CA Application Delivery Analysis

User Guide

10.1

ca technologies

This Documentation, which includes embedded help systems and electronically distributed materials, (hereinafter referred to as the “Documentation”) is for your informational purposes only and is subject to change or withdrawal by CA at any time.

This Documentation may not be copied, transferred, reproduced, disclosed, modified or duplicated, in whole or in part, without the prior written consent of CA. This Documentation is confidential and proprietary information of CA and may not be disclosed by you or used for any purpose other than as may be permitted in (i) a separate agreement between you and CA governing your use of the CA software to which the Documentation relates; or (ii) a separate confidentiality agreement between you and CA.

Notwithstanding the foregoing, if you are a licensed user of the software product(s) addressed in the Documentation, you may print or otherwise make available a reasonable number of copies of the Documentation for internal use by you and your employees in connection with that software, provided that all CA copyright notices and legends are affixed to each reproduced copy.

The right to print or otherwise make available copies of the Documentation is limited to the period during which the applicable license for such software remains in full force and effect. Should the license terminate for any reason, it is your responsibility to certify in writing to CA that all copies and partial copies of the Documentation have been returned to CA or destroyed.

TO THE EXTENT PERMITTED BY APPLICABLE LAW, CA PROVIDES THIS DOCUMENTATION “AS IS” WITHOUT WARRANTY OF ANY KIND, INCLUDING WITHOUT LIMITATION, ANY IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, OR NONINFRINGEMENT. IN NO EVENT WILL CA BE LIABLE TO YOU OR ANY THIRD PARTY FOR ANY LOSS OR DAMAGE, DIRECT OR INDIRECT, FROM THE USE OF THIS DOCUMENTATION, INCLUDING WITHOUT LIMITATION, LOST PROFITS, LOST INVESTMENT, BUSINESS INTERRUPTION, GOODWILL, OR LOST DATA, EVEN IF CA IS EXPRESSLY ADVISED IN ADVANCE OF THE POSSIBILITY OF SUCH LOSS OR DAMAGE.

The use of any software product referenced in the Documentation is governed by the applicable license agreement and such license agreement is not modified in any way by the terms of this notice.

The manufacturer of this Documentation is CA.

Provided with “Restricted Rights.” Use, duplication or disclosure by the United States Government is subject to the restrictions set forth in FAR Sections 12.212, 52.227-14, and 52.227-19(c)(1) - (2) and DFARS Section 252.227-7014(b)(3), as applicable, or their successors.

CA Technologies Product References

This document references the following CA Technologies products:

- CA Application Delivery Analysis
- CA GigaStor
- CA GigaStor Connector
- CA Multi-Port Monitor
- CA Performance Center
- CA NetQoS Performance Center
- CA Standard Monitor
- CA Virtual Systems Monitor

Contact CA Technologies

Contact CA Support

For your convenience, CA Technologies provides one site where you can access the information that you need for your Home Office, Small Business, and Enterprise CA Technologies products. At http://ca.com/support, you can access the following resources:

- Online and telephone contact information for technical assistance and customer services
- Information about user communities and forums
- Product and documentation downloads
- CA Support policies and guidelines
- Other helpful resources appropriate for your product

Providing Feedback About Product Documentation

If you have comments or questions about CA Technologies product documentation, you can send a message to techpubs@ca.com.

To provide feedback about CA Technologies product documentation, complete our short customer survey which is available on the CA Support website at http://ca.com/docs.

Contents

Chapter 1: Welcome to Application Delivery Analysis 11

Chapter 2: Using the Management Console 31

Chapter 3: Using the Operations Page 41

Chapter 4: Using the Incidents Page 53

Chapter 7: Using the Optimization Page 117

Chapter 8: Sharing Information from Report Pages and Views 129

Chapter 9: Troubleshooting 133

Chapter 1: Welcome to Application Delivery Analysis

This section contains the following topics:

Introduction

CA Application Delivery Analysis, the end-to-end performance monitoring module in the CA Performance Center (CA PC) and CA NetQoS Performance Center (CA NPC), tracks and measures application response time without desktop or server agents. End-to-end performance monitoring enables you to do the following:

- Gauge how well your network delivers services to end users
- Obtain the best overall view of what happens on the network

CA Application Delivery Analysis monitors TCP application packets from the network, into the data center, and out again and lets you measure Network Round Trip Time, Server Response Time, Data Transfer Time, and much more.

The management console separates response time into application, network, and server delay components to help you rapidly troubleshoot network performance bottlenecks and maintain application performance. Automated processes measure and analyze application performance for all TCP/IP user transactions. These processes then compare response times against thresholds and automatically investigate problems as they occur. Network and Operations groups now have the key diagnostic data to quickly solve performance problems.

CA Application Delivery Analysis also monitors availability and application performance so that you have a consistent set of service-quality metrics for your internal users and external service providers, even if there are no formal Operational Level Agreements (OLAs) in place.

Measure Performance

To monitor end-to-end performance, the management console uses:

- *Baselines* or averages, calculated by the management console, to enable you to see historically normal performance in your network. Use baselines to compare an application/server combination’s performance during a specific period to a historical average of past performance. Crossing a baseline does not necessarily indicate a problem.
- *Thresholds* tell the management console when performance exceeds acceptable limits. The management console includes default thresholds that can be adjusted by a management console administrator to a percentile-based value (a sensitivity factor) or millisecond values. When a performance threshold is crossed, the management console creates an incident. A management console administrator can create actions and notifications for incident responses to help identify problem sources.

By comparing thresholds with baselines, you can see how the thresholds compare to the historical mean and use this information for capacity planning. The management console uses thresholds to rate data. A threshold violation indicates an existing or potential problem.

Baselines

The baseline is the historical norm for performance. The management console automatically calculates hourly baselines and adjusts them for hour of day, day of week, and day of month. The activity of the past week is weighted most heavily so that the baselines adapt properly to systemic changes. These baselines are highly granular so that they reflect the impact of the business cycle. Every combination (metric, network, server, application) receives its own calculation for each hour of the day. Baselines let you contrast actual performance to that which is expected at that time, thereby providing context to the definition of "normal."

The management console calculates baselines for a particular network-server-application combination for each of the 10 baselined metrics. It does not average baselines; therefore, reports do not show baselines for aggregations. The management console includes baseline information in performance detail charts. It does not use baselines to determine thresholds.

To see a baseline on a chart, select the network, server, application, and baselined metric. Summary charts do not contain baselines. The Operations page shows plotted baseline values on performance charts to help you compare current performance with the normal values. The Performance Scorecard on the Management page shows baseline values in tabular format.

The management console shows baselines on the Operations: Metrics Details views when you enable baselines on the Chart Settings page.

Thresholds, Incidents, and Incident Responses

A *threshold* is the upper boundary of acceptable performance. Thresholds are important for these reasons:

- Thresholds enable the management console to rate data.
- Thresholds contribute to incident creation and the resultant incident responses and investigations that enable timely troubleshooting and problem resolution.

Thresholds exist by default for each application. The management console administrator sets thresholds for network and server performance. For each metric, you define a Minor (yellow) threshold, a Major (orange) threshold, and the minimum observations required to show the performance degradation.

The management console calculates default thresholds. It does not use baselines to calculate thresholds, it computes thresholds and baselines independently from the same data. Administrators can change thresholds to make them more or less sensitive to performance changes.

When thresholds are crossed, the management console creates incidents and launches configured incident responses. You can configure incident responses for each network, server, application, or monitoring device incident type. The management console includes a default incident response for each incident type.

The management console administrator can set up actions and notifications in response to each threshold that is met or violated for a specified time period. For a given incident response, the Administrator can specify an action or notification to occur in the following circumstances:

- When performance meets or exceeds the Minor threshold
- When performance meets or exceeds the Major threshold

The management console opens an incident for each condition. If an action exists for the condition, the management console launches that action automatically.

Remember the following details about incidents, incident responses, and actions:

- The management console creates incidents when thresholds are crossed.
- Incidents do not automatically trigger actions unless the following conditions are met:
    - The appropriate incident response is associated with the network, server, application, or monitoring device. You set the association when you create the incident response.
    - The associated incident response has associated actions. Default incident responses do not have associated actions.
    - The violation exceeds minimum severity and duration criteria.
- It is possible to have incidents without actions.

- It is possible to launch actions without incidents. An action associated with an incident is called an investigation, which you can launch manually or schedule.

Incidents and incident responses are useful for troubleshooting in the following ways:

- Incidents maintain a record of conditions when a problem occurs.
- Incident responses automatically gather information that help you troubleshoot a problem when it occurs and thereby reduce the mean time to repair (MTTR).

An incident response is a set of zero or more actions. An action can be a notification or an automatic investigation. For information about the types of incident responses you can create, see the *Administrator Guide*.

Investigations

If an incident response does not gather enough information to help you resolve the problem, you can launch an immediate or scheduled investigation to gather additional information to further troubleshoot the problem. An investigation is an action taken by the management console to gather diagnostic information about the cause of a threshold violation. Your management console user account must have an associated role that lets you perform investigations.

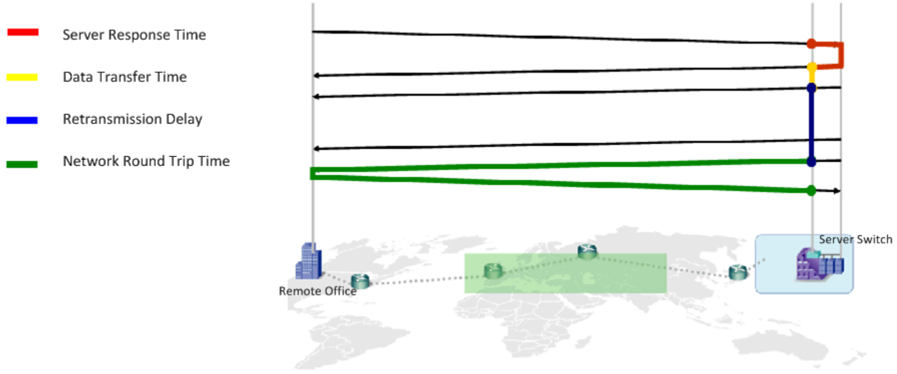
An investigation can take one of the following actions:

- Application connection time
- Packet capture
- Ping response time
- SNMP server query
- SNMP router query
- ICMP trace route
- TCP trace route

Note that the management console administrator must configure the SNMP communities that can be investigated. For more information, see the *Administrator Guide*.

Response Time

The management console includes measurements for response times, which are averaged over 5-minute intervals, including.



Note: Definitions of these response times differ for WAN-optimized transactions.

Server Response Time

The time it takes for a server to send an initial response to a client request or the initial server "think time." Increases in the Server Response Time generally indicate the following:

- A lack of server resources such as CPU, memory, disk, or I/O
- A poorly written application

Data Transfer Time

The time it takes to transmit a complete response measured from the initial to final packet. Data Transfer Time excludes the initial server response time and includes only Network Round Trip Time if there is more data to send than fits in the TCP window.

Retransmission Delay

The elapsed time between the original packet send and the last duplicate packet send. The management console reports Retransmission Delay as an average across observations and not just for the retransmitted packets. If one packet in a set of 10 requires 300 ms of retransmission time, the Retransmission Delay is reported as 30 ms (300 ms/10 packets).

Network Round Trip Time

The time it takes for a packet to travel across the network in both directions between the client and server.

Total Transaction Time

The time it takes to complete a TCP transaction or data request within a persistent TCP connection.

Effective Round Trip Time

Network Round Trip Time plus delays caused by retransmissions.

Key Concepts

To understand and use the data the management console collects and reports, you must become familiar with some key concepts and terminology.

The topics in this section explain most of the product terms that will be unfamiliar to new users, as well as some industry-standard terms.

Aggregation

An aggregation is a group of applications, servers, or networks you create to make comparison reporting meaningful. On the Engineering page, the management console treats an aggregation as one entity. The Operations and Incidents pages show data for each member in an aggregation separately.

An example of an aggregation is a group of IP subnets in a remote office where the management console analyzes performance metrics across individual client subnets that make up the remote office. It is good practice to make an aggregation for each remote office to allow comparisons.

Availability

The management console classifies a server or application as available if users have access to the application or server during a 5-minute period. The management console tracks availability metrics at the server and application port levels through passive data observation. When no user traffic is observed, the management console verifies availability by making active requests every 5 minutes.

The management console tracks server and application combinations for availability if you enabled availability monitoring. The management console does not track network availability.

Bubble Up

The management console sorts or "bubbles up" the worst performers to the top of the list on the Operations report pages. These pages show performance metrics for the worst-performing networks, servers, and applications.

Monitoring Device

A monitoring device monitors the response time of TCP sessions. The management console supports several types of monitoring devices.

Data Point

By default, the management console averages data over preset intervals to produce data points. For each time interval used to average data, the management console calculates a representative data point and shows this data in a chart or table. Each chart and table reports the time period used to calculate each data point.

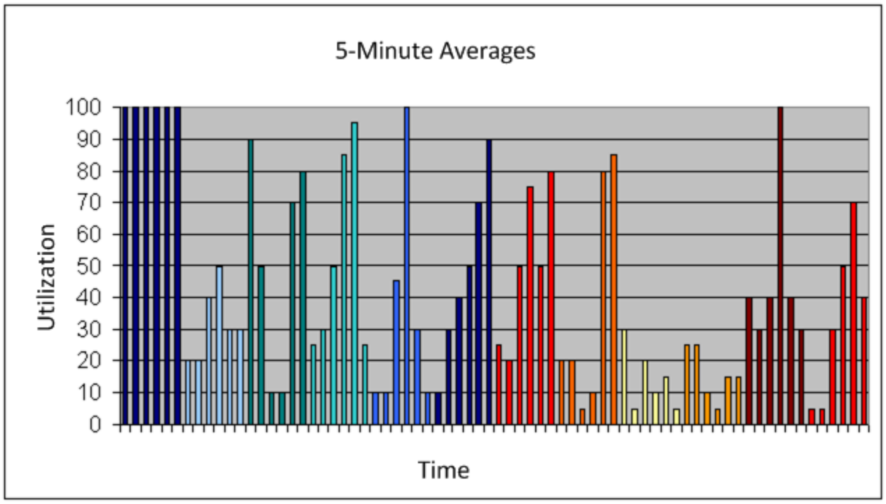
For views that use the time periods in the following table, the management console averages data using the following intervals:

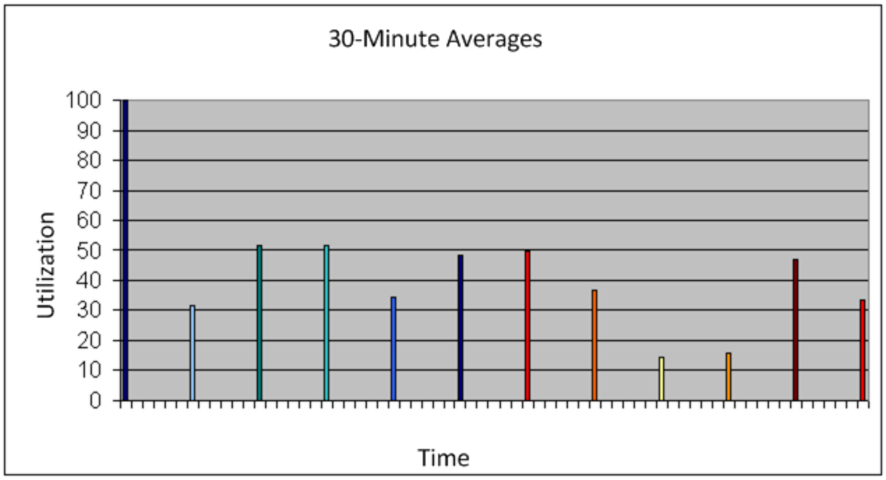
Time Period	Interval
Last hour	5 minutes
Last 8 hours	5 minutes
Last day	15 minutes
Last week	1 hour
Last month	6 hours

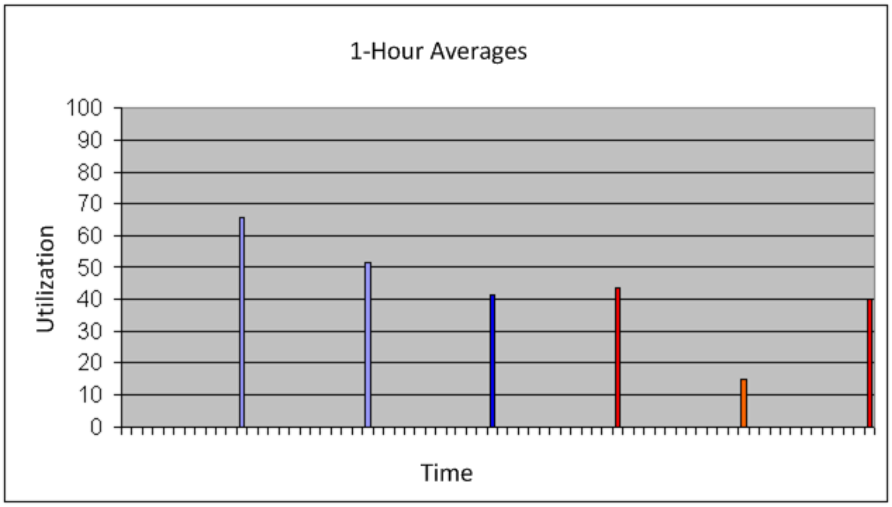
Configure reporting of data points as follows:

- Use shorter averaging intervals to analyze specific problems.
- Use longer averaging intervals to identify patterns and trends over time.

The following views show the transition of granular data points to curves used for trending and pattern identification as the averaging interval increases from 5 minutes to 1 hour.







Incident

An incident is an information record that the management console creates when it detects performance that meets or exceeds defined thresholds. You might need to analyze incidents to determine the event that caused the metric to rise. The management console reports incidents by assigned case numbers in the Incidents page.

If the management console is registered with the CA PC or the CA NPC, the management console synchronizes its incident status with events in the CA PC or CA NPC.

The management console stores incident records for as long as you have configured 5-minute data to store.

Incident Response

An incident response helps you troubleshoot a problem at the time that it occurs, and reduce the mean time to repair, assign incident responses to your business-critical applications, servers, and networks. Incident responses:

- Notify your team about a performance degradation.
- Actively investigate a problem to gather additional information that can help you identify the root cause of a performance degradation.

By default, the management console does not automatically launch an incident response.

Investigation

You can manually launch or schedule an investigative action for more information about the incident cause.

Maintenance Schedule

You can associate a maintenance schedule with each server that is monitored. The following rules apply to maintenance schedules:

- During the maintenance period, the management console flags data points and incidents for that server in the database and does not include them in reports.
- If you change a maintenance schedule, that change does not revise historical data.
- The management console still reports on server performance during a maintenance period, but does not open incidents for poor performance or low observation counts that occur during a maintenance period in OLA reports.
- When a maintenance schedule ends or begins, the management console closes open incidents and creates new incidents to reflect the current maintenance state.

Metrics

Metrics are parameters that the management console measures against a specific application on a specific server and network over a 5-minute reporting period. The management console also records the number of observations seen during the 5-minute period with the average value of the metric.

The management console reports the following metrics:

- Byte Volumes
- Connection Setup Time
- Data Rate (in bits/second) (to and from server); by packet (to and from server)
- Data Volume (in bytes) (to and from server); by packet (to and from server)
- Packet Volumes
- Percentile Throughput
- Composite Rate Per User
- Retransmission Delay
- TCP/IP Sessions - Complete
- TCP/IP Sessions - Open
- TCP/IP Sessions - Expired
- TCP/IP Session Times
- Users
- User Goodput

You cannot set thresholds for these metrics and the management console does not create incidents relative to these metrics. The metrics for which you can set thresholds are called *relative metrics*. The management console also reports on relative metrics.

Network Type

A network type is a classification of configured networks used to group networks of similar features and performance expectations. Use a network type to organize your enterprise for reporting, incident response generation, or service-level monitoring; for example:

- By bandwidth - 128K, T1, LAN
- By location - Midwest, HQ

You can use default the network types or create custom types.

Notification

A notification is a type of action associated with an incident response. A notification alerts a user or computer that a threshold has been met or exceeded for a specified period of time.

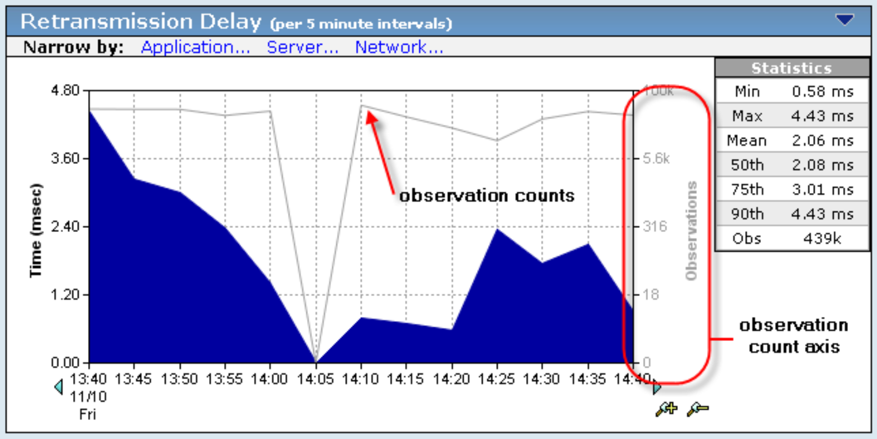
A notification can be one of the following:

- An email message sent to a specific person
- An SNMP trap sent to a receiver, such as a third-party monitoring platform

Observations

Observations are the number of monitored TCP transactions occurring during a specified time interval. Observations are opportunities for metric measurement. Specify a minimum number of observations for each metric to trigger an investigation and avoid notification unless a meaningful amount of traffic is present.

Observation counts appear in the management console views as a gray line. The axis on the right indicates the number of observations and is exponential, which includes count variations in the views.



Statistics	
Min	0.58 ms
Max	4.43 ms
Mean	2.06 ms
50th	2.08 ms
75th	3.01 ms
90th	4.43 ms
Obs	439k

Observation counts indicate utilization. The higher the observation count, the more heavily used the application, server, or network.

Use the observation count to determine an event’s significance. A high number of observations indicates that an event might impact many users, applications, or servers. If the number of observations for a server decreases concurrently with an increase in Server Connection Time, the increase in Server Connection Time might indicate that another operation, such as a backup, is running on the server.

Percentiles

Percentiles are percent calculations from a collection of 5-minute data points. The management console stores percentile data and uses the data in reports.

Rated Metrics

During a 5-minute reporting interval, if enough observations are noted, and a threshold is available for each combination of application, server, network, and 5-minute period, the management console rates the metric as Normal, Minor (yellow), or Major (orange). If the number of observations is insufficient or no threshold is available, the management console classifies the metric as Unrated.

Relative Metrics

Relative metrics have associated thresholds that the management console administrator can configure. The management console opens an incident when a threshold is crossed and calculates average measurements of relative metric parameters over a 5-minute period to report baselines.

Metrics described in the following sections are relative to the item they describe; for example, Network Round Trip Time relates to networks.

Network Metrics

Network metrics are relative to network performance.

Metric	Description
Network Round Trip Time	Time it takes for a packet to traverse the network.
Network Connection Time	Time it takes the client to confirm the server’s connection acknowledgment. Delay is probably caused by network latency.
Effective Round Trip Time	Network Round Trip Time plus delays caused by retransmissions.
Packet Loss Percentage	Ratio of retransmitted data to total data, percentage of data lost on the monitored network, and loss rate in packets per second. Part of the QoS view.

Server Metrics

Server metrics are relative to server performance.

Metric	Description
Server Response Time	Time it takes for the server to begin responding to a request.
Server Connection Time	Time it takes the server to acknowledge the initial client connection request.
Refused Sessions	Connection requests that were explicitly rejected by the server during the three-way handshake. Part of the Unfulfilled TCP/IP Session Requests view.
Unresponsive Sessions	Sessions where a connection request was sent, but the server never responded. Part of the Unfulfilled TCP/IP Session Requests view.

Understanding Combined Metrics

Combined metrics are relative to application performance and consist of both server and network metrics.

Metric	Description
Data Transfer Time	Time it takes to transmit a complete response as measured from the initial to final packet in a response. Data Transfer Time excludes the initial server response time and includes only Network Round Trip Time if there is more data to send than fits in the TCP window.
Transaction Time	Time elapsed from the moment a client sends the request (packet-level or transaction-level) to the point when the client receives the last packet in the response. The management console shows this type of response time data in the summary Response Time Composition: Average view in the Engineering page.

You can set a threshold for a relative metric. The management console calculates average measurements of relative metrics over a 5-minute period to report baselines.

Note: Metric definitions differ for WAN-optimized transactions.

Report Resolution

The default reporting resolutions for the management console reports are as follows:

- 5 minutes for up to 8-hour reports
- 15 minutes for longer than 8-hour and up to 24-hour reports
- 1 hour for longer than 1-day and up to 7-day reports
- 6 hours for longer than 7-day and up to 1-month reports

When you create custom time frames, select the report resolution from the following values:

- 5 minutes
- 15 minutes
- 30 minutes
- 1 hour
- 1 day

Use 5-minute intervals to view data over a time duration of fewer than 8 hours to report at the highest resolution. Use 1-day intervals when you have reporting durations greater than 1 month.

If you select a short interval for a long reporting period, the view might be difficult to read and trend because of the number of data points, and it might take a long time for the management console to generate.

Sample Resolution

The default sampling resolution for the management console reports is as follows:

- 5 minutes for 8-hour reports
- 15 minutes for daily reports
- 1 hour for weekly reports
- 6 hours for monthly reports

Operational Level Agreement (OLA)

Operational Level Management (OLM) includes procedures to ensure that adequate levels of service are delivered to all IT users in accordance with business priorities and at acceptable cost. By setting up and monitoring operational level agreements (OLAs) in the management console, you can determine whether levels of service are being met. View OLA reports on the Management page.

You would typically set an OLA on the following performance metrics:

- Server Response Time to quantify the data center's performance
- Network Round Trip Time to quantify the network infrastructure's performance
- Transaction Time to capture the user's experience with an application

Threshold Types

The management console automatically calculates thresholds from recent past performance by using percentiles and a sensitivity setting to classify each metric within a data sample.

Change thresholds by adjusting the sensitivity setting or by setting a percentage or milliseconds value.

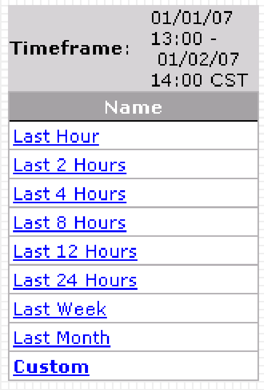
Threshold type	**Description**
None	(Off) Forces the management console to rate all data as U*nrated*. The management console does not create incidents for a metric with a threshold value of None.
Sensitivity	(Dynamic) Set a threshold of this type by using a percentage value. The management console factors this percentage value into the formula it uses to automatically calculate the threshold. A sensitivity setting of 200 causes traffic measuring over the 75th percentile to be rated Minor (yellow) or Major (orange). A lower sensitivity setting yields a higher threshold and thus results in fewer incidents.
Milliseconds	(Static) Set a threshold of this type to a specific a static value such as 100.

Throughput

The management console calculates throughput by dividing the bytes transferred by the elapsed time. The management console performs this calculation on a per-TCP transaction basis. The bytes transferred are from the server, not the total bytes. The throughput calculation is an appropriate measurement only for large transactions.

Timeframe

Use the Timeframe menu available through Settings on a report page to configure a view in the management console. The management console shows data for a time interval that ends at the current time. Select Custom to specify a time interval with an end date of other than today.



Timeframe: 01/01/07 13:00 - 01/02/07 14:00 CST

Name
Last Hour
Last 2 Hours
Last 4 Hours
Last 8 Hours
Last 12 Hours
Last 24 Hours
Last Week
Last Month
Custom

Log Into the Management Console

To start using the management console, access the server that hosts the console using the server name or IP address; for example:

http://<*IPAddress*>

The management console was designed for display in Microsoft Internet Explorer version 7 or 8 and Mozilla Firefox 3.6 and requires Adobe Flash Player. Contact the management console administrator for your login information.

View About Information

To view the management console version and revision information, click About on the menu bar.

- Version: Displays information about the installed version of the management console, the installation date and time, the HASP expiration date and time, and the configuration. It also includes a link to the product documentation.
- Revision History: Lists the history of installed versions and when they were installed.
- CA PC and CA NPC: Indicates the CA PC or CA NPC server address and product version number if you are attached as a data source to the CA PC or CA NPC.

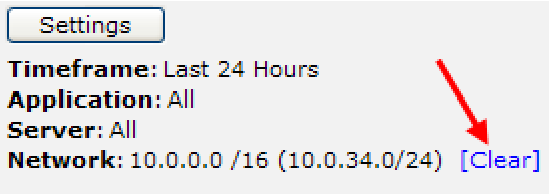
Navigation Tips and Tricks

As you become familiar with the management console interface, keep the following tips in mind.

- When you view report pages and views on the management console, navigate to reports by clicking networks, applications, metrics, incidents, and other listed items in the blue font to get detailed information.

- Watch for and click the link in any information boxes where you see this symbol: ❗. The management console uses the information boxes to guide you through processes and to prompt you for additional information.

- After you click a blue item in a report or view and examine the information, click [Clear] in the Settings area to return to the higher-level data or page.



Timeframe: Last 24 Hours
Application: All
Server: All
Network: 10.0.0.0 /16 (10.0.34.0/24)

- Hold your mouse over items such as controls or portions of views to see helpful ToolTip information. Hold your mouse over a bar representing an interface in a view to get details about the interface such as its router, description, and speed.

- Share information by printing, emailing a report page or a view, or saving a report page or view to a spreadsheet. To print, email, or export an entire report page, use the buttons at the top of the report page. To print, email, or export a view, use the blue gear menu (⚙) at the top of a view.

- Sort the data in views in ascending or descending order by clicking a column heading to change the sort order.

- If you add the management console as a data source in the CA Performance Center or CA NetQoS Performance Center, you can access the attached Performance Center from the management console by clicking the NPC link in the upper-right corner of the management console. For information about using the CA Performance Center or the CA NetQoS Performance Center, see the CA Performance Center or CA NetQoS Performance Center documentation.

Chapter 2: Using the Management Console

AITempty

This section contains the following topics:

Single Sign-On

Single Sign-On is the authentication scheme used by CA PC, CA NPC, and all supported data sources, including CA Application Delivery Analysis. Once a user is authenticated to CA PC or CA NPC, users can navigate among the CA PC, CA NPC, and registered data sources without signing in a second time.

Single Sign-On is installed automatically. A link to the CA PC or CA NPC which is attached, appears in the upper-right corner of the management console.

Edit Your User Profile

Edit your user profile, for example, to specify your local time zone. When registered as a data source with the CA PC or CA NPC, your changes are automatically synchronized within 5 minutes.

Follow these steps:

1. Click your user name on the menu bar.

   The User Profile page opens.

2. Change the following information:

   Password

   Type your new password.

   Confirm Password

   Retype your password.

   Email Address

   Type your email address.

   Time Zone

   Select your local time zone. The management console offsets report data to this time zone. The default time zone is CST6CDT, which is GMT minus 6 hours during Central Standard Time and minus 5 hours during Central Daylight Time. Use specific time zones instead of the Etc version of your time zone where possible because the Etc version does not account for daylight saving time and the offset is the opposite of what you would expect; that is Etc/GMT+4 is four hours behind GMT, not 4 hours ahead.

   Home Page

   Select the default page that you want to display when you access the management console.

3. Click Apply and then click OK.

Collected Data, Report Pages, and Views

Use the management console reports to determine whether a performance issue originates in the network infrastructure, server, or application. The management console generates reports using a combination of the following variables:

- Network
- Server
- Application
- Timeframe
- Metric

If you know the data you want to analyze, you can easily generate a report to show the required information.

The management console enables you to isolate performance problems to one or a combination of the elements listed in the following table.

Element	Possible Source of Performance Problem
Network infrastructure	Latency introduced by these: ■ Circuits ■ Traffic congestion ■ Routers ■ Switches
Server infrastructure	Latency introduced by these: ■ CPU processing ■ Memory I/O ■ Disk read and writes
Application architecture	Writing large data requests in many small packets for transmission

Determine the origin of performance problems using the following method to review and analyze the management console data:

1. Identify the application showing performance problems using Performance maps and Application Detail views.
2. Isolate the response-time components contributing to the performance problem. The colors in the Response Time Component view identify each component making up the total time.
3. Investigate the contributing factors to the performance problem by examining the Traffic, Sessions, Trends, Response Size, QoS, and Statistics pages.

   For more information on identifying and resolving performance issues, see Troubleshooting (see page 133).

