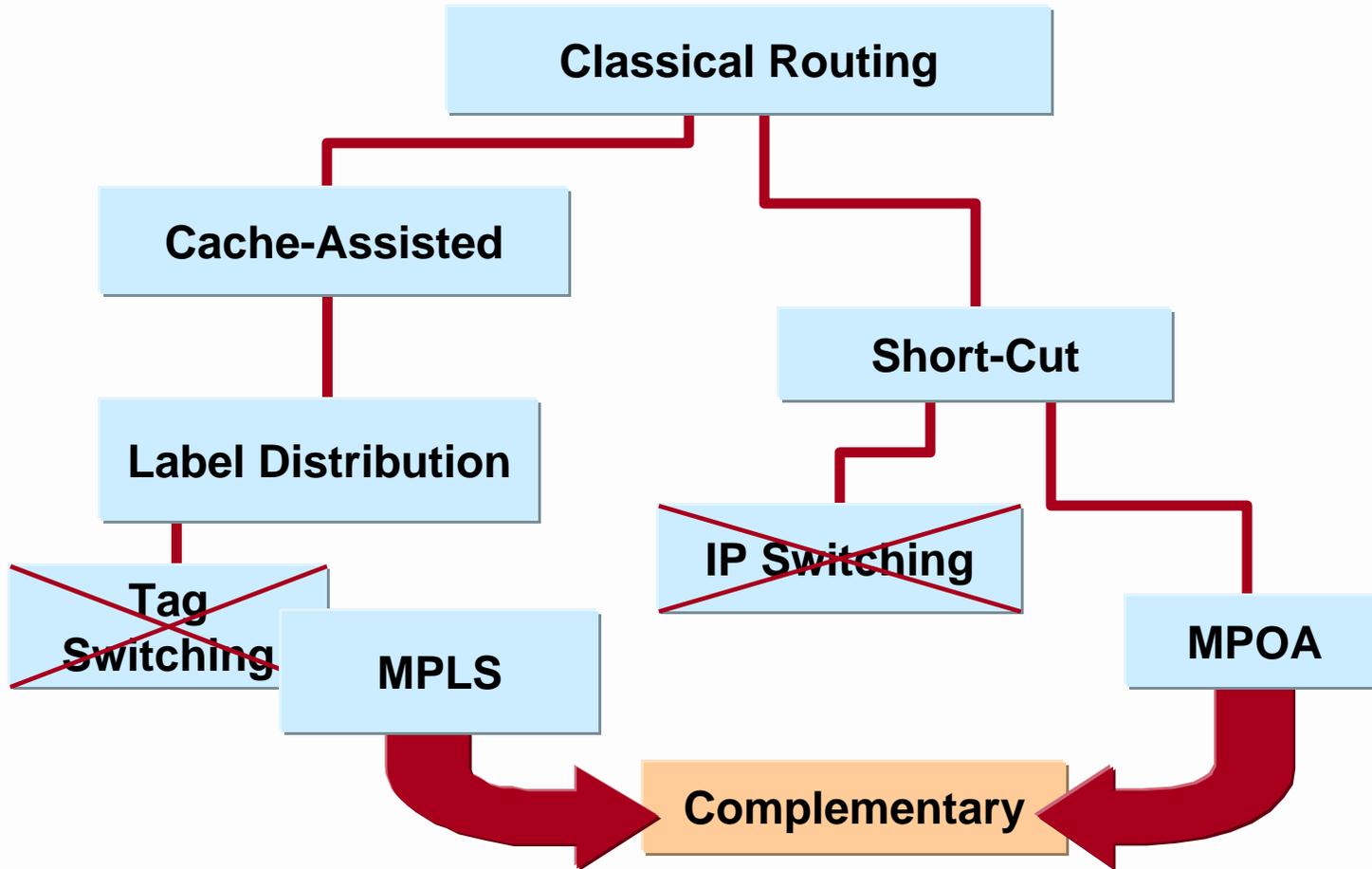


Ipsilon is now no more, having been acquired by Nokia, and the IP Switching technology has been dropped.

Perhaps the best thing to come out of the IP Switching initiative was to get MPOA moving faster through the Forum. Today MPOA is a fully approved standard, with implementations available from all of the major ATM vendors.

IP Switching still has some marketing momentum, but presumably will die out naturally as a result of MPOA's standardisation.

Routing Evolution



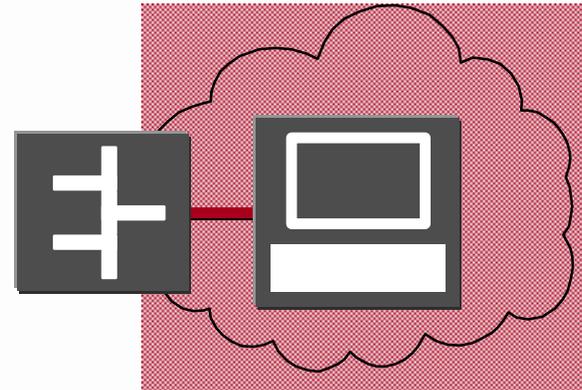
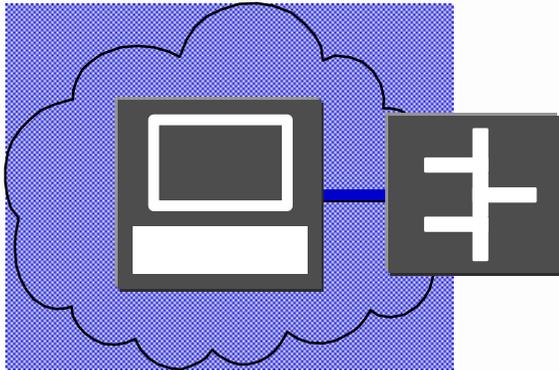
Some observers question the direction “tag switching or MPOA?”. In this respect they are missing an important point.

Tag Switching (or in its standard form, MPLS) is a routing acceleration technique that can be used over frame-based infrastructures such as Frame Relay or PPP.

MPOA is a short-cut technique that operates over a connection-oriented infrastructure such as ATM. MPLS, because it is based around frame-based store-and-forward techniques, will never offer the same performance improvement as MPOA. However, the two techniques are complementary in networks such a PPP access protocol over an ATM backbone.



Classical Routing



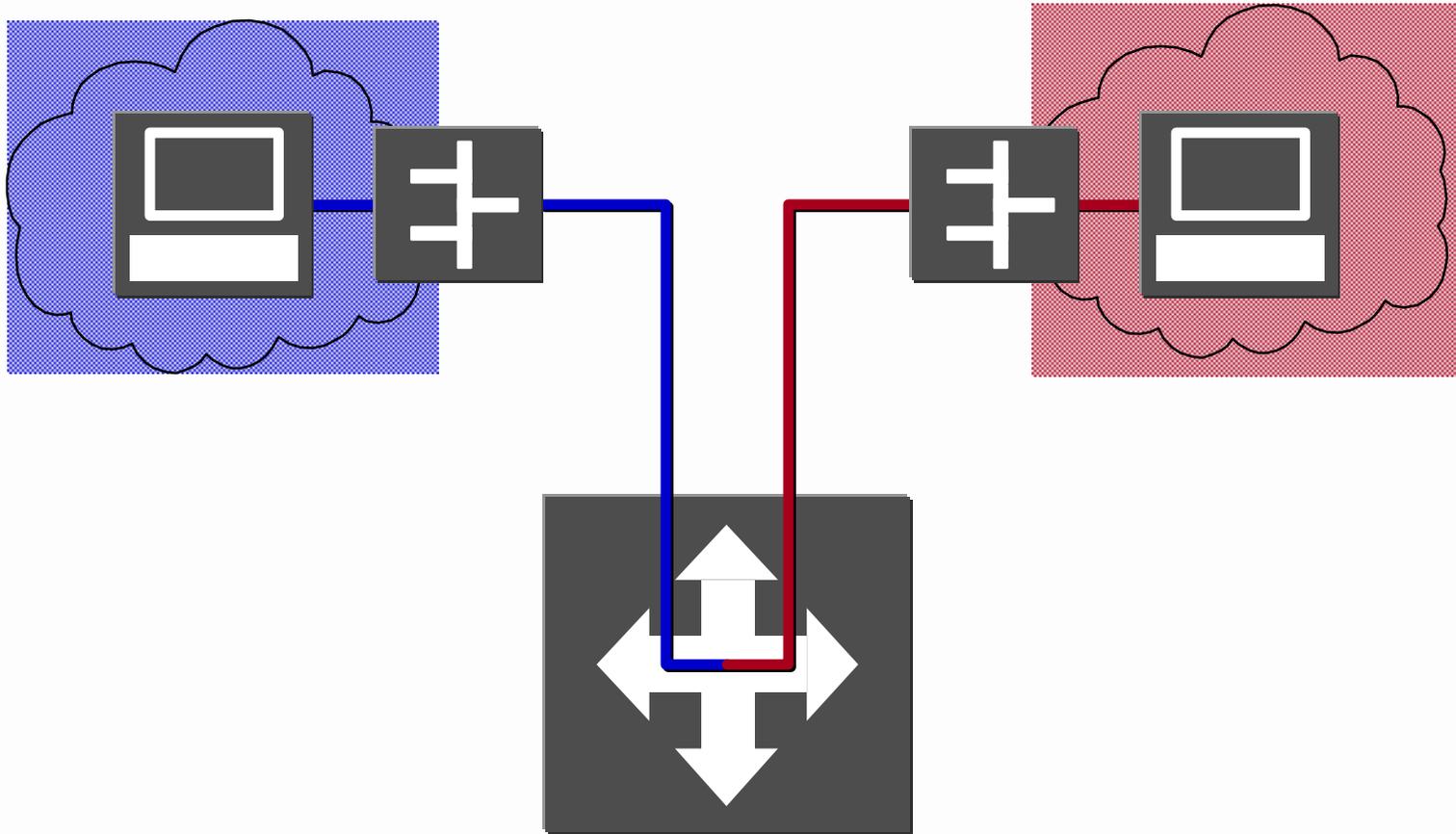
Let's look at these techniques in more detail, starting with Classical Routing.

Here we see two LAN subnets. In theory these could be any Network Layer protocol, but in reality all of the developments in this direction are based around IP.

For these examples I'll just use the Red and Blue subnet identifiers.

Here we see two end systems connected to basic shared media hubs.

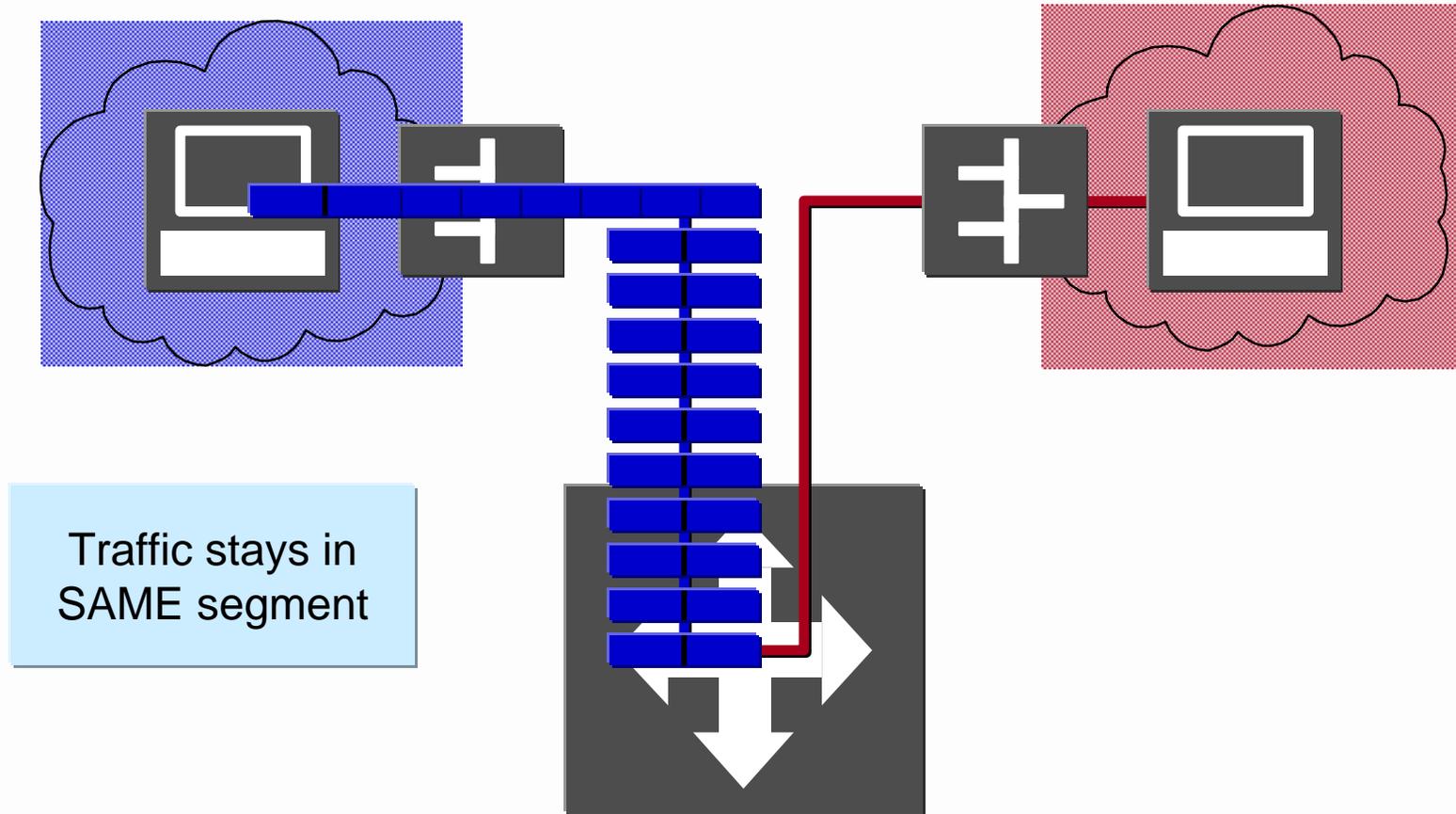
Classical Routing



The addition of a router to this network provides two essential elements.

