MADGE

SWITCHING TO ATM

Token Ring Switch Evaluation Guide

Authors, David Messina and Simon C Gawne

November 1995

Inside Front Cover Blank

Introduction

The emergence of Token Ring switching as a local area network (LAN) internetworking technology is causing organizations to re-examine the structure and design of their Token Ring networks. Although a number of vendors now offer Token Ring switches, the functionality and performance provided varies considerably. Companies opting for a switch-based network architecture must be confident that the chosen switching platforms serve the network users' current needs, while providing a migration path to other high speed technologies to address future demands.

This guide is intended for use by LAN administrators and other technical personnel who are currently planning to implement Token Ring switches in their networks and require guidelines on how to evaluate switches from different vendors. Initially, the guide investigates the distinct application needs that are driving the rapid adoption of Token Ring switches. Subsequently, the guide presents a set of key features and the functionality that a Token Ring switch must have to satisfy the application for which it was intended.

The guide then discusses the diversity of architectures available today in switches and how they relate back to the required features for the given Token Ring switch application. Ultimately, the guide focuses on switch performance, which is the most critical element of the evaluation process. The guide explains what is meant by switch performance and suggests ways of measuring the different performance parameters.

This guide will provide organizations with the information necessary for selecting the right switch for their specific needs.

Blank

Overview of Token Ring Switching

Token Ring switches are LAN internetworking devices which are used to interconnect multiple ring segments. The majority of switches are designed to offer high frame forwarding rates, thus enabling large amounts of data to be transferred between rings at a very rapid pace. Token Ring switches as layer 2 devices use the low-level information found in the frame header to determine the destination segment of each frame. Consequently, switches do not have the extensive processing overhead of layer 3 routers and can handle all types of data traffic including routable protocols such as TCP/IP and IPX, as well as non-routable protocols such as IBM NetBIOS and SNA. This simplicity means that switches are significantly lower in cost than equivalent routers.

Applications

Currently, two distinct types of Token Ring switches are starting to appear in the market: backbone switches and workgroup switches. Although these two breeds of internetworking devices may appear to be very similar, the applications they are designed for and the level of performance offered contrast considerably. The following two sections examine in detail their uses and the requirements that allow them to function in their defined roles.

Backbone Switches

Evolution from Distributed Backbones

Traditional Token Ring networks are based on a *distributed backbone* design in which multiple workgroup segments are interconnected via a shared 16Mbps Token Ring backbone. The shared backbone often runs between floors in a building and is frequently called a *vertical riser*. Each workgroup segment is linked to the backbone via a two port Token Ring source routing bridge, such as the Madge Smart Ringbridge or the IBM PC-based Bridge Program (Figure 1).

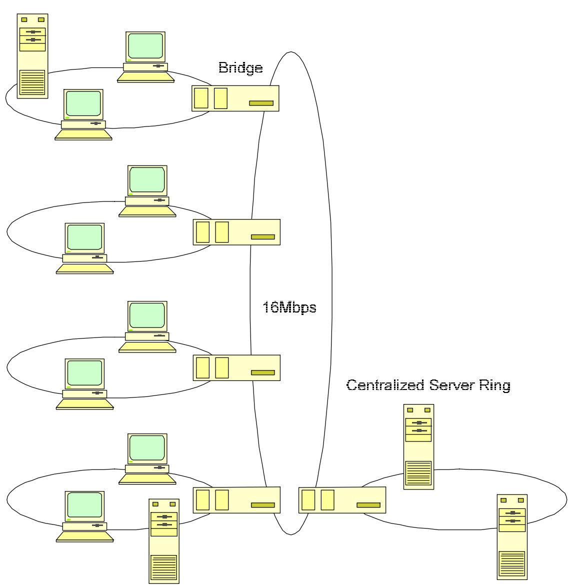

Figure 1: Distributed Backbone

Although this design has served the majority of Token Ring sites well for many years, it is increasingly becoming overstretched by the ever-growing presence of client/server applications and the resultant increase in network traffic. The 16Mbps backbone is an obvious bottleneck in the network, because it carries data between numerous workgroup rings which often are running at 16Mbps. In addition, the bridges used to connect the workgroup rings to the backbone have performance limitations primarily due to high latency, which impedes the flow of traffic from the workgroup rings to the backbone. Furthermore, it is difficult to manage and control the traffic flow, because the bridges are dispersed throughout the network and thus all over a building or campus.

To overcome some of these problems, organizations have been collapsing the LAN backbone into a central networking device such as a high performance layer 3 router.

This *collapsed backbone* architecture (Figure 2) has advantages over the distributed backbone architecture described above: for example, the backbone bottleneck is removed through the high internal throughput provided by the router. Additionally, traffic management and control is much easier to implement since all inter-ring traffic traverses a single device on the network.