Report Page Navigation

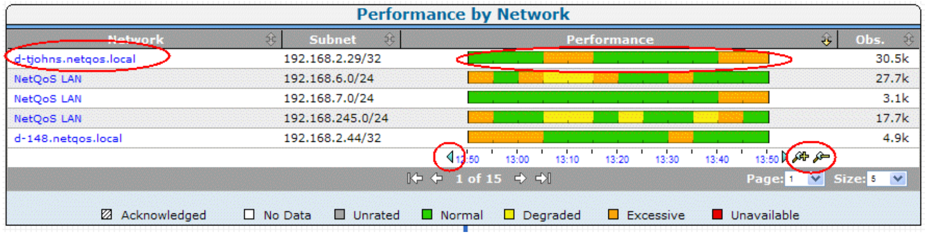
The tabbed pages correspond to functional roles in an IT department. For example, the Engineering pages let you quickly access reports that contain data of interest to a network engineer, starting with the Performance by Network report page. Each tabbed page presents data differently to facilitate troubleshooting, researching trends for optimization, or viewing summary reports.

Navigate to reports or select a view in the Show Me menu. Change the timeframe, selected metrics, applications, servers, and networks to show in a view or report by clicking Settings on the Operations, Incidents, Management, or Engineering, or Optimization views and report pages.

To filter data on a report page	See
Navigate to detailed information from the overview reports.	Navigate to a Report (see page 35)
Select a view in the Show Me menu.	Show Me Menu Navigation (see page 36)
Change settings for a view by clicking Settings.	Change Report Settings (see page 36)

Navigate to a Report

Access detailed information from the Operations reports by clicking links in the reports and other elements in the user interface as shown in the following sample Performance report.



Performance by Network

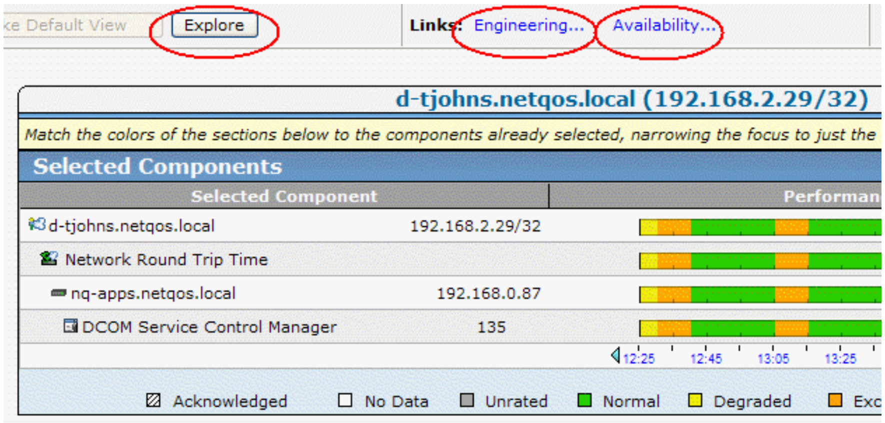
Network	Subnet	Performance	Obs.
d-tjohns.netqos.local	192.168.2.29/32		30.5k
NetQoS LAN	192.168.6.0/24		27.7k
NetQoS LAN	192.168.7.0/24		3.1k
NetQoS LAN	192.168.245.0/24		17.7k
d-148.netqos.local	192.168.2.44/32		4.9k

In the Performance report, view more information by doing the following:

- Clicking a network to view selected components for the network.
- Clicking arrows on the timeline to move the timeline backward or forward.
- Clicking the magnifying glasses to zoom in (narrow time frame) or zoom out (expand time frame).

You can also click a link to a related report. In the following example, clicking:

- Explore displays the associated Metric Detail reports.
- Engineering displays the Components reports for the selected components.
- Availability displays the Availability reports.



d-tjohns.netqos.local (192.168.2.29/32)

Match the colors of the sections below to the components already selected, narrowing the focus to just the

Selected Components

Selected Component		Performan
d-tjohns.netqos.local	192.168.2.29/32	
Network Round Trip Time		
nq-apps.netqos.local	192.168.0.87	
DCOM Service Control Manager	135	

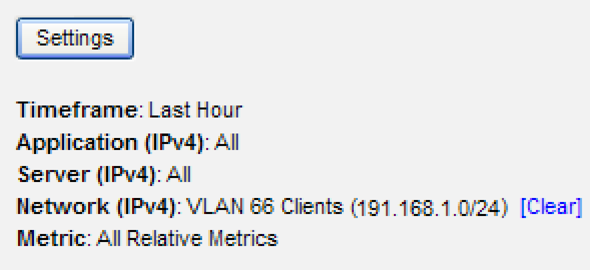
Show Me Menu Navigation

The six tabs at the top of the management console correspond to functional roles. Each tab presents data differently to facilitate troubleshooting, researching trends for optimization, or viewing summary reports.

On each tab, a Show Me menu lets you change your view of data on the page. On the Operations page, the Show Me menu lets you look at a data overview of a network, server, or application view.

Change Report Settings

The management console displays the current page and report settings below the Settings button at the top of each report page. In the following example, the management console shows data for the last hour for all applications and servers on the 191.168.1.0 network, and includes all relative metrics in the views and reports.



Timeframe: Last Hour
Application (IPv4): All
Server (IPv4): All
Network (IPv4): VLAN 66 Clients (191.168.1.0/24) [Clear]
Metric: All Relative Metrics

Click the Settings button to change the report settings:

Timeframe

Click the timeframe you want from the list, and choose whether to include data collected during schedule maintenance. Depending on the timeframe you choose, there can be a gap between the current time and the latest data in a report. For example, if you set the timeframe to a week, the reporting resolution is 60 minutes, so you may see up to a 55 minute gap.

- Less than or equal to 8 hours displays 5-minute increments
- Less than or equal to 16 hours displays 10-minute increments
- Less than or equal to 24 hours displays 15-minute increments
- Less than or equal to a week displays 60-minute increments
- Less than or equal to a month displays 6-hour increments

Metrics

Click to choose the metric or group of metrics you want. For example, from the Incidents page, you can choose All Server Metrics or a particular server metric.

Application/Server/Network Combination

Choose the combination of applications, servers and networks you want:

- Click the X button to clear an application, server, or network selection, and then make another selection.
- Click Select Group for All to apply a group filter from the CA PC or CA NPC to the applications, servers, and networks. To override the group filter, click the X button and then click the application, server or network you want.
- To search for an application, server or network, type the name you want and click Search.
- If you have created domains to separate duplicate IP traffic, choose the domain you want. This option filters the list of servers and networks to the corresponding domain. By default, the Default Domain applies to all servers and networks.

Drill to Details with the CA Multi-Port Monitor

The management console integrates with the CA Multi-Port Monitor to enable session-level visibility and packet export for session analysis, for example, in CA Observer Expert.

In the Operations, Incidents, or Engineering tab of the management console, the current reporting context, including the timeframe, network, server, and application settings, is applied to the drilldown views in the CA Multi-Port Monitor.

Unlike the CA Standard Monitor, which analyzes TCP sessions at the network level with a 5-minute granularity, the CA Multi-Port Monitor can analyze a TCP session between a server and a particular client at a 1-minute granularity. In addition, the CA Multi-Port Monitor enables you to analyze traffic volume for all traffic observed by the CA Multi-Port Monitor, including TCP and non-TCP traffic.

To follow a troubleshooting path from a management console report to a default analysis in the CA Multi-Port Monitor, we recommend that you start with a relatively narrow timeframe, such as one hour, selected in the management console report settings. Any filtering you apply to the report, such as narrowing the data to a single network, server, or application, remains in place after drilldown to the default view in the CA Multi-Port Monitor and can help direct your troubleshooting efforts toward the right area of the network.

Follow these steps:

1. Click Settings in the Operations, Incidents, or Engineering tab.

   The Settings dialog box opens.

2. Configure the report settings and click OK.

   To enable the drilldown to automatically select the logical port that observes the server traffic, you must configure the report settings in the management console to select a particular server.

3. Click Session Analysis.

   If the Session Analysis button is disabled, configure the report settings to choose a server that is monitored by the CA Multi-Port Monitor.

   The CA Multi-Port Monitor opens to the Analysis Menu displays response time metrics based on the context of the report settings.

   For more information about working with the Analysis Menu, see the CA Multi-Port Monitor product documentation.

   Tip: If you configure the report settings to show All Servers, and the CA Multi-Port Monitor is configured with more than one logical port, the Session Analysis dialog box prompts you to choose the logical port that observes the server traffic you want to analyze. Select the appropriate logical port and click OK to open the CA Multi-Port Monitor.

Change Report Format from a Chart to a Table

You can change a report format from a chart to a table or a detailed table or change a table to a chart. You can also change some report pages from a chart to a larger chart.

Follow these steps:

1. Click the Engineering page.
2. Click the blue gear menu (⚙) on the top-right corner of a chart and click Table.

   Note: Some charts let you select Larger Chart as the display format.
3. The management console shows the view in tabular format. To return to the chart view, click the blue gear menu (⚙) and click Charts.

Interpret Data in Report Pages and Views

If you have enough observations, and if a threshold is available for each combination of applications, server, network, and 5-minute period, the management console classifies the metric using one of these ratings:

Acknowledged

Indicates a management console user acknowledged an incident. When this happens, the management console marks data the incident covers as Acknowledged. The management console automatically marks future data covered by an acknowledged incident as Acknowledged.

No Data

Indicates no data is available.

Unrated

An *Unrated performance rating*, indicated by a gray severity state on the Operations page of the management console, means that either there is insufficient past data to establish a threshold (two full business days of data are needed), or there were not enough observations to exceed the minimum observations threshold.

Normal

Indicates the metric value is between zero and the Minor threshold.

Minor

Indicates the metric value exceeds the Minor threshold.

Major

Indicates the metric value exceeds the Major threshold.

Unavailable

Indicates the application on a server is not running (unavailable). This rating appears when you click All Server Metrics in Settings. This rating is only applicable to a user-defined application where the management console administrator has assigned servers to the application.

Chapter 3: Using the Operations Page

This section contains the following topics:

Use Operations Report Pages

Use the Operations page to explore and troubleshoot performance problems. A *performance map* bubbles up the worst-performing network, server, and applications to the top of the page. As you click these items, the report focus narrows.

Typically, end users report the following performance problems to operations centers and Support teams:

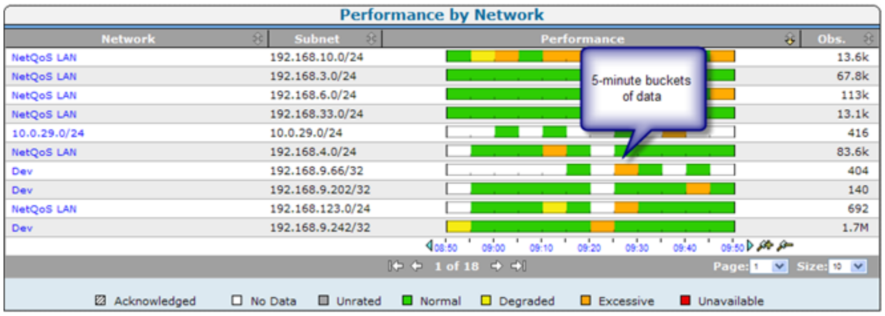
- Performance issue with an application
- Performance issue with a network
- Performance issue with a server

The performance maps collect the following data about performance issues:

- Name of application and server exhibiting performance problem
- Location and IP address of user, local, or remote office location to identify the network
- Time the performance problem first occurred
- History of performance problem and whether it reoccurred
- Other performance issues related to the problem

The Operations page shows performance maps by network, server, or application as a horizontal, color-coded bar chart that compares overall performance elements against defined thresholds. The color coding corresponds to a data rating.

Increments on the performance bar represent the average value for a five-minute period.



Performance by Network

Network	Subnet	Performance	Obs.
NetQoS LAN	192.168.10.0/24		13.6k
NetQoS LAN	192.168.3.0/24		67.8k
NetQoS LAN	192.168.6.0/24		113k
NetQoS LAN	192.168.33.0/24		13.1k
10.0.29.0/24	10.0.29.0/24		416
NetQoS LAN	192.168.4.0/24		83.6k
Dev	192.168.9.66/32		404
Dev	192.168.9.202/32		140
NetQoS LAN	192.168.123.0/24		692
Dev	192.168.9.242/32		1.7M

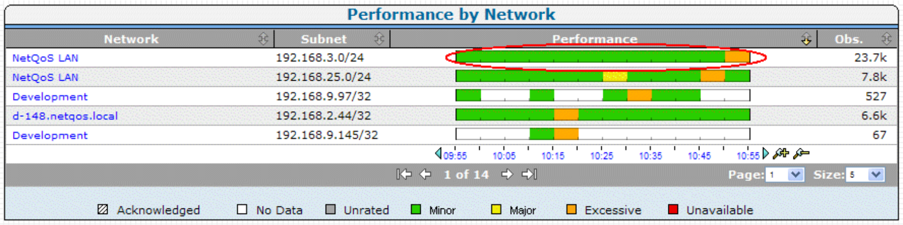
5-minute buckets of data

More information:

Email a Report Page (see page 130)
Print a Report Page (see page 131)
Change Report Settings (see page 36)
Export a Report to a File (see page 129)
Navigate the Operations Report Pages (see page 43)

Navigate the Operations Report Pages

To view the networks, servers, and applications that are related to a performance degradation, click one of the worst-performing components at the top of a view.



Performance by Network

Network	Subnet	Performance	Obs.
NetQoS LAN	192.168.3.0/24		23.7k
NetQoS LAN	192.168.25.0/24		7.8k
Development	192.168.9.97/32		527
d-148.netqos.local	192.168.2.44/32		6.6k
Development	192.168.9.145/32		67

For example, if you click the performance bar for the NetQoS LAN network, the management console displays the network metrics along with the affected servers and applications. To view servers and applications that are not affected by the Minor network performance degradation, click to expand the Unrelated link. The Unrelated folder includes components that are rated as Normal, are Unrated, or do not have any data.

Note that the management console displays an icon in front of a component to indicate its type. The following icons do not indicate the status of a component:

Icon	Description
🖧	Network
▬	Server
▣	Application

View Details for a Component

Investigate a performance problem by viewing the details for a component. Details include Related Metrics, Impacted Metrics, and Impacted Users.

Follow these steps:

1. Click the Operations page.
2. Click Explore under the Show Me menu.
3. Click the following tabs to investigate performance and change performance boundaries.

   Related Metrics

   This tab shows other metrics related to the problem. Use the information on this tab to help diagnose the problem. The chart for each metric includes the following:

   - Related statistic or averages for the metric. Statistics is the minimum, maximum, and mean; 50th, 75th, and 90th percentiles; and number of observations. Averages is the number of bytes or packets to and from the server.
   - Baseline performance, which you can use to compare against the actual data.

   Impacted Metrics

   This tab shows other metrics impacted by the problem. The chart for each metric includes the following:

   - Related statistic or averages for the metric. Statistics is the minimum, maximum, and mean; 50th, 75th, and 90th percentiles; and number of observations. Averages is the number of bytes or packets to and from the server.
   - Baseline performance, which you can compare against the actual data.
   - Right and left arrows that enable you to scroll backward or forward in time.
   - Grid that enables you to center the view.

   Impacted Users

   This tab shows the users impacted by the performance problem. For each user, the tab shows the IP address, subnet mask, and host name.

   Edit Thresholds

   This tab enables you to edit the thresholds that define the performance ratings. For more information about setting thresholds, see the *Administrator Guide*.

View Baselines

Baselines help you determine whether a performance condition is normal. The management console includes the following parameters in the baseline calculations. Configure these parameters by setting the associated database parameter.

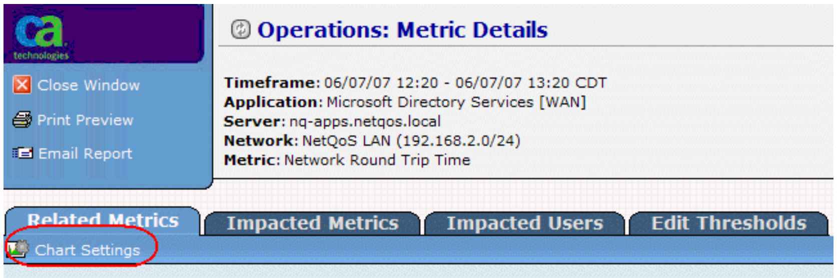
Parameter	**Default**	**Database parameter**	**Description**
Days Back Factor	7	maxDaysBack	The maximum number of days back from which to derive the baseline. The default is the past week.
Weeks Back Factor	12	maxWeeksBack	The maximum number of weeks back from which to derive the baseline for day of the week. The default is the last quarter.
Months Back Factor	6	maxMonthsBack	The maximum number of months back from which to derive the baseline for day of the month.

Follow these steps:

1. Click the Operations page.
2. Click a network, server, or application.
3. Narrow the selection by selecting a metric, network, application, or server.
4. Click Explore under the Show Me menu.

   The Operations: Metric Details page opens.

5. Click Chart Settings.



Operations: Metric Details

Timeframe: 06/07/07 12:20 - 06/07/07 13:20 CDT
Application: Microsoft Directory Services [WAN]
Server: nq-apps.netqos.local
Network: NetQoS LAN (192.168.2.0/24)
Metric: Network Round Trip Time

   Chart Settings opens.

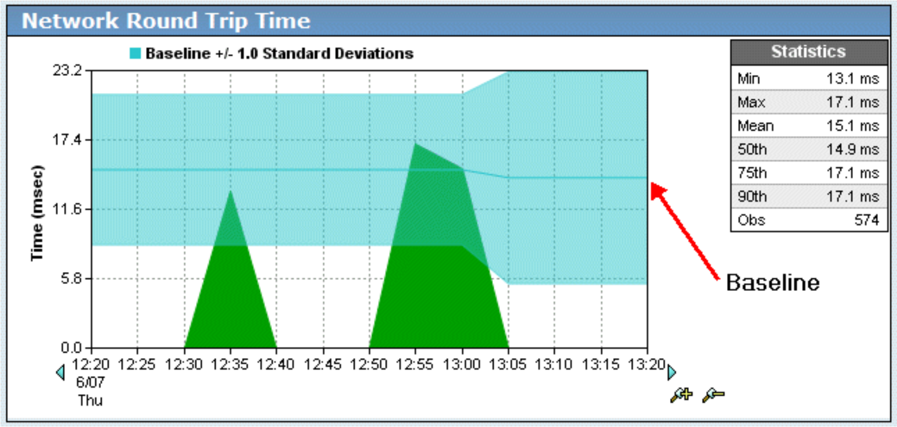
6. Configure the Primary Axis settings:

   a. Click Metric and Baseline

   b. Select a baseline from the Show box.

   c. Click OK.

   The detail view shows the selected baseline.



Statistics	
Min	13.1 ms
Max	17.1 ms
Mean	15.1 ms
50th	14.9 ms
75th	17.1 ms
90th	17.1 ms
Obs	574

View Incidents for a Component

The Incidents page lists the incident number, target, application, severity, time, and duration.

Follow these steps:

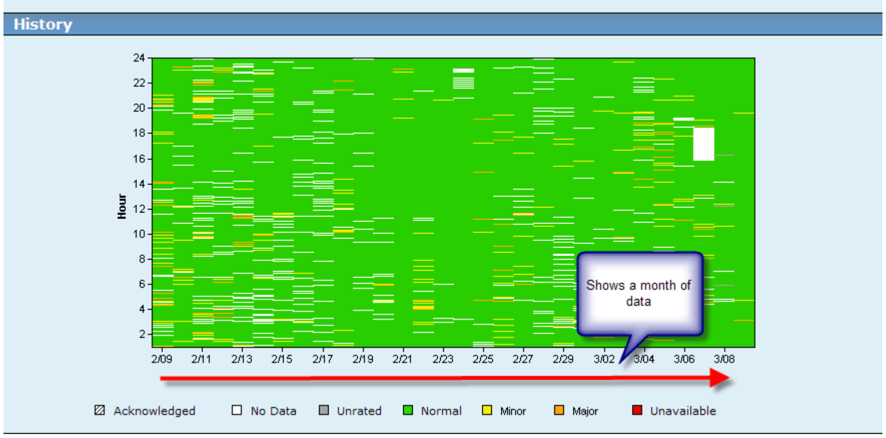
1. Click the Operations page.
2. Click the performance bar for the network, server, or application for which you want to see incidents.
3. Click Incidents in the Show Me menu, to view an Incidents page for the selected component.

View Historical Data for a Component

View historical data for a network, server, or application to analyze performance behavior over time.

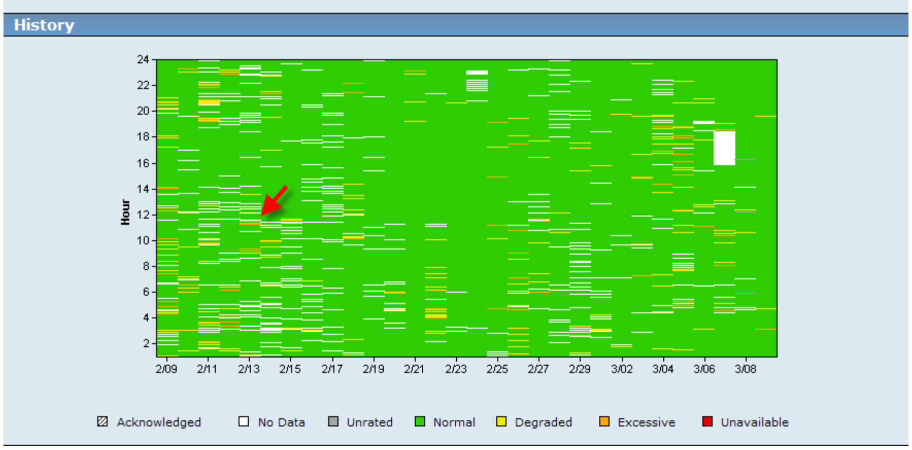
Follow these steps:

1. Click the Operations page
2. Click the performance bar for the network, server, or application for which you want to see historical data.
3. Click History in the Show Me menu. A chart displays data from the previous month for the selected item.



Shows a month of data

4. Click a specific data point in the chart to display the Selected Components view for that hour and date.



Troubleshoot an Application Performance Problem

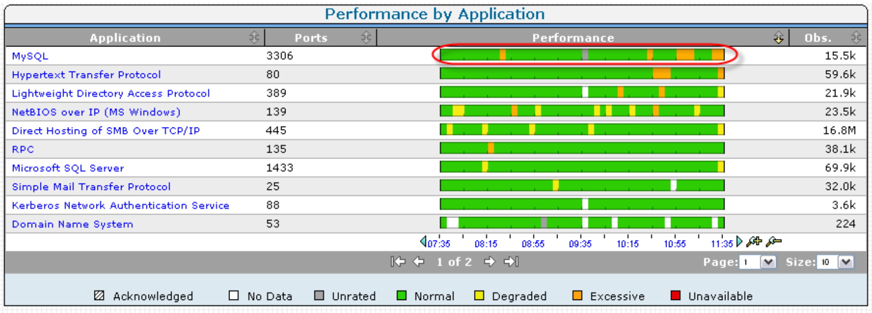
Use this procedure to troubleshoot a performance problem with an application. You can use the same procedure to investigate a problem with a network or server by clicking a link to narrow by metric, network, or server.

A user from corporate headquarters reports a performance problem with an application.

Follow these steps:

1. Click the Operations page.
2. Click Applications in the Show Me menu.
3. Determine if the application in question has bubbled to the top of the Performance by Application list.

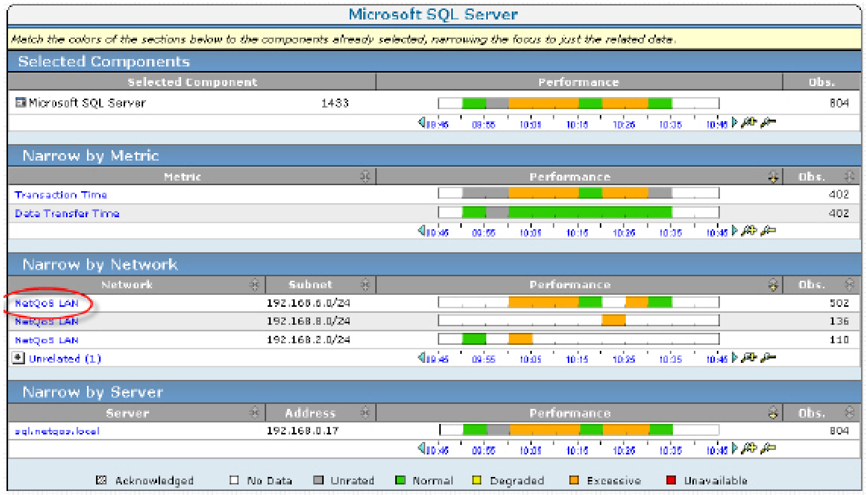
   If the application does not appear at the top of the list, select a larger Size setting. If the application does not appear on the list, the management console might not monitor it.
4. Click the performance bar next to the application name to show details about the application.

   Performance by Application

Application	Ports	Performance	Obs.
MySQL	3306		15.5k
Hypertext Transfer Protocol	80		59.6k
Lightweight Directory Access Protocol	389		21.9k
NetBIOS over IP (MS Windows)	139		23.5k
Direct Hosting of SMB Over TCP/IP	445		16.8M
RPC	135		38.1k
Microsoft SQL Server	1433		69.9k
Simple Mail Transfer Protocol	25		32.0k
Kerberos Network Authentication Service	88		3.6k
Domain Name System	53		224

   A detail page shows related metrics, networks, and servers you can use to navigate into data related to the problem.
5. (Optional) If you know the network or server that hosts the application, click the appropriate performance bar to narrow the scope of data.



Microsoft SQL Server

Match the colors of the sections below to the components already selected, narrowing the focus to just the related data.

Selected Components

Selected Component		Performance	Obs.
Microsoft SQL Server	1433		804

Narrow by Metric

Metric	Performance	Obs.
Transaction Time		402
Data Transfer Time		402

Narrow by Network

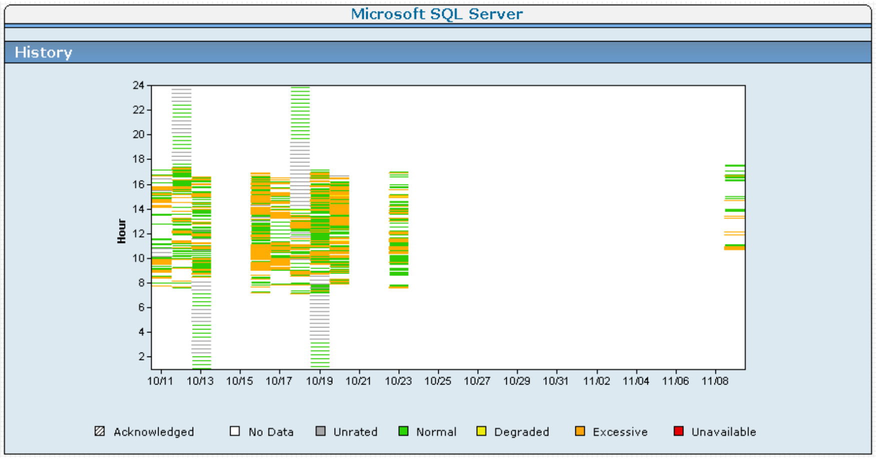
Network	Subnet	Performance	Obs.
NetQoS LAN	192.168.6.0/24		502
NetQoS LAN	192.168.8.0/24		136
NetQoS LAN	192.168.2.0/24		110
⊞ Unrelated (1)			

Narrow by Server

Server	Address	Performance	Obs.
sql.netqos.local	192.168.0.17		804

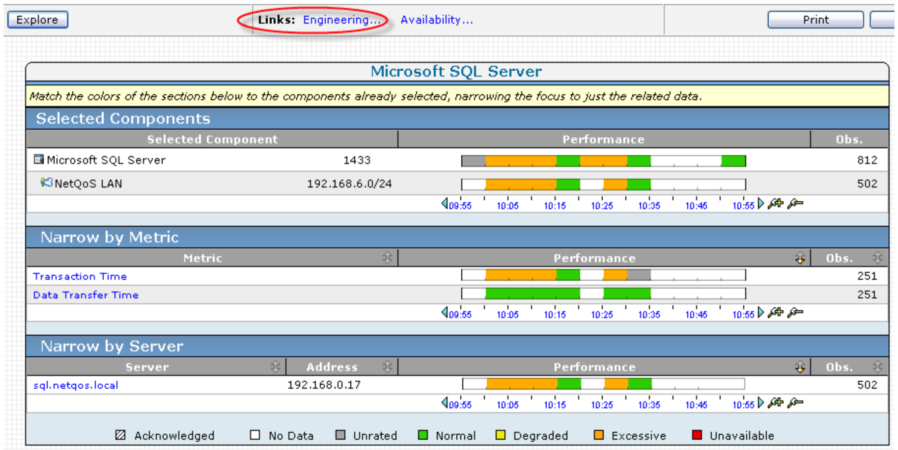
6. Click History in the Show Me menu.

7. Review the application history on the History page to identify and note systematic patterns of unavailability or compromised performance.



8. Click Performance in the Show Me menu, and then Explore to begin troubleshooting.

9. Perform more exhaustive troubleshooting by clicking the Engineering link at the top of the detail page.



Microsoft SQL Server

Match the colors of the sections below to the components already selected, narrowing the focus to just the related data.

Selected Components

Selected Component		Performance	Obs.
Microsoft SQL Server	1433		812
NetQoS LAN	192.168.6.0/24		502

Narrow by Metric

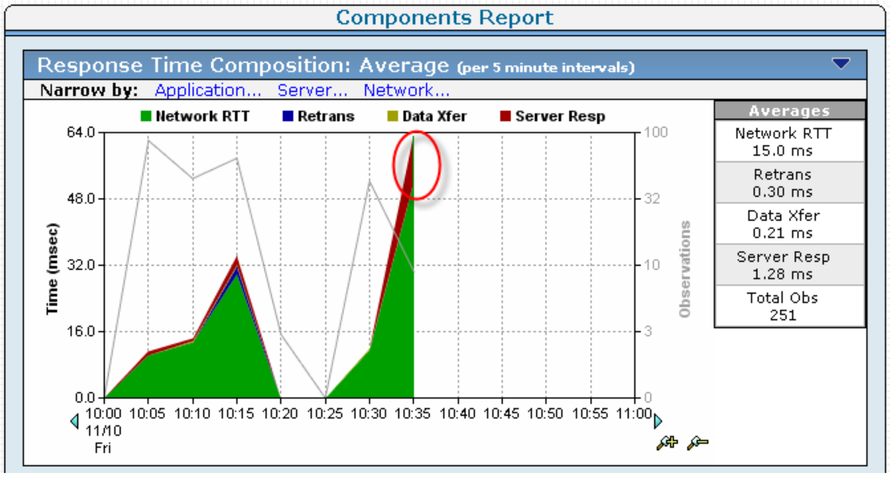
Metric	Performance	Obs.
Transaction Time		251
Data Transfer Time		251

Narrow by Server

Server	Address	Performance	Obs.
sql.netqos.local	192.168.0.17		502

▨ Acknowledged □ No Data ■ Unrated 🟩 Normal 🟨 Degraded 🟧 Excessive 🟥 Unavailable

10. Review the Response Time Composition: Average view on the Components Report page to pinpoint the time the reported issue occurred.



Averages
Network RTT 15.0 ms
Retrans 0.30 ms
Data Xfer 0.21 ms
Server Resp 1.28 ms
Total Obs 251

11. Identify which metric is the significant contributor to the performance problem at that point. If the metric that sees an increase in response time is:

    - Server Response Time (SRT), troubleshoot an increase in Server Response Time (see page 142)

    - Network Round Trip Time (NRTT), troubleshoot an increase in Network Round Trip Time (NRTT) (see page 147)

    - Retransmission Delay, troubleshoot an increase in Network Round Trip Time (NRTT) (see page 147)

    - Data Transfer Time, troubleshoot an increase in increase in Data Transfer Time (see page 151)

If the metrics report acceptable performance with respect to the historical timeline, the problem might be with the user’s computer.

Chapter 4: Using the Incidents Page

This section contains the following topics:

Use the Incidents Page

The management console creates an incident or record of information each time a threshold is crossed or a Operational Level Agreement (OLA) is not met. If the management console administrator has configured incident responses, the management console automatically launches action to collect more information or notify someone about the performance issue. If the incident response does not include enough information to help you resolve a problem, you can set up a troubleshooting investigation to gather more data.

The Incidents page lists sequentially numbered incidents. Incident reports show details of the related performance degradation and have a maximum time window of 24 hours, which you can shift to include the time of interest. Because incidents open and close according to incident creation rules, a set of consecutive incidents might represent an extended period of poor performance.

The management console stores historical incident records as long as it stores the 5-minute data. For information about changing the retention period for 5-minute data, see the *Administrator Guide*.