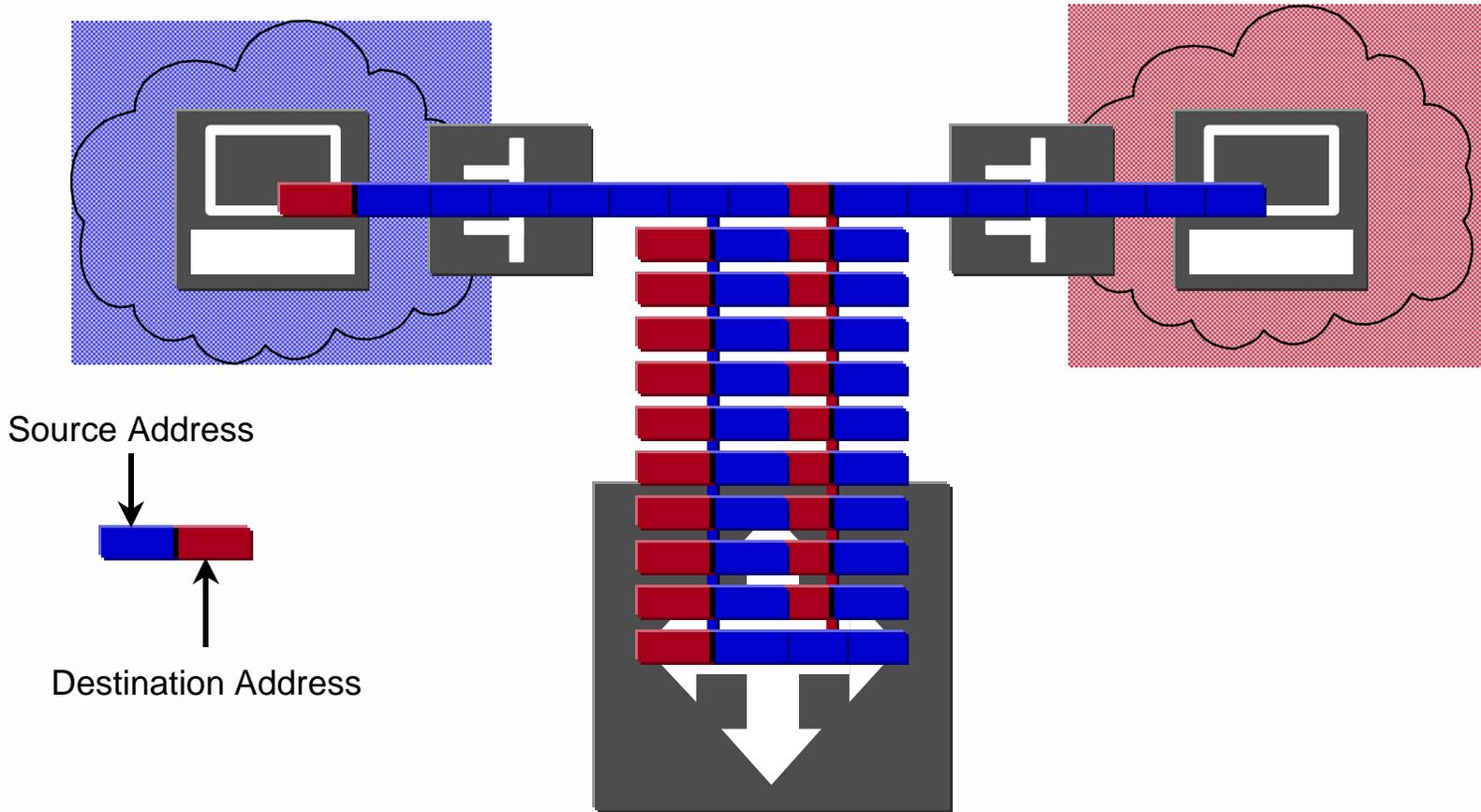
Physical connectivity between the cable hubs, and logical connectivity between the separate subnet IDs.

Classical Routing: Intra-segment



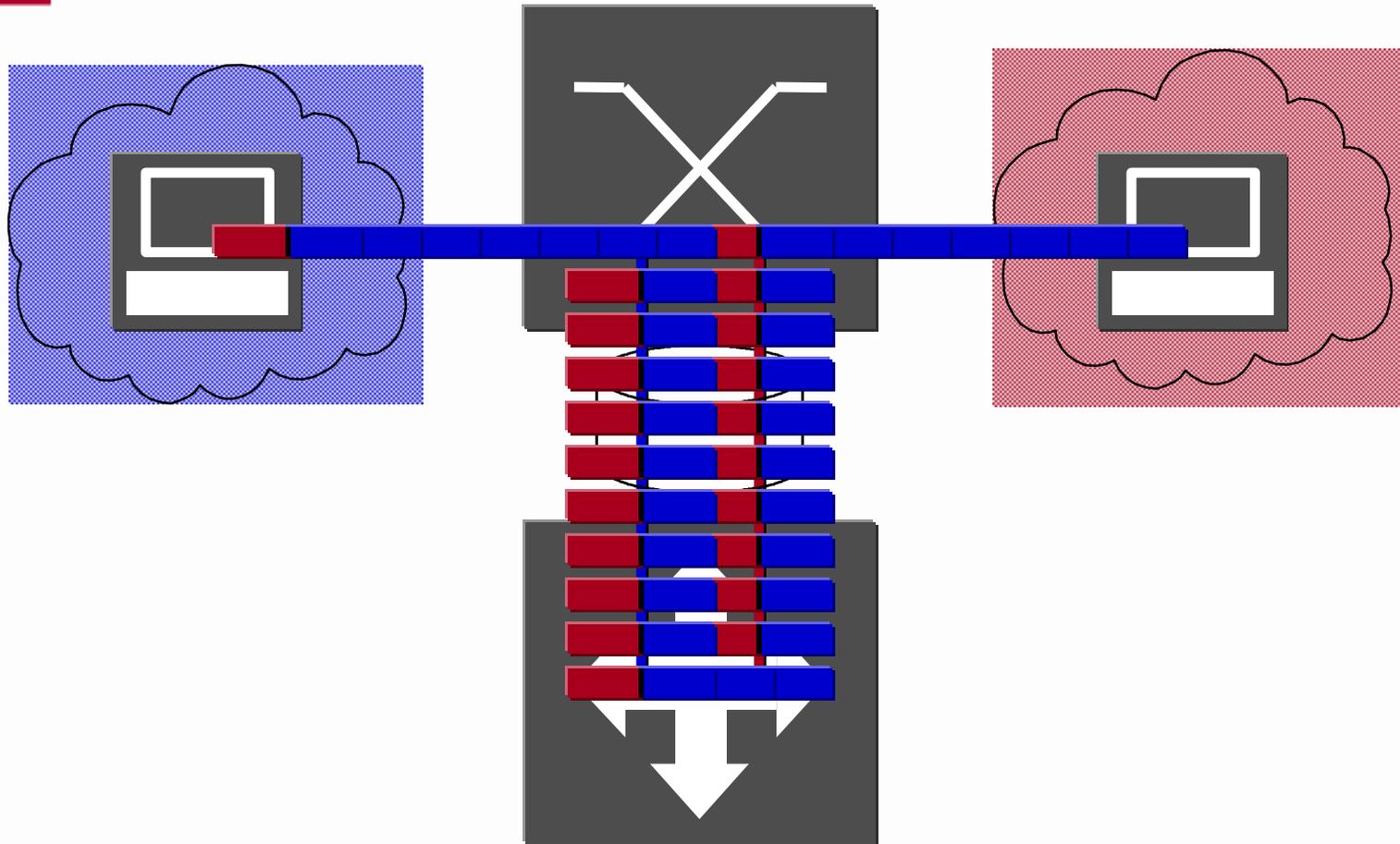
Traffic that operates within a subnet (intra-segment traffic) is kept in check by the router (ie. providing scalability by limiting broadcast spread).

Classical Routing: Intersegment



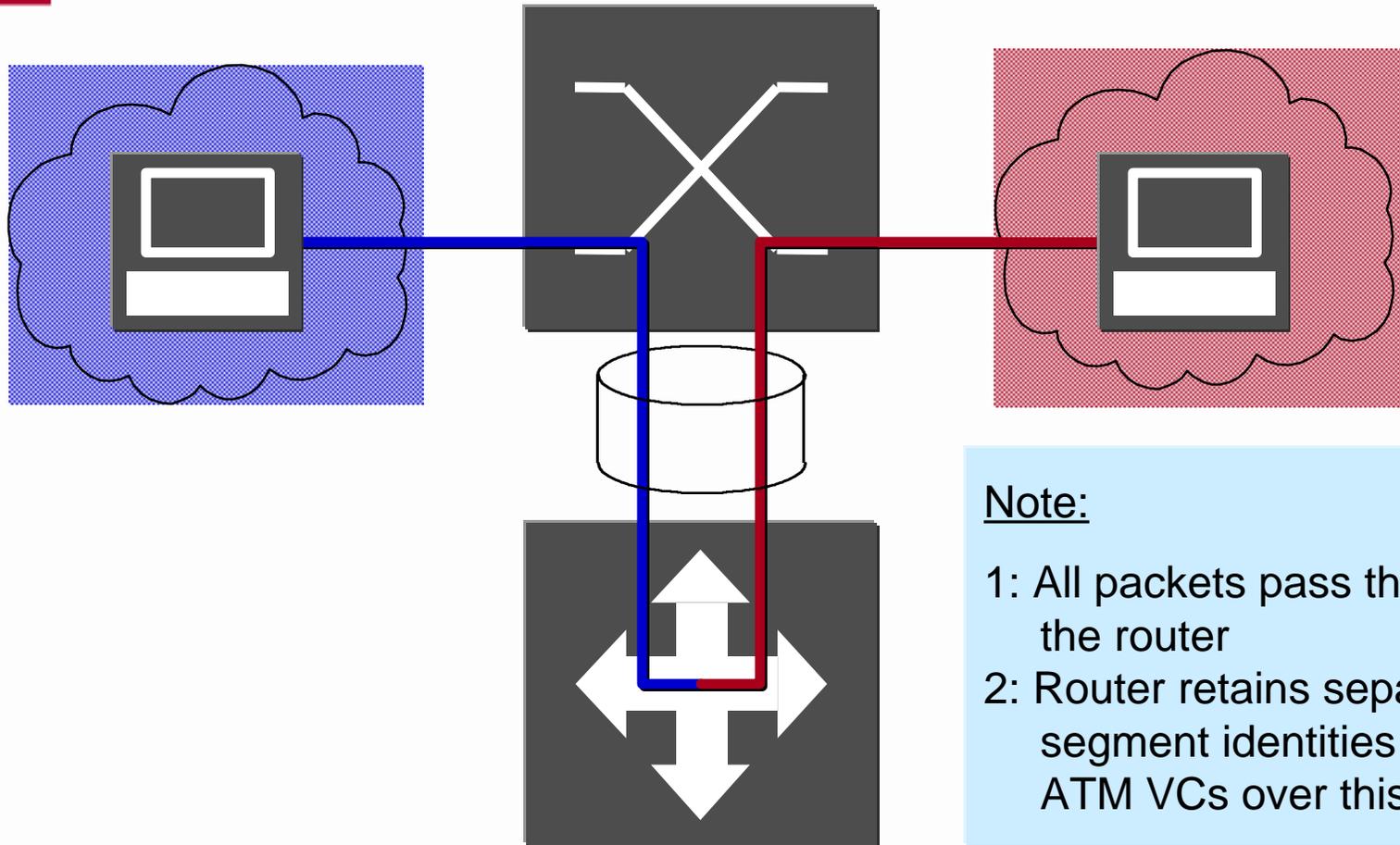
Only intersegment traffic actually passes through the router.