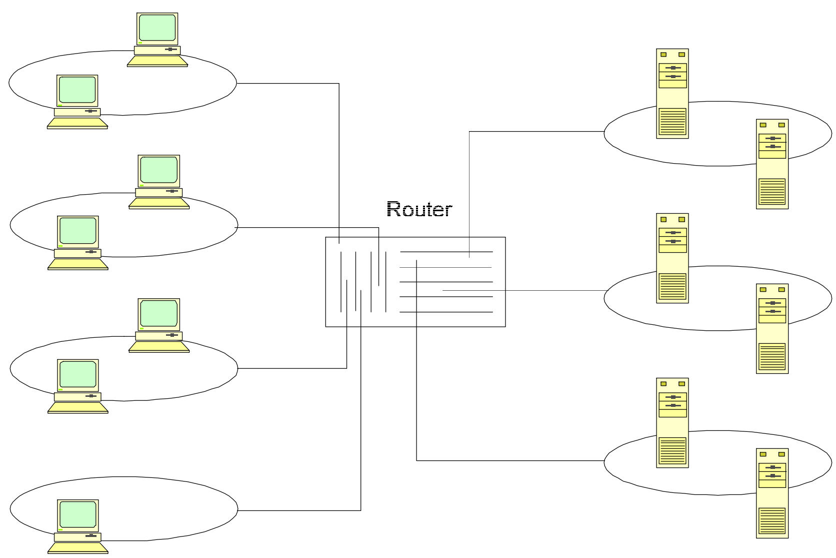

Figure 2: Collapsed Backbone

Unfortunately, this router-based network design still has some serious limitations. Routers typically exhibit even higher latency than bridges, and this can seriously impact the throughput that can be achieved between clients and servers. In addition, many Token Ring protocols, such as NetBIOS and SNA, are non-routable, which means they cannot be transported by a layer 3 device (to overcome this problem, many routers also offer layer 2 bridging). Finally, the large amounts of processing which must be performed in a router means that the cost of the solution can be extremely expensive.

Switched Collapsed Backbone

Collapsed backbone Token Ring switches perform a similar function to collapsed backbone routers for local area internetworking, except they operate at layer 2 rather than layer 3. This means Token Ring switches are able to carry all types of network traffic: routable protocols, such as TCP/IP and IPX, as well as non-routable protocols. The design of these switches is considerably less complex than that of routers, which allows the cost of a collapsed backbone switch to be less than half the cost of an equivalent router solution The right switch will also offer considerable performance benefits over the collapsed backbone router solution.

Characteristics of a Collapsed Backbone Switch

The backbone is the core of any Token Ring LAN because it represents the point at which all data communication converges. Workgroup rings must interact with centralized resources in a timely and reliable fashion or else the efficiencies provided by the network are lost. Any collapsed backbone switch design must demonstrate a full appreciation of the critical role that it will be serving within the LAN infrastructure. The following are the mandatory features for a Token Ring switch to competently fill that role:

- **High Performance:** High performance needs to be the cornerstone of a backbone switch, since all inter-ring traffic traverses it. A low performance switch will hamper network activity and degrade network responsiveness. Information on how to measure switch performance can be found further on in this guide.

- **Non-blocking:** Since the primary function of a backbone switch is the interconnection of multiple ring segments, it is important that the switch be able to support multiple data transfers between pairs of rings simultaneously. If a switch is able to handle traffic flows at full wire speed on all ports simultaneously, and if the throughput between any pair of ports is unaffected by the loading on other ports of the switch, then the switch is said to be “non-blocking.” A non-blocking switch is required for backbone applications to guarantee that the network delivers consistent response times under all load conditions.

- **Reliability:** The backbone switch as stated before is at the point of internetwork convergence. For the sake of the entire network, maximum up-time is essential. Ultimate reliability can be achieved with a product that allows for a redundant configuration, which removes the risk of a single point of failure. If a switch incorporates the IBM Spanning Tree algorithm, it will be able to support a back-up path if, for whatever reason, the primary one should fail. That back-up path can be provided either by another switch port or by a pre-existing source route bridge.

- **Centralized Server Support:** One of the objectives of backbone switching is the consolidation of resources in the network center. Servers are one of these key components earmarked for centralization. Centralized servers will most likely experience increased demand, and the switched network design must allow for this.

If a switch permits the direct attachment of nodes, a server can benefit from a dedicated bandwidth of 16Mbps. Direct attachment provides an upgrade path to standards based full duplex Token Ring, which provides the server with an aggregate bandwidth of 32Mbps. Full duplex should be supported on all ports for maximum flexibility and large server farm applications.

- **Traffic Monitoring:** Employment of backbone switching leads to the consolidation of network traffic in one location. The result is a unique opportunity to monitor the state of network traffic on a ring by ring basis. A switch must be integrated into a solid network management platform that has the ability to analyze traffic patterns.

- **Broadcast Control:** Traditional Token Ring networks have had their backbone performance severely hampered by the proliferation of broadcast packets, which are an integral part of the internetwork communication process. Broadcast control refers to any type of mechanism aimed at limiting enterprise-wide distribution of broadcast frames. Such control features attempt to target broadcasts to only concerned parties.

- **Scalability:** As the size of the backbone network grows, a switch must be able to expand along with it. Scalability itself is an obvious pre-requisite, but the options for expansion are just as important. Users differ in their architectural and technological philosophies, which should consequently be taken into account by the switch design. Gradual expansion paths allow for user discretion. Full duplex Token Ring as a first step to inter-connecting switches offers growth at no incremental expense. When higher speeds become necessary, options are an important factor. FDDI provides a solution for organizations that seek a mature technology. ATM will be for those ready to make the move to the next generation of networking technology.

- **Upgradability:** Switches are purchased with longevity in mind and, therefore, assurance that a switch will adapt with the network is fundamental. Vendors vary in their product philosophies along a scale between a hardware and a software focus. Understanding the switch vendors philosophy will provide insight into the staying power of a switch. Companies that subscribe to feature enhancements through software are more likely to insulate products from obsolescence.

- **ATM Connectivity:** Backbone Token Ring switches represent a considerable technology investment. Cost is an aspect of that investment, but so is the strengthened architectural commitment to Token Ring. The availability of an ATM interface allows the Token Ring infrastructure longevity and smooth integration with its natural successor.

- **Source Routing:** For a switch to smoothly integrate into the current network infrastructure source routing should be considered a mandatory feature. If the chosen switch is to replace a distributed bridge architecture, source route support will allow the network configuration to remain the same.