Incidents, Incident Responses, and Investigations

An *incident* is a record of information created when a performance threshold is crossed. *Thresholds* are boundaries of acceptable performance behavior, which exist by default for each application. Administrators can change thresholds to make them more or less sensitive to performance changes.

When thresholds are crossed, the management console creates incidents with assigned sequential case numbers and reports these incidents on the Incidents page. If the management console administrator configured incident responses and associated them with the violated thresholds, the management console launches one or more automatic responses. If you need more information to resolve a problem, you can launch an investigation to troubleshoot the problem.

The management console closes an incident if the following conditions are true:

- One full hour of acceptable performance has passed.
- The server in question enters a maintenance window.
- The incident is 24 hours old. If the problem persists, the management console opens a new incident.

When the management console is registered with the CA PC or CA NPC, and the management console detects a new Minor or Major condition, the management console opens an incident and CA PC or CA NPC opens a corresponding event.

- If the Minor or Major condition that triggered a management console incident exhibits one clock-hour of normal performance, the management console closes the incident, and the CA PC or CA NPC clears the corresponding event. Afterwards, if the Minor or Major incident condition returns, the management console opens a new incident and CA PC or CA NPC opens a corresponding event.
- If a management console incident remains Open for 24 hours, the management console automatically closes the incident regardless of its condition. If the Minor or Major incident condition still exists, the management console opens a new incident and CA PC or CA NPC increments the count of the corresponding event. Otherwise, after a synchronization delay of approximately 10 minutes, the CA PC or CA NPC clears the event.

To enable the CA PC or CA NPC to clear events that are associated with a management console incident where a server was offline, if the management console reports a full hour of No Data for a Minor or Major incident condition, CA PC or CA NPC clears the corresponding event. If the management console detects a new Minor or Major condition after a server is brought back online, the management console opens an incident and CA PC or CA NPC opens a corresponding event.

In CA PC or CA NPC, if a user closes an event that corresponds to a management console incident, the incident status in the management console changes to Acknowledged. After the incident condition exhibits one clock-hour of normal performance, the management console automatically Closes the incident.

View Incidents

The management console shows sequentially numbered incidents for networks, servers, and applications on the Incidents page and shows incidents for monitoring devices on the Administration page.

Incident reports show details of the related performance degradation and have a maximum time window of 24 hours. You can shift that window to include the time of interest. Because incidents open and close according to incident creation rules, a set of consecutive incidents might represent a single extended period of poor performance.

When registered with the CA PC or the CA NPC, the management console synchronizes its incidents with the registered CA PC or CA NPC:

- If you acknowledge an incident in the management console, the event status for the corresponding incident is updated automatically in CA PC or CA NPC.
- If you acknowledge an event in the Event Manager that was the result of a management console incident, in the management console, the corresponding incident status updates automatically.

For more information about how the CA PC or CA NPC synchronizes management console incidents, see the CA PC or CA NPC product documentation.

Note: Historical incident records are stored as long as the 5-minute data is stored. You can configure the retention period on the Administration page.

View Incidents by Monitoring Device

Monitoring device incidents are displayed on the Administration page.

Note: To view monitoring device incidents, your user account must have the Administrator product privilege. For more information, see the *Administrator Guide*.

View Incidents by Network, Server, or Application

The management console opens a Network or Server incident when the corresponding threshold for a Network, Server, or Combined metric is exceeded during a 5-minute interval. For example, if the threshold for Data Transfer Time is exceeded, the management console can create a Network or Server incident for the application.

Follow these steps:

1. Click the Incidents page to display a list of incidents.
2. To view incidents of interest:
    1. Select an option in the Show Me menu to display corresponding incidents.
    2. Click Settings to specify additional filter criteria for the incidents you want, such as a particular timeframe or network.
    3. Filter the list of incidents by choosing from following options:

        Incident State

        Specifies the incident state or states you want.

        Minimum Severity for

        Specifies the minimum incident severity you want and the minimum amount of time that the incident condition must persist. Note that Unavailable is the highest severity.

        View By

        Specifies display options for the list of incidents.

View Investigations Related to Incidents

Follow these steps:

1. Click the Incidents page
2. Click Overview in the Show Me menu.

    The incident lists open.
3. Click a link in an incident list to see the incident details.

    The incident detail page opens.
4. Click Investigations in the second Show Me menu.

    The investigations related to the incident appear if any exist.

View Incident Details

Use the Incident lists to drill into Network or Server incident details.

Follow these steps:

1. Click the Incidents page.
2. If the View By filter on the page is configured to display incident counts, click the performance bar in the Incident Count column to view a list of incidents.

   The list of incidents is filtered based on your selection.
3. Click an incident number to view its incident details.

   The incident details page opens.
4. Narrow the information by selecting a metric, server, and application.

   The management console lists the following information for the incident:

   Number

   Unique number that identifies the incident.

   Network

   Network source of the incident if this is a network-related incident.

   Server

   Server source of the incident if this is a server-related incident.

   Severity

   Color-coded rating of the incident (see page 40).

   Time Frame

   Total duration of the incident.

   Investigations

   Link to related investigations.

   Status

   Open or Closed.

   Selected Components

   Lets you select components to narrow the data.

   Performance bar

   Uses color-coded segments to show where problems exist.

   Obs.

   Number of observations.

Explore Incidents

Use the Explore button under the Show Me menu to view related metrics, impacted metrics, impacted users, or edit thresholds.

Follow these steps:

1. Click the Incidents page.
2. Click Overview, Applications, Servers, or Networks to view incidents for the selected component.
3. Click the selected component to view incident details.

   The Explore button is no longer grayed out.
4. Click Explore under the Show Me menu.
5. The Explore dialog box is displayed. The:

   Metrics tab

   Displays related metrics

   Impacted Metrics

   Displays impacted metrics.

   Impacted Users

   Displays a list of client IP addresses that accessed the application during the specified timeframe.

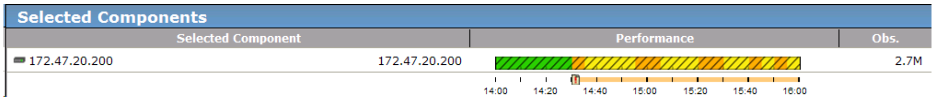
   Edit Thresholds

   Edits the performance thresholds for the selected metric.

Acknowledge Incidents

Acknowledging an incident reduces its priority in reports and indicates that you reviewed the incident. You can also unacknowledge an acknowledged incident to raise its priority in reports.

When you acknowledge the incident, it has the effect of acknowledging the entire lifetime of the incident, from top-of-hour to top-of-hour. In the following example, the incident detail page shows hash marks on the time line to indicate the incident status is Acknowledged from 14:00 to 16:00.



Selected Components

Selected Component		Performance	Obs.
172.47.20.200	172.47.20.200		2.7M

Follow these steps:

1. Click Acknowledge on the detail page for an incident.
2. Select an option on the Acknowledgement Confirmation page and click OK:
   - Acknowledge Incident: Select this option to mark the incident as acknowledged.
   - Unacknowledge Incident: Select this option to mark the incident as unacknowledged.

Compare Incidents

Use the Compare page to look for patterns across the networks, servers, or applications that might be related, such as signs of degradation. Click a particular network, server, or application to explore the details. You can also view incident details to see whether there are related investigations.

Comparing a:

Network incident

Compares the network performance related to the incident against similar networks.

Server incident

Compares server performance related to the incident against similar servers.

Follow these steps:

1. Click the Incidents page.
2. If the View By filter on the page is configured to display incident counts, click the performance bar in the Incident Count column to view a list of incidents.

   The list of incidents is filtered based on your selection.
3. Click an incident number to view its incident details.

   The incident details page opens.
4. Click Compare in the Show Me menu to compare the performance of the selected network or server to all networks or servers during the specified time frame.

View Incident History

View incidents that occurred over the last 30 days with the same target as the currently selected incident.

Follow these steps:

1. Click the Incidents page.
2. If the View By filter on the page is configured to display incident counts, click the performance bar in the Incident Count column to view a list of incidents.

   The list of incidents is filtered based on your selection.
3. Click an incident number to view its incident details.

   The incident details page opens.
4. Click History in the Show Me menu.

   The Incident History list is displayed.
5. Click an incident number to see its details.

View Network and Server Incidents by Application

The Incidents by Application report lists the network and server incidents for an application.

Incident report details show the related performance degradation, and have a maximum time window of 24 hours. You can shift that window to include the time of interest. Because incidents open and close due to incident creation rules, a set of consecutive incidents might represent a single extended period of poor performance.

To navigate to the Incidents by Application report, click Applications on the Incidents page.

View Server Incidents

The Incidents page lists server incidents by application. The incident report shows details of the related performance degradation and have a maximum time window of 24 hours. You can shift that window to include the time of interest. Because incidents open and close according to incident creation rules, a set of consecutive incidents might represent a single extended period of poor performance

To navigate to the Server Incidents page, click Servers on the Incidents page.

View Network Incidents by Application

The Incidents page lists network incidents by application. The incident report shows shows details of the related performance degradation and have a maximum time window of 24 hours. You can shift that window to include the time of interest. Because incidents open and close according to incident creation rules, a set of consecutive incidents might represent a single extended period of poor performance.

To navigate to the Network Incidents page, click Networks on the Incidents page.

Use the Investigations Report

An *investigation* is an action that the management console automatically initiates as part of an incident response or that is manually scheduled or launched by a user with Administrator product privileges. An investigation can be one of the following types:

Investigation Type	**Description**
Application Connection Time	Determines the time it takes to connect to an TCP/IP application port, which includes time for the server to respond with a connection acknowledgment.
Packet Capture	Enables and investigates a filtered capture of the particular server, application port, and network experiencing a problem.
Performance via SNMP	Investigates server or router performance-related information collected using the Simple Network Management Protocol (SNMP).
Ping Response Time	Measures the time it takes to receive a ping reply after sending a TCP ping request.
Ping Response Time vs. Packet Size	Measures the time it takes to receive a ping reply for TCP ping requests (data packets) of various sizes. Helps track excessive delays and lack of connectivity at various packet sizes.
Trace Route	Records the path and each hop between the management console and end points to detect latency and routing issues.

View Investigations

Use the Investigations Report to view details about an investigation, change the report settings, or delete an investigation.

Follow these steps:

1. Click the Incidents page.
2. Click Investigations.

   The Investigations Report page is displayed.
3. Click Settings to filter the list of incidents.
4. Click a timestamp link to view details about the results of an investigation.

Launch and Schedule Investigations

Set up an investigation to launch immediately or on a scheduled date and time.

When launching an investigation manually, you must disable your pop-up blockers, otherwise the investigation will be blocked from running. If you run an investigation and do not see a pop-up showing the status of the investigation as it runs, then your pop-up blocker is most likely still enabled, and your investigation did not run.

Follow these steps:

1. Click the Incidents page.

   Click Investigations in the Show Me menu.
2. The Investigations Report page opens.
3. Click Launch in the Show Me menu.

   The Investigation Types page opens.
4. Launch or schedule an investigation by performing one of the following tasks:

   - Click Launch next to the investigation type that you want to launch immediately.
   - Click Schedule next to the investigation type that you want to schedule.

   A Settings page appropriate for the type of investigation you selected opens. For information about each Settings page, see the following sections.

Application Connection Time Investigation

An *Application Connection Time investigation* determines the time it takes to connect to an TCP/IP application port, which includes time for the server to respond with a connection acknowledgment.

When scheduling this investigation, the time zone you specify for the notification is also used to schedule the investigation.

Specify the following settings to launch or schedule the Application Connection Time investigation:

Target of Investigation

Specifies the target you want to investigate, such as a server, based on its target type. To investigate a group of targets, choose Server Aggregation from the Target Type list.

Investigation Options

Specifies the following options:

Investigate from

Specifies the monitoring device from which you want to launch the investigation. Choose a monitoring device that can communicate with the server or device you want to investigate.

Application

Specifies the application to investigate.

Samples

Specifies the number of data samples, from 1 to 10, to observe during the investigation. A sample is a single statistical measurement.

Timeout (seconds)

Specifies the number of seconds, from 1 to 10, that must elapse before the investigation logs an application time out.

Notification Options

Notifies the specified email recipients about the results of the investigation. Note that the investigation is scheduled using the specified time zone for the email notification.

Schedule

The scheduling options only appear if you chose to schedule the investigation. Schedule the investigation to run at a particular time on a particular date or schedule the investigation on a weekly or monthly basis.

Packet Capture Investigation

A *Packet Capture investigation* performs a filtered capture of the particular server, application port, and network experiencing a problem.

When scheduling this investigation, the time zone you specify for the notification is also used to schedule the investigation.

Specify the following settings to launch or schedule a Packet Capture investigation:

Target of Investigation

Specifies the server you want to investigate.

Capture Filters

Specifies the following options:

Application

Specifies the application traffic you want to capture. To capture all application traffic on the target server, click All. To specify a custom port range, click Custom Application and specify the beginning and ending port numbers.

Network

Specifies the client networks for which you want to capture the target application traffic.

Click All to specify all client networks or click Specific Network to specify a particular network, and then click the Network link to choose the network you want.

Investigation Options

Specifies the following options:

Capture Period

Specifies the maximum time period, from 30 seconds to 30 minutes, to capture packets.

Maximum File Size

Specifies the maximum size, from 10 MB to 100 MB, for the packet capture file.

Bytes Per Packet

Specifies the number of bytes per packet to capture when capturing packets with a CA Standard Monitor. Choose between Header Only and 8192 bytes. Note that Header Only captures the MAC (Layer 2), IP (Layer 3), and TCP (Layer 4) header information.

This option is not applicable when the packets are captured by a CA GigaStor or CA Multi-Port Monitor.

Notification Options

Notifies the specified email recipients about the results of the investigation. Note that the investigation is scheduled using the specified time zone for the email notification.

Schedule

The scheduling options only appear if you chose to schedule the investigation. Schedule the investigation to run at a particular time on a particular date or schedule the investigation on a weekly or monthly basis.

Performance via SNMP Investigation

A *Performance via SNMP investigation* queries a server or router for performance-related information using the Simple Network Management Protocol (SNMP).

When scheduling this investigation, the time zone you specify for the notification is also used to schedule the investigation.

Specify the following settings to launch or schedule a Performance via SNMP investigation:

Target of Investigation

Specifies the target you want to investigate, such as a server, based on its target type. To investigate a group of targets, choose Server Aggregation from the Target Type list. To SNMP poll a router for performance information, the management console administrator must add the router as a network device. For more information, see the *Administrator Guide*.

Investigation Options

Specifies the following options:

Investigate from

Specifies the monitoring device from which you want to launch the investigation. Choose a monitoring device that can communicate with the server or device you want to investigate.

Sample Period (seconds)

Specifies the amount of time, from 5 to 300 seconds, to wait between samples for calculating rates.

Retries

Specifies the number of attempts, from zero to 4, to get a response.

Timeout (seconds)

Specifies the amount of time, from 1 to 30 seconds, that must elapse before the management console considers a server to be timed out.

Notification Options

Notifies the specified email recipients about the results of the investigation. Note that the investigation is scheduled using the specified time zone for the email notification.

Schedule

The scheduling options only appear if you chose to schedule the investigation. Schedule the investigation to run at a particular time on a particular date or schedule the investigation on a weekly or monthly basis.

Ping Response Time Investigation

A *Ping Response Time investigation* measures the time it takes to receive a ping reply after sending a TCP ping request.

When scheduling this investigation, the time zone you specify for the notification is also used to schedule the investigation.

Specify the following settings to launch or schedule a Ping Response Time investigation:

Target of Investigation

Specifies the target you want to investigate, such as a server, based on its target type. To investigate a group of targets, choose Server Aggregation from the Target Type list.

Investigation Options

Specifies the following options:

Investigate from

Specifies the monitoring device from which you want to launch the investigation. Choose a monitoring device that can communicate with the server or device you want to investigate.

Packet Size

Select from 32 to 8192 for the size in bytes of the test packets.

Samples

Select from 1 to 10 data samples to observe during the investigation. A sample is a single statistical measurement.

Timeout (seconds)

Select from 1 to 10 seconds that must elapse to consider a server timed out.

Notification Options

Notifies the specified email recipients about the results of the investigation. Note that the investigation is scheduled using the specified time zone for the email notification.

Schedule

The scheduling options only appear if you chose to schedule the investigation. Schedule the investigation to run at a particular time on a particular date or schedule the investigation on a weekly or monthly basis.

Ping Response Time vs Packet Size Investigation

A *Ping Response vs. Packet Size investigation* measures the time it takes to receive a ping reply for TCP ping requests (data packets) of various sizes. Helps track excessive delays and lack of connectivity at various packet sizes.

A Ping Response Time vs. Packet Size Investigation produces a report about the minimum, maximum, and average packet round trip time. Unlike the other types of investigations, the management console does not automatically launch this investigation in response to an incident.

When scheduling this investigation, the time zone you specify for the notification is also used to schedule the investigation.

Specify the following settings to launch or schedule a Ping Response vs. Packet Size investigation:

Target of Investigation

Specifies the target you want to investigate, such as a server, based on its target type. To investigate a group of targets, choose Server Aggregation from the Target Type list.

Investigation Options

Specifies the following options:

Investigate from

Specifies the monitoring device from which you want to launch the investigation. Choose a monitoring device that can communicate with the server or device you want to investigate.

Largest Packet Size

Select from 512 to 8192 bytes for the size of the largest packet to investigate.

Sampling Type

Specifies an option for checking packet size:

- Linear. Checks each packet size up to the largest packet size specified. This option requires more samples.
- Doubling. Covers the range with fewer samples and with less coverage of packet sizes.

Samples at each Size

Select from 1 to 10 for the number of samples at each size.

Timeout (seconds)

Select 1 to 10 for the number of seconds that must elapse for a server to be considered timed out.

Notification Options

Notifies the specified email recipients about the results of the investigation. Note that the investigation is scheduled using the specified time zone for the email notification.

Schedule

The scheduling options only appear if you chose to schedule the investigation. Schedule the investigation to run at a particular time on a particular date or schedule the investigation on a weekly or monthly basis.

Trace Route Investigation

A *Trace Route investigation* records the path and each hop between the management console and end points to detect latency and routing issues.

When scheduling this investigation, the time zone you specify for the notification is also used to schedule the investigation.

Specify the following settings to launch or schedule a Trace Route investigation:

Target of Investigation

Specifies the target you want to investigate, such as a server, based on its target type. To investigate a group of targets, choose Server Aggregation from the Target Type list.

Investigation Options

Specifies the following options:

Investigate from

Specifies the monitoring device from which you want to launch the investigation. Choose a monitoring device that can communicate with the server or device you want to investigate.

Protocol

Specifies the protocol, either ICMP or TCP, to use for the Trace Route investigation. The only time the traceroute performance between ICMP and a given TCP port should vary is if QoS is actively giving priority of one over the other. In this case, TCP (if it is also marked by a network router as would be the true application traffic) might be a preferred method. ICMP is commonly used and accepted provided it is allowed to pass within a given architecture as sometimes it is blocked for security reasons.

Packet Size

Specifies the size of the packet, 32 to 8192 bytes, to investigate.

Retries

Specifies the number of tries, from 1 to 4, to get a response.

Route Searches

Specifies the number of attempts, from zero to 20, to find additional routes for the selected target.

Timeout (seconds)

Specifies the number of seconds, from 1 to 10, that must elapse to consider a server timed out.

Investigate Routers via SNMP

Specifies whether to perform an SNMP query on each network device for its performance details. For information about adding a network device to the management console, see the *Administrator Guide*.

Notification Options

Notifies the specified email recipients about the results of the investigation. Note that the investigation is scheduled using the specified time zone for the email notification.

Schedule

The scheduling options only appear if you chose to schedule the investigation. Schedule the investigation to run at a particular time on a particular date or schedule the investigation on a weekly or monthly basis.

Delete a Scheduled Investigation

Delete a manually scheduled investigation if the investigation is no longer useful. Deleting a scheduled investigation does not delete its investigation reports.

Follow these steps:

1. Click the Incidents page.
2. Click Investigations, Launch in the Show Me menu.

   The Investigation Types page opens.
3. Expand the investigation type to view the scheduled investigations.
4. Click ✖ next to the scheduled investigation you want to delete.

   The investigation is deleted.

Chapter 5: Using the Management Page

This section contains the following topics:

Introduction

On the Management page, view the Performance Scorecard and Operational Level Agreements (OLAs) to determine how well an application performed over time.

OLA reports help you direct your efforts toward performance improvement. As you make changes in these areas, you can see the OLA reports improve.

The management console includes two levels of OLA reports.

- Executive reports which summarize OLA compliance by application
- Detail reports which include information about OLA threshold percentage, results, and number of observations

Use the Performance Scorecard

The Performance Scorecard shows how well applications perform in the enterprise each month. The management console rates performance using the following color coding:

- Unrated (gray)
- Normal (green)
- Minor (yellow)
- Major (orange)

The management console sorts the performance ratings by the number of observations.

Follow these steps:

1. Click the Management page.
2. Click Performance Scorecard in the Show Me menu.

   The Application List page opens.
3. (Optional) Click Settings to change report settings, such as filtering options.
4. Click the color-coded performance bar or click the application name in the Application List to view application details by network (see page 75).

More information:

Email a Report Page (see page 130)
Print a Report Page (see page 131)
Export a Report to a File (see page 129)

View Application Details by Time

Follow these steps:

1. Click the Management page.
2. Click Performance Scorecard on the Show Me menu.

   The Application List page opens.
3. (Optional) Click Settings to change report settings, such as filtering options.
4. Click an application name or performance link to view performance details.
5. Click Time on the third Show Me menu to view application details by time.
6. Select to show by Observations or Percentages in the Show by list.
7. Click [Detail] to see the associated Components reports on the Engineering page.

View Application Details by Network

Follow these steps:

1. Click the Management page.
2. Click Performance Scorecard on the Show Me menu.

   The Application List page opens.
3. (Optional) Click Settings to change report settings.
4. Click an application name or performance link to obtain detailed information about which networks do not perform in the same way as their peers.
5. Click Network on the third Show Me menu to view application details by network.
6. Click an option in the Show by list to filter the list:
   - Network with Performance
   - Network with Averages
   - Network Type with Performance
   - Network Type with Averages
7. Click [Detail] to see the Performance by Network report on the Engineering page.

View Application Details by Server

Follow these steps:

1. Click the Management page
2. Click Performance Scorecard on the Show Me menu.
3. On the Application List page, click Settings to optionally change report settings.
4. Click an application name or performance link to obtain detailed information about which servers do not perform in the same way as their peers.
5. Click Server on the third Show Me menu to view details by server.
6. In the Show by list, select to show Server with Performance or Server with Averages.
7. Click [Detail] to see the Performance by Server report on the Engineering page.

Use Operational Level Agreements

Operational Level Agreement (OLA) reports track performance of the services that network and application end users receive and tell you how often the OLA specification is met. Performance OLA and Availability OLA reports help you determine and evaluate improvements, benefits, and issues you experience.

Use OLA definitions configured on the Administration page to track data and show information by time, network, or server to identify where OLAs are being violated. You can quickly gauge whether the network is in compliance. If it is not in compliance, click a link on the Engineering page and view the component views, performance maps, or application and server availability timeline reports to isolate the issue location. Use the Incidents page to explore what caused performance issues.

The management console administrator can set up OLAs for performance and availability. Follow these guidelines for OLAs:

- Before setting a performance OLA, the management console administrator must define your network types because performance OLAs are applied by defined network type. For information about defining network types, see the *Administrator Guide*.
- Before setting an availability OLA, the management console administrator must enable availability monitoring for the applications and servers for which the administrator is setting the OLA threshold. For more information, see the *Administrator Guide*.

Understand Operational Level Management

Operational Level Management (OLM) includes the disciplined, proactive methodology and procedures that ensure that adequate levels of service are delivered to IT users in accordance with business priorities and at acceptable cost. Operational Level Agreement (OLA) reports reveal whether appropriate levels of service are being met.

Set an OLA against the following metrics:

- Server Response Time to quantify data center performance
- Network Round Trip Time to quantify network infrastructure performance
- Transaction Time to capture an application’s end-to-end performance

The Management page displays the following types of reports:

- Performance OLA reports, which show whether your applications are meeting your OLA goals.
- Availability OLA reports, which show the big picture of compliance to application availability goals. Access daily and hourly details by clicking links in these summary reports.

Deploying OLAs effectively involves an iterative process of defining an OLA, monitoring compliance, improving performance, and refining the OLA at a lower level to increase awareness of the OLA and performance improvements.

Use the management console to monitor OLAs against a static threshold. You can monitor mission-critical applications for operational levels during business hours. Scheduled maintenance and other planned periods of atypical use should be excluded from OLA monitoring when the management console administrator configures OLAs. See the *Administrator Guide* for information about setting up OLAs.

When monitoring OLAs, separate data center metrics from network metrics. Not every metric is meaningful for every network; for example, on back-end applications, you might not want to see Network Round Trip Time. Use longer-term time views of current network performance when selecting OLA threshold values to eliminate transient spikes or dips. Exclude certain network types from the OLA by disabling them.

View OLA Reports

When viewing OLA reports, specify a timeframe that makes sense. We recommend the following timeframes:

Timeframe	Description
Daily	Detailed, short-term troubleshooting, IT department-focused, technical content

Timeframe	Description
Weekly	A summary with additional detail for anomalies or trends; also IT focused
Monthly	A summary for business unit and executive management
Quarterly	Broad operational level information with compliance and as input for planning

You can usually attribute OLA violations to one of these areas:

- Time; that is, a particular time of day or day of the week
- User group; for example, VPN user networks
- Servers; for example, Web servers 2 and 5

Use Performance Detail OLA Reports

The Performance Detail OLA reports include the application name, OLA 1 and 2 results, OLA 1 and 2 percentages, and number of observations. Before you set a performance OLA , add network types. Performance OLAs apply to defined network types.

Use the Performance OLA List

The Performance OLA List shows the current applications being monitored by the management console and gives you a specific view of the compliance metrics of an application for the last month.

Follow these steps:

1. Click the Management page.
2. Click Performance Detail OLA on the Show Me menu.

   The Performance OLA List opens.
3. Click an application to see its performance OLA details.

   In the OLA report, a ✓ indicates that the application met the configured performance OLA, and a ! indicates that it did not.
4. If the application does not meet the OLA, the red exclamation point is displayed. To view additional information:
   - Click the time link for the application to see hourly details.
   - Click 📈 to see the detail reports available through the Engineering page.
5. Click Show by to filter OLA compliance details by network or network type.

Use the Performance Detail OLA by Time Report

Follow these steps:

1. Click the Management page
2. Click Performance Detail OLA on the Show Me menu.

   The Performance OLA List opens.
3. Click an application.

   The default view of Performance Detail OLA by Time report is the Daily View for the last month. This report shows the results for each day of the previous month and specifies the days on which violations occurred against the OLA definitions.
4. (Optional) Click an individual date in the Daily View for the application to navigate into an Hourly View that shows you specific violation and observation data for each hour reported.
5. Click Settings at the top of the page to select a different metric for viewing the applications data.
6. Click 📈 to see the Components Report.

Use the Performance Detail OLA by Network Report

Follow these steps:

1. Click the Management page.
2. Click Performance Detail OLA in the Show Me menu.

   The Performance OLA List opens.
3. Click an application.
4. Click Network in the Show Me menu, .

   The Performance Detail OLA by Network report lets you quickly identify clients that are in violation of the thresholds set for applications. The list shows you the logical groups of clients sorted by IP address and subnet mask.

   When troubleshooting an application, use this report to navigate to the client or group of clients experiencing the performance problem.
5. Select Network or Network Type in the Show by box.

   The report refreshes to show violation data for all entries.
6. Click Settings at the top of the page to select a different metric for viewing the network data.
7. Click 📈 to see the Performance by Network report.

Use the Performance Detail OLA by Server Report

Follow these steps:

1. Click the Management page.
2. Click Performance Detail OLA on the Show Me menu.

   The Performance OLA List opens.
3. Click an application.
4. Click Server in the Show Me menu.

   The Performance Detail OLA by Server report shows servers that are in violation of application OLAs. Use this report to quickly navigate to the server or servers experiencing the performance problem.
5. Click Settings at the top of the page to select a different metric for viewing the server data.
6. Click 📈 to see the Performance by Server report.

Use Performance Executive OLA Reports

The Performance Executive OLA report shows the application name and indicators to show compliance or a violation.

Use the Performance Executive OLA List

The Performance Executive OLA report is a high-level summary report with minimum detail. It focuses on your OLA compliance status according to OLA criteria that you can set.

Follow these steps:

1. Click the Management page.
2. Click Performance Executive OLA.

   The Performance OLA Executive List opens.

   In the reports, a ✓ indicates that the application met the configured performance OLA. A ! indicates that it did not.
3. Click an application to see details.

Use the Performance Executive OLA Summary

The Performance Executive OLA Summary report is a high-level summary report of a particular application with minimum detail. It focuses on your OLA compliance status according to OLA criteria that you set.

Follow these steps:

1. Click the Management page
2. Click Performance Executive OLA.

   The Performance OLA Executive List opens.

   In the reports, a ✓ indicates that the application met the configured performance OLA. A ❗ indicates that it did not.
3. Click an application to see summary details.
4. To view a daily view of OLA results by time, network, or server, click 📈.

Use Availability Detail OLA Reports

Before setting an availability OLA, enable availability monitoring for the applications and servers for which you are setting the OLA threshold.

Use the Availability OLA List

The Availability OLA List displays specific information about the OLA definition, which applications and servers are running, for what percentage of the time, and whether they meet OLA criteria. From the Availability OLA List, view the Availability OLA Definition Daily View or the Availability OLA Definition by Server report and set up OLA criteria.

To navigate to an Availability OLA report, click Availability Detail OLA on the Management page.

In the reports, a ✓ indicates that the application met the configured availability OLA and a ❗ indicates that it did not.

View the Availability OLA Definition Daily View

View and manage availability OLAs for user-defined applications.

Follow these steps:

1. Click the Management page.
2. Click Availability Detail OLA on the Show Me menu.

   The Availability OLA List opens.
3. Click a link for an application to see the Availability OLA Definition Daily View. The Daily View shows the application availability percentage by hour with the duration of downtime.
4. Click the time link for the application to see hourly details.
5. Click 📈 to see the Availability Timeline report for the application.

View the Availability OLA Definition by Server

Follow these steps:

1. Click the Management page.
2. Click Availability Detail OLA on the Show Me menu.

   The Availability OLA List opens.
3. Click a link for an application to see the Availability OLA Definition Daily View.
4. Click Server in the Show Me menu.

   The Availability OLA Definition By Server report shows the application availability by server, including the duration of downtime.
5. Click 📈 to view the Availability Timeline report.

Use Availability Executive OLA Reports

The Availability OLA Executive reports present high-level summaries of application and server compliance with OLA criteria.

Use the Availability OLA Executive List

The Availability OLA Executive List is a high-level summary report that tells you whether the application and server are running meet OLA criteria.

To navigate to the Availability OLA Executive List, click Availability Executive OLA on the Management page. Click the link for an application to see a summary report for the OLA definition.

View the Availability Executive OLA Summary

Follow these steps:

1. Click the Management page.
2. Click Availability Executive OLA.

   The Availability OLA Executive List opens.
3. Click a server link to see the Availability OLA Executive Summary.

   The Summary shows the server availability percentage.
4. Click 📈 to see the Daily View on the Summary report.

Chapter 6: Using the Engineering Page

Use the Engineering page to view and report in-depth performance metrics about the configured networks, servers, and applications. The Performance and Availability reports available from the Engineering page enable you to immediately identify the slowest applications on your network.

This section contains the following topics:

Use the Performance Maps

The Performance report pages appear by default on the Engineering page. Performance maps show the slowest networks, servers, and applications sorted by Transaction Time per application by using horizontal bar charts. Click an item in a chart to see detailed reports for that item.

Navigate the Performance Reports

On the Engineering page, in the Show Me menu, click Performance and then click a performance report:

Networks

Displays Performance by Network map with details about the worst performing networks at the top. Click a network to view the performance details report.

Servers

Displays Performance by Server map with details about the worst performing networks at the top. Click a server to view the performance details report.

Applications

Displays Performance by Application map with details about the worst-performing applications at the top. Click an application to view the performance details report.

For a WAN-optimized application, you can view the performance on each optimized network segment by finding the network segment. The management console appends the network segment to the application name, for example, SMTP [Client], SMTP [WAN], and SMTP [Server].

Tip: When analyzing both optimized and unoptimized performance data for the same application, keep in mind that the unoptimized application response time is faster because the data comes from local users in the data center.