Classical Routing Over ATM: **Intersegment**



If we migrate this concept to ATM, we see a one-armed router (ie. a single ATM physical connection that allows access to two or more logical subnets).

Classical Routing Over ATM: Intersegment

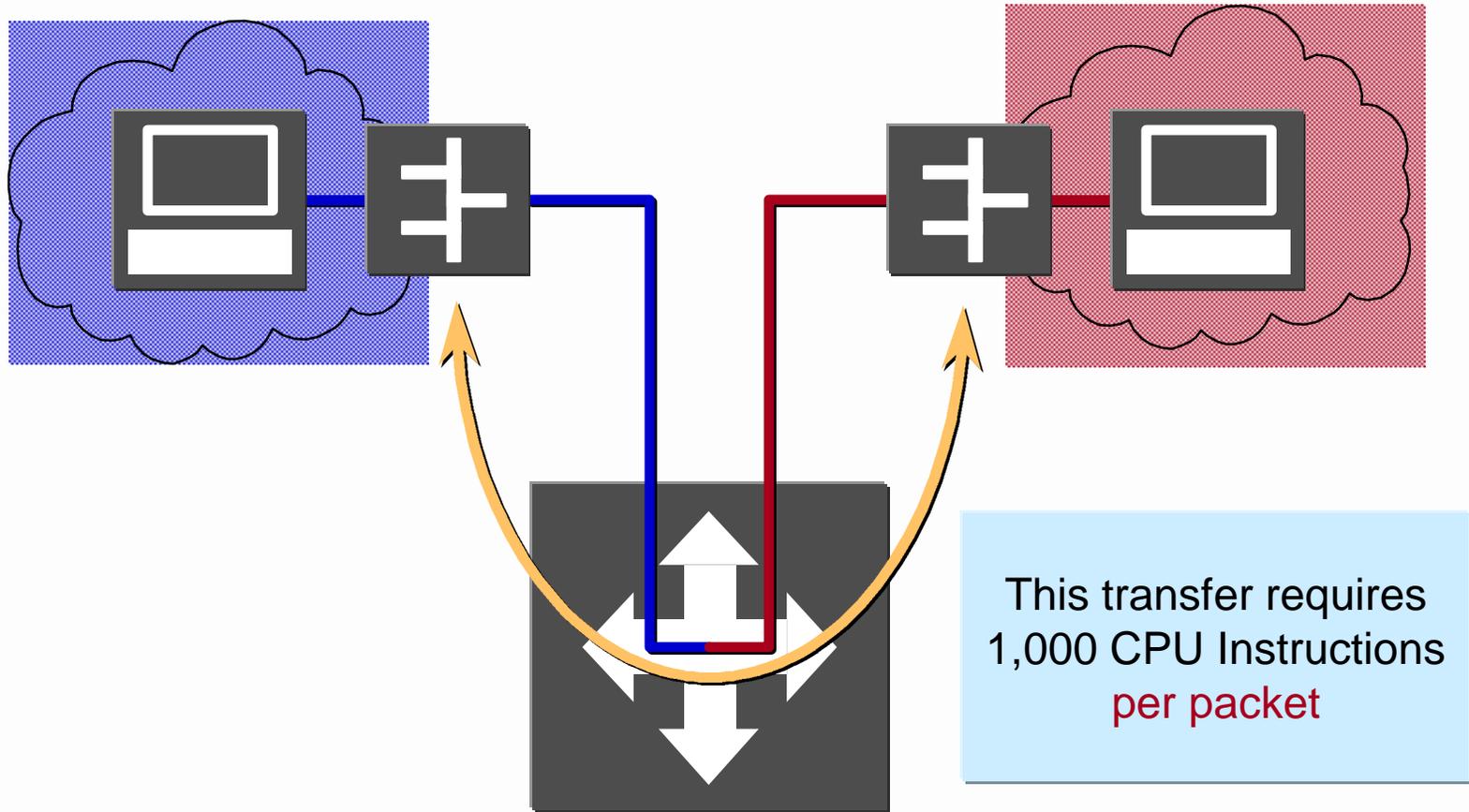


Note:

- 1: All packets pass through the router
- 2: Router retains separate segment identities using ATM VCs over this link

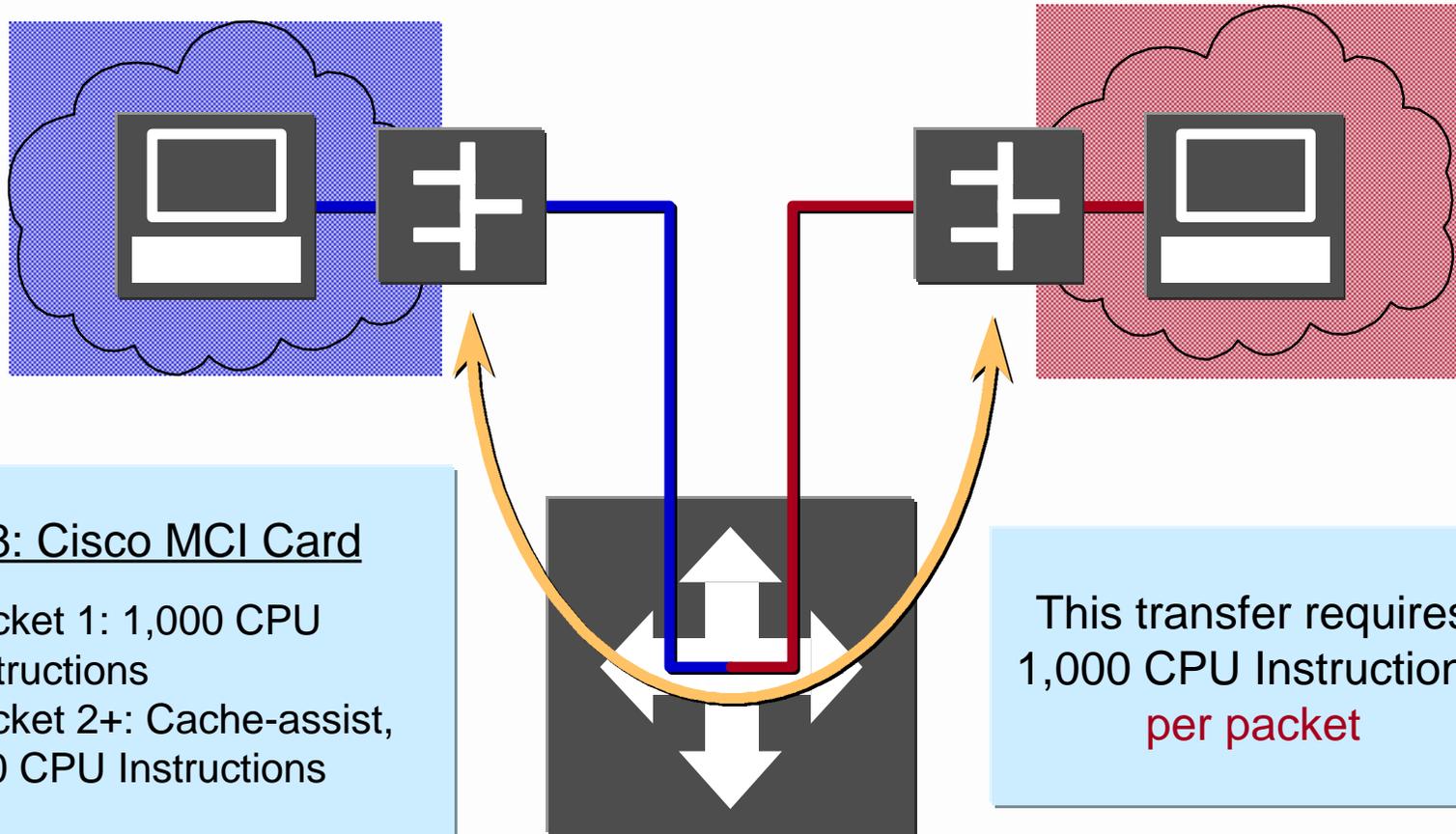
One-armed routing is the primary mechanism to interconnect LANE Emulated LANs on an ATM backbone today. As a classical technique, all packets pass through the router to get between subnets. The router maintains separate identities for the ELANs using ATM VCs over the physical connection.

Classical Routing: Cache-Assisted



As I mentioned earlier, a full routing decision requires around 1,000 machine instructions.

Classical Routing: Cache-Assisted



With the drop in RAM prices towards the end of the 1980s, it became feasible to provide large address caches in order to abbreviate the routing code path without compromising on the quality of the routing decision. This device fully routes the first packet, but then cached the source/destination IP addresses, allowing subsequent packets to be forwarded through a drastically shorter code path (around 100 instructions in the later releases). All modern routers make use of similar cached techniques, and the concept of a “fast path” and “slow path” through a router is now commonplace.

Caching and Performance Modes

- **Use of single-pass traffic filters (eg. Source/Destination Address Pairs)**
- **Use of regressive filter rules (logical AND, OR, XOR, NOT operations)**
- **Counting Operations (diagnostics, traffic management)**
- **Priority Queuing Mechanisms**
- **Additional Functions**
 - **Data Link Switching**
 - **NetBIOS Name Caching**
 - **Timer spoofing**

While caching works great for the majority of traffic, there are some issues with the technique.

Caching and Performance Modes

- **Use of single-pass traffic filters (eg. Source/Destination Address Pairs)**
- Use of regressive filter rules (logical AND, OR, XOR, NOT operations)
- Counting Operations (diagnostics, traffic management)
- Priority Queuing Mechanisms
- Additional Functions
 - Data Link Switching
 - NetBIOS Name Caching
 - Timer spoofing

Benchmark tests on routers have, for many years, included the idea of filters. A filter is a test applied to a packet as it passes through the router. The test could be “Does the packet have a source IP address of X?”. The action resulting from the test could be “forward the packet” or “drop the packet”.

Caching and Performance Modes

- Use of single-pass traffic filters (eg. Source/Destination Address Pairs)
- **Use of regressive filter rules (logical AND, OR, XOR, NOT operations)**
- Counting Operations (diagnostics, traffic management)
- Priority Queuing Mechanisms
- Additional Functions
 - Data Link Switching
 - NetBIOS Name Caching
 - Timer spoofing

Filters can be applied multiple times. Depending on the internal architecture of the router, these regressive tests may result in the packet passing through the filter routing many times.

Such filters are essential in a typical network installation.

Caching and Performance Modes

- Use of single-pass traffic filters (eg. Source/Destination Address Pairs)
- Use of regressive filter rules (logical AND, OR, XOR, NOT operations)
- **Counting Operations (diagnostics, traffic management)**
- Priority Queuing Mechanisms
- Additional Functions
 - Data Link Switching
 - NetBIOS Name Caching
 - Timer spoofing

One of the more demanding operations is for a filter to increment a given counter. This operation is essential in debugging environments, or for installations where traffic management and network accounting is used.

Caching and Performance Modes

- Use of single-pass traffic filters (eg. Source/Destination Address Pairs)
- Use of regressive filter rules (logical AND, OR, XOR, NOT operations)
- Counting Operations (diagnostics, traffic management)
- **Priority Queuing Mechanisms**
- Additional Functions
 - Data Link Switching
 - NetBIOS Name Caching
 - Timer spoofing

A popular endeavour today is to apply a priority mechanism to connectionless routed traffic. These schemes are an extension of a filter mechanism, where the filter rule is used to identify the traffic, and the action is to forward the traffic to a given outbound queue.

Caching and Performance Modes

- Use of single-pass traffic filters (eg. Source/Destination Address Pairs)
- Use of regressive filter rules (logical AND, OR, XOR, NOT operations)
- Counting Operations (diagnostics, traffic management)
- Priority Queuing Mechanisms
- **Additional Functions**
 - **Data Link Switching**
 - **NetBIOS Name Caching**
 - **Timer spoofing**

Modern routers are also required to perform a range of duties for which a connectionless device was never designed. A few years ago, Data Link Switching (DLSw, a technique to carry SNA over an IP backbone) was introduced. The initial implementations of DLSw were incredibly slow, and only an average level of improvement was ever achieved.

