This chapter includes the following topics:

- Business requirements overview
- Assessing demarcation points
- Recognizing your limits
- Meeting service needs
- Assigning the correct QoS system

CHAPTER 7

Beyond the Boundaries

You may only operate (own) part of your delivery mechanism, relying on third-party organizations to deliver your application to both your internal staff and your customers. This chapter discusses approaches to this situation, covering ways to work within these confinements.

It begins by discussing how and where to establish your demarcation points. This not only helps to define your management structure, it also expedites fault-condition resolution by having clearly defined areas of responsibility.

It then discusses using this demarcation point information to recognize when you have reached the limits of your responsibility in the continuing optimization and *application performance-monitoring* (APM) model.

Next, it discusses how to incorporate business criteria into the APM process and introduces the concept of business service management.

Considering the newly assigned demarcation points, the chapter revisits the concept of assigning the *quality of service* (QoS) mechanism learned in Chapter 5, "QoS and MPLS: Tools to Manage Application Performance," and Chapter 6, "Application Deployment."

Finally, it discusses the importance of reporting, ensuring that the information is aimed at the correct target audience, and that it is clearly presented for all parties across the delivery system.

Business Requirements Overview

As organizations increase their dependence on IT, business requirements become more critical and the adage "If you improve IT, you improve the business" can be applied. Business and IT alignment is critical to any IT department wanting to understand and report business contextual information for the delivered IT service, but it is often outside the context of traditional network management systems.

This situation is further complicated when you have to work within the confines of an outsourced network—core parts of the delivery system are under the control of a third-party organization. This may be as simple as a carrier who is providing point-to-point circuits, or down to comprehensive outsourcing contracts that include, for example, all application servers and client desktops.

The trick is just to apply the same rules and processes on smaller sections of your environment. As discussed, your first task is to define your objective. In this case, your objective is to clearly identify where responsibilities begin and end for relevant parties. You have then identified key transition parts in your overall delivery system.

Assessing Demarcation Points

Having established the profile of the application, its path through the network, and the pattern it leaves behind, you can start to assign demarcation points for monitoring performance. These fall where the responsibilities of the customer and IT service provider (in-house or outsourced) start and stop.

Where you draw the lines depends on the following:

- Business arrangement (such as outsourcer or internal service provider)
- Application-delivery requirements

In an ideal world, no demarcation lines would exist. Application support would be delivered on an end-to-end basis, with the application treated as a whole, and a single entity would be responsible for every aspect of the delivery process, from the server through the network and subsequently on to the client and the desktop.

In the majority of cases, this is not achievable, because many different skill sets are involved in each step of the delivery process, and this may not be the most optimum method from a business perspective.

Figure 7-1 illustrates a simple network scenario, including a WAN, access circuits, and LAN equipment and services. From this schematic, you can start to set demarcation points for network responsibility.

Figure 7-1 shows an example of defining boundaries for the measurement of application delivery, making it possible to develop a clear set of procedures for performance deterioration.

The demarcation point represented by the inner circle (circle number 1) is that of the traditional telecommunication services, where responsibility ends at the edge of the network just as it enters the customer tail (access) service.