Workgroup Switches

Token Ring is a deterministic topology that allows only one node access to the ring at a given moment in time. This structure, in conjunction with a 16Mbps transmission rate, has provided the workgroup with stable performance without much threat of congestion. Token Ring has been isolated from the congestion problems that have paralyzed Ethernet workgroups.

Initially, network congestion in a Token Ring network is a direct result of the backbone architecture. As previously discussed, the adoption of a switched Token Ring backbone will alleviate that problem. Workgroup switching is a solution that organizations will adopt for Token Ring, but it may be sometime after backbone switching has been implemented. The following is an example of how a workgroup switch would be used.

Consider a 16Mbps workgroup ring comprised of a 100 users all vying for the available bandwidth. The bandwidth per user is 16/100 or 0.16Mbps per user, which in this example is leading to poor workgroup performance. If the ring is segmented into four smaller rings of 25 users each, the bandwidth per user is 16/25 or 0.64Mbps per user —the bandwidth per user has increased by a factor of 4. A workgroup switch would be used to interconnect these newly-created, smaller high performance workgroup segments.

Workgroup switching employed in this manner has brought new life to congested Ethernet workgroups. Although Ethernet's transmission rate is 10Mbps, the effective transmission rate is around 4Mbps. This disparity is due to Ethernet's communication method which allows for multiple stations to send data onto the network at the same time. The shortcoming of this approach is that packets will collide and, after a time out, stations are forced to re-transmit. The more workstations on an Ethernet network the more likely that these collisions will occur.

Switches have been rapidly adopted for Ethernet workgroups because switches provide a means to microsegment the workgroup at an attractive price point. By considerably reducing the number of nodes vying for the network, the rate of collision has been drastically reduced and the effective bandwidth has increased. It should be noted that these new workgroups will most likely require additional hubs to manage them. As a natural progression, some workgroup switch implementations involve direct attachment of individual workgroup nodes, which guarantees that stations have access to their own network at wire speed.

Token Ring workgroups will experience the same changes as Ethernet segments, but the process will be more gradual Improving backbone communication will be of primary importance as that involves any-to-any communication. Workgroup communication is simpler, affecting only a local branch of the network involving a pre-defined flow between multiple clients and a few servers. The distinct applications of switches results in significant differences in required features. The workgroup switch features are detailed below.

Low Cost: The low cost per port of a workgroup switch is due to the simplicity of its application. Workgroup switches are not loaded with features that ensure reliability, guarantee high performance and control congestion. Backbone switches do have these features, which is where the cost disparity lies. Low cost is demanded for workgroup switches, because customers will purchase many more of these than backbone devices.

Manageable Across the Network: Introducing a workgroup switch into a network adds an additional layer of networking equipment to the infrastructure. The network manager should be able to maintain his view of the workgroups even though he is far removed from the local activity.

Transparent Bridge Support: Shrinking the size of Token Ring workgroups can add to some administrative problems if transparent bridge support is not available. Transparent bridging will allow workgroups to be made into smaller sizes, but still appear to the network as if they are on the same ring. No new ring numbers need be assigned. Furthermore, transparent bridging will help to minimize any issues with hop count limitations as the number of logical rings remains constant.

High Speed Local Server Connection: High performance workgroups often involve interaction with local servers that handle the specific requirements of their workgroups. Local server support should be available to help recreate this set-up in a switched environment. The workgroup switch configuration will accelerate traffic to the server and as such the server needs a higher speed connection. This could be full duplex Token Ring or ATM depending on the performance demands at a given point in time.

High Speed Uplink: Workgroup switches rely upon other devices for interconnection. A workgroup switch introduced alone into the infrastructure will not have a dramatic effect on enterprise performance. The switch must be adopted in conjunction with an alteration to the network backbone. The workgroup switch, as an aggregate of small 16Mbps segments desiring communication with centralized server resources, needs a high speed uplink to handle the aggregate traffic. For example, a high speed connection to an ATM backbone would complement a Token Ring workgroup quite well.

Switch Architectures

The performance of a switch is highly dependent on its architectural design. Sometimes the architecture can be the primary differentiator among switches. What follows are discussions of contrasting architectural features. The argument is particularly relevant to backbone switching, where performance is of the utmost importance.

Cut-through and Store-and-Forward Switches

Traditional routers and bridges are *store-and-forward* devices, which means a frame has to be completely received from the network before it can be processed to determine its destination. Receiving the whole frame introduces a significant delay, known as *latency*, which impacts network performance. Token Ring switches have started to appear on the market based on the same store-and-forward designs as bridges and routers.

In contrast to store-and-forward switches are *cut-through* switches, which begin processing a frame as soon as the first 20-30 bytes have been received. Information in the frame header is analyzed and the destination port identified almost instantaneously. A connection is then made between the input and output ports, and the frame starts transmitting onto the destination ring while it is still being received from the source ring.

The total delay or latency introduced by the cut-through switch can be as low as 50 microseconds, or approximately 2% of an equivalent store-and-forward device for 4K frames. Low latency is one of the important metrics that define a high performing switch. Although a switch may be high performing according to other metrics, it cannot offer the following four benefits without low latency.

- Users can communicate across the switch with the same performance as if they were attached to the same ring.
- Servers can be moved from individual workgroup rings to a central location on the network without any noticeable drop in performance or response times.
- Users can easily move between rings and still attach to the same servers without affecting response times.
- The network can carry latency sensitive applications such as video and voice.

These are essential attributes of a switched backbone and distinguish it from a bridged or routed design. Cut-through is well suited to Token Ring, because low latency comes without any sacrifices. Although cut-through precludes error checking, bad frames are so rare that their presence in a Token Ring network is inconsequential.