Caching and spoofing functions are also required in some environments.

Caching and Performance Modes

- **Use of single-pass traffic filters (eg. Source/Destination Address Pairs)**
- **Use of regressive filter rules (logical AND, OR, XOR, NOT operations)**
- **Counting Operations (diagnostics, traffic management)**
- **Priority Queuing Mechanisms**
- **Additional Functions**
 - **Data Link Switching**
 - **NetBIOS Name Caching**
 - **Timer spoofing**

Cache Miss!!!



The bottom line is that modern network requirements have essentially redefined the concept of what a router is. Many of the assumptions made by the simple technique of router acceleration are invalid in real world networks. These invalid assumptions may lead to situations in which routers experience “cache misses”, forcing traffic to move from the fast path to the slow path in the router, and so red-lining the load on the router processors.

What About “Giga-Routers”?

10 Meg Ethernet = 14,000 pps

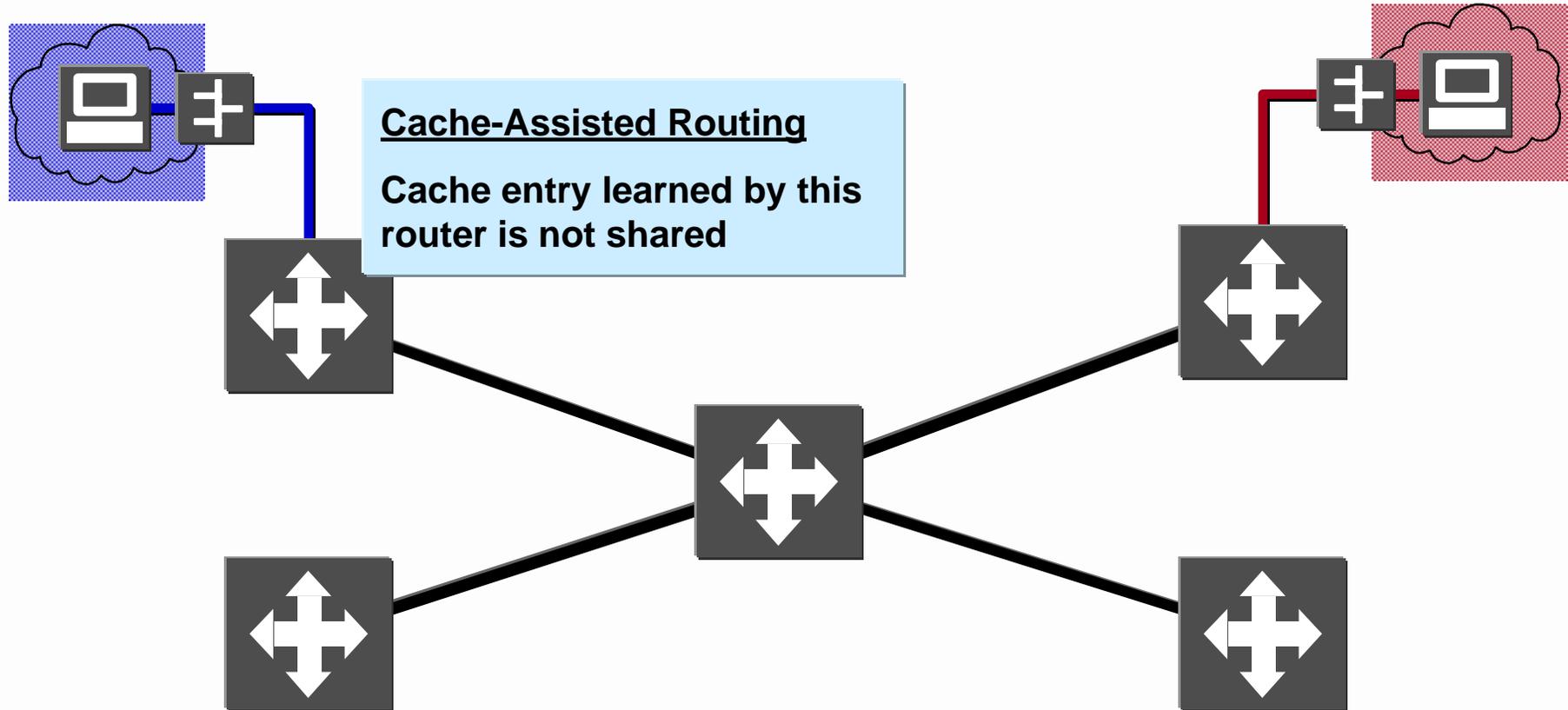
100 Meg Ethernet = 140,000 pps

1 Gig Ethernet = 1,400,000 pps
Full Duplex = 2,800,000 pps!!!

The advent of higher transmission speeds for Ethernet has also pushed the requirement for packet forwarding performance.

A single Gigabit Ethernet connection into a router could pass almost three million packets per second in full duplex mode. Of course, these are 64 byte packets and this kind of traffic pattern doesn't happen in real life. However, manufacturers of “Giga Routers” are claiming “wire speed” for these boxes. How do they do it? Basically these devices are extensions of cache-assisted routing. In this generation of box, more of the forwarding decision can be handed off, and complex ASIC chips do the rest. At the time of writing, not one of these devices has been benchmarked using traffic filters (an integral part of a router configuration), so the real-world performance is still unproven.

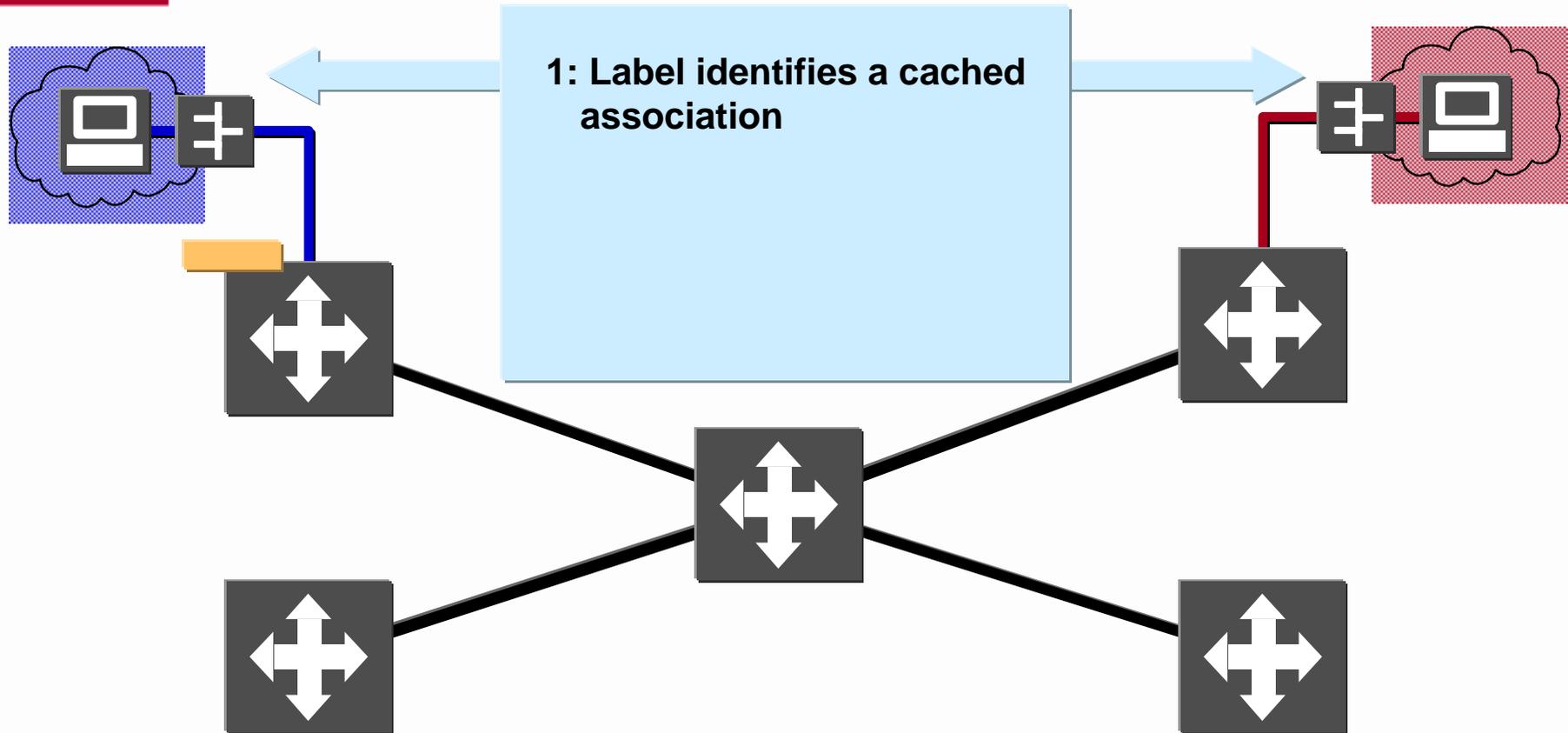
Classical Routing: Cache Distribution



Another disadvantage of caching is that each router must make the “discovery” on its own.

Cache distribution allows other routers to share in this advantage. A cache entry learned by the router in the top left of this backbone...

Cache Label Distribution Protocol

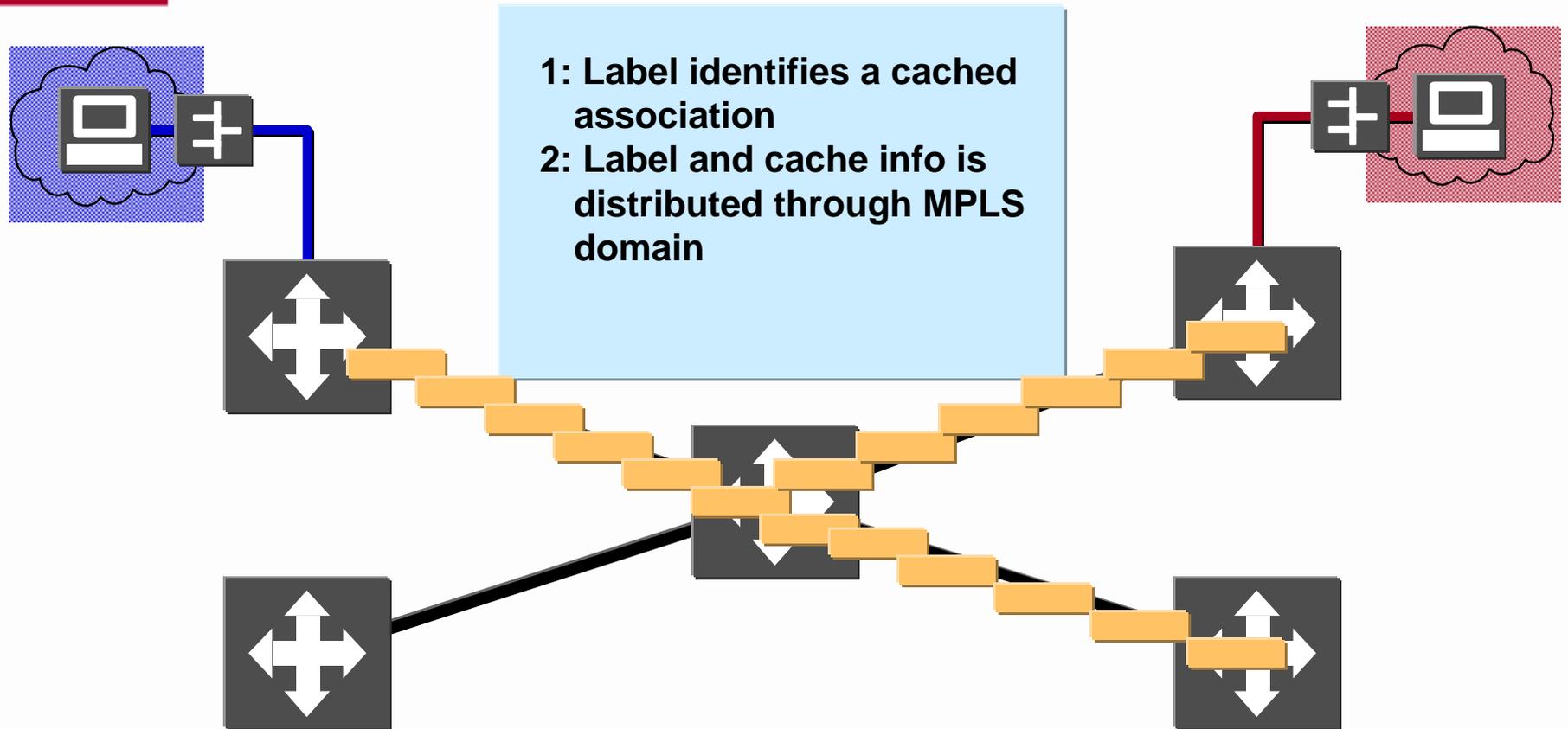


...can be "labelled" to indicate the flow identity, with the associated source and destination identifiers.