More information:

Email a Report Page (see page 130)
Print a Report Page (see page 131)
Change Report Settings (see page 36)
Change Report Format from a Chart to a Table (see page 39)
Export a Report Page to a CSV File (see page 129)
Export a View to a CSV File (see page 129)
Export a View to an XML File (see page 130)
Use Network Maps (see page 87)

Use Network Maps

Network maps show the slowest networks sorted by Transaction Time per application. The performance maps for networks use horizontal bars to show the slowest networks.

To view the Performance by Network map, click the Engineering page and then click Performance, Networks in the Show Me menu. In the Performance by Network report, click a network to see a detailed report.

A network is made up of forwarding points and buffers connected by circuits over varying distances. Congestion occurs within the forwarding points and buffers. When a network is substantially taxed by applications, users notice a decrease in performance.

Network Delay

Quantify network delay with this equation:
[Serialization Delay] + [Queue Delay] + [Routing/Switching Delay] + [Distance Delay] + [Protocol Delay] = Network Delay

Where:

Serialization Delay

Is the amount of time required to transmit a frame one bit at a time.

Queue Delay

Is amount of time frames wait inside network buffers until transmission on the network interface. Queue Delay is a function of Bandwidth/Utilization.

Routing/Switching Delay

Is the amount of time necessary for a network node to determine the next hop of a frame/packet and forward that frame to the outbound interface. Can be impacted by switching path, node resources, and policies (such as ACLs).

Distance Delay

Is the amount of time for an optical or electrical signal to travel across a path between two end points.

Protocol Delay

Is the amount of latency introduced by network related communication algorithms such as Carrier Sense Multiple Access/Collision Detection (CSMA/CD, which is legacy Ethernet), Carrier Sense Multiple Access/Collision Avoidance (CSMA/CA) and Request to Send/Clear to Send (RTS/CTS), which are (wireless access points), or delayed acknowledgements (TCP) of up to 200 milliseconds.

Report Workflow

Follow these steps:

1. Click the Engineering page.
2. Click Performance, Networks on the Show Me menu.
3. Click a network in the Performance by Network table.
4. View the primary indicators by clicking Components and looking at the Network Round Trip Time and the Retransmission Delay reports. Click Sessions and look at the Connection Setup Time report.
5. View the secondary indicators by clicking Traffic and looking at the Data Volume and Data Rate reports. Click Sessions and look at the TCP/IP Sessions reports.

Use Server Maps

Server maps show the slowest servers sorted by Transaction Time per application. The performance maps for servers use horizontal bars to show the slowest servers.

To view the Performance by Server map, click the Engineering page and then click Performance, Servers in the Show Me menu. In the Performance by Server report, click a server to see a detailed report.

Performance Indicators

A server is composed of these subsystems:

- CPU
- Memory
- I/O (network and disk)

When a subsystem is substantially taxed by an application, the users notice a decrease in performance.

An increase in CPU utilization could indicate many transactions, large queries, other processes unrelated to applications being monitored, TCP checksum calculations, and so forth.

An increase in memory utilization might indicate large data sets resident in memory, increases in the number of users or sessions, increases in the number of processes, failure to de-allocate memory (memory leak), and so forth.

An increase in I/O indicates increases in data writes to disk or network, increases in the number of users or sessions, memory pages to and from disk, and so forth.

Report Workflow

Follow these steps:

1. Click the Engineering page.
2. Click Performance, Servers on the Show Me menu.
3. Click a server in the Performance by Server table.
4. View the primary indicators by clicking Components and looking at the Server Response Time report. Click Sessions and look at the Connection Setup Time report.
5. View the secondary indicators by clicking Traffic and looking at the Data Volume and Data Rate reports. Click Sessions and look at the Unfulfilled TCP/IP Session Requests report and the TCP/IP Sessions reports.

Use Application Maps

On the Engineering page, the Performance by Application map uses horizontal bars to show the slowest applications. The report also shows the number of observations per application. Click an application link to view detailed reports for the selected application.

An application typically has two parts:

- Daemon running on a server or servers
- Client running on a user's computer

Applications that are not optimized for network transmission can cause a decrease in performance; for example, they could ping pong data across the network, they could open multiple consecutive TCP sessions instead of a persistent TCP session, or they could use a small application/TCP window size for high latency WAN links.

Follow these steps:

1. Click the Engineering page.
2. Click Performance, Applications in the Show Me menu.
3. Click an application in the Performance by Application table.
4. View the primary indicators by clicking Components and looking at the Data Transfer Time report. Click Response Size and look at the Data Transfer Time by Response Size report.
5. View the secondary indicators by clicking Traffic and looking at the Data Volume and Data Rate reports. Click Sessions and look at the TCP/IP Sessions reports.

View Performance Detail Reports

On the Engineering page, the management console produces views using 5-minute data and averages that data to produce 15-minute data:

- 5-minute data can be reported for up to 8 hours.
- 5-minute data can be reported for up to 24 hours.

Follow these steps:

1. Click the Engineering page.
2. Click Performance in the Show Me menu and then click:
   - Networks to display performance map by network.
   - Servers to display performance map by server.
   - Applications to display performance map by application.

   A performance map is displayed.
3. Click an item in the performance map to view its detail reports.

   By default, the Components report is displayed.
4. Select the detail report you want to view from the third Show Me menu.
   - Components reports (see page 90)
   - Traffic reports (see page 98)
   - Sessions reports
   - Response Size reports (see page 104)
   - QoS reports (see page 105)
   - Statistics reports (see page 109)
   - Trends reports (see page 112)

   Each performance detail report shows a collection of views related to that subject with a summary report and component views below it. See the following sections for more information.

Use the Components Reports

The Components detail reports show metric details for each selected application, server, and network.

Tip: To filter a detail report by a particular application, server, or network, click the links at the top-left corner of a report or click the Settings button.

Response Time Composition: Average

The Response Time Composition: Average report shows end-to-end response time stacked with the components that make up the total time:

- Network RTT: Network Round Trip Time
- Retrans: Retransmission Time
- Data Xfer: Data Transfer Time
- Server Resp: Server Response Time

This report shows the number of observed TCP transactions over the reporting time period.

A transaction is a single request and a single server response, one period of data transfer, one or more acknowledgments, and observed latency caused by retransmitted packets.

Look for	Might Indicate
Gradual increases in all values	Typical growth of network traffic over time
Spike in measurements	Anomaly to investigate further
One component of response time spiking while others remain constant	Anomaly to investigate further
An increase in the number of observations coupled with increasing Network Round Trip Time	Overutilization of links
High Data Transfer Time and low Network Round Trip Time	Server overutilization
High Retransmission Delay	Packet loss
Increasing observations	More application use
Increasing observations and a corresponding increase in Server Response Time	Overloaded server

Server Response Time

The Server Response Time report shows the amount of time a server takes to start responding to a request made by a client.

Tip: When viewing the Server Response Time data in tabular format, the bottom row of the table lists Average/Total statistics for each column. However, the Average/Total statistics are calculated from the Mean values. For example, the Average/Total value for the Min column is actually the minimum of the values in the Mean column - not the average value of the Min column.

This value can be affected by server speed, application design, and volume of requests.

Look for	Might Indicate
Server Response Time increases while the number of observations increases, then look for increases in the following metrics during the same period: ■ Data volumes ■ Data Transfer Time ■ TCP sessions that are open	Strong indication that the server is overloaded.
Server Connection Time increases while the number of observations decreases	Check whether other applications running on this server are monitored by the management console. If so, check for increased observations that correspond to increased Server Response Times for these applications to identify the culprit application process or processes on the server.
No other monitored applications on the server and the number of observations decreases while the Server Response Time increases	Another application might be affecting the Server Response Time; for example, a backup program running on the server can increase response times for concurrently active applications. Check for patterns of increases in Server Response Time and decreases in observations over time and, if necessary, contact other operations teams or use a network protocol analyzer to determine what application running on the server is causing the increase in Server Response Time.

Look for	Might Indicate
High Server Response Times	One of the following problems with a server: ■ Inadequate processing power ■ Insufficient available memory ■ Slow hard drives ■ Hanging processes ■ Misconfigured NIC settings ■ Presence of poorly written applications, such as database lookups that use large queries or poor indexing
Server Response Time varies consistently	Large periodic data volumes.
Applications send a response to indicate work has started immediately after receipt of a request and Server Response Time is always low	Actual delay in the user experience is exposed in the data transfer rate component.

Data Transfer Time

The Data Transfer Time report shows the time it takes to transmit a complete response as measured from the initial to final packet. Data Transfer Time excludes the initial server response time and includes only Network Round Trip Time if there is more data to send than fits in the TCP window.

Factors affecting this time are response size, available bandwidth, network latency because of distance, some server processing, and interactions such as the number of round trips, and individual packet size between the application and the network.

Data Transfer Time is related to the number of network round trips required to deliver all data and the delay per round trip.

Look for	Might Indicate
Increase in Data Transfer Time	One of the following conditions: ■ Application is poorly written ■ TCP/IP transmission window is not set large enough to allow the server to send a continuous stream of information Correlate a significant increase in Data Transfer Times with data volumes to determine if the increase is caused by larger amounts of data being transferred on the network or if there is a problem.
Variations in Data Transfer Time by subnet	Lack of bandwidth available to the region, retransmissions required to certain regions, and different uses of the application.

Retransmission Delay

Retransmission Delay is the additional delay in the Network Round Trip Time caused by packets needing retransmission.

The data shown is an average across all observations, not the actual retransmission time for one transaction. Given a total of five transactions, four with no retransmissions, and one with a 5-second Retransmission Delay, the view shows a 1-second Retransmission Delay.

The management console calculates Retransmission Delay by observing duplicate packets in the network from the monitoring device's vantage point next to the server. A monitoring device can see packets retransmitted by the server because of data losses in the server-to-client direction along the network path. These observations are included in Retransmission Delay. When data loss occurs in the client-to-server direction (for example, in the network path before reaching the server), the monitoring device cannot observe such packet loss and that delay is not included in the Retransmission Delay metric. The management console includes this associated delay with retransmission in the Network Round Trip Time metric, which includes the server response and client acknowledgment. A delay in client acknowledgment caused by unseen Retransmission Delay increases the NRTT value. This metric does not reveal the impact of losses on the Data Transfer Time because of TCP congestion.

Retransmissions can cause a given session to reduce the transaction time, but the speed of the bytes across the line remains constant unless the path changes or congestion increases.

Look for	Might Indicate
Retransmission Delay increases as the number of observations increases	The network might not be able to handle the higher load. Changing the buffer allocation or routing strategies might alleviate short-term spikes.
Retransmission Delay varies in a consistent pattern	Problem in the link to the region. Correlate the high retransmission periods to other information sources to determine what else is happening on the network at those times. If certain subnets indicate a greater ratio of Retransmission Delay to Network Round Trip Time delay than other regions, there might be a problem in the link to that region. If you suspect that a particular network is experiencing excessive delay because of retransmission, view the Retransmission Delay for the network in question and compare it to the average Retransmission Delay for all networks.

Retransmission Delay for the network is high compared to the average for all networks	Normal behavior as in a wireless environment.
A consistently high Retransmission Delay	Network cannot carry the required load or a device might be malfunctioning.
High values	Dropped packets caused by network congestion, misconfigured load balancing, redundant paths or routes, and network error conditions.

Network Round Trip Time

Network Round Trip Time is the amount of time it takes for a packet to travel round trip between the server and clients on a network, excluding latency from retransmissions.

Application and server processing times are excluded when calculating this value. The management console continuously refines the Network Round Trip Time by looking at the TCP acknowledgments for all application traffic, not just connection setup times, to build the most accurate model.

Use the Network Connection Time as a baseline for carrier latency and comparison to NRTT times.

Look for	**Might Indicate**
Increases in NRTT that correspond with increases in the number of observations	Insufficient bandwidth between the client hosts and the server for the volume of data being transmitted by the application. Increases in NRTT that do not correspond to increases in the number of observations might indicate this: ■ Another application is consuming the available bandwidth between the remote client hosts and the application server. ■ The carrier’s network switched to the protected or alternate path. ■ Some error condition exists in the network.
Number of observations increases with an increase in NRTT	Root source of the delay increase. See the Data Volume and TCP sessions views to determine if there was a corresponding increase in data, TCP sessions, or both.

Look for	Might Indicate
Number of observations decreases with the increase in NRTT	Other monitored applications between the monitoring device and the remote clients might be responsible for the apparent reduction in bandwidth.
When comparing subnets in the same office location, such as the same building, you should see a minimal difference in NRTT for the same application	A difference greater than 10 ms could be caused by one of the following: ■ Problems in the LAN architecture ■ Incorrectly configured switch or NIC port settings ■ Error conditions in the network ■ Utilization differences along the physical paths within the LAN
Users in remote offices at varying distances from the server and operating over different WAN links to the server experience different latency and NRTT	Provisioned bandwidths, link utilization patterns, and access technologies; for example, ATM versus IP VPN.

Effective Network Round Trip Time

Effective Round Trip Time is composed of Network Round Trip Time plus delays caused by retransmissions for a single transaction.

This metric is valuable because it reflects latency the user actually experiences. The Administrator can set an incident threshold for this metric to detect performance degradation in networks because of retransmissions.

Effective Round Trip Time is closer to the end-user network experience than Network Round Trip Time because Network Round Trip Time and Retransmission Delay are both included. Compare it to Network Round Trip Time and Retransmission Delay to see how each component contributes in relation to the other.

Traffic Reports

The Traffic detail reports for networks, servers, and applications on the Engineering page include the following metrics:

- Response Time Composition: Average
- Data Volume (in bytes)
- Data Volume (in packets)
- Data Rate (in bits/second)
- Data Rate (in packets/second)

Tip: To filter a detail report by a particular application, server, or network, click the links at the top-left corner of a report or click the Settings button.

Response Time Composition: Average

The Response Time Composition: Average report shows end-to-end response time stacked with the components that make up the total time:

- Network RTT: Network Round Trip Time
- Retrans: Retransmission Time
- Data Xfer: Data Transfer Time
- Server Resp: Server Response Time

This report shows the number of observed TCP transactions over the reporting time period.

A transaction is a single request and a single server response, one period of data transfer, one or more acknowledgments, and observed latency caused by retransmitted packets.

Look for	Might Indicate
Gradual increases in all values	Typical growth of network traffic over time
Spike in measurements	Anomaly to investigate further
One component of response time spiking while others remain constant	Anomaly to investigate further
An increase in the number of observations coupled with increasing Network Round Trip Time	Overutilization of links

High Data Transfer Time and low Network Round Trip Time	Server overutilization
High Retransmission Delay	Packet loss
Increasing observations	More application use
Increasing observations and a corresponding increase in Server Response Time	Overloaded server

Data Volume (in bytes)

Data Volume (in bytes) indicates the total number of application-layer bytes seen on the network.

Data Volume reports can expose the impact of high data transfer rates. Investigate whether an unusually high volume of data transmission is the cause of a elevated Network Round Trip Time.

Use this report to do the following:

- Identify anomalies on your network caused by excess traffic.
- Eliminate excess traffic as a possible cause of performance problems.
- Determine peak usage for a specific application, group of applications, or networks for capacity planning.

Data Volume (in packets)

Data Volume (in packets) indicates the total number of packets seen on a monitored network. The report includes zero-byte packets such as TCP acknowledgments.

Use this view with the Data Volume view to get an idea of the average packet size traversing your network. This is useful in identifying problem areas or attacks because many small packet sizes might indicate a denial of service attack.

Data Rate (in bits/second)

Data Rate (in bits/second) displays the data rate in bits per second (bytes/second times 8) over a specified time period.

Because this view compares the rate of data traffic from and to the server, use this view to plan capacity on the server.

Look for	**Might Indicate**
High bit rate	Overloaded subnet

Look for	Might Indicate
Low bit rate	Network or application issue

Data Rate (in packets/second)

Data Rate (in packets/second) displays the data rate in packets per second over a specified time period.

Look for	Might Indicate
High packet rate	Stressed routers, switches, and firewalls resulting in excessive packet discards
High packet rates for short periods	Can be mitigated by allocating sufficient buffer space in affected devices. Be careful when increasing buffer capacity because delay-sensitive applications might be adversely affected. It is better to discard a voice packet than to delay it in a buffer.
Sustained high packet rates	Might need to upgrade the device capacity.
As with routers, switches, and firewalls, the monitoring device might discard packets during short periods of high packet rates.	SPAN port could be dropping packets.
Unusually low packet rate	Network issue.
High packet rate coupled with a low bit rate	Security attack.

Sessions Reports

Use the TCP/IP Sessions Report to look for significant application and server problems related to the number of Unresponsive or Refused Sessions and for a general survey of the number of unique sessions and the length of those unique sessions. This report does not include HTTP sessions from any Web applications.

Tip: To filter a detail report by a particular application, server, or network, click the links at the top-left corner of a report or click the Settings button.

Response Time Composition: Average

The Response Time Composition: Average report shows end-to-end response time stacked with the components that make up the total time:

- Network RTT: Network Round Trip Time
- Retrans: Retransmission Time
- Data Xfer: Data Transfer Time
- Server Resp: Server Response Time

This report shows the number of observed TCP transactions over the reporting time period.

A transaction is a single request and a single server response, one period of data transfer, one or more acknowledgments, and observed latency caused by retransmitted packets.

Look for	**Might Indicate**
Gradual increases in all values	Typical growth of network traffic over time
Spike in measurements	Anomaly to investigate further
One component of response time spiking while others remain constant	Anomaly to investigate further
An increase in the number of observations coupled with increasing Network Round Trip Time	Overutilization of links
High Data Transfer Time and low Network Round Trip Time	Server overutilization
High Retransmission Delay	Packet loss
Increasing observations	More application use
Increasing observations and a corresponding increase in Server Response Time	Overloaded server

Connection Setup Time

Connection Setup Time is the time it takes to establish a TCP session between the client and server before data transfer can begin. The network component of this view should be approximately the same as the Effective Round Trip Time.

View the two components of Connection Setup Time separately:

- Server Connection Time (SCT) is the time from the initial SYN packet being received from the client until the server sends out the first SYN/ACK.
- Network Connection Time (NCT) is the time between the SYN/ACK being sent from the server until the ACK completing the three-way handshake is received.

Look for	Might Indicate
Connection Setup Time is significantly longer than SCT and NCT	Probably caused by server or LAN problems. ■ Compare the Connection Setup Time to the Network Round Trip Time, Retransmissions, and the Server Setup Time to determine whether these views show similar patterns. Typically, Connection Setup Time and Network Round Trip Time increase concurrently though the increase is not linear. ■ Increases in NRTT and Data Transfer Time while Connection Setup Time remains constant might indicate data loss in the direction of client to server. Suspected data loss can be confirmed by the use of a sniffer on the client network to view retransmissions.
Connection Setup Time is significantly longer than SCT and NCT	■ A spike in the Connection Setup Time correlated with a spike in the Network Round Trip Time indicates that delay is occurring on the network because of an increase in traffic volume and insufficient bandwidth, errors in the network, or an increase in latency caused by a carrier switch to an alternate path. ■ A spike in the Connection Setup Time alone could indicate that the server is stressed because the CPU is overloaded or the TCP/IP session limit is exceeded.

TCP/IP Sessions

The TCP/IP Sessions report shows the number of unique connections between the clients and the server and shows Open, Complete, and Expired TCP/IP sessions.

The management console calculates Expired and Complete TCP/IP sessions during the 5-minute monitoring period. Open sessions are the number of sessions still open at the end of a monitoring period. Open sessions might become Expired or Complete during a subsequent reporting interval. The management console classifies a session as Expired if it detects no packets in 15 minutes.

Look for	Might Indicate
Multiple expired sessions that are not completed.	Too many expired sessions left open can cause the server to hang. Servers can support only a maximum number of simultaneous connections.

Unfulfilled TCP/IP Session Requests

Unfulfilled TCP/IP Session Requests represents the number of unique, unsuccessful connections between the clients and server, including:

- Refused sessions. A *refused session* occurs when a connection request was explicitly rejected by the server during the three-way handshake.
- Unresponsive sessions. An *unresponsive session* occurs when a connection request was sent, but the server never responded.

Look for	Might Indicate
A significant number of Refused Sessions	Demand has started to stress one or more servers. ■ The server might be too busy from normal system requests or from a malicious attack. ■ The application license running on the server might have exceeded the maximum number of allowed users or sessions.

TCP/IP Session Times

TCP/IP Sessions Times shows the duration of each user session.

Look for	Might Indicate
Many short sessions from the same users	Application might benefit from connection pooling.

Look for	Might Indicate
Sessions are long and there is little traffic	Application might be keeping sessions open unnecessarily.
Peak usage for a server is a problem one week	Increase in the number of open sessions has been increasing in past weeks. Check the Trends view described in the next section.

Response Size Reports

Response Time is a measure of Data Transfer Time by size of response transferred. Connection Setup Time is not included.

Comparing the times it takes to transmit responses of various sizes lets you see whether performance issues are related to the application or network.

Tip: To filter a detail report by a particular application, server, or network, click the links at the top-left corner of a report or click the Settings button.

Response Time Composition: Average

If you see high application response times and high Network Round Trip Times on this view with high data volume and high data transfer rates on Traffic views, Response Size might be a contributing factor to slow performance. High Server Response Time rates could also be sufficient cause to look at Response Size.

More information:

Response Time Composition: Average (see page 91)

Data Transfer Time by Response Size

This report shows the average Data Transfer Time by response size. Generally, the larger the data size, the longer the transfer time.

Look for	Might Indicate
Limited or specific Response Size.	A specific transaction type is causing slow end-to-end response time.
Volume of larger responses.	Use the Data Transfer Time by Response Size view to determine the volume of larger responses, which might be causing delay on the network.

Look for	Might Indicate
The gray observation counts line for each response size in this view indicates the number of responses of each size.	The observation counts can indicate application design issues such as an application that responds with small-volume data transfers and uses no larger data transfers.

Average (Large) Data Transfer Times

The Average Data Transfer Times reports show the average Data Transfer Time by response size in KB for small, medium, and large responses.

Look for the following in these views:

- Comparing the response time of one response size over time indicates whether changes in application response time are caused by changes in the amount of data returned. If the response times of each response size remain fairly constant but the application runs slower, the problem might be caused by a change in user behavior; for example, the user might be requesting more of the larger objects. Use the gray observation line to determine how many transactions fall into each size category.
- The relative response time indicates whether the application is waiting for the request to be analyzed or the response to be transmitted.

QoS Reports

High traffic volume and high data transfer rates combined with slow response times for a given response size might prompt you to consult the Quality of Service report. This report also includes detail into specific users of your application in the Users view.

Tip: To filter a detail report by a particular application, server, or network, click the links at the top-left corner of a report or click the Settings button.

Response Time Composition: Average

The Response Time Composition: Average report shows end-to-end response time stacked with the components that make up the total time:

- Network RTT: Network Round Trip Time
- Retrans: Retransmission Time
- Data Xfer: Data Transfer Time
- Server Resp: Server Response Time

This report shows the number of observed TCP transactions over the reporting time period.

A transaction is a single request and a single server response, one period of data transfer, one or more acknowledgments, and observed latency caused by retransmitted packets.

Look for	Might Indicate
Gradual increases in all values	Typical growth of network traffic over time
Spike in measurements	Anomaly to investigate further
One component of response time spiking while others remain constant	Anomaly to investigate further
An increase in the number of observations coupled with increasing Network Round Trip Time	Overutilization of links
High Data Transfer Time and low Network Round Trip Time	Server overutilization
High Retransmission Delay	Packet loss
Increasing observations	More application use
Increasing observations and a corresponding increase in Server Response Time	Overloaded server

More information:

Response Time Composition: Average (see page 91)

Users

The Users view shows the number of unique IP addresses, subnets, or both on the monitored network during the given time period.

Look for large numbers of concurrent users accessing the server.

Packet Loss Percentage

The Packet Loss Percentage view shows the percentage of data lost on the monitored network and the loss rate in packets per second.

The monitoring device calculates Packet Loss Percentage by observing the ratio of retransmitted data to total data within the network from its vantage point next to the server. The monitoring device can see packets retransmitted by the server because of data losses in the server-to-client direction along the network path. When data loss occurs in the client-to-server direction (in the network path before reaching the server, for example), the monitoring device cannot observe such packet loss and that delay is not included in the Packet Loss Percentage.

Look for the following in this view.

Look for	**Might Indicate**
Concurrent values greater than 1%.	High data loss.
The view shows only packet loss while the data table shows packet and byte loss.	Packet loss causes the data to be retransmitted, which causes TCP to decrease the window size, which slows the transfer time of subsequent packets. Retransmitted packets increase the load on the network infrastructure. Retransmitted bytes indicate the amount of bandwidth taken up by the retransmitted data.

User Goodput

The User Goodput view captures the client experience. It shows the number of good bytes transmitted (minus retransmissions) divided by the amount of time required to transmit this information.

When you download data from the Web, the browser typically computes the throughput for you as you download the file. Throughput is the number of bytes transferred divided by the period of activity. Typically, you would use this throughput value to gauge the effectiveness of your network connection and view the calculated User Goodput by subtracting retransmitted bytes from the throughput because retransmitted bytes artificially inflate throughput.

If you have a throughput measurement of 100 kbps, but half of the download consists of retransmissions, the User Goodput is 50 kbps; therefore, you benefit from only half of the throughput.

User Goodput and Throughput are computed only during periods of active transmission. The monitoring device averages the Traffic Rate over 5-minute intervals. Much of that time could be silent. There is little relationship between User Goodput/Throughput and Traffic Rate.

To illustrate this point, consider the following scenario. If the only transfer during a 5-minute interval is a 50 KB document downloaded without retransmissions in 1 second, these metrics apply:

- User Goodput and Throughput are 400 kbps, which is calculated as follows:

  50 KB * (8 bits/byte)/1 second = 400 kbps

- Traffic Rate is 1.3 kbps, which is calculated as follows:

  50 KB * (8 bits/byte)/300 second = 1.3 kbps

User Goodput is useful only if there is significant data transfer and is used for bulk transfers, not interactive transactions. Throughput is computed using the same formula as User Goodput (bytes divided by Data Transfer Time) except that it includes retransmitted packets. Throughput is always greater than or equal to User Goodput.

Look for	**Might Indicate**
Dips in User Goodput	Network congestion. Use User Goodput as a network metric only when there is traffic being transferred. few bytes, whether transmitted slowly or quickly, might not be useful.
Poor Goodput	Data Transfer Time is hampered by poor server or application performance.

More information:

Throughput (see page 27)

Composite Rate Per User

The Composite Rate Per User view shows the volume of traffic transmitted, divided by the time interval, divided by the number of unique users in that interval.

The Composite Rate Per User view is affected by the address space used on the network. With Class A or B addresses, the Composite Rate Per User view shows results for individual subnets rather than individual clients. Class C addresses are different; the Composite Rate Per User view reflects each client.

Use this view to do the following:

- Help predict the impact of new users on your applications.
- Pre-deploy testing to determine the anticipated application load on various network links.

Statistics Reports

The Statistics report page shows variations in response time. Use this report page for impact analysis. The Response Time Composition: Average view shows the total response time. The Response Time Composition: Standard Deviation view shows the amount of variation that exists in the response time measurements.

Tip: To filter a detail report by a particular application, server, or network, click the links at the top-left corner of a report or click the Settings button.

Response Time Composition: Average

The Response Time Composition: Average report shows end-to-end response time stacked with the components that make up the total time:

- Network RTT: Network Round Trip Time
- Retrans: Retransmission Time
- Data Xfer: Data Transfer Time
- Server Resp: Server Response Time

This report shows the number of observed TCP transactions over the reporting time period.

A transaction is a single request and a single server response, one period of data transfer, one or more acknowledgments, and observed latency caused by retransmitted packets.

Look for	**Might Indicate**
Gradual increases in all values	Typical growth of network traffic over time
Spike in measurements	Anomaly to investigate further
One component of response time spiking while others remain constant	Anomaly to investigate further
An increase in the number of observations coupled with increasing Network Round Trip Time	Overutilization of links
High Data Transfer Time and low Network Round Trip Time	Server overutilization
High Retransmission Delay	Packet loss
Increasing observations	More application use
Increasing observations and a corresponding increase in Server Response Time	Overloaded server

Response Time Composition: Standard Deviation

The Response Time Composition: Standard Deviation metric quantifies the amount of variation that exists in the response time measurements.

Look for	Might Indicate
All users have similar behavior (a low Standard Deviation) and the degree of variation is changing over time.	An application with good average response time but high Standard Deviation might perform poorly for a significant number of users.
Regular patterns in the Standard Deviation.	Changes in application use over time or changes in the client mix. Variations in the links to networks can cause higher Standard Deviation in the network times. When only a few samples are available, the variation tends to be higher than when many data points are available.
Consistently high Standard Deviation.	Values are too disparate for meaningful analysis; for example, comparing data including a fast application and a slow application would give a high variation in the user experience between the applications. In this case, the data should be filtered by application to achieve more meaningful results. A consistently high Standard Deviation might also indicate a high degree of variation in the application behavior.

Percentiles

Use the Server Response Time Percentiles, Data Transfer Time Percentiles, Retransmission Delay Time Percentiles, and Network Round Trip Time Percentiles views on the Engineering page to find out what percentage of transactions are impacted by a given issue. If the network problem appears only in the 90th percentile view, the issue is confined to only the slowest transactions. If the issue appears in the 75th or 50th percentile views, the problem involves a greater portion of the transactions.

A few users might complain of slow application response while most users have normal response times. The Network Round Trip Time Percentiles view might indicate this situation with a 90th percentile line much higher than the 75th or 50th percentile lines. You can see that about 10% of the transactions have poor performance while the remaining 90% have better results. Perhaps users at remote sites or users of particular transaction types have poor experience.

If the 90th, 75th, and 50th percentile lines are close together, you can conclude that most users were seeing similar results without much variation wherever they were located and whatever transaction types they were executing.

Trends Reports

The Trends views on the Engineering page show specific time period, average end-to-end response time data. The report title indicates the granularity of the interval.

Tip: To filter a detail report by a particular application, server, or network, click the links at the top-left corner of a report or click the Settings button.

Trend Report	Interval
One Hour Response Time Composition: Average	5 minutes
Eight Hour Response Time Composition: Average	5 minutes
Daily Response Time Composition: Average	15 minutes
Weekly Response Time Composition: Average	1 hour
Monthly Response Time Composition: Average	6 hours

Use these views as follows:

- Consult the Trends Report views with a spike in traffic volume when investigating high Data Transfer Rate and Network Round Trip Time or for variations in end-to-end response time.
- Identify patterns of congestion through elevated response times. Is the congestion periodic, persistent, or sudden?
- Check for significant variation from previous traffic trends. The Trends views are helpful to identify periods of high network utilization.
- Compare daily data with other days of the week or view how the number of observations compares for different days.

Use Availability Reports

The management console defines application availability as observed successful TCP transactions during time slices or a response to a request to the application port on the server. The management console gathers information at the application port level at 5-minute intervals.

The management console questions availability if one of the following conditions is true during an observed 5-minute interval:

- Fewer than 10 observations
- Greater than 10% refused sessions

If necessary, the management console checks availability by performing these steps:

1. Tests the application by attempting to connect on the defined application port. For applications defined by a port range, the management console tries connecting on the first eight ports in the range.
2. If the management console receives no response from the application, the application is rated Unavailable.
3. Optionally, the management console pings the server to verify its status.
   - If the server acknowledges the ping request, the management console considers the server available.
   - If the server does not acknowledge the ping request, the management console considers the server unavailable.
4. If the application or server is unavailable, the management console opens a server incident.

These two scenarios point to different problems:

- An application is not running, but the server that hosts it is running.
- The TCP port is locked only for this application, such as port 80 for Web.

View Application and Server Availability Reports

Follow these steps:

1. Click the Engineering page.
2. Click Availability on the Show Me menu.

   The Application Availability or Server Availability report opens.

3. Click the color-coded performance bar and click Time to view an availability timeline report.
4. Click the Arrow menu and select Aggregations to display an availability report for aggregations only.
5. Click the Arrow menu and select Table to display an availability report in table format.

More information:

Email a Report Page (see page 130)
Print a Report Page (see page 131)
Change Report Settings (see page 36)
Export a Report Page to a CSV File (see page 129)
Export a View to a CSV File (see page 129)

View the Availability Timeline Report

Follow these steps:

1. Click the Engineering page.
2. Click Availability on the Show Me menu.

   The Application Availability or Server Availability report opens.

3. Click the link for an application or server.
4. Click Time in the Show Me menu to view the Timeline Summary and Availability Timeline reports.

View Incidents Related to Availability

To view incidents reported during a specified time period, click Related Incidents on the Availability report page. An Incidents report lists incidents reported during the Availability report time period.

More information:

Use the Incidents Page (see page 53)

Use List Reports

The List reports on the Engineering page let you view the networks, servers, and applications monitored by the management console.

Learn about List Reports

On the Engineering page, click Lists to view lists of networks, servers, and applications. From these lists, navigate into performance maps for specific networks, servers, or applications.