Non-Blocking and Blocking Switches

In the broadest sense, a non-blocking device is capable of supporting multiple simultaneous data transfers between pairs of rings. However, the term non-blocking is more specifically used to describe how a switch performs in a heavily loaded network environment or when a particular output port is blocked because another user is transferring data onto that ring segment. If, in these situations, the switch performance drops or communications across the switch are halted then the switch is said to be blocking. A blocking switch can seriously impact network performance in certain high performance applications such as the network backbone.

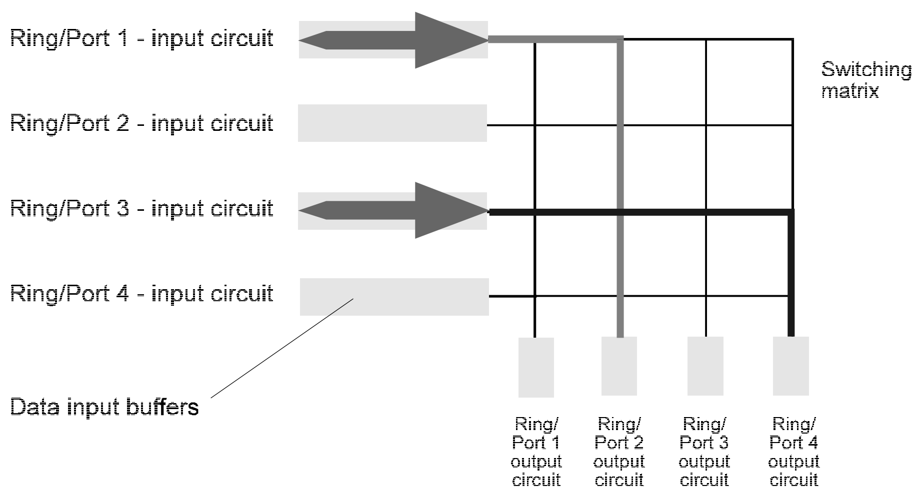

Figure 3: Matrix Switch

One example of a switch that is blocking under these configurations is an input buffered matrix switch (Figure 3). As shown, the matrix switch has four Token Ring interfaces (ports 1 to 4) connected to rings 1 to 4. Each interface has a receive path and a transmit path. The input path has a block of memory which is used to buffer frames if the output port is busy. Consider a frame coming in from ring 1 destined for ring 2. If ring 2 is not busy, the switch will automatically setup a connection across the matrix between ports 1 and 2 allowing data to flow between the two rings. If, at the same time, another frame arrives at port 3 destined for ring 4, the switch will simultaneously carry this frame.

Now consider a slightly more complex situation as shown in Figure 4. The switch is carrying a large frame from ring 1 to ring 2. Simultaneously, a frame is received from ring 3 also destined for ring 2; since ring 2 is busy, the frame is buffered in the input memory on port 3 until ring 2 becomes free. If a second frame is received from ring 3, but this time destined for ring 4 (which is not busy), the frame still has to be buffered on port 3 until the first frame is cleared. This is because the memory on port 3 is a FIFO (First In First Out) memory, and frames cannot jump the queue. Hence, port 3 becomes *blocked* for all further data transmissions until the first frame is cleared.

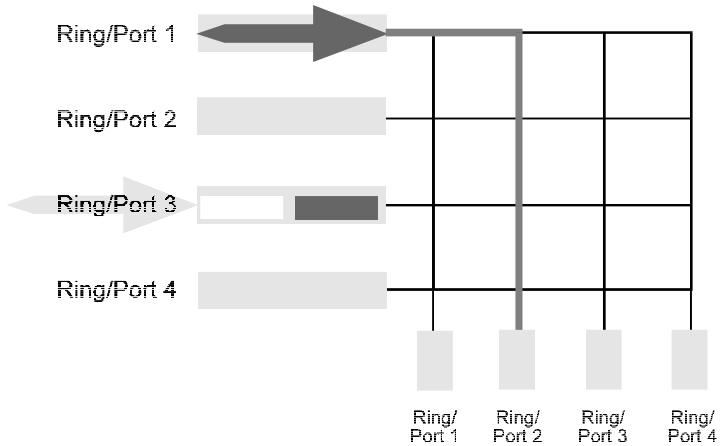

Figure 4: Blocking Example

The effect on the performance of the network can be significant in heavily loaded environments: users will experience noticeably longer response times and the network performance will diminish. Therefore, a blocking switch such as the one just described should not be used in environments where a high level of robustness and performance is required.

In contrast, a shared memory switch is designed to be non-blocking even in such heavily loaded environments. A shared memory switch has a central memory block which can be accessed by any of the ports at any time (Figure 5). Consider a frame being received from ring 1 destined for ring 2. The frame is copied to some location X in the shared memory; at the same time port 2 is able to access the data (even as it is being written by port 1) and transmit onto ring 2. If, at the same time, a frame is received from ring 3 destined for ring 4, the same process occurs: the frame is copied from port 3 to some location in memory, say Y, where it is accessed by port 4 and transmitted onto ring 4.