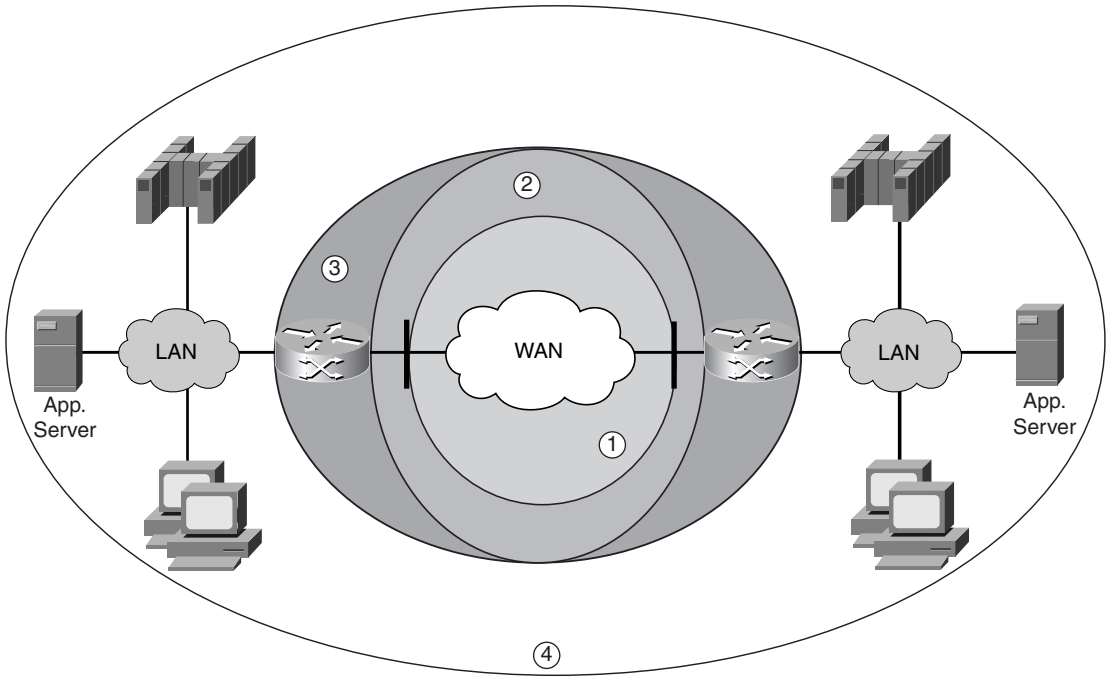
This demarcation point is useful from a carrier perspective, but has little relevance for end users, because they just see the network as a transmissions path.

In circle number 2, the telecommunication's responsibility has been taken a little further to cover all aspects of the WAN delivery up to the point of delivery at the customer premises.

Circle number 3 expands the responsibility and covers the point at which the LAN interface is presented to the customer site.

Finally, circle number 4 covers the client/server LAN area, concentrating on the delivery to the relevant hosts.

Figure 7-1 *A Simple Example of Demarcation Points*



Each of these areas can be delivered as either an individual or a combination approach to the delivery model. For example, you may want to just purchase a link from a carrier. In that scenario, you simply apply the demarcation point at the end of the link, define metric collection at that point, and effectively monitor and report on your application delivery relevant to that endpoint. (For details, see Chapter 8, "Monitoring the Delivery.")

You may then want to assign more of the responsibilities to different departments within your own organization. For example, the network delivery team assumes responsibility for the delivery mechanism from where they pick up delivery from the carrier, to the point at which it is delivered to the interface of the internal switch. Responsibility then moves to the LAN or server administrators.

For example, Acme Electronics is an electronics retail firm, and they have recently moved to sell their goods online (in addition to their traditional storefront operations). Their business objectives are straightforward. They intend to migrate 80 percent of their sales revenue from their retail outlets to the web-based offering. In addition, they want to ensure that each customer will have to wait no longer than 8 seconds for online confirmation of an order.

Acme's network has recently been outsourced, and at the time they moved to the web service, they had a network carrier providing connection between their web server and the warehouse database system.

Acme's responsibilities include the web server, warehouse database system, associated LAN systems, and connection to the Internet. The carrier's responsibility includes the carriage of data between the web server and the warehouse database system.

After you have defined these areas of responsibility, you can start to assign equipment within each of these areas to the responsible parties. You may break an area down to whole equipment (such as a router or switch) or down to a specific interface. This demarcation is sometimes made easier by understanding who owns the equipment. In an outsourced agreement, for example, all network infrastructure may be supplied by the outsource company; therefore, demarcation points for that path or process clearly reside where the outsource equipment meets the customer equipment. It should be added that, in these scenarios, the cabling is quite often overlooked and can cause a contention if the fault is found to reside between the customer and outsourced equipment. Therefore, any agreement should clearly state cabling and interface conditions.