View Networks, Servers, and Applications

From the Engineering page, view lists of networks, servers, or applications. From these lists, click a link in the performance maps for a specific network, server, or application to view metrics specific to the selected component.

On the Engineering page, click Lists and then click Networks, Servers, or Applications to view a report.

In the Network List, Server List, or Application List, click a link on the map to view the performance report.

If you want to	Do this
View the Performance by Network report for the network related to an application or server.	Click the Networks link in the Map By column (if available) on the Server List or Application List page.
View the Performance by Server report for the server related to an application or network.	Click the Servers link in the Map By column on the Network List or Application List page.
View the Performance by Application report for the application related to a server.	Click the Applications link in the Map By column on the Network List or Server List page.
Display a list report for aggregations.	Click the Arrow menu and select Aggregations.

More information:

Email a Report Page (see page 130)
Print a Report Page (see page 131)
Export a Report Page to a CSV File (see page 129)

Chapter 7: Using the Optimization Page

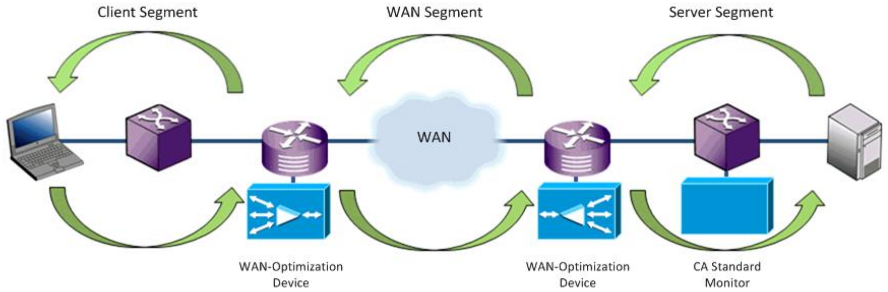
This section contains the following topics:

Learn About Optimized Transactions

This section describes WAN-optimized transactions as delivered through Cisco WAAS or Riverbed Steelhead, and helps you to interpret data for optimized transactions shown in the management console reports on the Optimization page.

WAN-optimization solutions consist of an optimization device on either side of a WAN. The WAN-optimization devices optimize the TCP connection between a client and server by breaking a single TCP connection into these three segments:

- Client to WAN-optimization device at the edge
- WAN-optimization edge device to WAN-optimization core device
- WAN-optimization core device to server



WAN-optimization data is available for the three TCP segments created by the WAN-optimization solution. Because it is now monitoring from multiple monitoring points for a single optimized application-server-network combination, the management console generates a separate set of metrics for each of the three segments and treats each set as a separate application. Application behavior along all three segments is analogous to that of a three-tier application with the source and destination ports and addresses remaining the same throughout the tiers. The Optimization page of the management console shows data for WAN-optimized transactions.

Monitor Optimized Transactions

The Engineering page reports WAN-optimized applications by creating a separate application for each network segment and appending the network segment to the application name. The management console reports identify segments when they exist. In each view, the management console shows the application name and its associated segment in the following format:

<ApplicationName> [*<Segment>*]

The *<Segment>* is Client, Server, or WAN; for example:

HTTP [Client]

The [*<Segment>*] label does not appear if no traffic from segmented applications was collected.

View the Optimization Page

The Optimization page opens when the management console monitors application traffic on the WAN network segment. To view the Optimization page, you must have a role that lets you access the Engineering page.

The Optimization page reports the optimized application experience from the client's perspective.

The Settings parameters on the Optimization page are similar to the Engineering page with the following exceptions:

- The selection for servers defaults to All Servers. There is no filtering option for servers.
- The available applications include only WAN-optimized applications.
- Available networks include only those from which segmented applications saw Client segment data.
- You cannot filter a particular metric or sampling interval.

Navigate the Optimization Report Pages

Use the Show Me menu to navigate the Optimization report pages. From either report, you can drill into the Components report to view performance details for a particular application:

Performance

Displays a performance map for each Client segment application. The Client Experience for Optimized Transactions view reflects the true client experience. The measurements come from the remote Cisco WAE device or the remote CA Standard Monitor that monitors the remote Steelhead appliance.

Bandwidth Reduction

Displays volume metrics for each WAN segment application. The Bandwidth Reduction view shows the total bandwidth reduction, in bytes, between the optimized WAN segment and either the data center Server segment or the branch Client segment. You can choose to compare Total Bytes or From Server Bytes.

To report the Client segment in a Steelhead environment, a CA Standard Monitor must monitor the branch Steelhead appliance.

Tip: Click Make Default View to display the current report by default.

View Performance Detail Reports for Optimized Transactions

Use the Components report to view performance details for optimized transactions and to compare the effects of WAN optimization. From the Optimization page, drill into a particular application to display the Components report.

The following table summarizes the performance detail views and the segments from which they are derived:

View	Application Segment Shown	Reference
Response Time Composition	Client	Response Time Composition: Average (see page 122)
Server Response Time	Server	Server Response Time (see page 123)
Network Round Trip Time	WAN	Network Round Trip Time (see page 124)
Retransmission Delay	WAN	Retransmission Delay (see page 95)
Packet Loss Percentage	WAN	Packet Loss Percentage (see page 96)

View	Application Segment Shown	Reference
Data Rate (in bits/second)	WAN	Data Rate (in bits/second) (see page 126)
Data Rate (in packets/second)	WAN	Data Rate (in packets/second) (see page 126)
Data Volume (in bytes)	WAN	Data Volume (in bytes) (see page 126)
Data Volume (in packets)	WAN	Data Volume (in packets) (see page 126)

In each view, the management console shows the application name and its associated segment in the following format:
<ApplicationName> [*<Segment>*]

The *<Segment>* is Client, Server, or WAN.

The Components report provides performance detail views for optimized transactions. The views average 5-minute data to produce 15-minute data, and use:

- 5-minute data for views of up to 8 hours.
- 15-minute data for views of up to 24 hours.

More information:

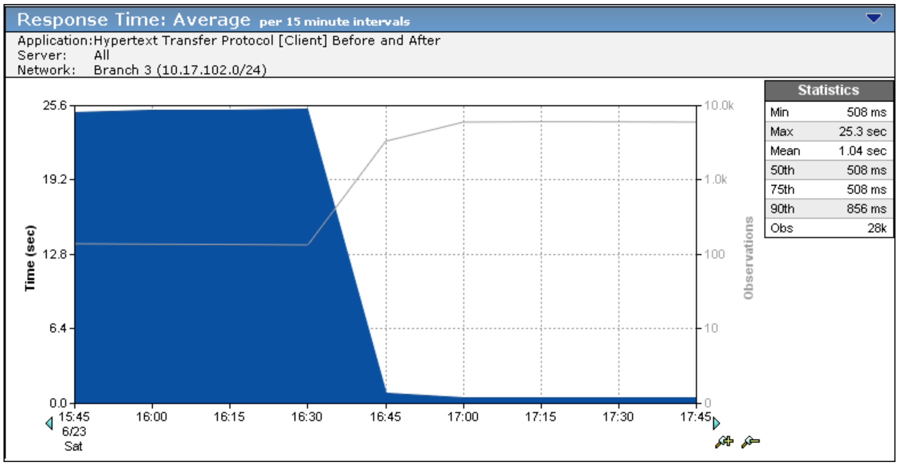
Compare the Effects of WAN Optimization (see page 127)

Response Time Composition: Average

The Response Time Composition: Average view shows Transaction Time for the Client application segment. This view reflects the true client experience. The measurements come from:

- The remote Cisco WAE device.
- The remote CA Standard Monitor that monitors a branch Steelhead appliance. If a CA Standard Monitor does not monitor the Client segment at a branch location, the view for that client network is empty.

This example view shows the before-and-after impact of WAN-optimization on an HTTP application. Users from this subnet experience a 25-fold decrease in response time and a 50-fold increase in transactions.



Response Time: Average per 15 minute intervals

Application: Hypertext Transfer Protocol [Client] Before and After
Server: All
Network: Branch 3 (10.17.102.0/24)

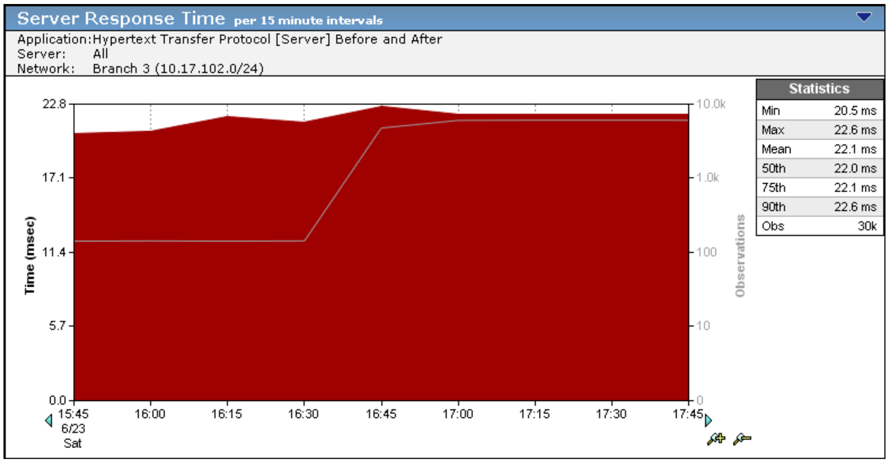
Statistics	
Min	508 ms
Max	25.3 sec
Mean	1.04 sec
50th	508 ms
75th	508 ms
90th	856 ms
Obs	28k

A transaction is a single request and a single server response, one period of data transfer, one or more acknowledgments, and observed latency caused by retransmitted packets.

Server Response Time

The Server Response Time view shows the time a server takes to respond to an application request. This view displays information for applications in the server segment. Where SPAN data exists, the management console uses it to populate this segment; otherwise, the management console uses FlowAgent data from the server side of the data center WAN-optimization device.

The following example view, which accompanies the Response Time: Average view, shows the same 50-fold increase in transactions for this HTTP application with no appreciable change in Server Response Time.



Server Response Time per 15 minute intervals

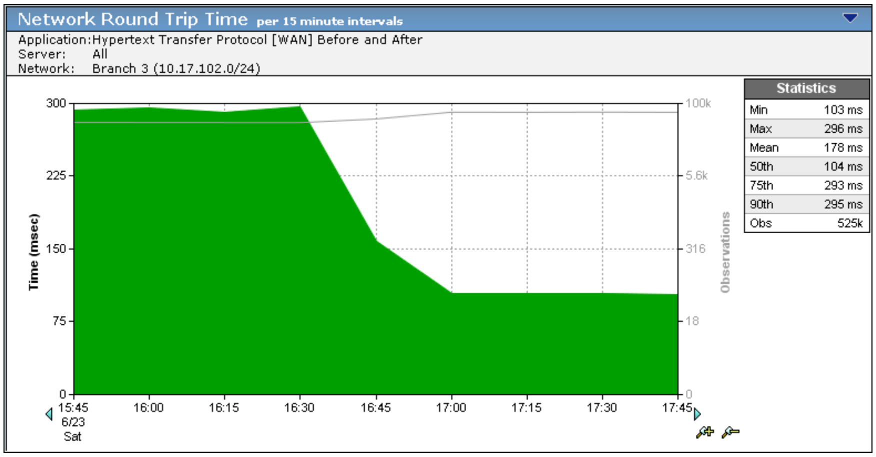
Application: Hypertext Transfer Protocol [Server] Before and After
Server: All
Network: Branch 3 (10.17.102.0/24)

Statistics	
Min	20.5 ms
Max	22.6 ms
Mean	22.1 ms
50th	22.0 ms
75th	22.1 ms
90th	22.6 ms
Obs	30k

The Response Time: Average view shows a significant improvement in data center efficiency. The Server Response Time view quantifies the benefit of data center offload whenever client-side caching is active. The Server Response Time value can be affected by server speed, application design, and volume of requests.

Network Round Trip Time

The Network Round Trip Time view shows the time it takes for a packet to travel the entire round trip between the local and remote WAN-optimization devices on a network, excluding latency caused by retransmissions. The following example view shows a significant improvement in Network Round Trip Time over the WAN when WAN-optimization is activated because queuing delays are reduced or eliminated.



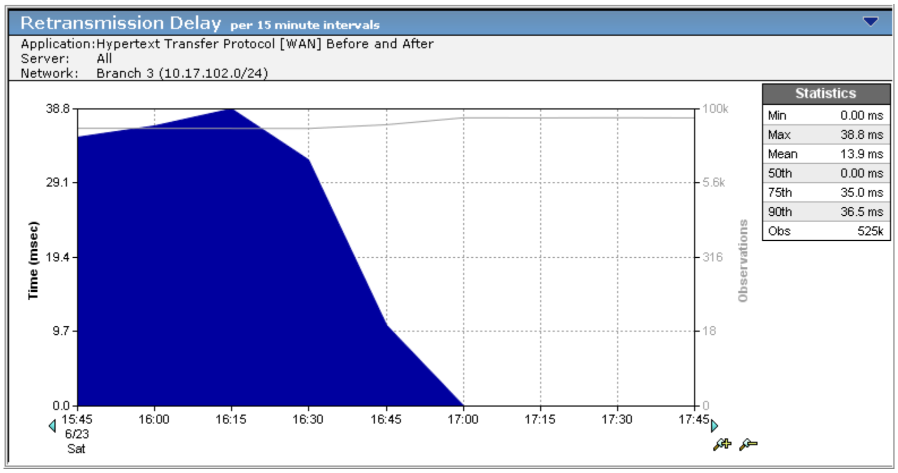
Network Round Trip Time per 15 minute intervals

Application:Hypertext Transfer Protocol [WAN] Before and After
Server: All
Network: Branch 3 (10.17.102.0/24)

Statistics	
Min	103 ms
Max	296 ms
Mean	178 ms
50th	104 ms
75th	293 ms
90th	295 ms
Obs	525k

Retransmission Delay

Retransmission Delay is the additional delay in the Network Round Trip Time caused by packets needing retransmission. The following example view shows that retransmissions over the WAN for this HTTP application cease when WAN-optimization is active.



Retransmission Delay per 15 minute intervals

Application:Hypertext Transfer Protocol [WAN] Before and After
Server: All
Network: Branch 3 (10.17.102.0/24)

Statistics	
Min	0.00 ms
Max	38.8 ms
Mean	13.9 ms
50th	0.00 ms
75th	35.0 ms
90th	36.5 ms
Obs	525k

The data is an average across all observations, not the actual retransmission time for one transaction. Given a total of five transactions, four with no retransmissions, and one with a 5-second Retransmission Delay, the view displays a 1-second Retransmission Delay.

Retransmission Delay is calculated by observing duplicate packets within the network from its vantage point next to the server. The monitoring device can see packets retransmitted by the server because of data losses in the server-to-client direction along the network path. These observations are included in the Retransmission Delay. When data loss occurs in the client-to-server direction (in the network path before reaching the server, for example), the monitoring device cannot observe such packet loss and that delay is not included in the Retransmission Delay metric.

The management console includes this retransmission delay in the Network Round Trip Time metric, which includes the server response and client acknowledgment; therefore, it is a delay in client acknowledgment caused by unseen Retransmission Delay that increases the Network Round Trip Time value. This metric does not reveal the impact of losses on the Data Transfer Time because of TCP congestion.

Retransmissions can cause a given session to reduce the transaction time, but the speed of the bytes across the line remains constant unless the path changes or congestion increases.

Packet Loss Percentage

The Packet Loss Percentage view shows the percentage of data lost on the monitored network and the loss rate in packets per second.

The monitoring device calculates Packet Loss Percentage by observing the ratio of retransmitted data to total data within the network from its vantage point next to the server. The monitoring device can see packets retransmitted by the server because of data losses in the server-to-client direction along the network path. When data loss occurs in the client-to-server direction (in the network path before reaching the server, for example), the monitoring device cannot observe such packet loss and that delay is not included in the Packet Loss Percentage.

Data Rate (in bits/second)

The Data Rate (in bits/second) view displays the data rate in bits/second (bytes/second times 8) across the WAN over a specified time period.

Data Rate (in packets/second)

The Data Rate (in packets/second) view shows the data rate in packets per second for an application over the WAN for a specified time period.

Data Volume (in bytes)

The Data Volume (in bytes) view indicates the total number of application-layer bytes seen on the network.

Data Volume (in packets)

The Data Volume (in packets) view indicates the total number of packets seen on a monitored network. Zero-byte packets, such as TCP acknowledgments, are counted in this view.

Compare the Effects of WAN Optimization

Use the Components report to compare the effects of WAN optimization on response time and volume metrics. The management console does not automatically select the time period for you. Browse to the corresponding time period where WAN optimization started or stopped to compare the response time difference.

To compare the effects of WAN optimization, filter the Components report to an application on a particular network.

Follow these steps:

1. From the Optimization page, click Performance in the Show Me menu.
2. Click Settings to specify the application and network you want.
3. Click OK.

   The Client Experience for Optimized Transactions view displays the application.
4. Click Performance, Components in the Show Me menu.
5. Click the Show Before and After Optimization option.

   The component reports now provide both unoptimized application performance data from the server SPAN and WAN-optimized performance data.
6. Click the left and right arrows to browse to the corresponding time period where WAN optimization started or stopped to compare the response time difference. The management console does not automatically select the time period for you.

Detect Spillover Traffic

When there is spillover for a combination that should be fully optimized (that is, a session that should be optimized is not optimized because the WAN-optimization device cannot take another session at that moment), SPAN measurements for the spillover sessions go to the [Server] segment.

No measurements for the spillover sessions affect the [Client] or [WAN] segments. Server Response Time [Server] is accurate, even though it contains measurements for spillover and optimized sessions. All metrics on the Optimization page are accurate. Network Round Trip Time [Server] shows an increase because it no longer gets 100% local ACKs. The spillover measurements show actual Network Round Trip Time out to the client. RetransDelay[Server] and PacketLossPct[Server] might also show increases.

In a properly sized WAN-optimization deployment, there should be little to no spillover. If you find a consistent difference between off hours and high-load hours for Network Round Trip Time [Server], you might have spillover. The solution is to increase the capacity of the WAN-optimization deployment. If all networks show the increase, the data center WAN-optimization device might lack adequate capacity. If only a few networks show the increase, those branch WAN-optimization devices might lack adequate capacity.

Another clue to spillover is when more sessions are reported on the [Server] segment than on the [Client] segment. This happens when non-optimized sessions for a combination that has at least some optimization are reported in the [Server] segment rather than in the parent application where they belong.

Chapter 8: Sharing Information from Report Pages and Views

This section contains the following topics:

Export a Report to a File

Export report data to a CSV file. You can export data from a particular view or all views on a report page to a CSV file or an XML file.

Export a Report Page to a CSV File

Export data from all views on a management console page to a CSV file. For example, you can export all the view data from the Operations page to a CSV file.

Follow these steps:

1. Click a report page.
2. Click Export.

   The File Download dialog box opens.
3. Click Save to save the report as a CSV file or click Open to view the CSV file.

Export a View to a CSV File

Export data from a view on the Engineering page to a CSV file.

Follow these steps:

1. Click the Engineering page.
2. Click the blue arrow ⚙ menu on the view you want and select Export to CSV.

   The File Download dialog box opens.
3. Click Save to save the view as a CSV file or click Open to open the CSV file.

Export a View to an XML File

Export data from a view on the Engineering page to an XML file.

Follow these steps:

1. Click the Engineering page.
2. Click the blue arrow ⚙ menu on the view you want and select Export to XML.

   The XML-formatted data is displayed.

Email a Report Page

Email a report page in HTML format to specified recipients. When you email a report page, you can send the report immediately or schedule it for sending later.

Follow these steps:

1. Click a report page.
2. Click Email.

   Send E-Mail Properties opens.

3. Specify the following settings and click OK:

   Email Properties

   Notifies the specified email recipients. Note that the email notification is scheduled using the specified time zone.

   Scheduling Options

   The scheduling options only appear if you chose to schedule the notification. Schedule the notification to run at a particular time on a particular date or schedule the notification on a weekly or monthly basis.

Print a Report Page

Use the Print Preview button to format a report page for printing, and then use the web browser to print the report.

Follow these steps:

1. Click a report page.
2. Click Print Preview.

   The management console displays the report page in a browser window.
3. Use the Web browser to print the report page.

Chapter 9: Troubleshooting

This section contains the following topics:

Overview

Use the management console to:

- Quantify the performance of applications delivered over the enterprise network with operational level reporting
- Validate the impact of planned or unplanned changes with before-and-after analysis
- Solve performance problems faster by automatically identifying end-user issues, isolating the cause, and then gathering diagnostic data when incidents occur

The source of performance problems can be any one or a combination of these three elements of an implementation:

- Network infrastructure, such as latency introduced by circuits, traffic congestion, routers, and switches.
- Server infrastructure, such as latency introduced by CPU processing, memory I/O, or disk read and writes.
- Application architecture, for example, writing large data requests in numerous small packets for transmission.

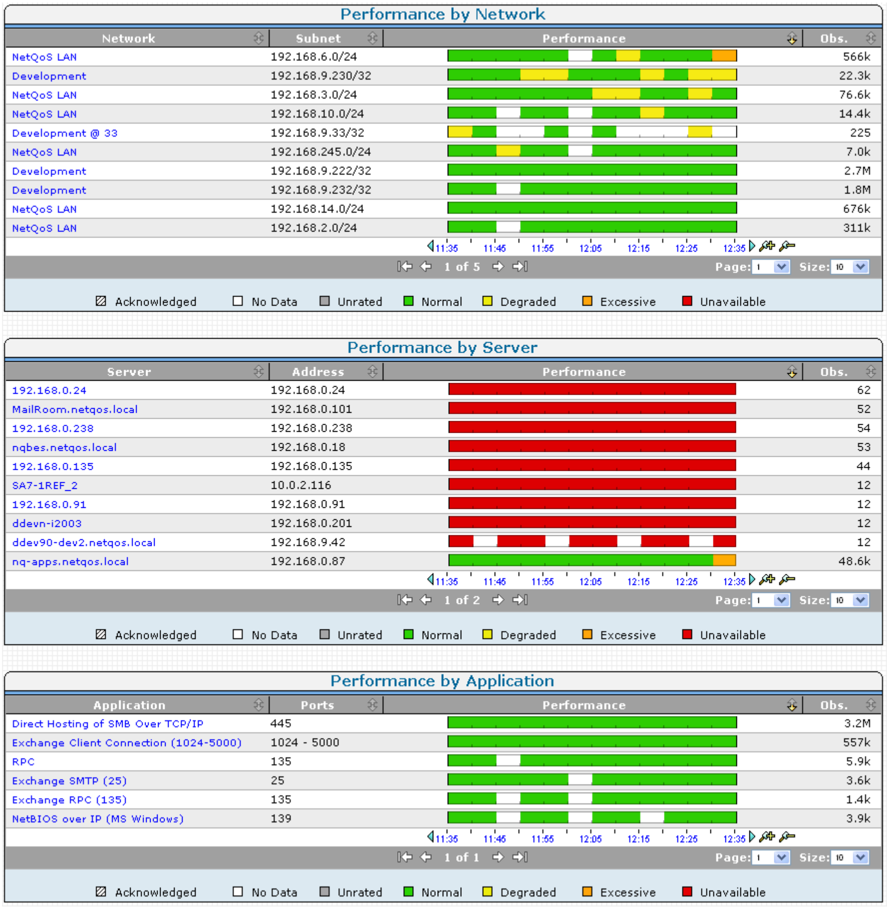
You can determine the origin of performance problems using the following method to review and analyze management console data:

1. Identify the application showing performance problems using Performance Maps and Application Detail views.
2. Isolate the time component contributing to the performance problem. The colors in the Response Time Component view identify each component making up the total time.
3. Investigate the contributing factors to the performance problem by examining the Traffic, Sessions, Trends, Response Size, QoS, and Statistics pages.

Use the Operations Page

The Operations page displays horizontal bar charts that compare overall performance elements to a threshold. The worst-performing monitored networks, servers, and applications are listed in descending order on each view.

You can quickly see the areas of concern in the enterprise.



Performance by Network

Network	Subnet	Performance	Obs.
NetQoS LAN	192.168.6.0/24		566k
Development	192.168.9.230/32		22.3k
NetQoS LAN	192.168.3.0/24		76.6k
NetQoS LAN	192.168.10.0/24		14.4k
Development @ 33	192.168.9.33/32		225
NetQoS LAN	192.168.245.0/24		7.0k
Development	192.168.9.222/32		2.7M
Development	192.168.9.232/32		1.8M
NetQoS LAN	192.168.14.0/24		676k
NetQoS LAN	192.168.2.0/24		311k

11:35 11:45 11:55 12:05 12:15 12:25 12:35
1 of 5
Page: 1 Size: 10
▨ Acknowledged □ No Data ■ Unrated 🟩 Normal 🟨 Degraded 🟧 Excessive 🟥 Unavailable

Performance by Server

Server	Address	Performance	Obs.
192.168.0.24	192.168.0.24		62
MailRoom.netqos.local	192.168.0.101		52
192.168.0.238	192.168.0.238		54
nqbes.netqos.local	192.168.0.18		53
192.168.0.135	192.168.0.135		44
SA7-1REF_2	10.0.2.116		12
192.168.0.91	192.168.0.91		12
ddevn-i2003	192.168.0.201		12
ddev90-dev2.netqos.local	192.168.9.42		12
nq-apps.netqos.local	192.168.0.87		48.6k

11:35 11:45 11:55 12:05 12:15 12:25 12:35
1 of 2
Page: 1 Size: 10
▨ Acknowledged □ No Data ■ Unrated 🟩 Normal 🟨 Degraded 🟧 Excessive 🟥 Unavailable

Performance by Application

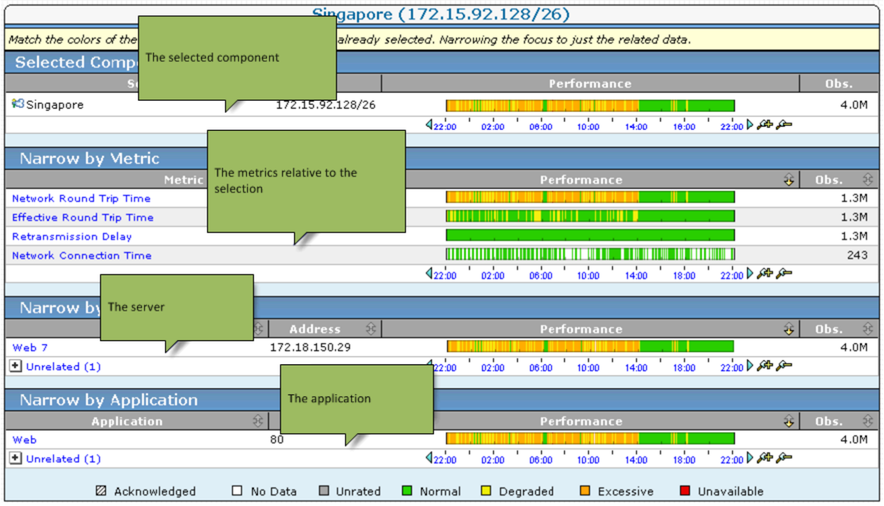
Application	Ports	Performance	Obs.
Direct Hosting of SMB Over TCP/IP	445		3.2M
Exchange Client Connection (1024-5000)	1024 - 5000		557k
RPC	135		5.9k
Exchange SMTP (25)	25		3.6k
Exchange RPC (135)	135		1.4k
NetBIOS over IP (MS Windows)	139		3.9k

11:35 11:45 11:55 12:05 12:15 12:25 12:35
1 of 1
Page: 1 Size: 10
▨ Acknowledged □ No Data ■ Unrated 🟩 Normal 🟨 Degraded 🟧 Excessive 🟥 Unavailable

To isolate the component and metric contributing to the performance problem, click one of the worst-performing items at the top of its view.

In the following example, the Singapore network provides information about:

- The metrics and related components of interest
- Threshold-based color patterns indicating the affected items
- Performance patterns that match the selected component



Singapore (172.15.92.128/26)

Match the colors of the *already selected. Narrowing the focus to just the related data.*

The selected component

The metrics relative to the selection

The server

The application

Selected Comp

S		Performance	Obs.
Singapore	172.15.92.128/26		4.0M

Narrow by Metric

Metric	Performance	Obs.
Network Round Trip Time		1.3M
Effective Round Trip Time		1.3M
Retransmission Delay		1.3M
Network Connection Time		243

Narrow by

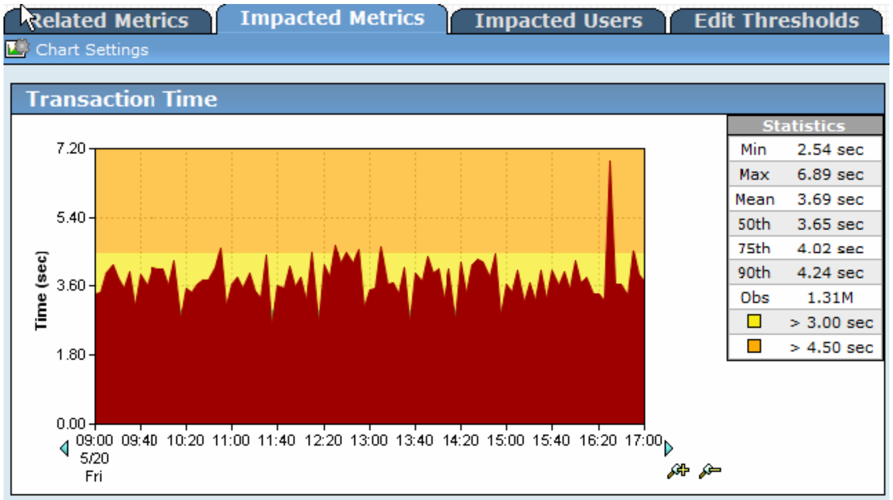
	Address	Performance	Obs.
Web 7	172.18.150.29		4.0M
⊞ Unrelated (1)			

Narrow by Application

Application		Performance	Obs.
Web	80		4.0M
⊞ Unrelated (1)			

▨ Acknowledged □ No Data ▨ Unrated 🟩 Normal 🟨 Degraded 🟧 Excessive 🟥 Unavailable

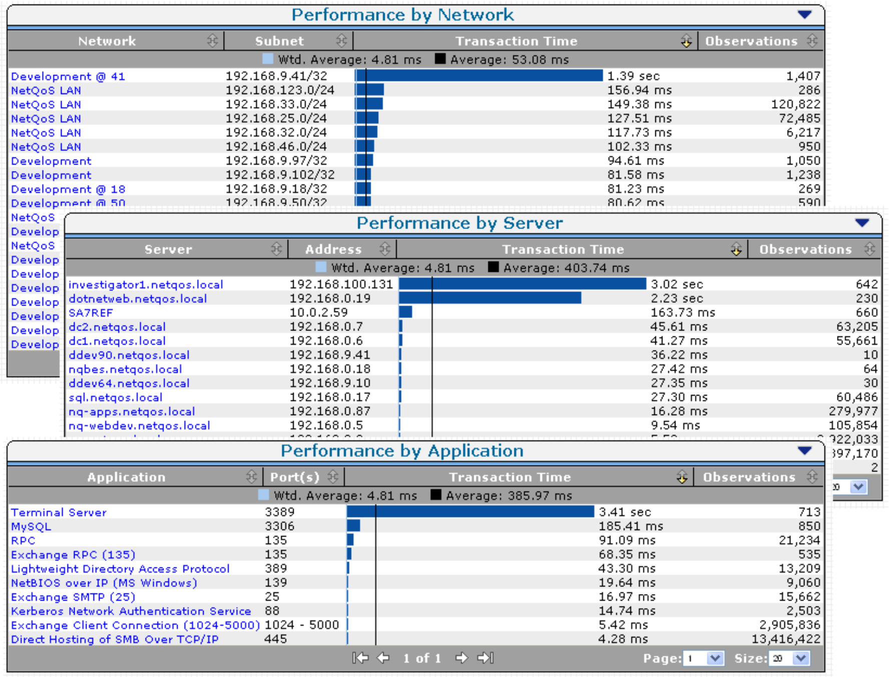
To investigate the performance further, select the matching performance profiles and then click Explore to see more details:



Statistics	
Min	2.54 sec
Max	6.89 sec
Mean	3.69 sec
50th	3.65 sec
75th	4.02 sec
90th	4.24 sec
Obs	1.31M
🟨	> 3.00 sec
🟧	> 4.50 sec

Use the Engineering Page

The Performance Maps display in the Engineering page. The format is a horizontal bar chart comparing various elements -- Total Transaction Time per configured application sorted by the longest times (slowest applications) on top, Network Round Trip Time per configured network sorted by the longest times (slowest networks) on top, and Server Response Time per configured server sorted by the longest response times (slowest servers) on top.