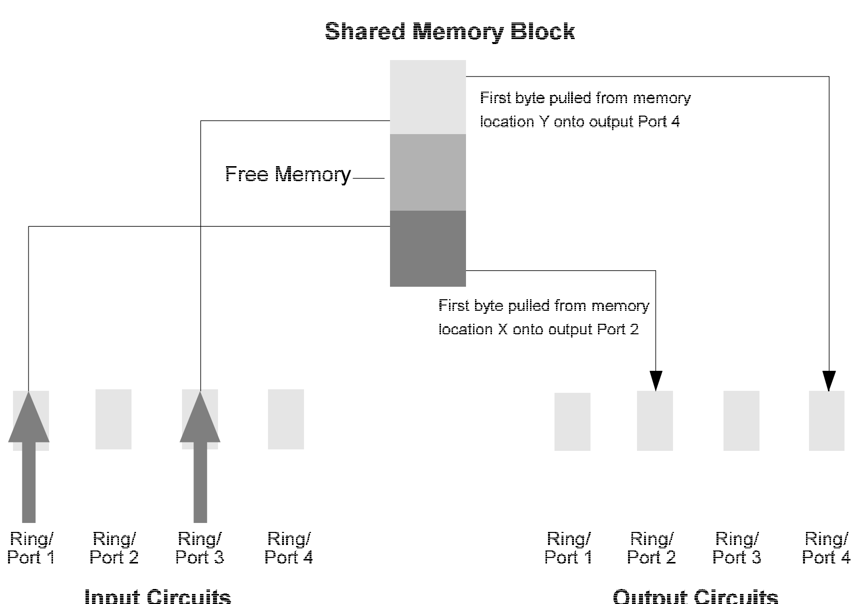

Figure 5: Shared Memory

Now consider the more complex situation where ring 1 is transmitting a large frame to ring 2 (via location X in memory), and at the same time ring 3 wants to transmit data to ring 2 (Figure 6). Port 3 copies the frame to some location Y in memory where it is stored until ring 2 becomes free. If a frame is sent on ring 3 destined for ring 4 (which is not busy), port 3 will copy the data to another location in memory Z, where port 4 can simultaneously access it and transmit it onto ring 4. Even though the first frame from port 3 is stored in memory waiting for port 2 to become free, the switch is able to continue to forward data from ring 3 to other available rings.

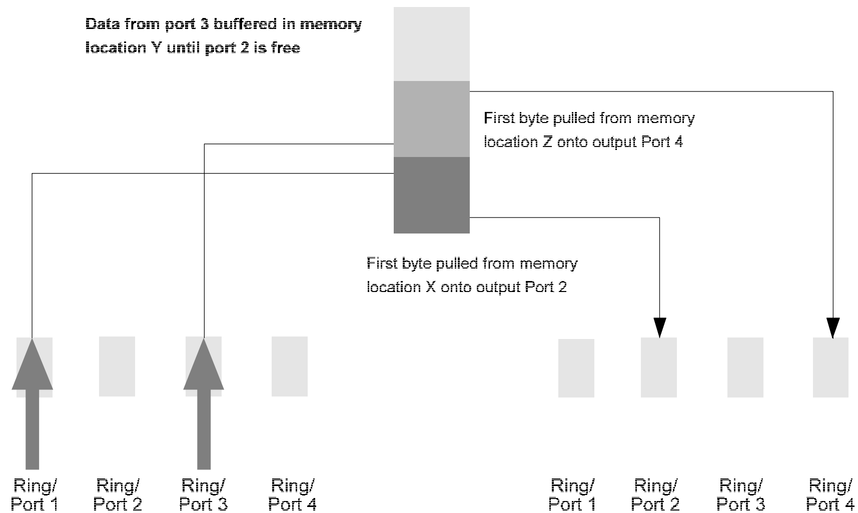

Figure 6: Shared Memory with Same Destination LAN

As a consequence of the central pool of memory, busy ports can make use of memory not being utilized by other ports. The shared memory switch is a non-blocking switch which is optimized for high performance network applications.

Switch Performance

Many organizations are considering Token Ring switching, because of their network performance problems. Fortunately, of all the evaluation criteria the guide has detailed, performance is the most readily quantifiable. In this section, the guide recommends methods for measuring the different performance metrics that collectively define switch performance. The three key metrics are: latency, throughput and packet loss rate (PLR). These parameters and the methods discussed for recording them are used by many network consulting firms, such as The Tolly Group and Strategic Network Consultants, Inc..

The performance tests focus on the two types of frames that are typically generated on Token Ring networks:

- *small frames:* generally used for data acknowledgment and protocol-specific functions
- *large frames:* used for carrying actual data across the network

Although this section is entitled switch performance, the steps that are delineated here can be taken for other internetworking devices as well. Testing a switch in comparison to an organization's current network topology would also be a worthwhile endeavour.

Latency

The term latency is used to describe the delay introduced into traffic flow across the network by individual components. The factors that contribute to the overall latency of a device are dependent on its particular architecture.

The measured latency of a cut-through device will amount to the time it takes to receive the first 20-30 bytes of the frame, where the destination address can be read and the output port discovered. Conversely, a store-and-forward device has a latency comprised of multiple components, including:

- the time taken to receive the whole frame
- the time taken to process the frame to determine its destination
- any time involved in copying data across the backplane between ports
- the transmit set-up time on the destination port

The latency of a store-and-forward device can vary greatly. This variability is largely a result of the size of the frame being received; the bigger the frame the longer it takes to be received into a switch. This frame reception time defines the minimum latency that a store-and-forward device can achieve. For a 28byte frame that time is 14 microseconds, whereas for a 4Kbyte frame the minimum latency is 2,100 microseconds (See Appendix A for formula). The latency differences between cut-through and store-and-forward devices are most prominent at the larger frame sizes.

Often, vendors promoting store-and-forward devices quote latency figures that only include the time involved in processing a frame. The latency test is an extremely useful way of unearthing the real numbers.

Recommended Test Bed One

The structure of the test bed for latency is rather simple. A source node and ring are separated from a destination node and ring by a switch. The source and destination nodes must each be provided with a time stamping mechanism. The time stamping will be done at the point in time a frame starts leaving the source node and the time the frame starts being received by the destination node (Although hereafter we will refer to the testing in terms of a frame's time of entry and departure from the switch itself. The test bed generates an extremely close approximation of the true latency. The time stamping mechanisms are not perfect, because they do not reside in the switch. Any incremental delay introduced by the source node/ring and destination node/ring will be uniform across all products tested.).