Note that these identifiers can be the individual host IDs, but address aggregation can be used to summarise cache entries on the basis of subnet ID, or even using a CIDR address prefix.

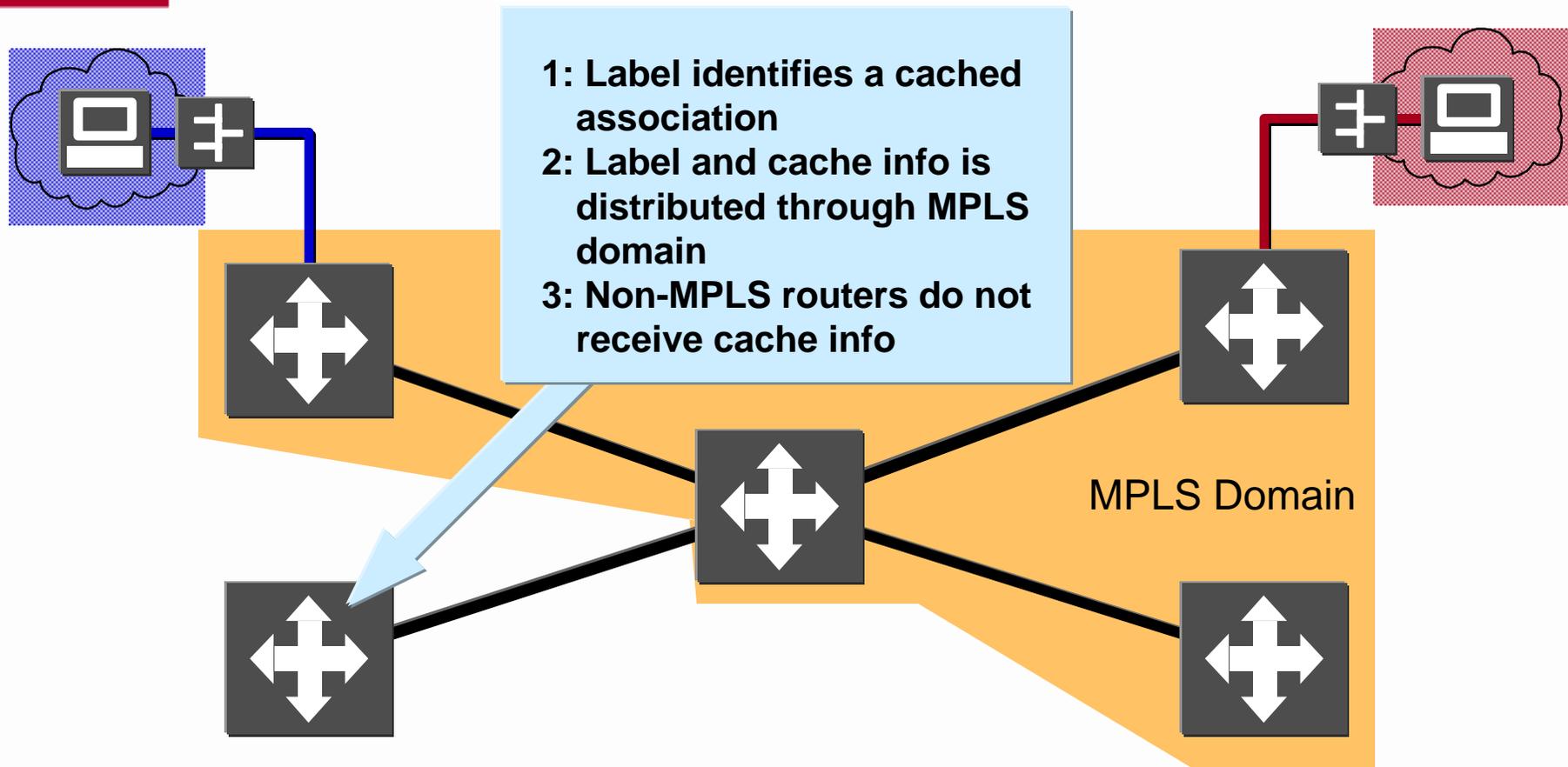
Note that Quality of Service (which is session-based, and therefore associated with a given host pair) is lost as soon as aggregation is enabled.

Cache Label Distribution Protocol



A label distribution protocol must be supported by all of the routers that take part in this process. These routers form a separate, accelerated domain (I've called it the MPLS domain).

Cache Label Distribution Protocol



The shaded area indicates the scope of this domain, with routers that do not recognise the MPLS labels acting outside of the domain. These units either fully route the packets they receive, or create cache entries by their own efforts.

Within the MPLS domain, we can imagine that the network has become "connection-oriented" (although the "C-word" is implicitly banned within the MPLS working group). This is necessary to achieve the goals of MPLS...

- **Allow routing tables to scale**
- **Enable differentiated services**
- **Accelerate routing performance**
- **Offer improved integration of Frame and Cell technologies**

The MPLS Working Group of the IETF has four primary goals.

- **Allow routing tables to scale**
- Enable differentiated services
- Accelerate routing performance
- Offer improved integration of Frame and Cell technologies

The first is to allow routing to scale to support the millions of hosts, and hundreds of thousands of routes required in today's Internet.

I've covered some of the reasons behind this in the tutorial on "Why Do We Need a New IP?".

- Allow routing tables to scale
- **Enable differentiated services**
- Accelerate routing performance
- Offer improved integration of Frame and Cell technologies

The next is to allow service differentiation across the Internet. The reason for this is very simple. There isn't a single ISP in the world today that's actually operating at a profit, and the only direction for margins right now is down. The trend is for bandwidth providers (the old-style Telcos) to acquire service providers, and so gain the ability to discount bandwidth further.

If there were a mechanism to differentiate service (currently the Internet offers only "best effort"), then this would be an opportunity to increase margins on these premium services (such as high bandwidth corporate Extranets, or Voice Over IP connections).

MPLS Working Group Goals...

- Allow routing tables to scale
- Enable differentiated services
- **Accelerate routing performance**
- Offer improved integration of Frame and Cell technologies

Of course we need routing performance to scale, as I've discussed already.

- Allow routing tables to scale
- Enable differentiated services
- Accelerate routing performance
- **Offer improved integration of Frame and Cell technologies**

Finally MPLS is the framework that allows a better integration of frame and cell technologies.

If IP continues to be the delivery protocol to the application, but ATM is the predominant WAN and LAN Backbone technology, then the two worlds will need to be tightly coupled.

Short-Cut Routing (eg. MPOA)

- **MUST** be implemented over a connection-oriented infrastructure (eg. ATM)
- Components are “Virtual Routers”, made up of **MPOA Client**, **MPOA Server** and ATM backbone
- Initial packets are **fully routed** (usually with cache-assistance)
- If sufficient packets in flow, then **short-cut** is established

The right hand side of my evolutionary tree is the idea of short cut routing. After the temporary distraction of IP Switching, it's now clear that MPOA is the only multi-vendor, standards-based technique for short-cut routing, so I'll focus on the specifics of this technology.

Short-Cut Routing (eg. MPOA)

- **MUST be implemented over a connection-oriented infrastructure (eg. ATM)**
- Components are “Virtual Routers”, made up of MPOA Client, MPOA Server and ATM backbone
- Initial packets are fully routed (usually with cache-assistance)
- If sufficient packets in flow, then short-cut is established

MPOA *must* be implemented over a connection-oriented infrastructure. This is its primary strength because it can avoid the issues of cache size, labelling formats and label distribution by using the *existing* virtual connection identifiers of ATM.

In addition, by *decoupling* the notion of network load from the CPU load of routing engines, it's possible to scale up the applied load faster than the rate of advance of CPU technology.

Short-Cut Routing (eg. MPOA)

- MUST be implemented over a connection-oriented infrastructure (eg. ATM)
- **Components are “Virtual Routers”, made up of MPOA Client, MPOA Server and ATM backbone**
- Initial packets are fully routed (usually with cache-assistance)
- If sufficient packets in flow, then short-cut is established

The components of an MPOA system are Virtual Routers. These are made up of MPOA Clients (MPC) devices at the edges of the network, and MPOA Servers (MPS) acting as the routing engines, all connected over a high performance ATM backbone.

Short-Cut Routing (eg. MPOA)

- MUST be implemented over a connection-oriented infrastructure (eg. ATM)
- Components are “Virtual Routers”, made up of MPOA Client, MPOA Server and ATM backbone
- **Initial packets are fully routed (usually with cache-assistance)**
- If sufficient packets in flow, then short-cut is established

The initial inter-segment packets are fully routed, with the usual local cache assistance.

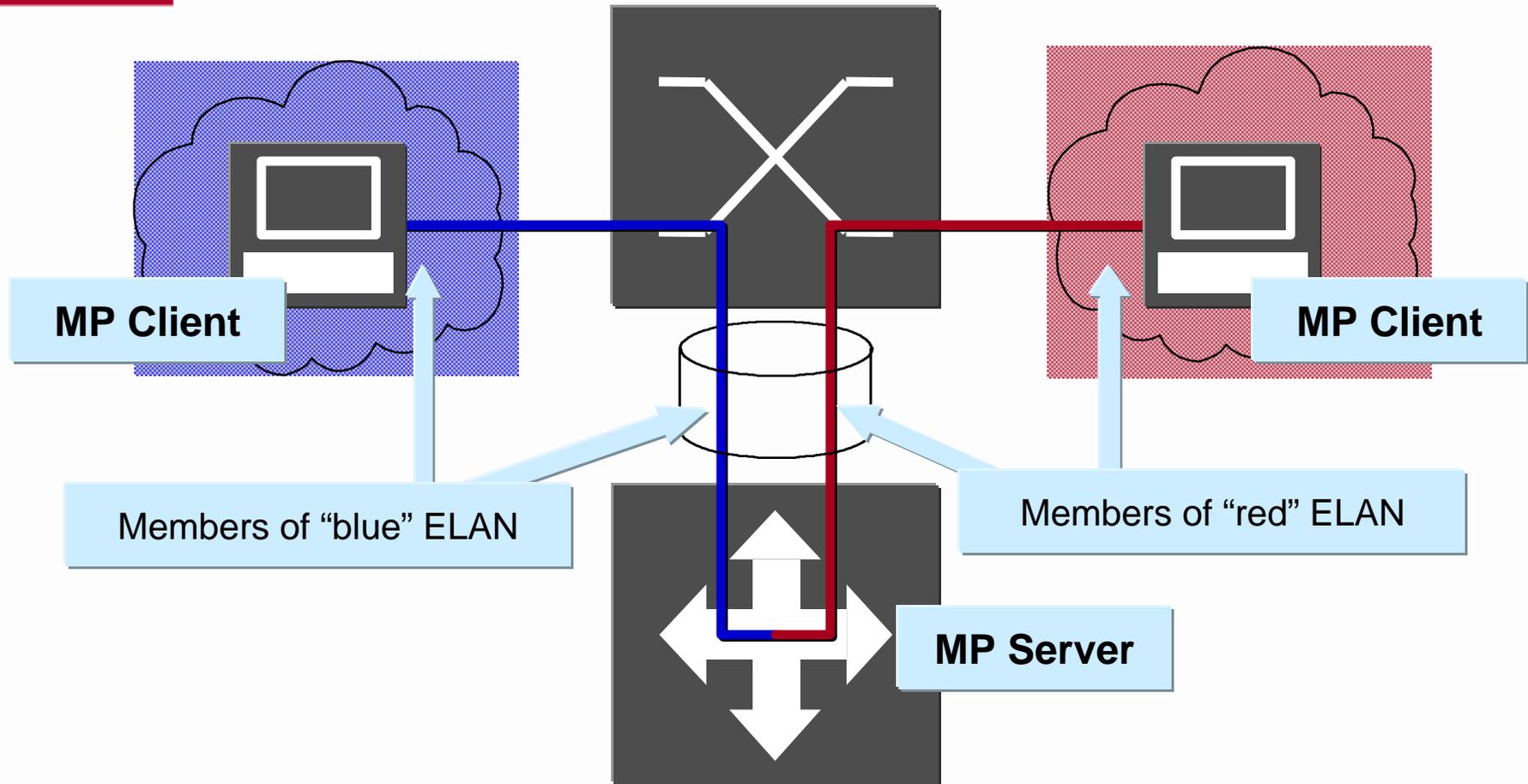
This is a vital first step because it is essential that the quality of the routing decision is maintained as we move towards a short cut technique.