Performance by Network

🟦 Wtd. Average: 4.81 ms ■ Average: 53.08 ms

Network	Subnet	Transaction Time	Observations
Development @ 41	192.168.9.41/32	1.39 sec	1,407
NetQoS LAN	192.168.123.0/24	156.94 ms	286
NetQoS LAN	192.168.33.0/24	149.38 ms	120,822
NetQoS LAN	192.168.25.0/24	127.51 ms	72,485
NetQoS LAN	192.168.32.0/24	117.73 ms	6,217
NetQoS LAN	192.168.46.0/24	102.33 ms	950
Development	192.168.9.97/32	94.61 ms	1,050
Development	192.168.9.102/32	81.58 ms	1,238
Development @ 18	192.168.9.18/32	81.23 ms	269
Development @ 50	192.168.9.50/32	80.62 ms	590
NetQoS			
Develop			
NetQoS			
Develop			
Develop			
Develop			
Develop			
Develop			
Develop			
Develop			

Performance by Server

🟦 Wtd. Average: 4.81 ms ■ Average: 403.74 ms

Server	Address	Transaction Time	Observations
investigator1.netqos.local	192.168.100.131	3.02 sec	642
dotnetweb.netqos.local	192.168.0.19	2.23 sec	230
SA7REF	10.0.2.59	163.73 ms	660
dc2.netqos.local	192.168.0.7	45.61 ms	63,205
dc1.netqos.local	192.168.0.6	41.27 ms	55,661
ddev90.netqos.local	192.168.9.41	36.22 ms	10
nqbes.netqos.local	192.168.0.18	27.42 ms	64
ddev64.netqos.local	192.168.9.10	27.35 ms	30
sql.netqos.local	192.168.0.17	27.30 ms	60,486
nq-apps.netqos.local	192.168.0.87	16.28 ms	279,977
nq-webdev.netqos.local	192.168.0.5	9.54 ms	105,854
			922,033
			397,170
			2

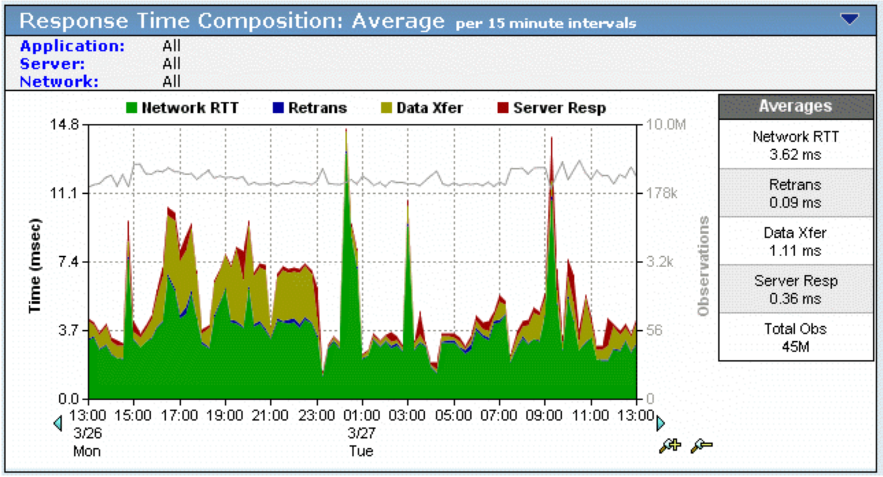
Performance by Application

🟦 Wtd. Average: 4.81 ms ■ Average: 385.97 ms

Application	Port(s)	Transaction Time	Observations
Terminal Server	3389	3.41 sec	713
MySQL	3306	185.41 ms	850
RPC	135	91.09 ms	21,234
Exchange RPC (135)	135	68.35 ms	535
Lightweight Directory Access Protocol	389	43.30 ms	13,209
NetBIOS over IP (MS Windows)	139	19.64 ms	9,060
Exchange SMTP (25)	25	16.97 ms	15,662
Kerberos Network Authentication Service	88	14.74 ms	2,503
Exchange Client Connection (1024-5000)	1024 - 5000	5.42 ms	2,905,836
Direct Hosting of SMB Over TCP/IP	445	4.28 ms	13,416,422

You can Isolate the component time contributing to a performance problem by clicking one of the items listed on the view. The page refreshes and additional information focused on your selection displays.

The following example shows the Response Time Composition: Average view:

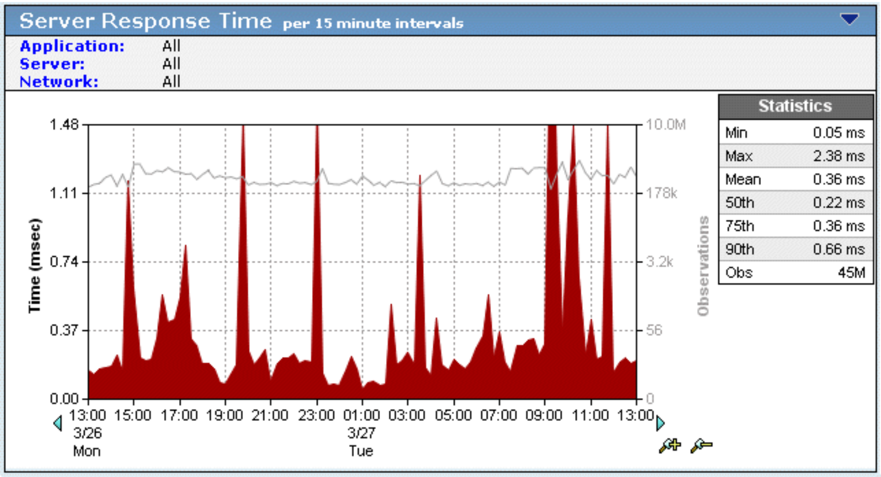


Response Time Composition: Average per 15 minute intervals

Application: All
Server: All
Network: All

Averages
Network RTT 3.62 ms
Retrans 0.09 ms
Data Xfer 1.11 ms
Server Resp 0.36 ms
Total Obs 45M

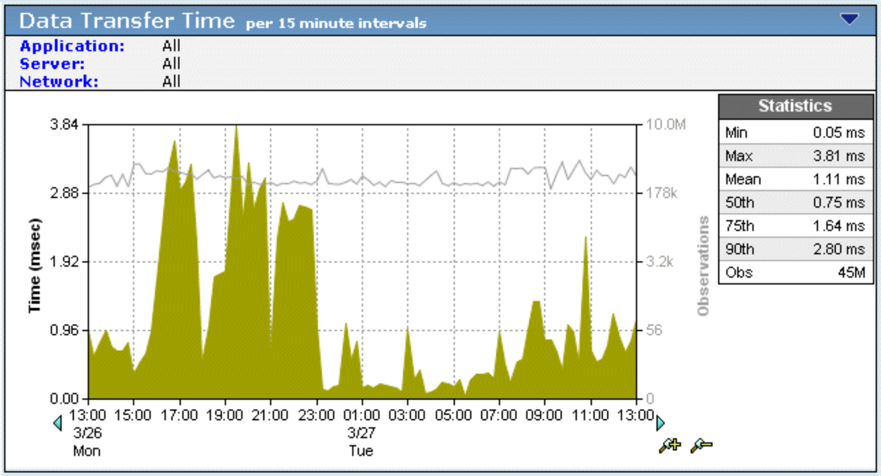
The components of the Response Time Composition: Average view are stacked beneath it, as partially shown in the following image:



Server Response Time per 15 minute intervals

Application: All
Server: All
Network: All

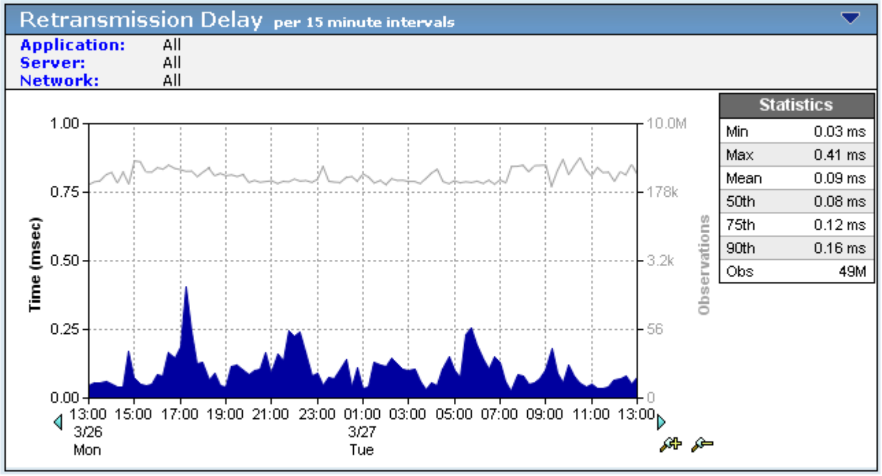
Statistics	
Min	0.05 ms
Max	2.38 ms
Mean	0.36 ms
50th	0.22 ms
75th	0.36 ms
90th	0.66 ms
Obs	45M



Data Transfer Time per 15 minute intervals

Application: All
Server: All
Network: All

Statistics	
Min	0.05 ms
Max	3.81 ms
Mean	1.11 ms
50th	0.75 ms
75th	1.64 ms
90th	2.80 ms
Obs	45M



Retransmission Delay per 15 minute intervals

Application: All
Server: All
Network: All

Statistics	
Min	0.03 ms
Max	0.41 ms
Mean	0.09 ms
50th	0.08 ms
75th	0.12 ms
90th	0.16 ms
Obs	49M

When viewing the stacked component view, the component with the most color overall or in any spike is the main contributor to the slow performance of the application at that time.

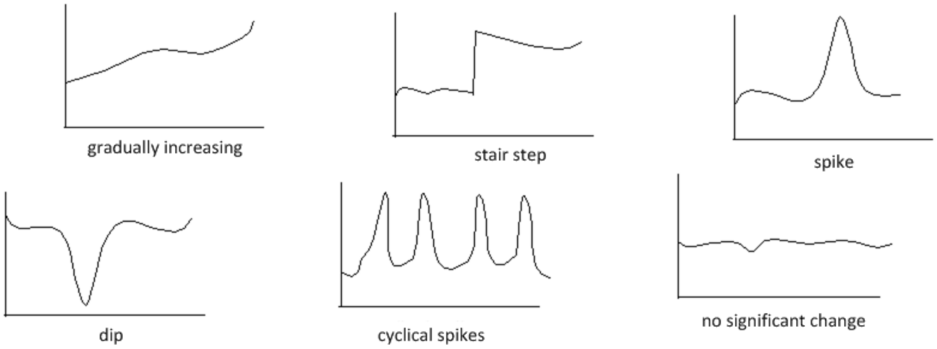
Identify the component contributing the most latency to the overall application performance, and then scroll down to the individual component view to begin investigating the extent of the performance issue.

Investigate the extent of the performance issue by following these general guidelines:

1. Start with Total Transaction Time.
2. Drill down to the stacked views - use Daily views created with data points at 15-minute intervals. If you cannot see performance events at 15-minute intervals, lower the interval to 5 minutes.
3. View spike events as Last Hour or 8-Hour views for 5-minute intervals for high data resolution.
4. Correlate a spike or color pattern observed in one chart with any similar spikes or dips or color pattern that occurs at the same time to other metrics.
5. Open multiple Web sessions to facilitate comparisons. Also use the Zoom In - Zoom Out features to find large-scale patterns in particular times of day, days of week and month related issues.

General Troubleshooting

Use the management console to narrow the source of an issue to a problem with the network, server, or application. A suggested approach to interpreting the data you see in a series of reports is described in this section. Note the change in metric views that follow these indicative patterns:



Investigate to find which component of the metric is causing this change.

Determining the Cause of a Problem

1. Click the Operations page, select one or more components (the relative network, server, application) and click Links: Engineering to see your choices reflected on the Engineering page.
2. Click Components in the Show Me menu.
3. Begin your investigation with the Components report. This report is composed of various response time views stacked on top of each other to show how each contributes to the overall response time. The components in the Components report are:

   - Response Time Composition: Average
   - Server Response Time
   - Data Transfer Time
   - Retransmission Delay
   - Network Round Trip Time
   - Effective Network Round Trip Time

   Given a point in time on the view, the component color dominating the view indicates the component contributing the most delay to the overall application response time.

   Considered independently:

   - High values of Server Response Time indicate there could be issues with the server
   - High values of Data Transfer Time typically indicate that the application is the source of the problem
   - High values of NRTT or Retransmission Time typically indicate problems in the network
4. After you identify the component of response time that is contributing the most delay to the Component report, look at the individual view for that component by scrolling down the page of charts.
5. In the individual components view, notice how many observations were recorded when the issue occurred. The number of observations helps to identify the problem source: server, network, or application.

   The number of observations with respect to what is "normal" helps you to understand:

   - The significance of the event -- higher observation counts typically indicate a significant impact on the users of the application being analyzed, while lower values indicate the impact on end users was limited and/or not enough data points exist to analyze the issue.

- The relevant contribution of the application of interest -- lower observation counts might indicate that the performance event was not caused by the application(s) being analyzed.

6. Follow the relevant process to confirm these results and to identify the source problem area as being the server, the network, or the application. In rare cases, a performance problem might have it source in more than one of the possible contributors, but that is unusual.

Consider these possibilities:

- Finding 1: A positive finding indicating that the problem source is the server, the network, or the application. For example, the data shows the server was the issue.
- Finding 2: A negative finding indicating that the problem source was not the network, server, or application. For example, the data clearly shows the network was not the issue.
- Finding 3: A negative finding indicating the problem source was not the final remaining network, server, or application. For example, the data further shows the application was not the issue.

The findings demonstrate that the server can be ruled in as the culprit while the network and application can be ruled out as not being culprits. With such a three-pronged finding, one might rest confident in identifying the root culprit and proceeding with remediation.

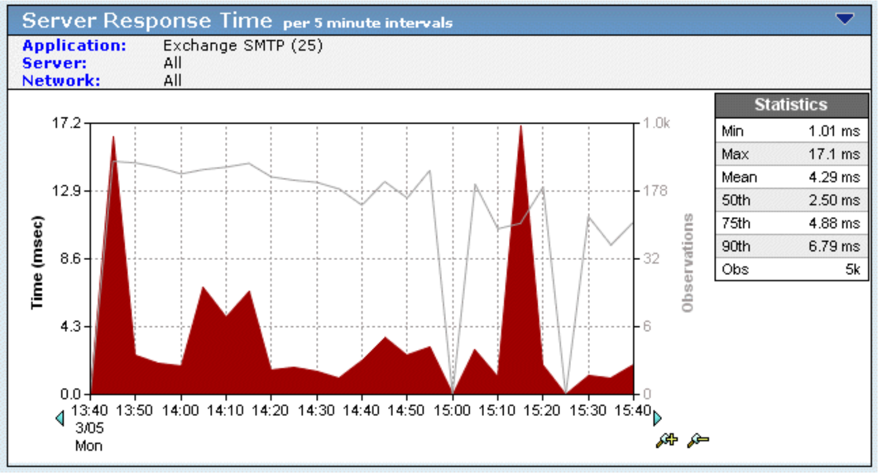
Note: Leave your mouse cursor at the point or interest in the Response Time Component view and use the down arrow keys to scroll down the page, bringing the lower views into view under the mouse cursor. As you move down the page with your arrow keys, the mouse pointer maintains its exact position on the screen, thus permitting you to easily align the given point on the Component view with the views that are lower on the page.

Increase in Server Response Time

An increase in Server Response Time is correlated with observation counts -- both an increase or decrease is considered in the following analysis.

Increase in Server Response Time and Observations Count

An increase in the Server Response Time and in the number of Observations is a strong indicator that a performance problem is associated with the server. This conclusion can be reinforced by correlating it with other relative data.



Server Response Time per 5 minute intervals

Application: Exchange SMTP (25)
Server: All
Network: All

Statistics	
Min	1.01 ms
Max	17.1 ms
Mean	4.29 ms
50th	2.50 ms
75th	4.88 ms
90th	6.79 ms
Obs	5k

Performance problems associated with a server should be visible across all network sets and aggregations -- both local network sets, including users in the same building as the data center, and remote network sets, including users across WAN connections.

If both the Server Response Time and number of Observations peak at the same point in time as the observed performance issue, review the following data sets for the same point in time:

- Traffic -- [Data Volume, Data Rate]

  Check whether the Data Volumes/Rates increased. Servers work harder when writing higher data volumes to the network. Abnormal increases in data volumes that coincide with increases in Server Response Time indicate a server having difficulty keeping up with demand.

- Sessions -- [Connection Setup Time, TCP/IP Sessions, Unfulfilled TCP/IP Session Requests, TCP/IP Session Times]

  Check whether there is a concurrent increase in Server Connection Setup Time that could indicate the OS kernel increased the amount of time that it took to respond to new session requests.

  Check whether the number of TCP/IP sessions increased by a significant number, say greater than 10%. Additional TCP sessions and accompanying application requests require more resources from the server and tax its horsepower.

  Check whether there was an abnormal increase in the number of Unfulfilled TCP/IP requests. A high increase is a significant indicator the hardware resources of the server are overburdened.

- QoS -- [Users, Packet Loss Percentage, User Goodput, Composite Rate Per User]

Check whether there is a significant increase in the number of users. Increases in the number of users increases the demand on server resources. The point when a certain number of users cause the Server Response Time to degrade can be interpreted as a future proactive point for upgrading servers or load balancing the application among similarly-configured servers.

- Statistics -- [Response Time Composition: Standard Deviation, Server Response Time Percentiles, Data Transfer Time Percentiles, Retransmission Delay Time Percentiles, Network Round Trip Time Percentiles]

  Check whether there was an increase in the standard deviation for Server Response Time and/or Percentiles. This could indicate inconsistent and sporadic performance by the server as seen in more "outlying" data points that are at significantly varying distances from the average, and is a strong indicator of server-based issues.

Increase in Server Response Time and Decrease in Observations Count

An increase in the Server Response Time while the number of Observations decreases is a possible indicator of two very different events:

- Another application on the server is responsible for the increase in Server Response Time. This application might or might not be monitored by management console.
- The application service is unreliable due to instability on the server in CPU, Memory, or NIC. This might result in sporadic loss of service or ultimately a complete service outage.

If the Server Response Time spikes while the number of Observations dips at the same time as the observed performance issue, complete the following steps to determine which event caused the degradation in performance:

Determine the Cause of a Spike in SRT and Dip in Observation Count

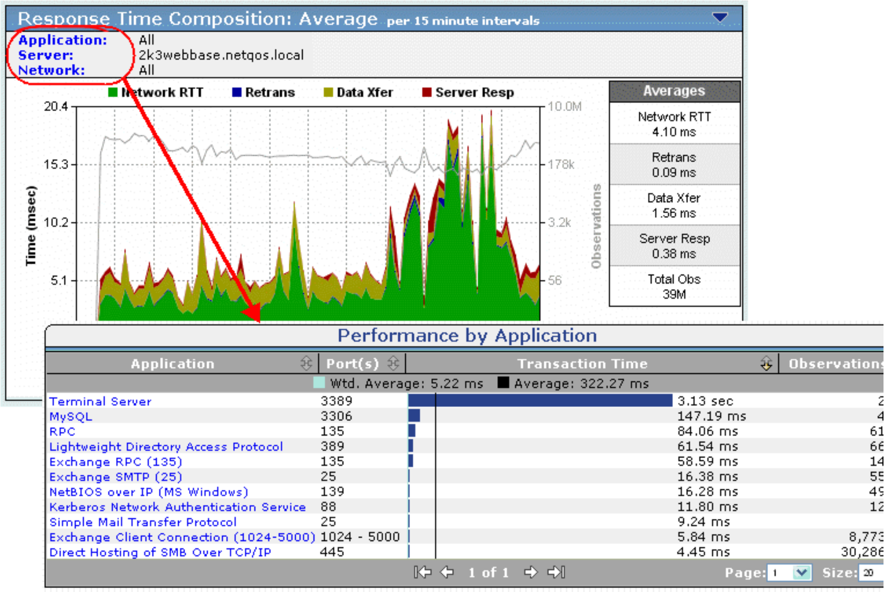
First determine if another application is active on the server.

1. Click the Engineering page.

2. Select the following Settings from any Response Time view:

   - Application -- All
   - Server -- select the name of the server
   - Network Set -- All

3. Click the blue hypertext Application in the header of the Response Time Composition view to see all applications that are being monitored the server:



Response Time Composition: Average per 15 minute intervals

Application: All
Server: 2k3webbase.netqos.local
Network: All

Averages

Averages
Network RTT 4.10 ms
Retrans 0.09 ms
Data Xfer 1.56 ms
Server Resp 0.38 ms
Total Obs 39M

Performance by Application

Wtd. Average: 5.22 ms ■ Average: 322.27 ms

Application	Port(s)	Transaction Time	Observation
Terminal Server	3389	3.13 sec	2
MySQL	3306	147.19 ms	4
RPC	135	84.06 ms	61
Lightweight Directory Access Protocol	389	61.54 ms	66
Exchange RPC (135)	135	58.59 ms	14
Exchange SMTP (25)	25	16.38 ms	55
NetBIOS over IP (MS Windows)	139	16.28 ms	49
Kerberos Network Authentication Service	88	11.80 ms	12
Simple Mail Transfer Protocol	25	9.24 ms	
Exchange Client Connection (1024-5000)	1024 - 5000	5.84 ms	8,773
Direct Hosting of SMB Over TCP/IP	445	4.45 ms	30,286

1 of 1 Page: 1 Size: 20

If another application appears in the resulting Performance Map, repeat these steps to determine if it is the problem source for the performance issue.

The key is an increase in the number of observations for the particular application at the same time the performance problem was reported.

4. If no other application appears in the resulting Performance Map, use the back arrow to return to the Engineering page. Click Trends on the horizontal menu and check whether this performance event demonstrates a pattern over the past weeks and month. If a pattern appears you can use a protocol analyzer at the specific time in question on a projected recurring time and date to identify the problem source application.

   Examples of applications that create issues for the primary application on a server are:

   - Backups performed at night
   - Upgrades to anti-virus software

- Agents used by other teams to measure the server capacity or performance

Backup processes are typically scheduled through the backup software agent for recurring time periods. Therefore, there should be a pattern visible on the Trends views showing cyclical spikes of Server Response Time at the same time every 24-hour period, every other 24-hour period, or whatever the backup schedule dictates.

Upgrades to anti-virus definitions are typically done on a weekly basis, or in emergency situations on an "as needed" basis. Consult with your anti-virus software vendor and/or desktop/security team to determine the automated update release schedule. Occasionally, application development teams might install third party agents or encode performance agents into their software. Review change notifications to determine if any changes to the software on the server have been made just prior the start of the performance issue.

Determine if the server/application service has become unreliable.

1. Set up thresholds in the management console to launch an investigation when Server Response Time exceeds accepted values. The management console automatically gathers relevant information such as CPU and memory utilization.
2. Review of server system logs can reveal errors and other events that might be affecting the stability of the application.
3. Check the switch port facing the server and the server NIC to ensure that it is set for the correct duplex and speed settings from the following table, and that it is free from errors.

Server	**Switch**	**Result**
Auto	Auto	Full duplex, auto speed
Auto	Manual	Half duplex, manual speed
Manual	Auto	Half duplex, manual speed
Manual - Full	Manual - Full	Full duplex, manual speed (Assumes same speed, 10 Mbps, 100 Mbps, 1000 Mbps, is set on both ends)

Increase in Network Round Trip Time (NRTT)

Network Round Trip Time can be defined by the following equation:

NRTT = S_Delay + Q_Delay + R/SW_Delay + D_Delay + P_Delay

Where:

S_Delay

Serialization Delay - [(Frame size * 8)/(Access Rate)]

Q_Delay

Queue Delay - dependent on utilization and S_Delay

R/SW_Delay

Routing/Switch Delay - typically no greater than 1 ms per hop

D_Delay

Distance Delay - propagation delay due distance traveled. Typically 5ìs/km for fiber, 5.56ìs/km for copper, 3.3ìs/km satellite

P_Delay

Protocol Delay - delay added by transmission or higher level protocols

For example: CSMA/CD for shared Ethernet

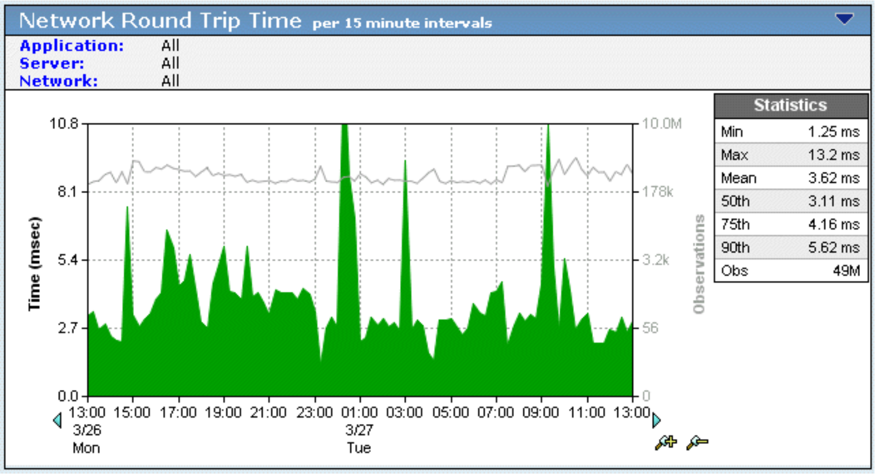
Generally, increases in NRTT associated with an application are caused by an increase in any of the variables listed above. However, the typical reasons for an increase in NRTT are listed in the order they typically occur:

- Increase in Q_Delay caused by increased utilization of a circuit, thus deep network queues
- Increase in D_Delay because of carrier or enterprise fail-over to protected/redundant path that is longer in distance
- Increase in R/SW_Delay because of network errors
- Increase in S_Delay because of enterprise failover to redundant path with lower bandwidth

To determine if a network issue is limited to a single remote site, contrast and compare spikes in NRTT of the affected remote site with other sites that are comparable in terms of distance from server, bandwidth, and user count. If the NRTT increases or spikes among multiple sites at the same point in time, the issue could be carrier related or might be caused by instability in the routing protocol.

Increase in NRTT and Observations Count

An increase in the NRTT and in the Observations count is a strong indicator that a performance problem is based on an application utilization of the network. The strength of this indicator can be reinforced by correlating it with other corresponding data points to build a complete finding that the problem source is the network.



Application: All
Server: All
Network: All

Statistics	
Min	1.25 ms
Max	13.2 ms
Mean	3.62 ms
50th	3.11 ms
75th	4.16 ms
90th	5.62 ms
Obs	49M

If both the NRTT and number of Observations peak at the same point in time as the observed performance issue, review the following data sets for the same point in time:

- Components -- [Retransmission Delay]

  Check whether the length of retransmissions increased. Retransmissions indicate that network queues are filling at a rate faster than they can be emptied, thus incurring packet drop and related TCP retransmissions

- Sessions -- [Connection Setup Time, TCP/IP Sessions]

  Check whether there is a concurrent increase in Network Connection Setup Time. This increase indicates the three-way TCP handshake is being delayed on the network because of queue depth being increased by other pre-established sessions within the network.

  Check whether the number of TCP/IP sessions increased by a significant number (greater than 10%). Additional TCP sessions and accompanying application data require more bandwidth.

- Traffic -- [Data Volume, Data Rate]

  Check whether the Data Volumes/Rates increased. Higher volumes of data on the network increase queue depth and related delay. Abnormal increases in data volumes that coincide with increases in NRTT indicate a network having difficulty keeping up with demand.

- QoS -- [Users]

Check whether there is a significant increase in the number of users. Increases in network utilization typically coincide with increases in numbers of users. The point at which a certain number of users cause the NRTT to degrade can be interpreted as a future proactive point for upgrading network bandwidth for other similar sites.

- Statistics -- [Response Time Composition: Standard Deviation, Network Round Trip Time Percentiles]

  Check whether there was an increase in the standard deviation for NRTT and/or Percentiles. This increase indicates inconsistent and sporadic performance by the network as evidenced in more "outlying" data points (such as, points that are at significantly varying distances from the average), and is a strong indicator of network based issues.

Increase in NRTT and Decrease in Observations Count

An increase in the NRTT while the Observations count decreases is a possible indicator of two very different events:

- Another application on the network is responsible for the increase in NRTT. This application might, or might not, be monitored by the management console.
- The application service has become unreliable because of instability on the network (link failures, routing divergence, STP divergence, carrier failover, and other errors). This might result in sporadic loss of service or ultimately a complete service outage.

If the NRTT spikes while the Observations count dips at the same point in time as the observed performance issue, complete the following actions to determine which event mentioned above caused the degradation in performance.

Determine the Cause of a Degradation in Performance

Determine if another application is active on the network:

1. Click the Engineering page.
2. Select the following Settings from any Response Time view:
   - Application -- All
   - Server -- All
   - Network Set - the relevant aggregation, such as the remote site aggregation as shown below
3. On the Response Time Composition view, click the blue hypertext Application link to see all applications that are being monitored by the management console on this server.

   If another application appears in the resulting Performance Map, complete these steps again for this application to determine if it is the problem source of the performance issue.

   The key is an increase in the number of observations at the same time a performance issue was reported.
4. If no other application appears in the resulting Performance Map, use the back arrow to go back to the main Engineering page. Select Trends from the horizontal menu and check whether this performance event demonstrates a pattern over the past weeks and month. If a pattern appears you need to use historic NetFlow data, IP accounting, or protocol analyzer data to determine if an application was saturating the network at the times in question. If no historical data is available, you can manually review the time in question on a projected recurring time and date to identify the problem source application.

   Examples of applications that create issues for the primary application on a network are:
   - Replication or backup data between remote sites
   - Large file transfers between servers
   - Users streaming large amounts of data on subrate WAN links (< T1)
   - Anti-virus upgrades across subrate WAN links

Determine if network has become unreliable:

1. Check the switch port facing the server and the server NIC to ensure that it is set for the correct duplex and speed settings (see the following table) and is free from errors.

Server	Switch	Result
Auto	Auto	Full duplex, auto speed

Server	Switch	Result
Auto	Manual	Half duplex, manual speed
Manual	Auto	Half duplex, manual speed
Manual - Full	Manual - Full	Full duplex, manual speed (Assumes same speed is set on both ends)

2. Determine if the network has become unreliable because of configuration issues or network errors by setting up thresholds in the management console to launch an investigation when NRTT exceeds accepted values. The management console automatically gathers relevant information from routers and switches.
3. Review router and switch logs, interface errors, and change records to discover any events that might be affecting the stability of the network, such as routing divergence and circuit errors.

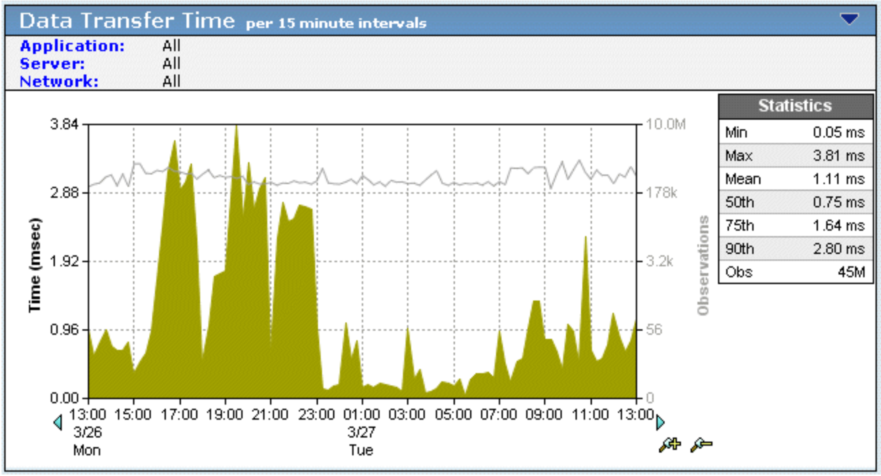
Increase in Data Transfer Time

An increase in Data Transfer Time is correlated with observation counts -- both an increase or decrease is considered in the following analysis.

Increase in Data Transfer Time and Observations Count

An increase in the Data Transfer Time and in Observations count is a good indicator that a performance problem is associated with the application.

The strength of this indicator can be significantly increased by correlating it with corresponding data points to build a complete finding that the problem source is the application.



Data Transfer Time per 15 minute intervals

Application: All
Server: All
Network: All

Statistics	
Min	0.05 ms
Max	3.81 ms
Mean	1.11 ms
50th	0.75 ms
75th	1.64 ms
90th	2.80 ms
Obs	45M

Performance problems associated with an application vary by site depending on available bandwidth and latency caused by distance. It is not uncommon for a subset of application users located at the same site to complain about application performance. You can query across applications to see if the spike in Data Transfer Time can be correlated to applications crossing the same network path, or originating from the same server. If so, the problem might be in the network or server.