When a device is selected for calculating latency, it is advised that its time stamping procedure be investigated. The large majority of devices on the market today that measure latency have not been adjusted to address the advent of cut-through technology. These devices are optimized to measure store-and-forward latency, which would be acceptable in an evaluation that only involves store-and-forward devices.

These products measure latency as the difference between the time at which the final bit of a frame enters the switch and the time the first bit of a frame leaves the switch. In other words, only the processing time of this store-and-forward device is being measured. This measurement will not work for a cut-through switch, because the first bit of a frame leaves the switch before the last bit leaves its source node. In this scenario, the measured latency would be less than zero. If a proper evaluation is to be made between store-and-forward and cut-through switches, the chosen product must be able to stamp time for the first bit into the switch and the first bit out of the switch.

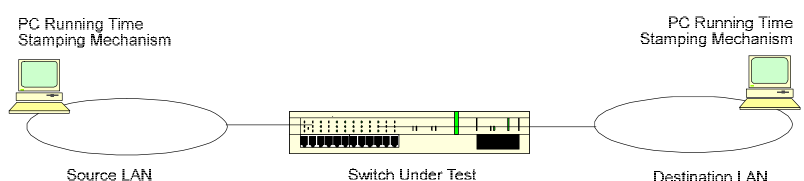

Figure 7: Latency Measurement Test

One product that supports the first-bit-in, first-bit-out method is the LANPharaoh by Azure. The LANPharaoh test bed structure conforms to the description above (Figure 7). LANPharaoh interfaces with the source and destination nodes and is responsible for generating the test frame, time stamping and calculating the latency by subtracting the differences in time. The LANPharaoh or any other device used should run the test at two small frame sizes (28 and 64 bytes) and at two large sizes (1,024 and another approximating 4,000 bytes). The final latency numbers should be an average of multiple tests (at least five for each frame size) done for each frame size. If there is a large variance between the low and high measurements for a given frame size for a given switch, there should be concern about its performance predictability.

If any of the switches being evaluated are modular, an additional battery of tests is suggested. The reason is that some switches' architectures produce variable performance depending on whether switching is done within modules or between modules. After testing latency between two rings on the same module, the next step would be to test latency between LANs on distinct modules.

Another test used to evaluate the effects of architecture on latency is the running of two concurrent data streams instead of just one. If a switch is truly non-blocking, the latency should be the same for both data streams and should be the same as for the single stream. To properly complete this test will require four Token Ring LANs instead of two.

Recommended Test Bed Two

If time stamping mechanisms are not available for direct latency measurements, an alternative test is possible. This test bed involves the use of Novell's PERFORM3. This utility is not capable of measuring latency, but instead will help evaluate its overall effect on the system.

Initially, PERFORM3 should be used to assess the performance of a single client, with NETX loaded, reading files (set to the maximum size, which is 65.535bytes) from a server. This test should be run for a number of frame sizes from small to large, from which a performance graph can be plotted. The results from PERFORM3 will be presented in the form of Kbyte/sec and will serve as a baseline.

The same test should then be run with a switch segmenting the client and the server into two different rings. In essence, the latency defines the responsiveness of server during the test. The switch closest to approximating the overall system performance of the same ring test is the switch with the lowest latency. As touched on before, the latency differences and thus overall system performance will be most apparent for large frame sizes.

As with the first test bed, it is advised that a modular switch be further tested with the client and server residing on different module cards.

Throughput

Throughput is another telling performance metric. Throughput is defined as the maximum frame forwarding rate that an internetworking device can achieve without losing data. Throughput and latency are two completely independent performance metrics. Latency is a measurement that looks at one isolated packet, whereas throughput evaluates a device's ability to handle aggregate packets. Together, they define the raw performance of a switch.

Vendors of switches and other internetworking devices often quote aggregate throughput numbers, which are based on theoretical maximums. Real-world tests will help to verify claimed throughput figures. Aggregate throughput rates often sound staggeringly impressive, but are highly dependent on the number of ports in the configured device. When comparing aggregate throughput rates, whether tested or theoretical, the numbers should always be divided by the number of ports, so that per port rates can be defined.

Another way that vendors can manipulate throughput figures is with the frame size used as the basis for quotation (See Appendix B for theoretical maximum calculations). The smaller the frame size quoted, the larger and more impressive the throughput number will be. Throughput numbers for large frames are rarely quoted directly for that reason.

The number of large frames at wire speed is substantially lower than that of small frames. Thus, large frames are much less taxing on the processing power of a switch. There is a much greater likelihood that a switch will achieve the theoretical maximum throughput at large frame sizes. Achieving the theoretical maximum throughput for large frames allows vendors to claim their products run at wire speed, even if they do not do so at small frames.

There is no guaranteed consistency among vendors, so be sure that throughput is being compared for the same frame size. Live performance testing will clarify any vendor created confusion.

Recommended Test Bed

The equipment required for this test can be segmented into two functions: forwarding and receiving. The devices that do the forwarding are responsible for the generation of frames and verification of frame sizes and rates on the transmitting LAN. The receiving devices simply handle the tallying of frames on the destination LANs. The receiving devices are used to gauge whether or not the switch has been able to process all the packets from the source LAN.

The frame generating function can be capably handled by network analyzers like the Wandel and Goltermann Technologies, Inc. DominoLAN or by frame generation software like the LANQuest FrameThrower. For smaller frame sizes, it might be necessary to have more than one frame generating device per LAN segment. As for frame reception, a network analyzer will most ably handle that task (Figure 8).