From Acme's perspective, this can be further defined, with the server administrators taking responsibility for the web server and database server (including software and hardware) and the network engineers taking responsibility for the LAN components of the network.

By taking your profile and reference point, and then matching it to your demarcation point, you are in a position to write the *service level agreement* (SLA), bearing in mind three basic rules:

- **Keep it simple**—If it is too complex, the SLA will never be agreed to or implemented.
- **Stick to business**—It is no good having a perfect SLA unrelated to the actual business objective.
- **Minimize overhead**—You are in business to make money, not administer an SLA.

When defining demarcation points, consider aligning with metric collection locations. This will be dealt with more in the "Defining Metrics" section of Chapter 8, but the concept should be considered at this point. The theory is as follows. A demarcation point is a strategic connection point in the conversation transmission. By observing specific conditions (metrics) at this point, you can gain a quick and inductive view of the application's performance. In a sense, you are looking for these demarcation points to become the natural aggregators of an application's performance issues.

One of Acme's business objectives is to have no user waiting for more than 4 seconds to receive confirmation of an order. For the order to be confirmed, the web server must receive notification from the warehouse database that the product is in stock. So in Acme's network, a strategic connection point is the warehouse-facing network connection on the web server. From here they can track the request from the web server to the warehouse database, and the response from the warehouse server to the web server.

Recognizing Your Limits

At some point, you will either reach the limit of your responsibility or maybe even the limit of your skill level.

At this point, you need to make a decision on the state of the application. Consider the following scenarios:

- **Scenario 1**—You have met the business delivery requirements.
- **Scenario 2**—You have optimized as far as you can given the infrastructure and application profile, but have not met the business delivery requirements.

In Scenario 1, you have achieved the objective of meeting the business delivery requirements. (In Acme's case, every order response from the warehouse server is less than 4 seconds.) You can consider the issue resolved. (As discussed, the optimization process is a constantly evolving process.)

In Scenario 2, you have effectively gone as far as you can go. You have applied all QoS strategies to improve the application architecture, but the application performance is still not meeting the business requirements. At this stage, you must take steps to escalate the situation further. How you do this depends on what level of optimization was achieved. You can determine this by studying the metrics collected at the relevant demarcation points.

Each metric reveals application performance status; this in turn indicates areas that need improvement, such as an increase in bandwidth or even server CPU speed. You can find details about how to collect and apply these metrics in Chapter 8, "Monitoring the Delivery," and in Chapter 10, "When Applications Fail."

In Acme's case, the response time received at the metric point reveals that the warehouse server is taking 7 seconds to respond. Investigation by Acme revealed that the requests are being received at the warehouse database in good time, but the database server is taking an excessive time to process the response.

Assuming these metrics do not bring up more areas for you to optimize, your decision must then be to escalate. The metrics will identify the area of responsibility (more specifically the third party or vendor) that needs to address some issues. By establishing this audience, you will be in a position to deliver a clear and concise message conveying the true impact to your application-delivery model, enabling the audience to make recommendations for improvement. You can then verify these using your virtual environment, as described in Chapter 2, "Understanding Your Business."

Acme was able to identify the most probable cause of concern to be the warehouse server architecture. This conclusion was reached by simple refection on their metric collection. The web server showed it was sending and processing the requests in time, and the speed across the network (when a response was received) showed no congestion, drops, or latency, so by elimination, the area of concern and responsibility was the warehouse database server. The database administrators confirmed that the database was optimally tuned, so the last area to address was the server architecture. This is a simplified example, but it helps to illustrate how the observations at specific demarcation points ultimately assist in identifying the most probable area of concern.