Short-Cut Routing (eg. MPOA)

- MUST be implemented over a connection-oriented infrastructure (eg. ATM)
- Components are “Virtual Routers”, made up of MPOA Client, MPOA Server and ATM backbone
- Initial packets are fully routed (usually with cache-assistance)
- **If sufficient packets in flow, then short-cut is established**

If enough packets are passed between a given source and destination identity, then a short cut can be established.
Here's how it works...

MPOA: Components

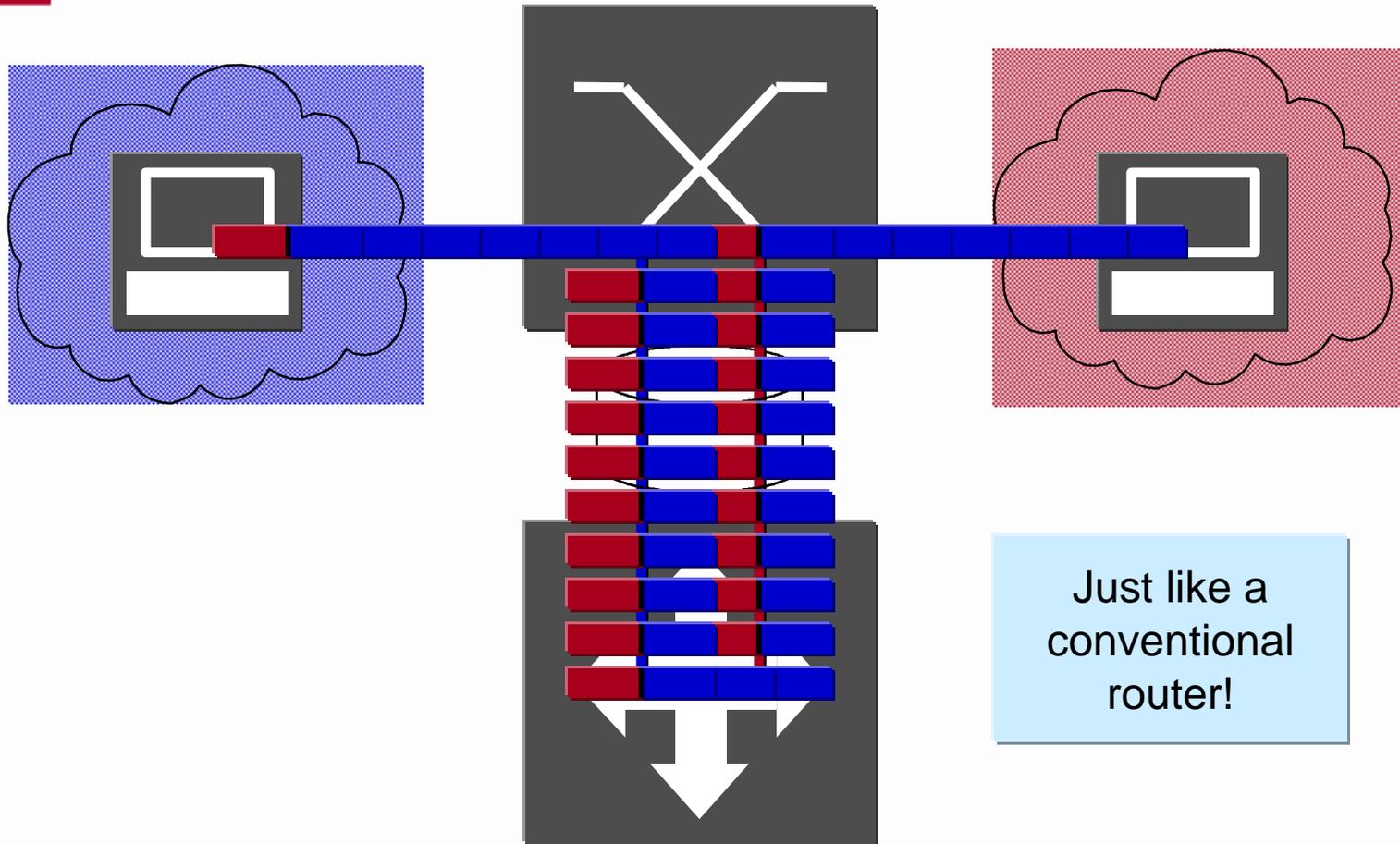


The components are laid out like this.

The diagram looks much the same as a conventional one-armed router network, as it should since MPOA is a logical migration from this form of interconnection.

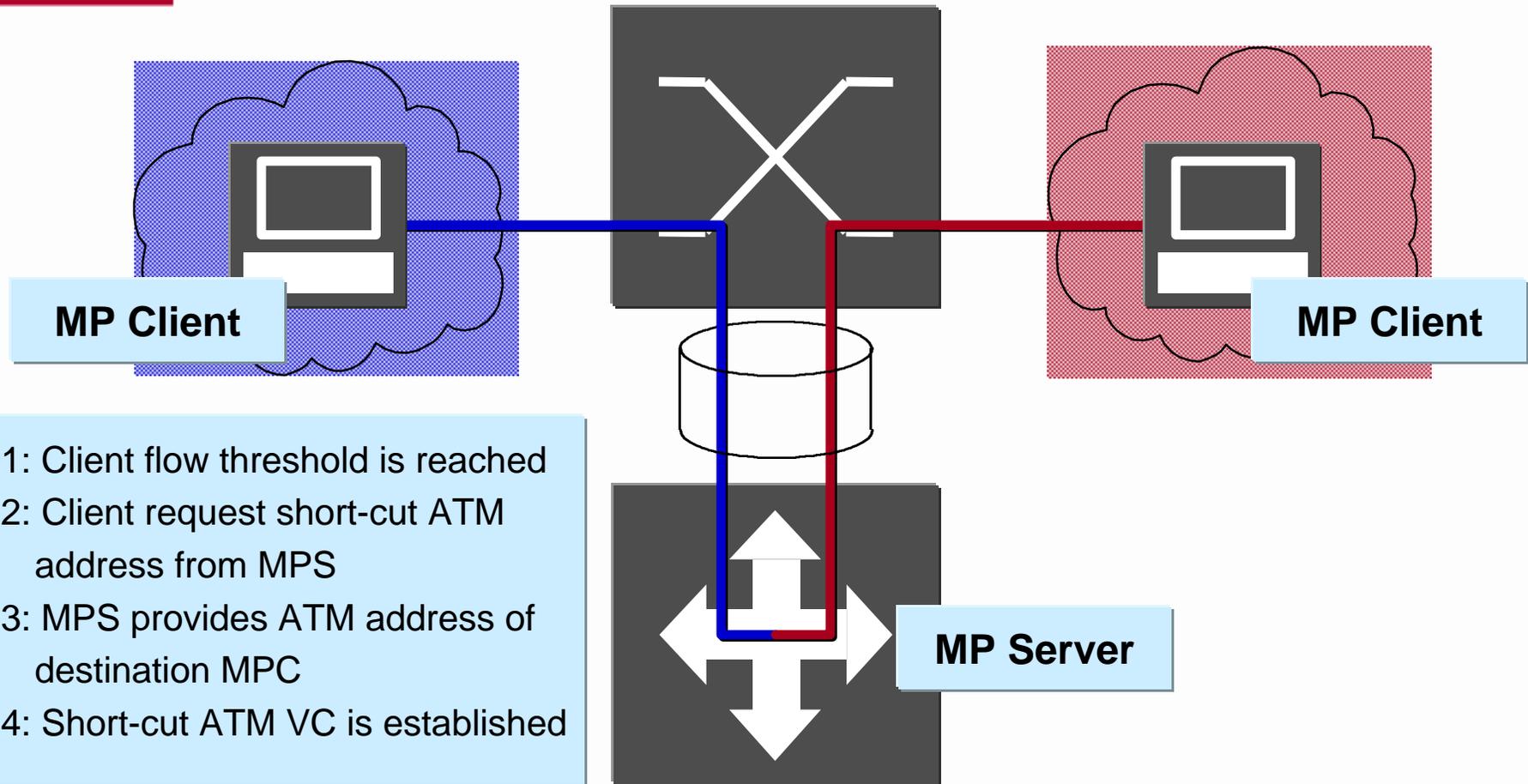
Note that the logical ATM Virtual Connections on the ATM link are associated with the Red and Blue ELANs, and use conventional LAN Emulation to communicate with the members of this ELAN.

MPOA: Before Short-Cut



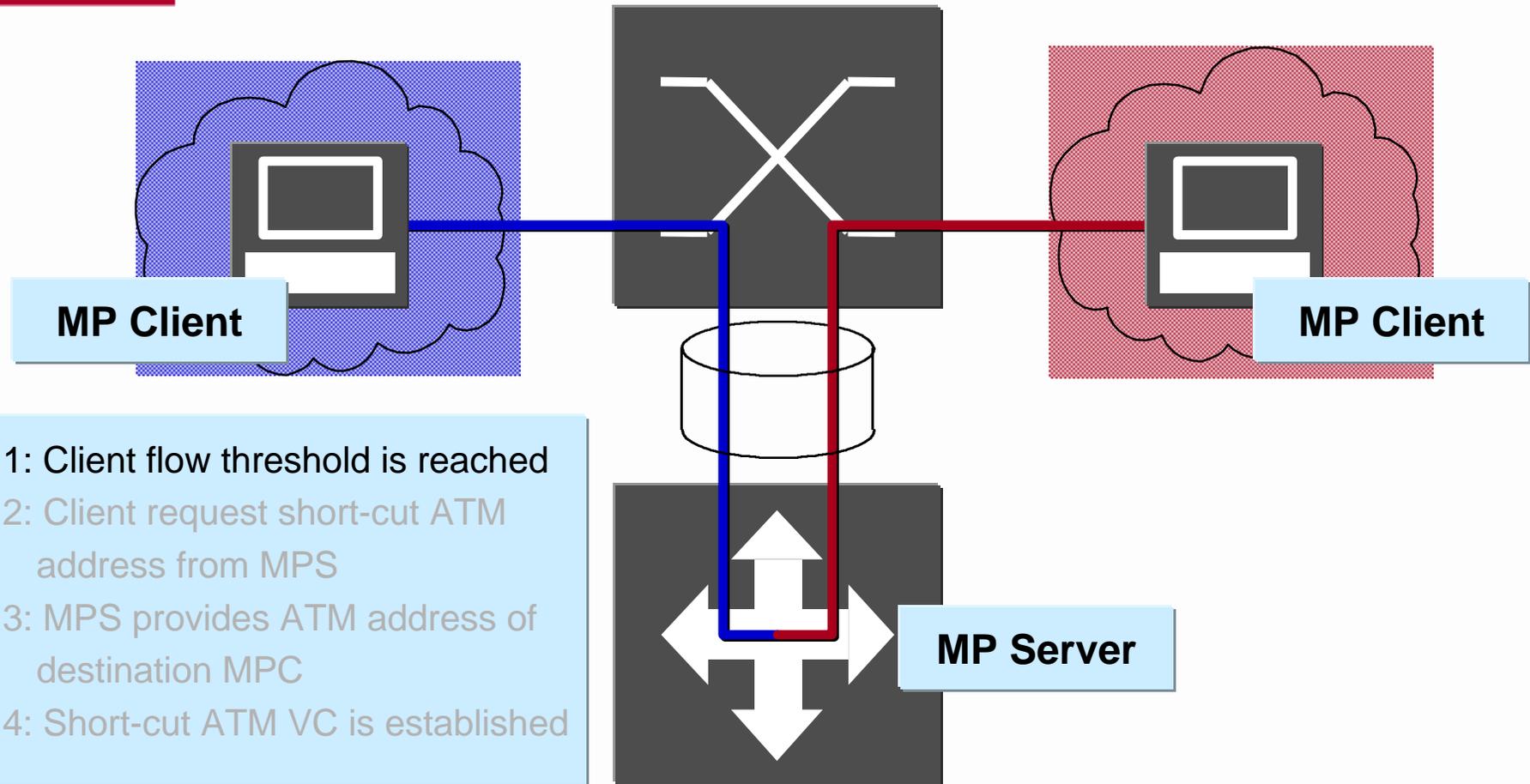
Before the short cut is established, MPOA looks *exactly* like a one-armed router.