If an increase in Data Transfer Time is isolated to a single application on a single server, and both the Data Transfer Time and number of Observations spike at the same point in time as the observed performance issue, review the following data sets for that time:

- Traffic -- [Data Volume, Data Rate]

  Check whether the Data Volumes/Rates increased. The application might not be writing effectively across slow WAN links or to sites that are far way in terms of distance.

- Components -- [Network Round Trip Time, Retransmission Delay, Server Response Time]

  Check whether the NRTT and Retransmission Delay remained somewhat constant as the volume increased. Somewhat constant and acceptable Network Round Trip Time coupled with insignificant packet drop indicates an application issue.

  Check whether Server Response Time increased. Increases in Server Response Time without increases in session or user counts, unfulfilled TCP session requests, and NRTT indicate the problem is the application.

- Sessions -- [Connection Setup Time, TCP/IP Sessions, Unfulfilled TCP/IP Sessions]

  Check whether there is a concurrent increase in Server Connection Setup Time. Such an increase indicates the OS kernel increased the amount of time that it took to respond to new session requests.

  Check whether the number of TCP/IP sessions increased by a significant number (greater than 10%). Additional TCP sessions and accompanying application requests require more resources from the server and tax its horsepower.

  Check whether Unfulfilled TCP/IP requests remained constant (optimally zero). An absence of unfulfilled TCP requests during active session establishment demonstrates server resources were available to users, and further relieves the server of responsibility for the performance issue.

- QoS -- [Users]

  Check whether there are a constant number of users before, during, and after the performance event. The absence of new users to the server at the time of the performance event tends to absolve server load as a factor.

- Statistics -- [Response Time Composition: Standard Deviation, Data Transfer Time Percentiles]

  Check whether there was an increase in the standard deviation for Data Transfer Time and/or Percentiles -- without an accompanying increase in NRTT or a significant increase in Server Response Time. Such an increase indicates inconsistent and sporadic performance by the application as evidenced in more "outlying" data points (such as, points that are at significantly varying distances from the average), and is a good indicator of application based issues.

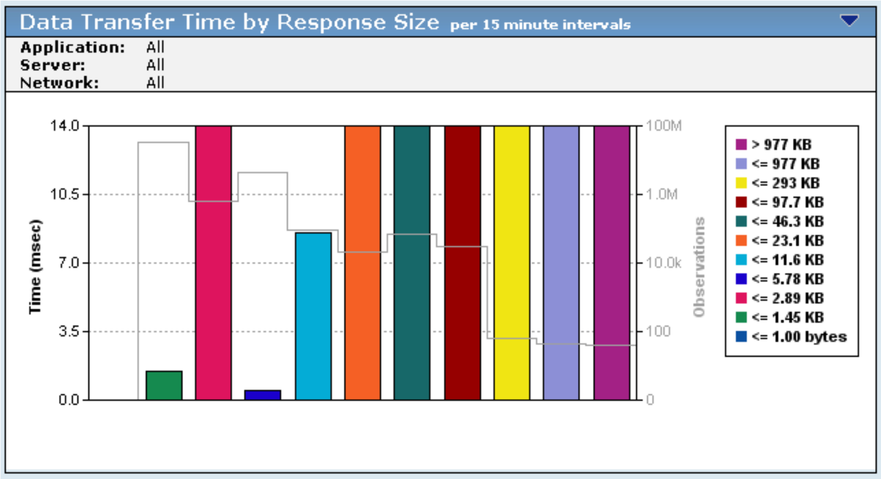
Data Transfer Time = 0 (Ping-Pong Application)

A "Ping-Pong" application sends single-packet responses to most or all requests; it is an application problem. The management console calculates the Data Transfer Time by subtracting the time stamp associated with the last application response to the client data request from the time stamp associated with the first application response to the client data request. If the timestamps are equal, the result is 0 because only one packet was sent by the application in response to the client request, and that this packet must be acknowledged at the application layer before the client can request the next packet.

If the amount of data needed to complete the total transaction numbers in the tens or hundreds of thousand bytes, this can cause significant throughput issues for the application because the client must wait one NRTT+ for each packet to be delivered and acknowledged by the application. If the client asked for all data in a single request, the application would respond by sending as many bytes in multiple packets at one time as the TCP active window allows it to send. This dramatically decreases the number of round trips necessary to transmit the data, and increases the performance of the application.

To verify this scenario when Data Transfer Time is equal to zero:

1. In the Show Me menu, click Response Size.
2. Scroll to the Data Transfer Time by Response Size view and determine if the number of data transfers less than 1.45 KBs, one segment or packet based on an Ethernet frame, is the most significant bar of the following view. This indicates that most of the packet contain a minimum of data:



Data Transfer Time by Response Size per 15 minute intervals

Application: All
Server: All
Network: All

Increase in Data Transfer Time and Decrease in Observations Count

An increase in the Data Transfer Time while the Observations count decreases is a typically an indicator that the problem is related to an issue on the server or network. If the Server Response Time increased along with the Data Transfer Time, review the server resources and application code.

If the NRTT and Retransmission Delay increased along with the Data Transfer Time -- review the network path and associated devices after isolating the problem network segments.

Operations

Typically end users and business management report performance problems to operations centers and support teams in one of three ways:

- There is a performance issue with an application.
- There is a performance issue with a network.
- There is a performance issue with a server.

The Operations page uses a simple and straightforward method for working directly with reports of performance problems. When performance problems are reported, it is important to gather the following problem information from the person reporting the issue:

- Name of application and the server exhibiting performance problem
- Location and/or IP address of end user, local or remote office location, so that the network they are on can be identified
- Time performance problem was first noticed, and if problem is continuing
- History of performance problem -- does it seem to be recurring--several times a day, same time on different days, always at the end of the month?
- Any other performance issues that the use believes might be related -- other applications or servers slow

Example 1: Performance Problems with an Application

A user from corporate headquarters reports they are experiencing a problem with an application.

Approach

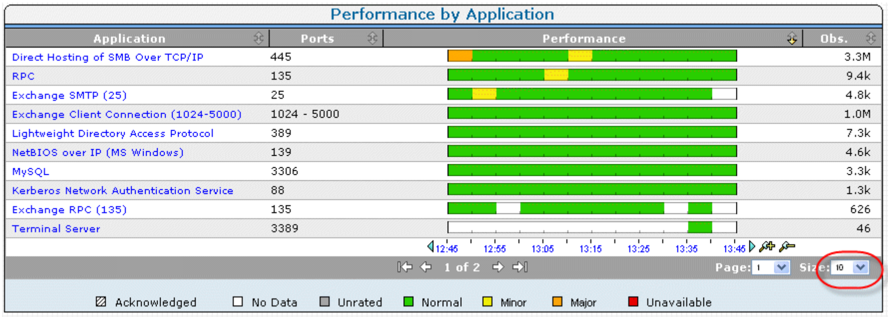
Gather the problem information from the user and create a trouble ticket.

1. Click the Operations page.
2. Click Applications in the Show Me menu.

   The Performance by Application page opens.

3. Determine if the reported application has bubbled up in the list to the top. Typically it is among the top ten. If the application does not appear in the top ten, change the Size setting to a larger value until the application appears in the list. If the application is not in the list, it might not be monitored by CA Application Delivery Analysis.

Equation 1: Worst-performing applictions bubble-up to the top of the list.



Performance by Application

Application	Ports	Performance	Obs.
Direct Hosting of SMB Over TCP/IP	445		3.3M
RPC	135		9.4k
Exchange SMTP (25)	25		4.8k
Exchange Client Connection (1024-5000)	1024 - 5000		1.0M
Lightweight Directory Access Protocol	389		7.3k
NetBIOS over IP (MS Windows)	139		4.6k
MySQL	3306		3.3k
Kerberos Network Authentication Service	88		1.3k
Exchange RPC (135)	135		626
Terminal Server	3389		46

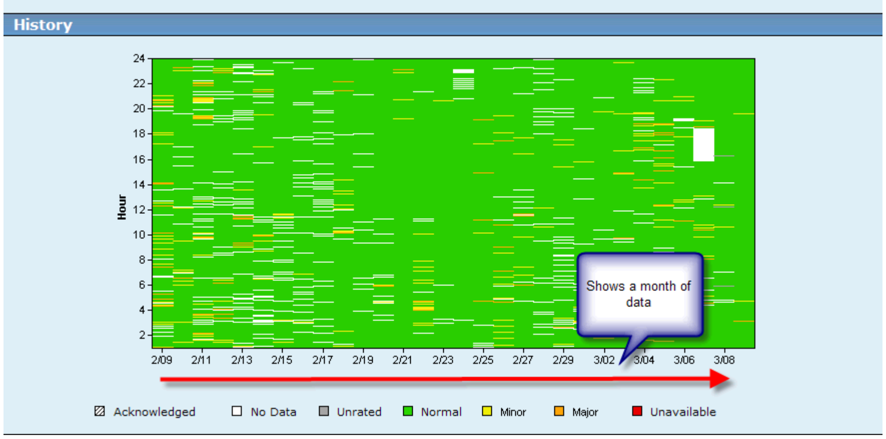
1 of 2 Page: 1 Size: 10

4. Click the performance bar to the right of the application name to display a detail page. From the detail page, you can narrow the scope of data by clicking a network or server metric.

5. If the network or server is known, select the appropriate performance bar to further narrow the scope.

6. In the Show Me menu, click History.

7. Review the history of the application to identify and note any systematic patterns of unavailability or performance degradation.



8. In the Show Me menu, you can also click Performance and then click Explore to begin troubleshooting. You can also do more exhaustive troubleshooting by clicking Links: Engineering at the top of the page to display the Response Time Component Delays page. Notice the selected application name and server or network name in the query selection box at the top of the page.

9. Review the Response Time Composition: Average view, and pinpoint the time that the reported issue occurred. Identify which component is the most significant contributor at that point in time.

   - If Server Response Time (SRT) dominates this time point, see Increase in Server Response Time (see page 142).

   - If Network Round Trip Time (NRTT) and/or Retransmission Delay dominates this time point, see Increase in Network Round Trip Time (NRTT) (see page 147).

   - If Data Transfer Time dominates this time point, see Increase in Data Transfer Time (see page 151).

   - If all the components are normal with respect to the historical timeline, the problem might be with the user's computer.

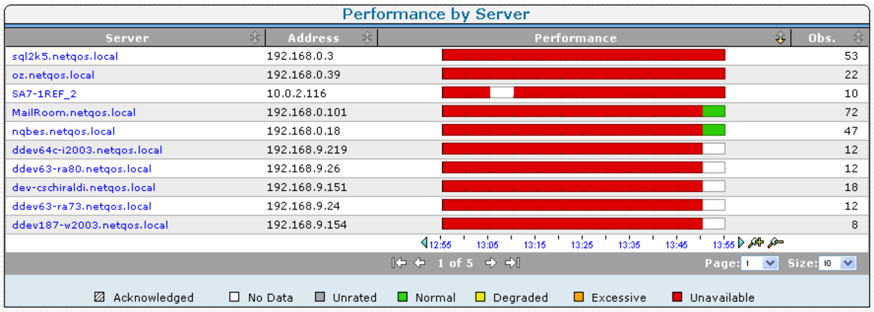
Example 2: Performance Problem with a Server

An end user from corporate headquarters reports they are experiencing a problem with a server.

Approach

Gather the problem information from the user and create a trouble ticket.

1. Click the Operations page
2. Scroll down to the middle of the page and review the Performance by Server view.
3. Determine if the reported server has bubbled up in the list to the top. Typically it is among the top ten. If the server does not appear in the top ten, click Settings and select the server from the Server List.



Performance by Server

Server	Address	Performance	Obs.
sql2k5.netqos.local	192.168.0.3		53
oz.netqos.local	192.168.0.39		22
SA7-1REF_2	10.0.2.116		10
MailRoom.netqos.local	192.168.0.101		72
nqbes.netqos.local	192.168.0.18		47
ddev64c-i2003.netqos.local	192.168.9.219		12
ddev63-ra80.netqos.local	192.168.9.26		12
dev-cschiraldi.netqos.local	192.168.9.151		18
ddev63-ra73.netqos.local	192.168.9.24		12
ddev187-w2003.netqos.local	192.168.9.154		8

▨ Acknowledged □ No Data 🟫 Unrated 🟩 Normal 🟨 Degraded 🟧 Excessive 🟥 Unavailable

4. Select the performance bar to the right of the server name to display a detail page.
5. If network or application is known, select the appropriate performance bar to narrow the scope.
6. In the Show Me menu, click History.
7. Review the history of the server to identify and note any systematic patterns.
8. In the Show Me menu, you can also click Performance, and Explore to begin troubleshooting. You can also do more exhaustive troubleshooting by clicking Links: Engineering at the top of the page to display the Response Time Component Delays page. Notice the selected server name and/or application or network name in the query selection box at the top of the page.
9. Review the first view, the Response Time Composition: Average view, and pinpoint the time on the view where the issue is reported to have occurred. Identify which component is the most significant contributor at that point in time.
   - If Server Response Time (SRT) dominates this time point, see Increase in Server Response Time (see page 142).
   - If Network Round Trip Time (NRTT) and/or Retransmission Delay dominates this time point, see Increase in Network Round Trip Time (NRTT) (see page 147).
   - If Data Transfer Time dominates this time point, see Increase in Data Transfer Time (see page 151).

- If all the components are normal with respect to the historical timeline, the problem might be with the user computer.

Example 3: Performance Problems with a Network

An end user from corporate headquarters reports they are experiencing a problem with a specific network.

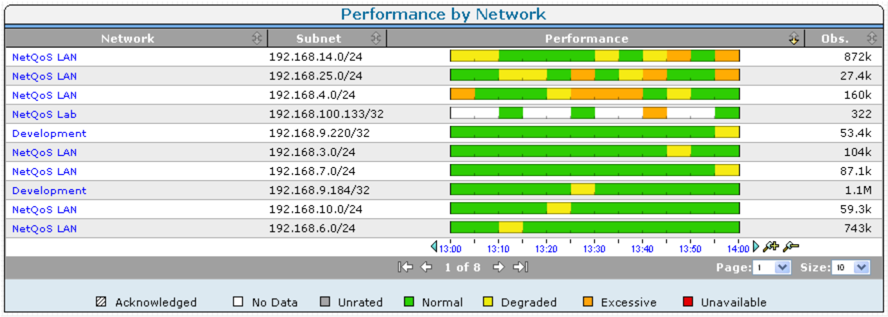
Approach

Gather the problem information from the user and create a trouble ticket.

1. Click the Operations page.
2. Click Networks in the Show Me menu.

   The Performance by Network view opens.

3. Determine if the reported network has bubbled up in the list to the Top N. If the network does not appear in the Top N, change the Size: setting to a larger value until the network appears in the list. If the network does not appear in the list, it might not be monitored by CA Application Delivery Analysis.



Performance by Network

Network	Subnet	Performance	Obs.
NetQoS LAN	192.168.14.0/24		872k
NetQoS LAN	192.168.25.0/24		27.4k
NetQoS LAN	192.168.4.0/24		160k
NetQoS Lab	192.168.100.133/32		322
Development	192.168.9.220/32		53.4k
NetQoS LAN	192.168.3.0/24		104k
NetQoS LAN	192.168.7.0/24		87.1k
Development	192.168.9.184/32		1.1M
NetQoS LAN	192.168.10.0/24		59.3k
NetQoS LAN	192.168.6.0/24		743k

4. Select the performance bar to the right of the network name to display the detail page. From the detail page, you can narrow the scope of data by clicking an application or server metric.
5. If application or server is known, select the appropriate performance bar to further narrow the scope.
6. In the Show Me menu, click History.
7. Review the history of the network to identify and note any systematic patterns or trends.
8. Select Links: Engineering at the top of the page to display the Response Time Component Delays page. You should notice the selected network name and/or server or application name in the query selection box at the top of the page.
9. Review the first view, the Response Time Composition: Average view, and pinpoint the time on the view the issue is reported to have occurred. Identify which component is the most significant contributor at that point in time.
   - If Server Response Time (SRT) dominates this time point, see Increase in Server Response Time and Observations Count (see page 143).
   - If Network Round Trip Time (NRTT) and/or Retransmission Delay dominates this time point, see Increase in NRTT and Observations Count (see page 148).

- If Data Transfer Time dominates this time point, see Increase in Data Transfer Time and Observations Count (see page 152).
- If all the components are normal with respect to the historical timeline, the problem might be with the user computer.

Investigations

You can initiate investigations in the following ways:

- Launch the investigation manually
- Schedule the investigation
- Create an incident response with an investigative action; this type of investigation is launched automatically when an incident is opened for the item with which the incident response is associated

To manually investigate a router, switch, or other SNMP-enabled device, request that the CA Application Delivery Analysis administrator add it as a network device.

Analyze Data Flow from a User through the Data Center

To help diagnose a user that is getting poor performance, you might want to launch a trace route investigation from the user device through the data center to highlight problem areas.

One way to do that is to navigate to the user device and launch a trace route from there. Using the management console, you can add the user device to CA Application Delivery Analysis and then do a trace route to that device.

Generate a Report of Data Flow from a User through the Data Center

Use the Operations page to look for a user on a poorly performing network.

1. Click the Operations page.
2. Click Networks in the Show Me menu.

   The Performance by Network page opens.
3. Click network name and then click Explore.

   The Operations Metric Details page opens.
4. Click the Impacted Users tab.

   The Users report appears.
5. Note the IP Address/Mask field that includes the /24 or /32 subnet mask.
6. Use the Administration page of the management console to add the corresponding network device to CA Application Delivery Analysis.
7. Use the Incidents page of the management console to launch a trace route investigation.

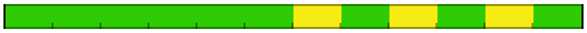
Identify Impacted Networks

Performance can degrade in an application and affect several networks accessing that application.

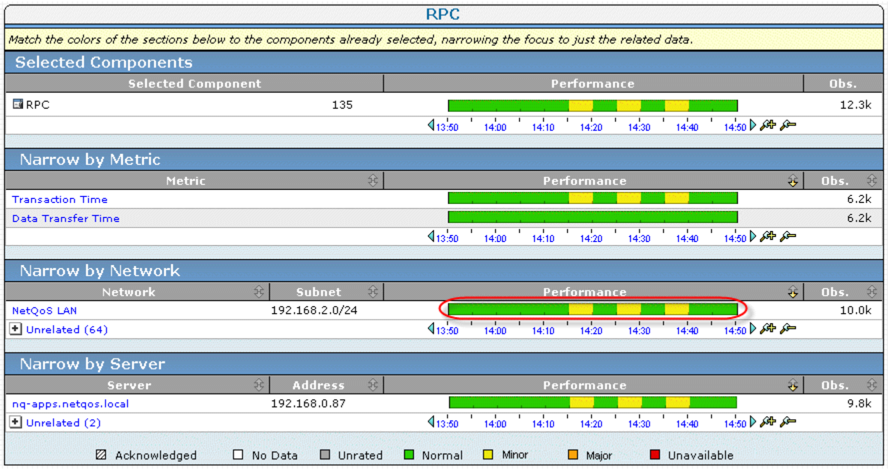
Application Performance

On the Operations page, the Performance by Application report displays the worst-performing applications reported to the management console. To display all networks affected by the performance degradation, click the application name at the top of the list.

As you look down the charts of servers and networks, you can select those whose ratings match the ratings in the Selected Component page. For example, if this is the view of the application performance:



Click this performance bar and view the Narrow by Network page on the subsequent page:



RPC

Match the colors of the sections below to the components already selected, narrowing the focus to just the related data.

Selected Components

Selected Component		Performance	Obs.
RPC	135		12.3k

Narrow by Metric

Metric	Performance	Obs.
Transaction Time		6.2k
Data Transfer Time		6.2k

Narrow by Network

Network	Subnet	Performance	Obs.
NetQoS LAN	192.168.2.0/24		10.0k
Unrelated (64)			

Narrow by Server

Server	Address	Performance	Obs.
nq-apps.netqos.local	192.168.0.87		9.8k
Unrelated (2)			

▨ Acknowledged □ No Data 🟫 Unrated 🟩 Normal 🟨 Minor 🟧 Major 🟥 Unavailable

You can easily see the networks involved.

Identify Impacted Networks from a Server Incident

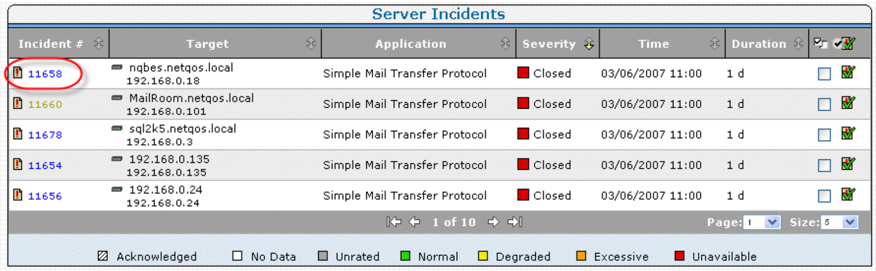
When a server incident occurs, the Incidents page displays it. To understand the impact the incident caused, you might want to know:

- What networks and user groups were affected?
- Which specific users were affected?
- How did the number of users of the system change when the Incident occurred?
- For a specific site, which users were affected?

Identify Users and Networks Impacted by an Incident

To see the networks and user groups affected, follow this procedure:

1. Click the link for an incident on the Server Incidents page to display more details about the incident:



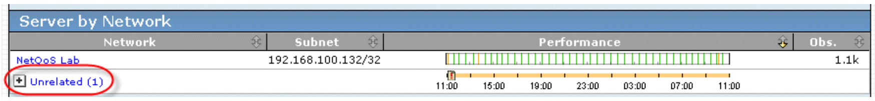
Server Incidents

Incident #	Target	Application	Severity	Time	Duration	
11658	nqbes.netqos.local 192.168.0.18	Simple Mail Transfer Protocol	🟥 Closed	03/06/2007 11:00	1 d	□
11660	MailRoom.netqos.local 192.168.0.101	Simple Mail Transfer Protocol	🟥 Closed	03/06/2007 11:00	1 d	□
11678	sql2k5.netqos.local 192.168.0.3	Simple Mail Transfer Protocol	🟥 Closed	03/06/2007 11:00	1 d	□
11654	192.168.0.135 192.168.0.135	Simple Mail Transfer Protocol	🟥 Closed	03/06/2007 11:00	1 d	□
11656	192.168.0.24 192.168.0.24	Simple Mail Transfer Protocol	🟥 Closed	03/06/2007 11:00	1 d	□

▨ Acknowledged □ No Data ■ Unrated 🟩 Normal 🟨 Degraded 🟧 Excessive 🟥 Unavailable

2. Click the plus icon next to the Unrelated link to see the unaffected networks:

Equation 2: Click the link to view unaffected networks.

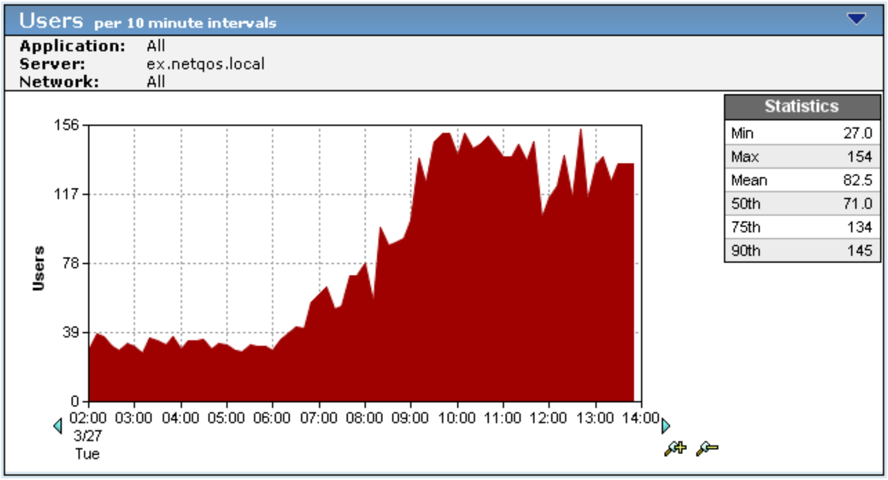


Server by Network

Network	Subnet	Performance	Obs.
NetQoS Lab	192.168.100.132/32		1.1k
⊞ Unrelated (1)			

3. To understand the specific users that were affected by the incident, click Explore.
4. Click the Impacted Users tab to see a list of the IP addresses and host names (if available) for all the users who were accessing the application/server at the time of the incident.
5. Close the Impacted Users report and click Links: Engineering to see the total number of users at the time the incident occurred and how the problem might have impacted the volume of users.

   The Components Report page opens.
6. Click QoS on Show Me menu.
7. Scroll down to the Users view to see a visual representation of the number of users over time.



Users per 10 minute intervals

Application: All
Server: ex.netqos.local
Network: All

Statistics	
Min	27.0
Max	154
Mean	82.5
50th	71.0
75th	134
90th	145

8. Click Settings and select a network to see the users on a specific network.

Determine if a Network Contributes to Poor Application Performance

Follow these steps:

1. Click the Operations page.
2. Click Applications in the Show Me menu.

   The Performance by Application report.
3. Click an application that shows poor performance. This application performance profile appears in the Selected Components section of the next page.
4. Click Settings. On the Settings page, under Metric, select All Network Metrics.
5. Look down the page to the Narrow by Networks display to view the networks that application is running on. If these networks show a similar degradation profile as the application, then they are contributing. If the networks are showing normal performance while the application was slow, then the networks were not contributing.
6. Alternatively, you can simply look at the Performance by Network report on the Operations Overview page to see the worst performing networks. If one of these is running the application of interest, which you can see if you click the network, then it is contributing to the poor performance of the Narrow by Application list.

Determine the Network Component to Degraded Application Performance

Complete these steps:

1. Click the Operations page.
2. Click Applications in the Show Me menu.

   The Performance by Application report.
3. Click an application that shows poor performance. This application performance profile appears in the Selected Components section of the next page.
4. Click Settings. On the Settings page, under Metric, select All Network Metrics.
5. Look down the page to the Narrow by Networks display to view the networks that application is running on. If these networks show a similar degradation profile as the application, then they are contributing. If the networks are showing normal performance while the application was slow, then the networks were not contributing.
6. Alternatively, you can simply look at the Performance by Network report on the Operations Overview page to see the worst performing networks. If one of these is running the application of interest, which you can see if you click the network, then it is contributing to the poor performance of the Narrow by Application list.

Determine the Location of Servers with Performance Issues

Server performance issues can be quickly seen on the Performance by Server view in the Operations page. Troubleshooting with the management console helps narrow the source of an issue to a single server. You can determine the:

- *recency* of the issue on the time axis of the view
- *duration* of the problem from the width of the colored segment
- *severity* of the issue by the color of the segment
- *pervasiveness* of the problem by how many networks display with a similar degradation profile in the corresponding Performance by Network view.

The root cause might be obvious from an investigation. The management console focuses investigation efforts so that you might use other specialized tools to find the root cause quickly.

To help locate servers in your enterprise network, a best practice is to name devices in a meaningful way within the management console, such as Third Bank servers, South Building Server C, and so on. Servers in server farms can contain a common identifier, Web servers should be named similarly, and so forth.

SNMP Queries Using Investigations

This section illustrates how to use SNMP performance investigations in the Incidents page.

Situation 1-- Server Incident

A Server Incident is identified in the Incidents page Overview page.

Overview

It is important to ascertain how long the Incident has been open. The longer an Incident continues, the more users that might be affected by the issue. On the Incidents page, click the incident to display its incident details. Review the data in the Duration field.

Also check the Severity field to understand the latest status of the incident. An incident rated Major should be examined as soon as possible. If the incident is Closed, then the performance degradation is no longer occurring and has reached previous normal performance levels for a significant period of time.

Click the incident number to see the Details page.

Incident Details

For analysis, first identify which metric is causing the incident by checking the Server by Metric section. SNMP performance queries are very useful when verifying degradation of Server Response Time metrics including unresponsive and Refused Sessions.

In the Show Me, list, click Investigations to view the associated SNMP performance investigations for the server or server groups.

The investigations associated with this incident are listed. The SNMP poll investigation type queries the server for the SNMP performance metrics using the host resources MIB. Click the View button to see the SNMP poll data. If there is no SNMP poll investigation in the list, launch one by completing the following steps.

SNMP Performance Investigation

It might be helpful to manually launch an investigation if:

- There is no SNMP poll for a server incident
- The investigation date is stale (older than a day)
- Additional servers need to be included in the analysis

More information:

Performance via SNMP Investigation (see page 67)

Analyze the SNMP Performance Data

After the SNMP performance report for the server displays, find the troubleshooting section that corresponds the type of server incident that occurred.

Server Response Time or Server Connection Time Incidents Analysis

For Server Response Time or Server Connection Time incidents, the server is slower to respond than normal to requests by the clients.

CPU

Review the CPU utilization for the server. If it is high, near 70 to 100 percent, then check the processes consuming the CPU. If there is a process consuming all the CPU, it is slower to respond. This might be caused by too many concurrent users or a large CPU load by the process. To reduce CPU utilization, you can restart the process. This would terminate any work done by the users. If the process is non-critical, stop the process. If the CPU utilization is low, then CPU is not an issue with this incident.

Memory

Review the memory utilization for the server; if it exceeds 70 percent, check the processes consuming memory. Slow response might be caused by too many concurrent users or a memory leak with the process. To reduce memory utilization, restart the process. If the process is non-critical, stop the process. If both CPU and memory are low, then typical hardware constraints are not an issue.

Interface Statistics

If both the memory and the CPU utilization are fine for the server, then it is possible the NICs are experiencing issues making the server response appear slower. Check the Interface statistics. First ensure that all the interfaces are listed properly and are connected. Verify that both the interface speeds and the IP Addresses for the NICs are reported correctly. Duplex mismatch errors would inhibit the ability of the interface to send and receive data. Duplex mismatch errors, cable problems, or CRC errors would indicate a problem with the NIC and its connectivity through the discards and errors fields.

Disk Space

Check the capacity of the drives. Servers slow down because of limited capacity for the data. If the server is out of disk space, its ability to process data is degraded. Additional other considerations require access to the server to check for fragmentation or I/O errors with bad segments. If any given drive has less than 10-15% free space, this can also be a problem for applications with large working data sets.

Server Refused or Unresponsive Sessions Incidents Analysis

If you see unresponsive or refused sessions (server running out of available threads) along with expired sessions it indicates that resources are not being released by the application.

For server refused or unresponsive session incidents it indicates that the server is not responding to requests for new sessions in a timely manner.

Top CPU Processes or Memory Processes

In analyzing the CPU resources, ensure that the processes responsible for hosting the application are running. For example, in the process list for a SQL Server sqlserver.exe should be listed in the processes for CPU or memory. The process should be running and not consuming all the CPU. If the CPU is high, near 70 or 100 percent, identify the process that is consuming the CPU. If this process is non-critical, stop the process or restart it. The process might be locked or in a poor state that is unnecessarily taking processing power away from serving critical applications.

If the memory utilization is over 80 percent, identify the process that is consuming the memory. If this process is not critical, stop the process or restart it. The process might be locked or in a poor state, unnecessarily taking critical memory resources away from the critical processes. If the process that is consuming the memory is critical, restarting the process should clear its memory cache and let it restart normally.

Disk Space

If the expected processes are not listed, they potentially might have failed because of disk space limitations. Check the drives to ensure that the server has enough disk space to launch the applications. Then log in to the server to see if the processes are running and are able to start. Restarting the services or processes to restore connectivity might be required.

Interface Statistics

If the processes are running and operating normally, then it is possible the NICs are experiencing issues making the server appear slower or unavailable. Check the interface statistics. Ensure that all the interfaces are listed properly and are connected.

When the SNMP data looks good, you can initiate an operational investigation, including:

- Review the number of active sessions, the Time_Wait, and other TCP/IP parameters on the server to ensure they are not exhausting TCP port availability.
- Compare Server Connection Time with other applications running between remote office and server to see if the increase is across the board.

Situation 2 -- Network Incident

You can also launch an SNMP investigation for routers and layer 3 switches in the infrastructure. For this situation, the SNMP performance investigation is launched using TCP Trace Route investigation.

When configuring the Trace Route investigation, under Investigation Options, be sure to set Investigate Routers via SNMP to Yes and add network devices for each router in the path.

More information:

Trace Route Investigation (see page 71)

Analyze the SNMP Trace Route Data

The report presents a route with multiple hops. For detailed analysis, select the performance statistics for the hop with the largest delay. Click the Report icon in the far right column. The SNMP performance report for the router opens in a new window. For network incidents, follow the checklist for determining the root cause of a network issue.