Meeting Service Needs

An accurately defined and monitored APM system can assist in the ultimate aim of effective delivery of the application, while providing sufficient data feedback into management tools that will ultimately assist in optimizing delivery.

Delivery criteria cannot be concerned solely with technical parameters, but must meet business objectives. The following three major factors differentiate an organization:

- Product or service offerings
- Customers
- Delivery to market

Together these factors make up the organization's unique value proposition. Each organization's mission is to create and maintain the systems and processes to support its value proposition.

Business systems and processes operate under a combination of time, cost, and quality requirements that make it essential to use IT. IT reliability and flexibility are often the critical make-or-break issues behind business initiatives. Companies across a broad spectrum of the economy rely on IT to create competitive advantage and improve productivity, with business initiatives such as supplier and supply-chain management, business and logistics systems, customer relationship management, and e-commerce. IT supports data-driven decision making and synchronizes real-time activity between the managers, operators, and users of the business systems and processes.

To provide a better service to its customers and users, an enterprise needs to evaluate and manage technology at a business process level. This provides a business-oriented view of the enterprise's information systems, translating the traditional metrics into demonstrable measures that make sense to the company's senior executives.

Key business services supplied by an enterprise to its customers quite often rely on a number of business processes. For example, acceptance of a customer's product order will probably entail the delivery of product information to a company's website, the checking of stock availability, and the process of credit authorization. Each of these three processes will rely on combinations of multiple technology resources, including network devices, system platforms, applications, and databases.

For an organization to ensure its product ordering is functioning correctly, it has to collect and relate a large amount of availability, fault, status, and performance information that extends across the whole array of management systems. The information has to be put into context across the relevant business processes and then consolidated across business units and possibly companies.

IT organizations were traditionally managed based on the technical capabilities of underlying infrastructure such as networks, systems, storage, and applications. This silo-based approach to management, where each of the technologies is managed separately, evolved over time as various technologies were developed.

Today, most organizations have recognized that this approach is no longer viable. It fails to take into account the interrelated nature of individual components, and most importantly does not adequately address the business priorities of the firm. In response, business executives and IT leaders have begun to demand a more holistic approach for managing IT resources. This has led to the development of systems that take in the technical metrics and traps and assess them relevant to the unique business criteria defined by each organization, as discussed in Part II, "Aligning the Network's Business and Technical Requirements." This approach is known as *business service management* (BSM).

BSM is the logical progression of network and systems management and fits nicely into the emerging requirement to report business value. BSM brings into line IT services with the business processes they enable, and then manages these services in a way that is consistent with business priority. Many benefits result from this, including the following:

- Investment in service delivery is appropriate to business benefit and overcapitalization is avoided.
- Problem resolution is prioritized in line with business requirements.
- Service-improvement programs are initiated only if there is a real payback to the business.
- Business process owners are aware of service disruption and can take action to mitigate the consequences to the business.
- Executives understand the value of their IT service provider.

To be accountable for the delivery of service, IT must manage simultaneously across business processes, management systems, and different business units, enterprises, and/or organizations. As the IT infrastructure has become so important to the enterprise's ability to conduct business, its complexity has increased dramatically, making it imperative for the business to understand the technology issues and consequences involved.

To gain this understanding, enterprises must view the infrastructure from a business service perspective. To achieve this, each line of business must be accessible with a degree of granularity that allows access to each individual technology element for diagnosis and repair—for example, access to an element manager to view infrastructure availability or an application analyzer to view the raw data of an application conversation. Lines of business services now tend to cut across the traditional management systems and business processes. They can also span organizations and divisions. An application such as supply-chain management can range across the functional organizations of sales, marketing, purchasing, manufacturing, and customer service.

A BSM system complements an enterprise's existing management systems to achieve this level of management sophistication. It operates as a top layer and provides a single, centralized method of command and control for the entire IT infrastructure.