MPOA: Sequence of Operation



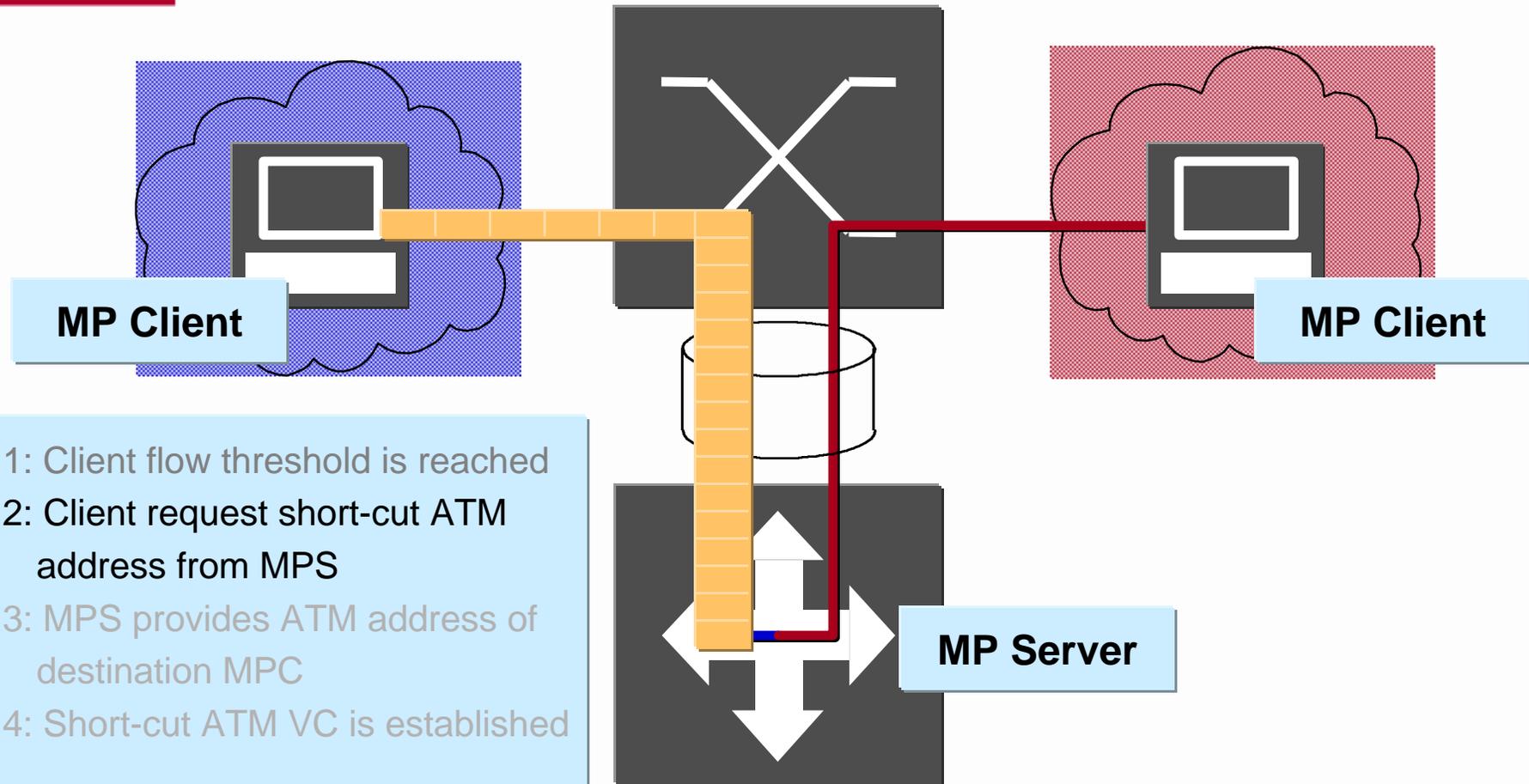
As the communication flow continues, a specific sequence of events that I've outlined here will take place.

MPOA: Flow Threshold Reached



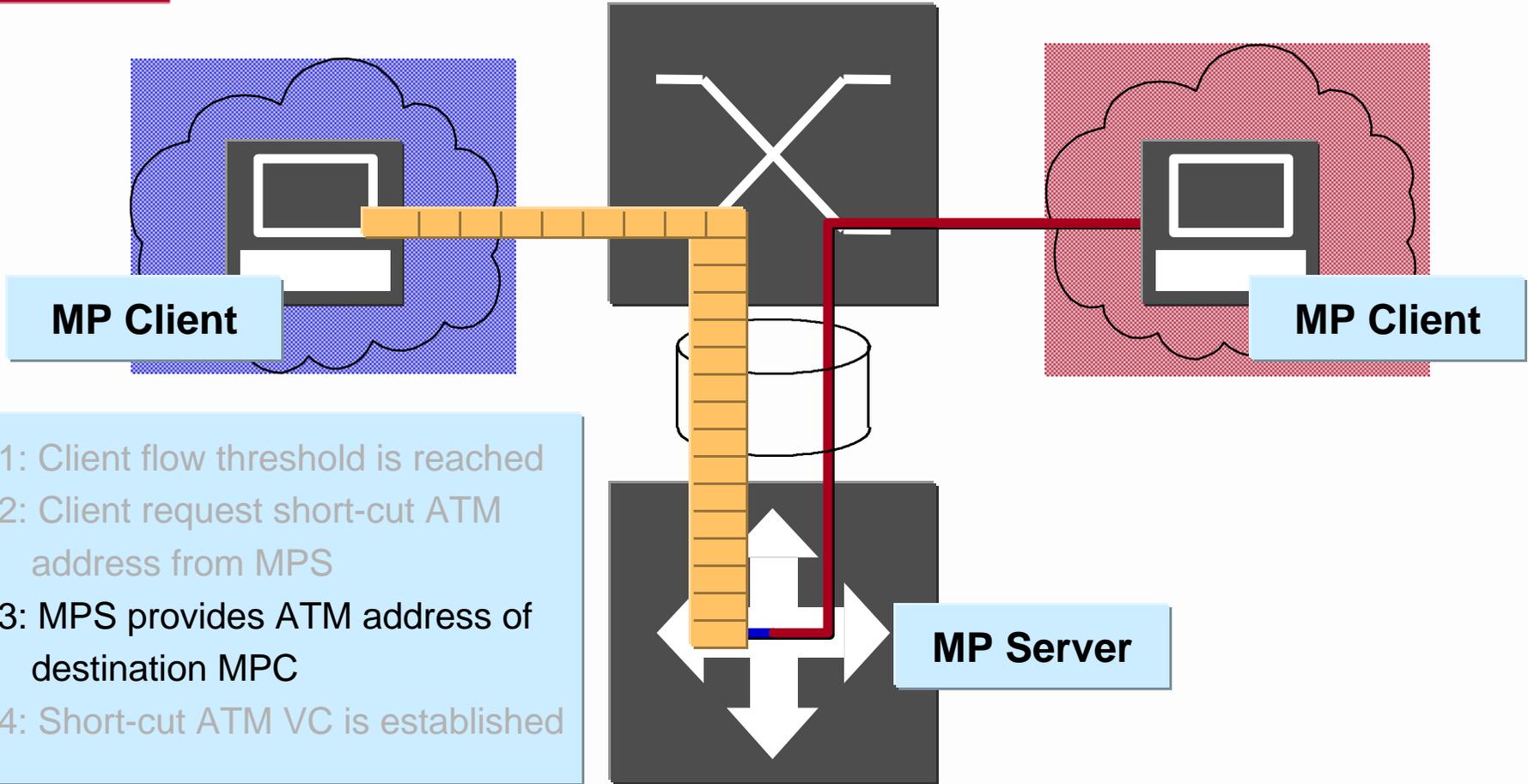
First, one of the MPC devices will “notice” that it has sent a sequence of packets to the same destination IP address. Inside the MPOA software for the MPC there will be a threshold (possible units for this threshold include number of packets, packets per second, etc.).

MPOA: MPC Request



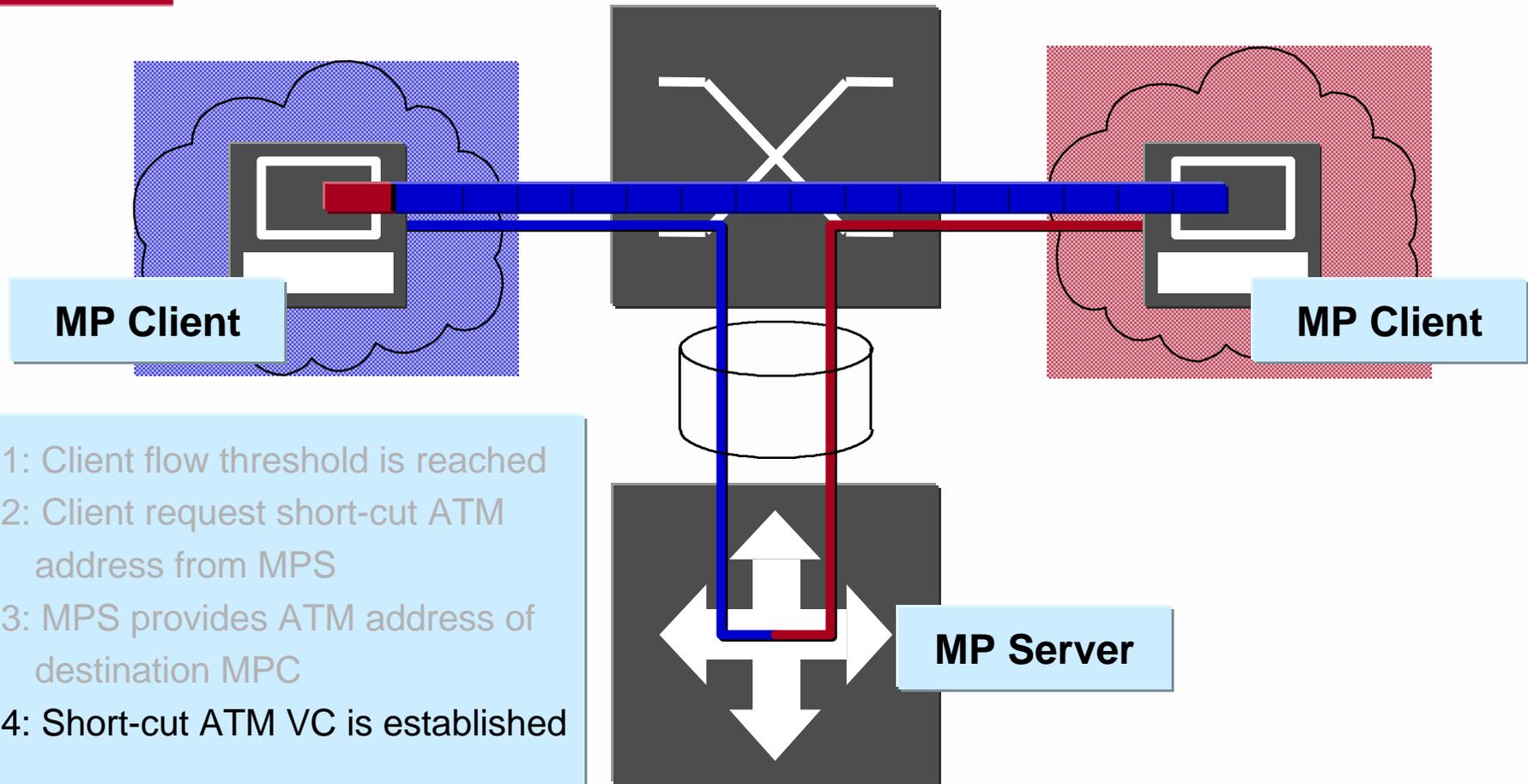
To make the short-cut, the MPC in the Blue ELAN needs the ATM address of the MPC in the Red ELAN. To get this address the MPC will ask the MPS.

MPOA: MPS Responds



The MPS receives this request and will check that a short-cut is allowed between these end points. If all is well, the MPS will respond with the ATM address.

MPOA: Short-Cut VCC Established



The Blue MPC can now make an ATM connection directly to the Red MPC and begin to transfer traffic over this connection.