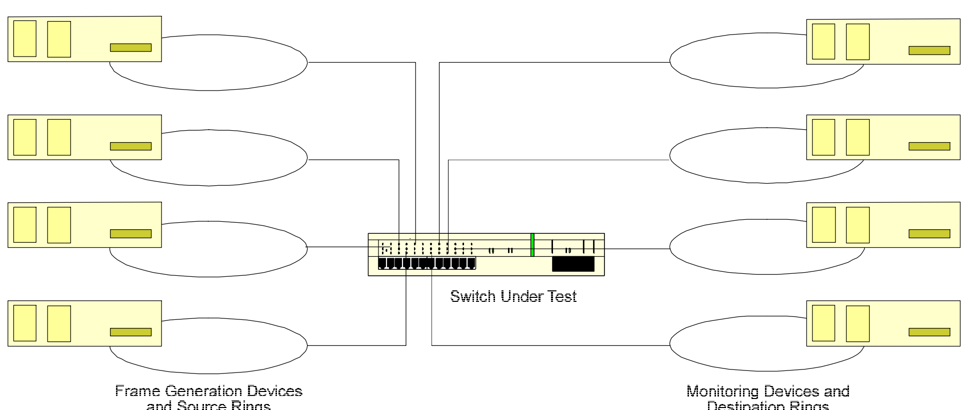

Figure 8: Throughput Test Bed

Throughput is assessed via a binary search procedure that begins at 50% of the theoretical maximum rate (wire speed). (The throughput testing should be done at the same sizes recommended for the latency testing.) After successful completion of this step (no packet loss), the rate should be increased using this formula:

$x+(100-x)/2=y$. Where x is the current percentage of the theoretical maximum and y is the next percentage to test. If the test results in packet loss, then reduce the percentage of the theoretical maximum according to the following formula:

$x-(x-z)/2=y$. Where x is the current percentage, z is the previous test percentage and y is the next percentage to test.

The tests should be done first for single data streams and subsequently for multiple data streams. However, testing of multiple data streams is limited by two factors. The first is the number of available ports per switch. For each pair of ports, there is one data stream available for testing. For fairness in testing, the maximum number of data streams supported by the switch with the smallest port density is what should be tested uniformly across all switches. The other limiting factor is the number of devices available for forwarding and receiving multiple streams of data.

If any of the switches being evaluated are modular, it would be prudent to test the best case and worst case throughput scenarios. In other words, single and multiple stream throughput should be tested intra-module as well as between modules.

Finally, special attention should be spent on testing switches for non-blocking. Most vendors with switches that have a blocking architecture do not openly advertise this fact, because they know it has serious negative performance limitations. The test for non-blocking involves four rings attached to a switch (Figure 9). Rings C and D have servers on them (which in practice will be two network analyzers). Ring B has one client (a client is a workstation running frame generation software) sending the maximum amount of data it can to the server on Ring D. Ring A has two clients. Client A1 sends data to C and client A2 data to D. If the switch is blocking, the performance between A1 and C will drop because A2's path to D will be blocked. Another means for uncovering the non-blocking or blocking nature of a switch is through calculating what is known as the packet loss rate.

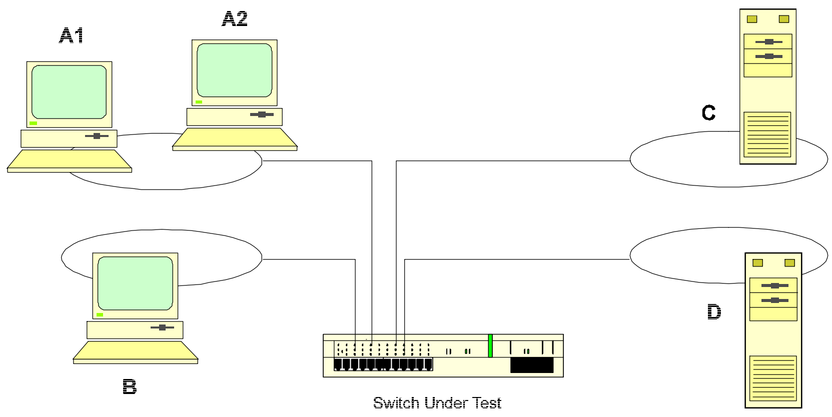

Figure 9: Non-Blocking Test Bed

Packet Loss Rate

The packet loss rate reports the percentage of packets that were originally forwarded to the switch but were not received onto the destination network within a pre-defined period of time. The packet loss rate is highly dependent on the buffering scheme employed by a tested switch. It is possible that the buffers might not be large enough for a fully loaded network or that the active congestion control mechanism is not responsive enough.

The packet loss rate is of great import if the switch being evaluated is for the network backbone. The adoption of switching technology for that application is inspired by current network congestion. Severe packet loss will be counter-productive to any performance improvements that the switch might provide to a network in terms of latency and/or in throughput. In addition to keeping the network from running at wire speeds, a switch with a high PLR will further contribute to network congestion with the re-transmission of a large number of lost frames.

Recommended Test Bed

PLR, like throughput, is tied to the theoretical maximum packet forwarding rate but is measured by a different procedure. The test starts with loading the network at the theoretical maximum rate for each combination of packet size and number of data streams. If the switch drops packets at 100% of the theoretical maximum it is stepped down 10% until the packet loss rate is zero for two successive trials.