This alignment is very simple in theory, and has been discussed at some length in Part II of this book; however, many organizations have been frustrated in their attempts to achieve it.

A major factor is that the IT department is not aware of business strategies at a detailed enough level. They do not know what business strategies to align with.

Business activities that require IT support appear to be practically thrown over the wall, because IT becomes involved only after decisions are made and it is implementation time. The IT department will receive an application rolled out on their environment (or to roll out) with little or no information pertinent to its delivery criteria or its criticality to the overall business model.

Without a detailed business strategy, IT's ability to plan well is extremely limited, and the best IT can do is to execute rapidly and attempt to architect flexible enough processes and systems to accommodate change with little or no notice.

For IT management to succeed in becoming a key participant in business strategies, there must be open communications with the business groups for such alignment. Only then will the IT department be able to develop detailed metrics for managing internal IT processes and tracking workload fluctuation, resource availability, and service quality pertinent to the business requirements.

Assigning the Correct QoS System

Where the entire end-to-end delivery system is under single control, you can use the simple approach to assigning QoS protocols and configurations discussed in the "Allocating Network Resources" section of Chapter 5. The implementation of this chapter's systems depends on how much control you have—for example, a priority set by your system at an edge could effectively be reclassified in the core, rendering your original configuration redundant. There is an element of cooperation and agreement dependent on your individual organization's unique arrangement with the third-party system.

QoS solutions are achieved through the use of traffic conditioners in the end-to-end network. The traffic-conditioning model implemented in RFC 2475, "An Architecture for Differentiated Services," includes classifier, policer (meter), marker, shaping, and dropping mechanisms. Cisco routers use many traffic-conditioning implementations to provide these functions. The methods used depend on the objective of the QoS policies implemented and the traffic present in the network.

The methodologies discussed so far in this book remain central to the overall system management and ultimate APM infrastructure. Central to these methodologies is the profile of the application. You must understand the application characteristics and dependencies, and then assign QoS based on a combination of those characteristics and the business delivery requirements.

The traffic-conditioning mechanisms used are as follows[1]:

- **Classifier**—Determines what classes packets belong to by inspecting various fields within the packet header including, but not limited to, source address, destination address, protocol identifier, source port, and/or destination port. The Cisco *Network-Based*

Application Recognition (NBAR; see Chapter 6) can be used to further classify traffic based on Layer 4 through Layer 7 information within the protocol data unit. Classification is most often done using access control lists, but NBAR can be leveraged to police file-sharing applications and worms, which typically do not use well-known port numbers.

- **Meter**—Measures the speed of a traffic stream. The metering process can be used to affect the operations of the marker and shaper/dropper. Traffic that conforms to the *committed information rate* (CIR) is treated differently than the traffic that exceeds the CIR. A dual leaky-bucket algorithm is used with policing mechanisms that meter traffic. This enables the administrator to configure a *committed burst* (Bc) and *excess burst* (Be) rate that can be controlled with different exceed and violate actions. Policing can be used to re-mark and/or drop traffic.
- **Marker**—Responsible for turning on prioritization bit values in the Layer 2 and/or Layer 3 headers as follows to signify the importance of traffic:
  - IP traffic employs the IP Precedence or DSCP fields.
  - Layer 2 traffic depends on the technology used.
  - Ethernet has three 802.1p priority bits in the 802.1 header that can be marked.
  - Frame Relay includes a *Discard Eligible* (DE) bit, which can be marked to signify that the frame should be dropped when the service provider experiences congestion.
  - ATM has a *Cell Loss Priority* (CLP) bit that is similar in use to Frame Relay.
- **Shaper**—Limits bandwidth used on a link by queuing traffic that exceeds the set rate. Shaping is implemented only on egress links and proves especially useful in Frame Relay hub-and-spoke architectures. Frame Relay hubs may be running at DS-3 speeds (44.736 Mbps), whereas small sites may only have a fractional DS-1 access pipe of 128 kbps. The central site (hub) can flood the remote site link with excess traffic and not use resources properly. Framing would shape the traffic stream from the central site to a maximum rate of 128 kbps to the remote site.
- **Dropper**—Discards out-of-profile packets. This mechanism proves very useful in metropolitan Ethernet environments where a service provider will give the customer a Gigabit Ethernet handoff, but only allow 1 Mbps of bidirectional traffic based on the customer contract. The service provider would have the benefit of instantly provisioning. Dropping mechanisms can be deterministic in the case of weighted random early detection (WRED) or undeterministic in the case of tail drop.