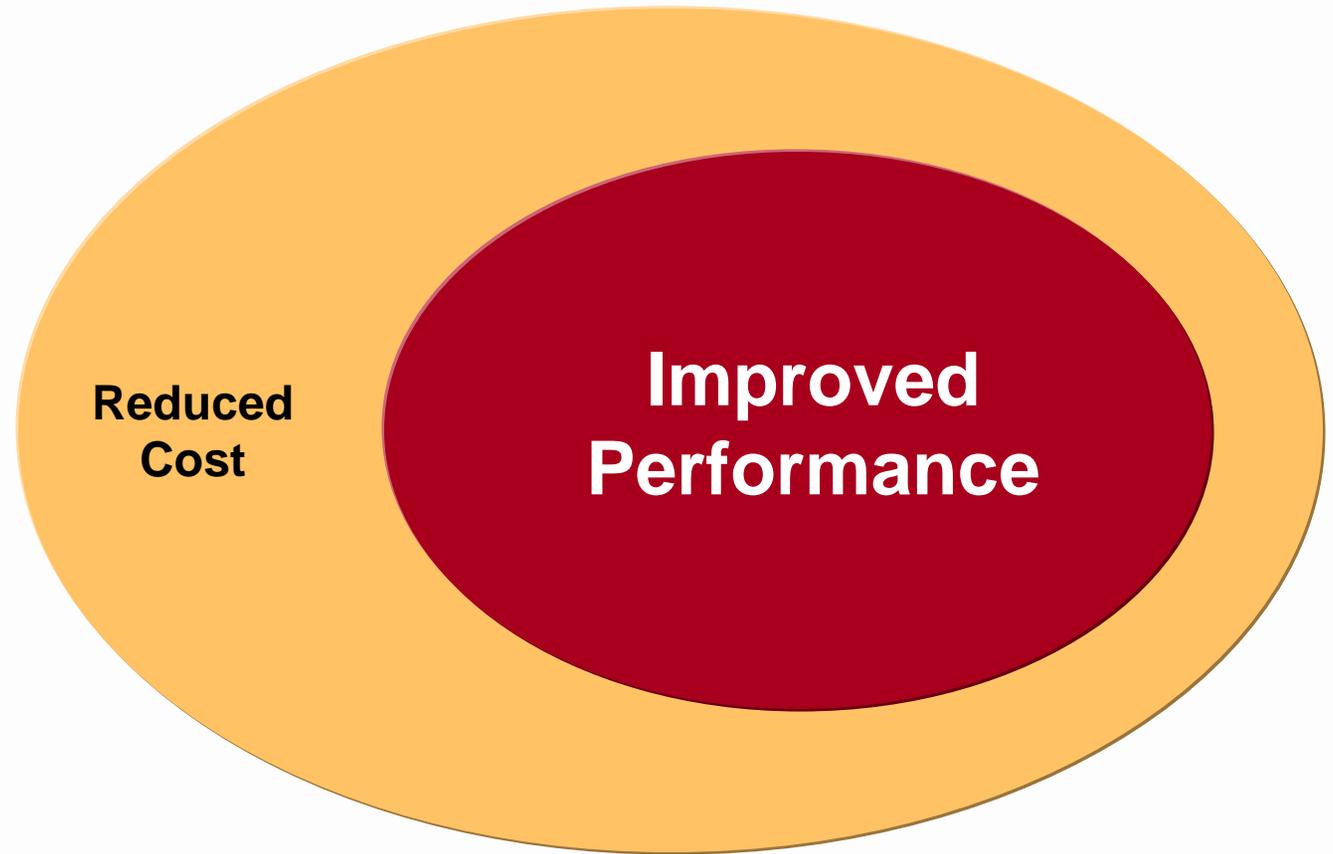
Note 1: While the short cut requests are in progress, traffic continues to flow through the “one-armed” router.

Note 2: The short-cut VCC will be held up until a certain inactivity period has elapsed. This period can be set up by the Network Administrator.

Note 3: The Red MPC ATM address will be cached by the Blue MPC, and the Blue MPC ATM address by the Red MPC.



The Result of MPOA...



The result of migration to MPOA is dramatic. At the edge of the network we no longer need super-computing routers to keep pace with our traffic levels and edge population. This dramatically reduces the cost of the edge equipment, without sacrificing our ability to implement hierarchy or scalability.

At the core of the network performance is improved to ATM switching levels.

- **Security Drawbacks**
 - **Full Network Layer and Transport Layer Filtering Possible**
 - **But once short-cut VCC is established, no more filtering**
- **Scalability Issues**
 - **One VCC per IP Source/Destination Pair**
 - **Will this scale to the WAN, ISP and Internet?**

It's particularly important to position MPOA v1 correctly. MPOA is designed to enable short cut routing *within an Enterprise backbone*.

Initial trials of MPOA have criticised its potential security drawbacks and its scalability limitations. Let's look at these issues separately.

- **Security Drawbacks**
 - **Full Network Layer and Transport Layer Filtering Possible**
 - **But once short-cut VCC is established, no more filtering**
- **Scalability Issues**
 - **One VCC per IP Source/Destination Pair**
 - **Will this scale to the WAN, ISP and Internet?**

The Multiprotocol Server in an MPOA environment operates as a true Network Layer Router prior to the short cut being established. Thus it's possible to apply the exact same filter mechanisms to the short-cut determination as would be the case with a conventional router. However, once the short cut has been set up using one set of criteria (eg. "Between a given source/destination pair for FTP only"), there's nothing to stop this VCC being used for another application (eg. Telnet). However, the same source and destination end points would be enforced.

- **Security Drawbacks**
 - **Full Network Layer and Transport Layer Filtering Possible**
 - **But once short-cut VCC is established, no more filtering**
- **Scalability Issues**
 - **One VCC per IP Source/Destination Pair**
 - **Will this scale to the WAN, ISP and Internet?**

Is this a problem? Very unlikely. Since MPOA is intended for use within an Enterprise, then it's unlikely that you'd want to apply such complex firewalling rules. If you did, then it would most likely be to a specific small part of the network (eg. The Accounts Dept.), and the simplest security policy would be to disallow any shortcuts *into* the Accounts Dept. (the folks inside Accounting could still activate shortcuts outbound, but this could be restricted to a well-known set of server addresses).

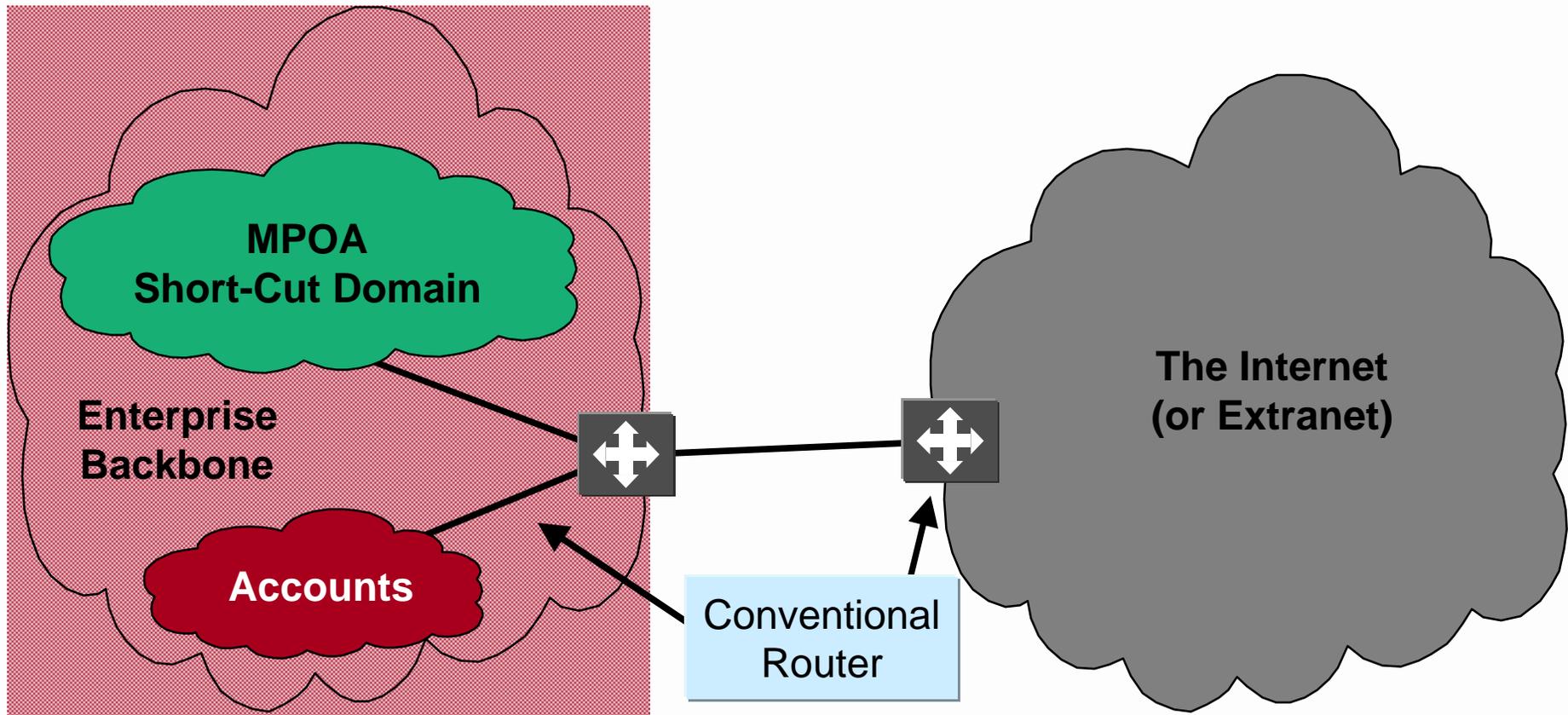
Administering such a policy is a breeze with MPOA since there is a centralised Policy Server (currently the LECS) in the network. Ultimately this Policy Server will be a Directory Server. In connectionless routed networks we'll have to wait until Directory-Enabled networking protocols become finalised before central policy management is possible.

- **Security Drawbacks**
 - Full Network Layer and Transport Layer Filtering Possible
 - But once short-cut VCC is established, no more filtering
- **Scalability Issues**
 - **One VCC per IP Source/Destination Pair**
 - **Will this scale to the WAN, ISP and Internet?**

Another misleading criticism of MPOA v1 has been its potential scalability problems if it's deployed beyond the Enterprise.

MPOA v1 sets up a discrete short cut VCC for every qualified IP Source/Destination pair. For a larger network this could lead to a concept called "VC Starvation", in which the ATM switches that make up the core of the network actually run out of resources to set up new VCs. For an Enterprise network it's unlikely that this will ever happen, even with straightforward MPOA v1. MPOA simply was not designed for deployment outside the Enterprise in its current form, and so the next version of MPOA (and the collaboration work with the MPLS Working Group) focuses on this scalability issue.

MPOA v1: Position It Correctly



To summarise these issues...MPOA short cuts offer significant advantages within the Enterprise Backbone. If the two concerns I've listed here become an issue, the short term solution (and the natural design evolution of the network) is to limit the scope of the short-cut domains using conventional routers.

The longer term solution can be thought of as "MPOA v2", or more likely MPLS, which will be a natural complement to MPOA. It's unlikely that such an MPLS solution (ie. one that includes connectionless backbones) will be commercially available before 2000.

- **Routing is a wonderful thing**
- A router is an expensive and low-performance way to do it
- Short-Cut Routing is currently the most advanced option to:
 - Reduce Cost of Routing
 - Improve Performance of Routing
 - Create a true Virtual Infrastructure
- MPOA and MPLS will be complementary routing technologies

So in summary, we know that routing is a wonderful thing.

- Routing is a wonderful thing
- **A router is an expensive and low-performance way to do it**
- Short-Cut Routing is currently the most advanced option to:
 - Reduce Cost of Routing
 - Improve Performance of Routing
 - Create a true Virtual Infrastructure
- MPOA and MPLS will be complementary routing technologies

However, a conventional router is an expensive, performance-limited way to do it. Stand-alone routers still have a place in remote access and low demand internetworks.

- Routing is a wonderful thing
- A router is an expensive and low-performance way to do it
- **Short-Cut Routing is currently the most advanced option to:**
 - **Reduce Cost of Routing**
 - **Improve Performance of Routing**
 - **Create a true Virtual Infrastructure**
- MPOA and MPLS will be complementary routing technologies

However, short-cut routing is the state-of-the-art for high demand backbones, and is specifically focused at reducing the cost of routing, maintaining performance, and allowing the creation of a truly virtual infrastructure.

- Routing is a wonderful thing
- A router is an expensive and low-performance way to do it
- Short-Cut Routing is currently the most advanced option to:
 - Reduce Cost of Routing
 - Improve Performance of Routing
 - Create a true Virtual Infrastructure
- **MPOA and MPLS will be complementary routing technologies**

In the near future we'll see MPOA and MPLS essentially converging to become complementary (or even identical) technologies within the extended Enterprise internetwork.



The End

This concludes the tutorial.

If you aren't viewing this tutorial on the FORE Systems' ATM Academy Site, then you can find additional tutorials at:

<http://academy.fore.com/>