Although the procedure for calculating the PLR is different than the other tests, the recommended environments are the same. The same devices that were used for doing the throughput tests can be used for the PLR test. The PLR should be tested in single and multiple stream combinations for both intra-module and inter-module (if applicable) switching. The PLR will be a strong indicator of the non-blocking/ blocking architecture of a given switch. If the switch suffers from severe packet loss it is said to be blocking. Directing multiple data streams at one output port, as was done in one of the throughput tests, will help to verify the stability of a switch. It is quite possible that a switch's buffering mechanism is designed to handle concurrent data streams but not streams that converge on the same output port.

The Right Switch

This guide was devised to assist organizations in the Token Ring switch evaluation process. It is not all-encompassing and only specifies the minimum requirements for a given switch. It understands that there will be additional features a switch must have to address a particular network's needs.

Performance evaluation has received a substantial amount of attention in this document primarily due to the complexity involved in conducting testing that is both accurate and informative. As mentioned earlier, expectations of performance should be tempered by application need. Superb performance on all metrics is not as important for a workgroup switch as is a solid price performance quotient.

On the other hand, a backbone switch should be expected to receive high marks for each performance parameter as they are all of equal import. A low latency switch with unimpressive throughput should be less appealing than a balanced switch that provides average latency and throughput. Of course, if a switch tests well for all performance metrics then selection is that much easier.

Even more important than switch to switch comparisons is the comparison between the leading switch candidate and the internetworking device it is poised to replace. If a switch cannot offer substantial gains in performance across all metrics, it may be hard to justify its purchase when there is already a large investment made in the internetworking device installed base. For instance, a store-and-forward switch may not be a cure for a network that currently has a high performance collapsed backbone router. The store-and-forward device can offer little if any improvement in latency and is not necessarily going to have higher throughput.

Switch performance has been a main topic herein however, the guide has tried to articulate that switch evaluation is not a one-dimensional look at performance alone. The other minimum requirements as stated are, by their very definition, elements that *must* be present in a switch if it is to perform its intended role. Consider a "high performance" backbone switch without any broadcast control features. The purported performance increase will never materialize, because the network will be ravaged by congestion producing broadcast storms.

Many of these required features can be specifically checked; for example, the availability of a desired uplink interface, or the observed hands-on usability of the network management. These feature tests used in conjunction with the performance testing should help organizations arrive at the point where they can identify the right switch for their needs.

Appendix A

The formula used to calculate the receive time of a data frame is as follows:

$$\frac{\text{Frame size (in bytes)*8bits}}{\text{Speed of the network in bits per second}}$$

As an example, consider a 4096byte frame on a 16Mbps Token Ring.

$$\frac{\text{4096bytes*8bits/byte}}{\text{16,000,000bits/second}} = \text{2,100 microseconds}$$

Appendix B

The formula used to calculate the theoretical maximum throughput is as follows:

$$\frac{\text{Speed of the network in bits per second}}{\text{[Frame size (in bytes)+Interframe overhead]*8bits}}$$

The interframe overhead includes 3bytes for the SDEL, EDEL and FS fields of a Token Ring frame. In addition, it includes the interframe gap. This interframe gap is dependent on the Token Ring interface used by a given switch. This information can be obtained from the switch manufacturer. The interframe gap for a Madge Smart Ringswitch is 5bytes. Thus, the total interframe overhead for a Madge Smart Ringswitch is 8bytes.

As an example, consider 28byte frames coming from a 16Mbps Token Ring which are processed by a Smart Ringswitch.

$$\frac{\text{16,000,000bits/second}}{\text{[28+8]bytes*8bits/byte}} = \text{55,556pps}$$

It is important to know whether or not the frame generation program used is measuring the size of the entire frame including the header. If the program is including the size of the whole packet, then the formula above applies. If the program only counts the size of the data, the size of the header has to be added to the size of the data. For a one hop test the header is 24bytes. If this header size is not added to the data size the theoretical maximum calculation will be completely wrong.

Consider a 28byte frame. If that frame size includes header information, the maximum throughput is as calculated above. However, if those 28bytes are just data, the total frame size is actually 52bytes. In this case, the theoretical maximum throughput is roughly a half less than for a complete 28byte frame.

References

1. Strategic Networks Consulting, Inc., The 1995 Switched Ethernet Evaluation Report, April 1995.

2. The Tolly Group, Test Report Number 5281-Madge Networks, Inc. Smart Ringswitch/Cisco Systems, Inc. Catalyst 1600, September 1995.

Inside Back Cover Blank

Amsterdam
Atlanta
Bangkok
Berlin
Boston
Brussels
Canberra
Cape Town
Chicago
Cologne
Copenhagen
Dallas
Denver
Detroit
Frankfurt
Hamburg
Hong Kong
Jakarta
Johannesburg
Kuala Lumpur
London
Los Angeles
Madrid
Melbourne
Minneapolis
Munich
Nashville
New York
Ontario
Oslo
Paris
Perth
Philadelphia
San Jose
Seattle
Singapore
Stockholm
Sydney
Tel Aviv
Tokyo
Toronto
Vancouver
Washington DC

Madge Networks

Europe, Middle East, Africa	Americas	Asia	Japan
Knaves Beech Business Pk	2310 North First Street	64-01 Central Plaza	Believe Mita
Loudwater, High Wycombe	San Jose	18 Harbour Road	43-16 Shiba 3-chome
Bucks HP10 9QZ	Calif. 95131-1011	Wanchai	Minato-ku, Tokyo 105
England	United States	Hong Kong	Japan
Tel: 44 1628 858000	Tel: 1 408 955 0700	Tel: 852 2593 9888	Tel: 81 3 5232 3281
Fax: 44 1628 858011	Fax: 1 408 955 0970	Fax: 852 2519 8022	Fax: 81 3 5232 3208

980-301-01 11/95