All nodes in an internetwork perform traffic forwarding, queuing, and congestion-avoidance (WRED) procedures. Other QoS mechanisms are used, depending on the node's physical location within the network. As a general rule, edge nodes (ingress and egress) perform classification, marking, policing, shaping (only on egress), and dropping. A core router's

main function in the network is to forward packets at high speeds. The CPU- and memory-intensive tasks associated with edge device QoS functionality would burden core routers.

You need to consider how the demarcation lines within your environment theoretically reshape your delivery topology, remembering QoS is deployed primarily at the enterprise LAN-WAN boundary. Based on the demarcation points, there may be a virtual shift in the delivery model.

Table 7-1 identifies various Cisco QoS mechanisms and their application to the QoS building blocks.

Table 7-1 *Cisco IOS Traffic-Conditioning Mechanisms*

Traffic-Conditioning Mechanism	Examples
Classification	Modular QoS CLI (MQC) IP to ATM (Class of service) NBAR QoS Policy Propagation over BGP (QPPB) Route maps ACCESS control lists
Marking	Committed access rate (CAR) Class-based marking QPPB Route maps
Metering	Weighted fair queuing (WFQ) Class-based WFQ (CBWFQ) Priority queuing (PQ) Custom queuing (CQ) Weighted round-robin (WRR) Modified-deficit round-robin (MDRR) CAR MQC policing Class-based low-latency queuing (LLQ)
Shaping	Generic traffic shaping (GTS) Frame Relay traffic shaping (FRTS) Virtual circuit (VC) shaping
Dropping	WRED Flow-based weighted random early detection (FRED) CAR

Summary

In a system with noncentralized control, your first objective after having established the profile of the application, its path through the network, and the pattern it leaves behind is to assign demarcation points for monitoring performance.

Demarcation points are strategic connection points in the conversation transmission. By observing specific conditions (metrics) at these points, you can gain a quick and inductive view of the application's performance. The demarcation points become the natural aggregators of an application's performance issues.

The collected metrics provide the background and data to assess when you have taken the optimization as far as possible within the confines of your control. They will then assist in reassigning responsibility or escalating requirements and issues to third-party organizations.

When assigning QoS protocols in a system with noncentralized control, consider the virtual boundaries imposed by the different areas of control (such as a carrier-controlled core network).

End Note

1 Hartmann, Dennis. Introduction to QoS. Global Knowledge. http://knowledgestorm.inc.com/search/viewabstract/inc/67554/index.jsp

References Used in This Chapter

- Cisco Systems, Planning for Quality of service—CiscoWorks Policy Manager, http://www.cisco.com/en/US/products/sw/cscowork/ps2064/products_user_guide_chapter09186a0080191f03.html
- RFC 1513, Token Ring Extensions to the Remote Network Monitoring MIB, S. Waldbusser, http://rfc.net/rfc1513.html, September 1993
- RFC 1757, Remote Network Monitoring Management Information Base, S. Waldbusser, http://rfc.net/rfc1757.html, February 1995
- RFC 2021, Remote Network Monitoring Management Information Base, S. Waldbusser, http://rfc.net/rfc2021.html, January 1997
- RFC 2457, An Architecture for Differentiated Services, S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, W. Weiss, http://rfc.net/rfc2475.html, December 1998