1. Check the CPU and Memory utilization of the device. If the either the memory or CPU usage is high, the router is being overloaded by a large configuration or burdened with excessive packet processing. Oftentimes CPU usage can rise if of a virus outbreak or DDOS attack on the device when it is being flooded with erroneous packets.

   Follow the troubleshooting process for the type of device to reduce CPU or memory utilization to return performance levels to normal.

2. Ensure that all interfaces are listed correctly and the status is Up. If the interface in question has failed over to a secondary interface, ensure that the speeds are the same. If the secondary interface has a lower speed, network latency can increase until the primary interface is returned to functioning properly.

3. Verify that the interface is not seeing any discards or errors. If the discards or errors are being reported, log into the router and troubleshoot the interface connectivity for issues relating to the media type of the interface. Discards and errors tend to indicate a problem with the connectivity between that interface and the connecting router or device.

4. Ensure that the interface utilization is not exceeding capacity. The interface ingress bits per second must be less than the circuit rating. If the utilization of the circuit is a problem, look into the traffic flows on the interface to identify what is causing the sudden congestion.

5. If the device is operating as expected, then check the remaining hops investigated via SNMP following the same workflow to identify and resolve any potential issues with the network latency between the end-user and the servers affected by the network incident.

Performance and Availability OLA Tracking

Operational Level Management (OLM) is the disciplined, proactive methodology and procedures used to ensure that adequate levels of operation are delivered to all IT users in accordance with business priorities and at acceptable cost. Operational Level Agreements (OLA) reports in the management console can reveal if levels of operation are met. Common performance metrics on which to set an OLA are:

- Server Response Time to quantify the performance of the data center
- Network Round Trip Time to quantify the performance of the network infrastructure
- Transaction Time to capture the end-to-end performance of an application

Because managers need high-level summary reports to answer broad questions such as "Are we meeting our goals?" the Performance Executive OLA reports are a bubble-up report that quickly give the answer to that question for each application. The Availability Executive OLA report shows the big picture of compliance to application availability goals. Daily and hourly details can be accessed by clicking into these summary reports.

Generate OLA Reports at a frequency that makes sense:

Frequency	Description
Daily	Very detailed, short-term troubleshooting, IT Department focused, technical content
Weekly	Summary with additional detail for anomalies or trends, also IT focused
Monthly	Summary for business unit and executive management; typically presented as a report card
Quarterly	Quantifies broad operational level with compliance and as input for planning

OLA violations can usually be seen one of these areas:

- Time, such as time of day or day of the week
- User group, such as VPN user networks
- Servers; for example, Web servers #2 and #5

OLA reports drive action for performance improvement. As you make changes in these areas you can watch the OLA reports improve.

Chapter 10: Analysis

This section contains the following topics:

Impact Analysis

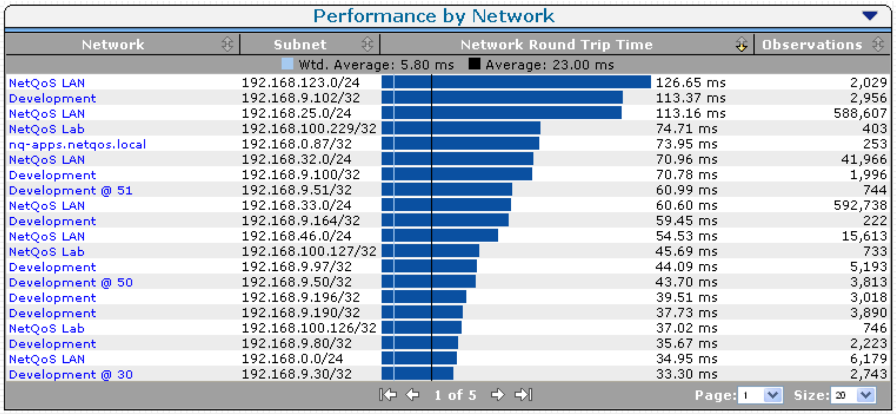
The management console gives IT staff the ability to measure and validate the impact of infrastructure changes. Two examples of the analysis technique follow.

Validate a QoS Policy Implementation

A QoS policy can improve response times for remote users by prioritizing critical application traffic. This example shows how to validate that it actually did. In this example, the policy was implemented yesterday afternoon just after lunch.

1. Click the Engineering page
2. Click Performance, Networks on the Show Me menu.
3. Click Settings and make the following selections:
   - Timeframe: Last 24 Hours
   - Metric: Network Round Trip Time
   - Application, Server, and Network: No Selection
4. Click OK.

   The following report appears:



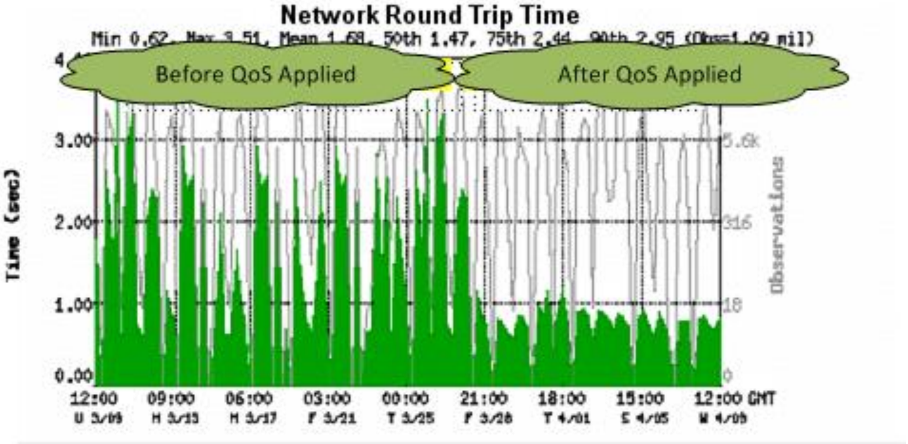
Performance by Network

🟦 Wtd. Average: 5.80 ms ■ Average: 23.00 ms

Network	Subnet	Network Round Trip Time	Observations
NetQoS LAN	192.168.123.0/24	126.65 ms	2,029
Development	192.168.9.102/32	113.37 ms	2,956
NetQoS LAN	192.168.25.0/24	113.16 ms	588,607
NetQoS Lab	192.168.100.229/32	74.71 ms	403
nq-apps.netqos.local	192.168.0.87/32	73.95 ms	253
NetQoS LAN	192.168.32.0/24	70.96 ms	41,966
Development	192.168.9.100/32	70.78 ms	1,996
Development @ 51	192.168.9.51/32	60.99 ms	744
NetQoS LAN	192.168.33.0/24	60.60 ms	592,738
Development	192.168.9.164/32	59.45 ms	222
NetQoS LAN	192.168.46.0/24	54.53 ms	15,613
NetQoS Lab	192.168.100.127/32	45.69 ms	733
Development	192.168.9.97/32	44.09 ms	5,193
Development @ 50	192.168.9.50/32	43.70 ms	3,813
Development	192.168.9.196/32	39.51 ms	3,018
Development	192.168.9.190/32	37.73 ms	3,890
NetQoS Lab	192.168.100.126/32	37.02 ms	746
Development	192.168.9.80/32	35.67 ms	2,223
NetQoS LAN	192.168.0.0/24	34.95 ms	6,179
Development @ 30	192.168.9.30/32	33.30 ms	2,743

This view of the daily Network Round Trip Times shows the networks with the highest latency. The 56 K and VPN connections typically exhibit the slowest response.

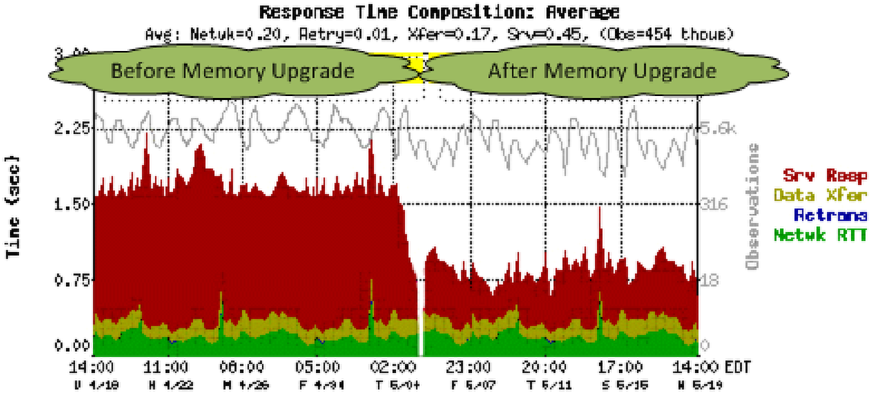
The QoS policy was implemented in the Pittsburgh, PA link. Effects of the QoS policy change are shown in the Network Round Trip Time view:



The time before the QoS policy being applied shows a marked contrast to the faster times after the policy was implemented.

Validate a Server Memory Upgrade

Slow response time from a multi-tiered application can cause confusion as to which element needs upgrading. Possible causes could be an overloaded back end database server or inadequate memory on the application server. Use the management console to determine the effective solution. This example looks at the result of upgrading the application server memory on total response time.



Trend Server Response Time

Server Response Time (SRT) is the amount of server "think time" that passes between the instant a server receives a client request packet, to the instant it puts the first response packet on the network. SRT is affected by the following:

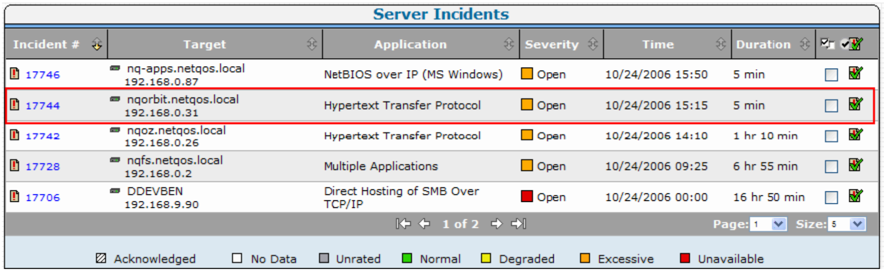
- Server hardware such as CPU power, available memory, Disk I/O, and NIC I/O
- Application behavior such as query and index optimization and application algorithms
- Hanging or malfunctioning processes
- Utilization, or processing power required by applications

Generally the faster the server hardware, the better written the application, and the lower the server utilization -- the lower the SRT. SRT values vary by server platform and application.

General rating of Server Response Time values are shown in the following table. Applications are single-tiered unless noted.

Application	Excellent	Good	Poor
Citrix	50ms	75ms	200ms
Citrix (two-tier)	90ms	125ms	200ms
CRM (2-tier)	70ms	90ms	200ms
HTTP (Java, 2-tier)	120ms	150ms	250ms
HTTP (no Java)	75ms	90ms	200ms
Lotus Notes	50ms	75ms	200ms
MS Exchange	50ms	75ms	200ms
MS SQL	60ms	90ms	150ms
MS Terminal Services	50ms	75ms	200ms
MS Terminal Services (2-tier)	90ms	125ms	200ms
Oracle	50ms	75ms	200ms
Other	75ms	90ms	200ms
Other (2-tier)	90ms	120ms	200ms

You can trend Server Response Time over time to determine long-term issues. A precipitating event might signal an opportunity for further analysis. Consider this server incident in the Incidents page:



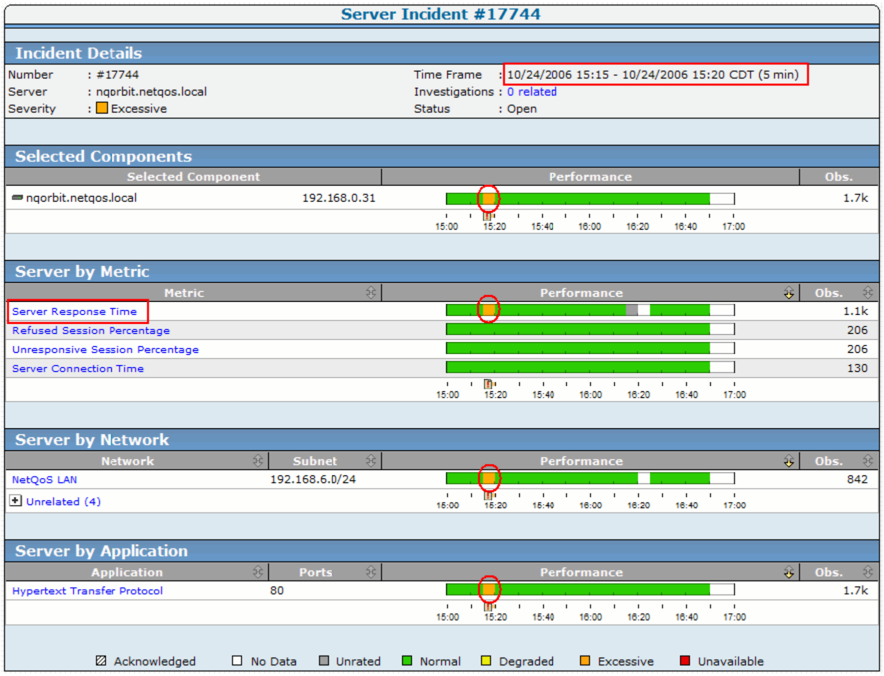
Server Incidents

Incident #	Target	Application	Severity	Time	Duration		
17746	nq-apps.netqos.local 192.168.0.87	NetBIOS over IP (MS Windows)	🟧 Open	10/24/2006 15:50	5 min	□	☑
17744	nqorbit.netqos.local 192.168.0.31	Hypertext Transfer Protocol	🟧 Open	10/24/2006 15:15	5 min	□	☑
17742	nqoz.netqos.local 192.168.0.26	Hypertext Transfer Protocol	🟧 Open	10/24/2006 14:10	1 hr 10 min	□	☑
17728	nqfs.netqos.local 192.168.0.2	Multiple Applications	🟧 Open	10/24/2006 09:25	6 hr 55 min	□	☑
17706	DDEVBEN 192.168.9.90	Direct Hosting of SMB Over TCP/IP	🟥 Open	10/24/2006 00:00	16 hr 50 min	□	☑

1 of 2 Page: 1 Size: 5

▨ Acknowledged □ No Data ■ Unrated 🟩 Normal 🟨 Degraded 🟧 Excessive 🟥 Unavailable

Click the link for the incident to display details of the metric that crossed a threshold and initiated the incident:



Server Incident #17744

Incident Details

Number : #17744
Server : nqorbit.netqos.local
Severity : 🟧 Excessive
Time Frame : 10/24/2006 15:15 - 10/24/2006 15:20 CDT (5 min)
Investigations : 0 related
Status : Open

Selected Components

Selected Component		Performance	Obs.
nqorbit.netqos.local	192.168.0.31		1.7k

Server by Metric

Metric	Performance	Obs.
Server Response Time		1.1k
Refused Session Percentage		206
Unresponsive Session Percentage		206
Server Connection Time		130

Server by Network

Network	Subnet	Performance	Obs.
NetQoS LAN	192.168.6.0/24		842
⊞ Unrelated (4)			

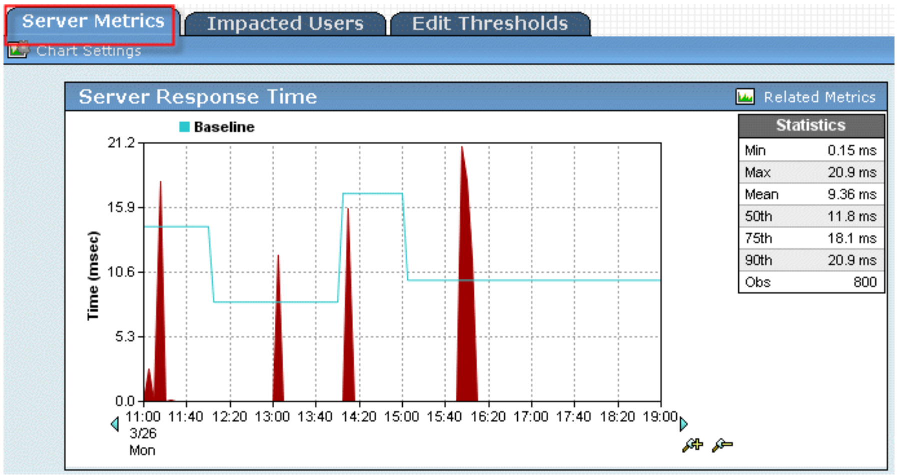
Server by Application

Application	Ports	Performance	Obs.
Hypertext Transfer Protocol	80		1.7k

▨ Acknowledged □ No Data ■ Unrated 🟩 Normal 🟨 Degraded 🟧 Excessive 🟥 Unavailable

The 5-minute period when the incident opened is easy to see, and the metric that caused it was the Server Response Time.

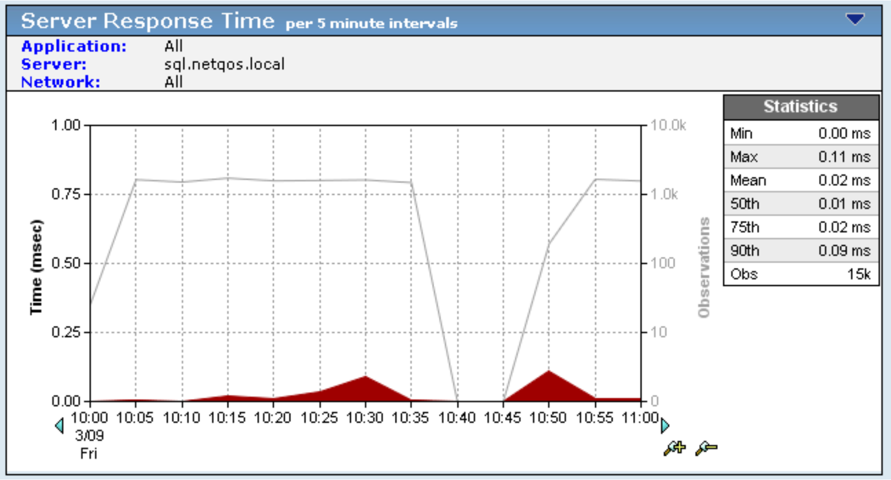
Click Explore in the header to see the following detail view. The time frame shown on this view is the same as on the previous page. You might want to see if there is any pattern in a longer-term view of this server and metric. Longer-term views are available in the Engineering page. You can leave this window open for reference.



Statistics	
Min	0.15 ms
Max	20.9 ms
Mean	9.36 ms
50th	11.8 ms
75th	18.1 ms
90th	20.9 ms
Obs	800

Click the Engineering link in the header to display the following view. This view is of the same time frame, and therefore looks the same as the view from the Incidents page.

To change the time frame to a longer term, click Settings at the top of the page.



Application: All
Server: sql.netqos.local
Network: All

Statistics	
Min	0.00 ms
Max	0.11 ms
Mean	0.02 ms
50th	0.01 ms
75th	0.02 ms
90th	0.09 ms
Obs	15k

You can also investigate the other metrics related to the incident from any report page in the Engineering page. A high metric such as Refused Sessions indicates that the server is operational, but is too busy to answer a request.

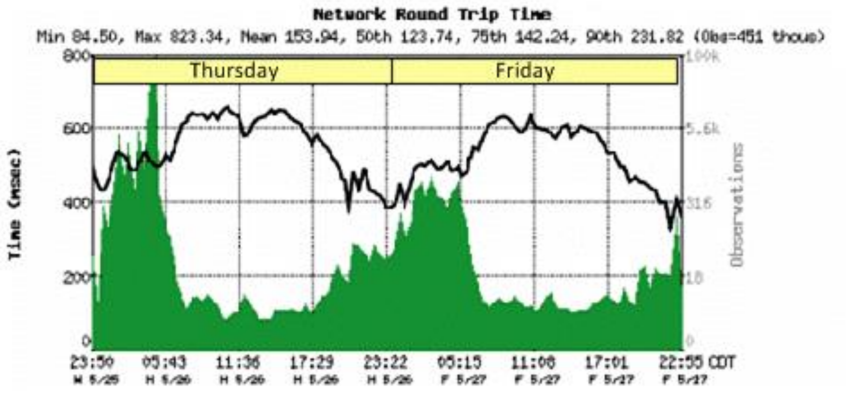
Application Performance and Volume Trends

Performance views display observation counts as well as metric measurements. Response time views use a linear time scale on the left and logarithmic observation count scale on the right y-axis. Even a slight increase or decrease in the observation count is significant because it is on the logarithmic scale.

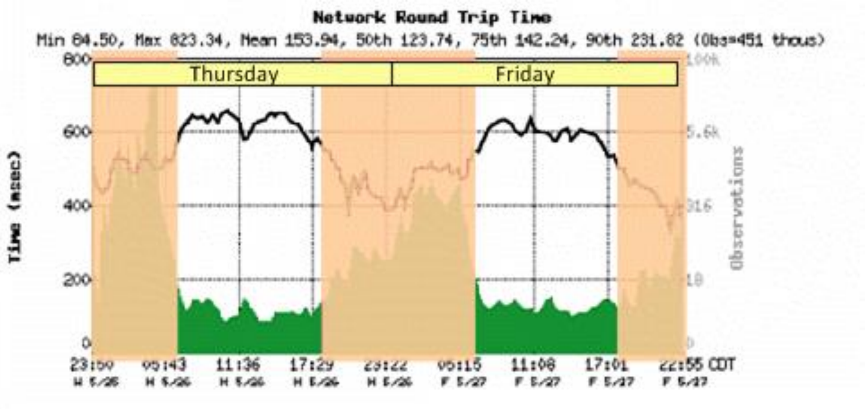
Performance experienced by the user as Network Round Trip Time (NRTT) and volume vary together as seen on the performance views in this section. To see the trends in metrics, vary the time period of the views. This section discusses the analysis of performance metrics and volume trends.

Trends in Short-Term Views

You can see similar patterns from operations during business days on the following performance views:

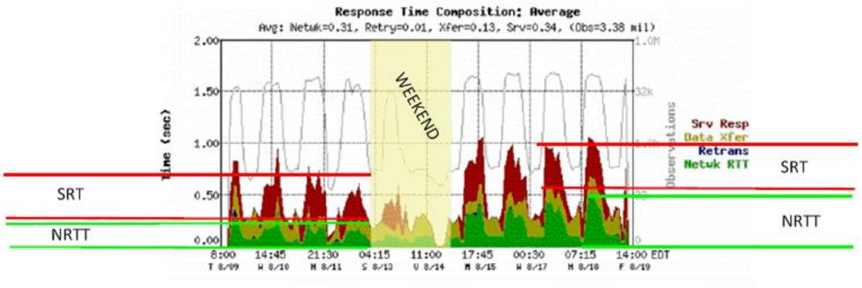


High observation counts correspond to low NRTT during peak hours as indicated:



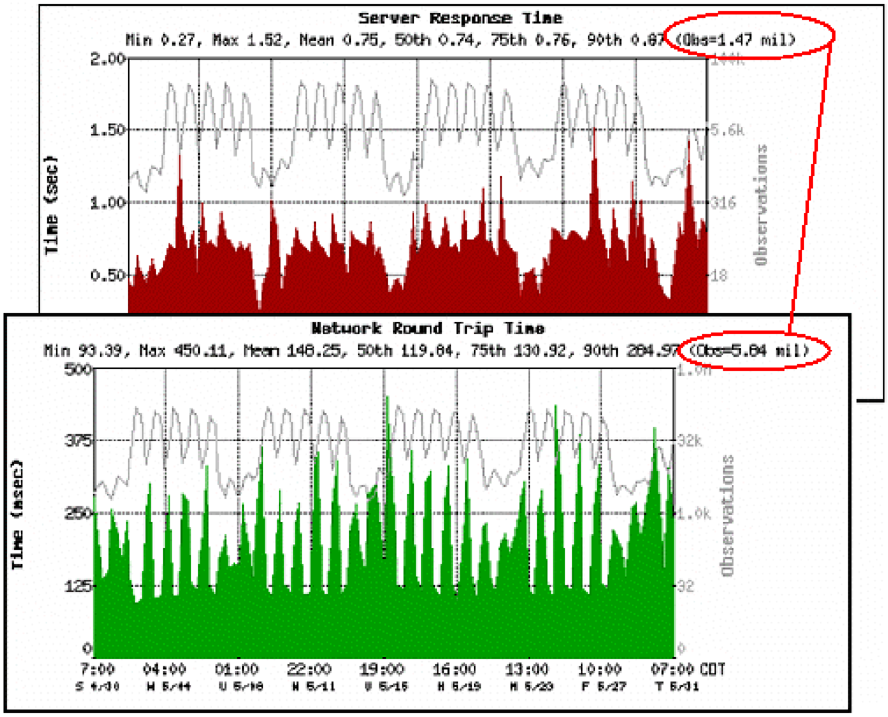
During non-work hours the opposite pattern occurs. The explanation is that overseas users are accessing the applications during nighttime hours, and those circuits experience high latency WAN access. Baselines calculated for time of day and day of week reflect these normal patterns.

You can also see weekly patterns in the following 10-day view. High volume causes longer NRTT and SRT during peak usage. The NRTT (green) component is greater in the second week than in the first, while the SRT component remains constant. This application has a fairly consistent observation pattern over the time period also. Increase in Response Time the second week is probably because of a network change or a variation in the use of other applications:

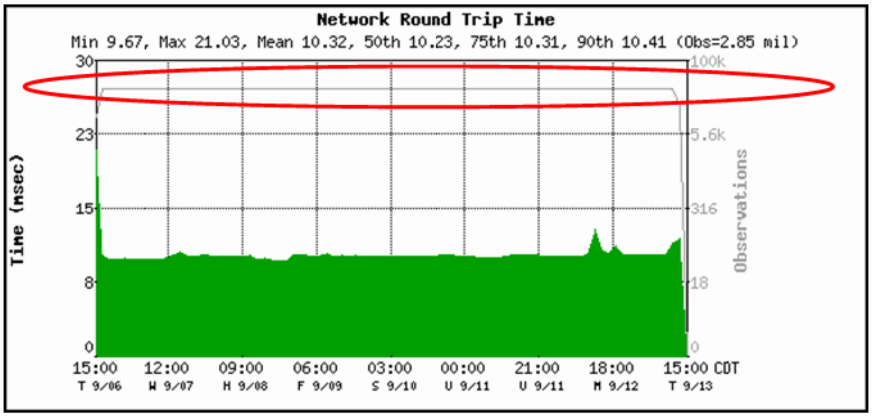


Looking at the monthly patterns, you can see trends that could have implications for capacity planning. In the following example, you can see almost a 4:1 ratio of NRTT observations to SRT observations.

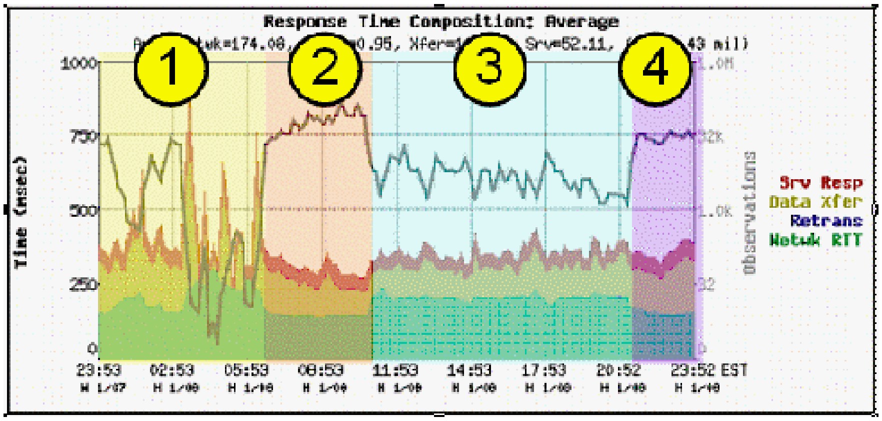
This application averages almost four round trips per TCP transaction. Applications designed to use a high number of turns might be impacted by network degradation more than those with fewer turns.



A flat observation count on a performance view suggests a batch process or an active agent:



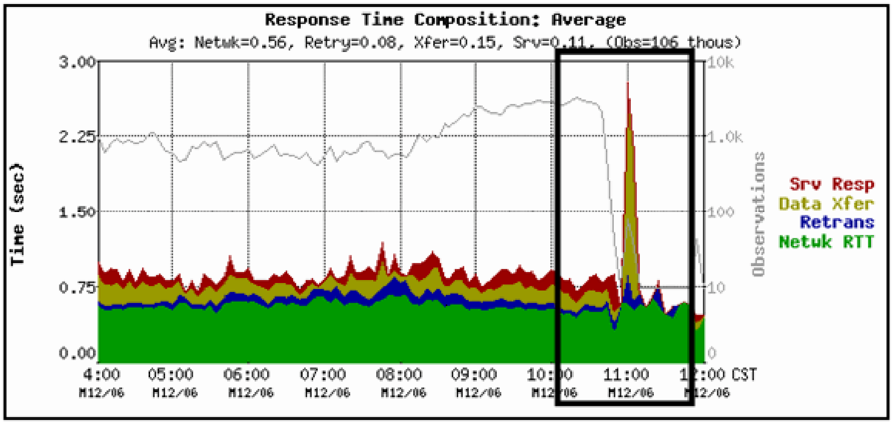
The following view shows the daily pattern for the package tracking application of a shipping company.



The patterns occur according to the daily work flow:

1. Between late night and early morning, the batch window offline processing and backups occur.
2. While loading trucks in the morning, packages are scanned. This creates the high observation counts in this part of the pattern.
3. The delivery operation is varied, and shows a steady system response.
4. The last section reflects the truck unloading activity.

The following view is an example of application failure. The observation count drops off dramatically.

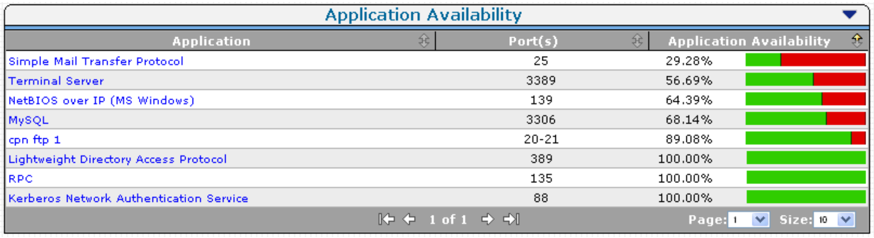

Response Time Composition: Average
Avg: Netwk=0.56, Retry=0.08, Xfer=0.15, Srv=0.11, (Obs=106 thous)
Time (sec)
3.00
2.25
1.50
0.75
0.00
10k
1.0k
100
10
0
Observations
Srv Resp
Data Xfer
Retrans
Netwk RTT
4:00
M12/06
05:00
M12/06
06:00
M12/06
07:00
M12/06
08:00
M12/06
09:00
M12/06
10:00
M12/06
11:00
M12/06
12:00 CST
M12/06

Unavailable Status: Analyzing Availability Metrics and Incidents

If you enabled availability monitoring, the management console monitors both application and server availability. When there is an interruption in service, it could be caused by one or the other. This section discusses the method to explore that data.

1. Click the Engineering page.
2. Click Availability on the Show Me menu.

   The Application Availability report opens.

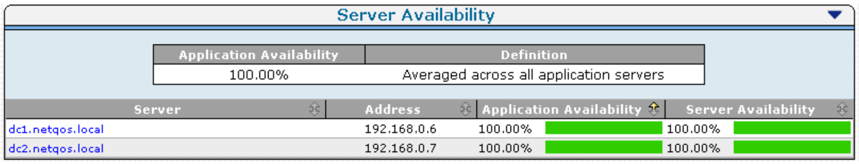


Application Availability

Application	Port(s)	Application Availability
Simple Mail Transfer Protocol	25	29.28%
Terminal Server	3389	56.69%
NetBIOS over IP (MS Windows)	139	64.39%
MySQL	3306	68.14%
cpn ftp 1	20-21	89.08%
Lightweight Directory Access Protocol	389	100.00%
RPC	135	100.00%
Kerberos Network Authentication Service	88	100.00%

3. Click an application to see the servers that host the application.

   The Server Availability report opens.



Server Availability

Application Availability	Definition
100.00%	Averaged across all application servers

Server	Address	Application Availability	Server Availability
dc1.netqos.local	192.168.0.6	100.00%	100.00%
dc2.netqos.local	192.168.0.7	100.00%	100.00%

4. Click the performance bar in the Application Availability column to see the Availability Timeline report.

   Note: When an application is 100% available (green), the Server Availability view can show some servers as not 100% available. For an application to be available, is it not necessary for every server in the farm to be available. The CA Application Delivery Analysis administrator specifies the number of servers in a farm that must be available for the application server to be considered available.

5. Click the Related Incidents link to see associated incidents from this view.

Use the Performance Scorecard

On the Management page, click Performance Scorecard. The resulting Application List page shows how applications perform in the enterprise each month. The Application List page rates performance using the following color coding:

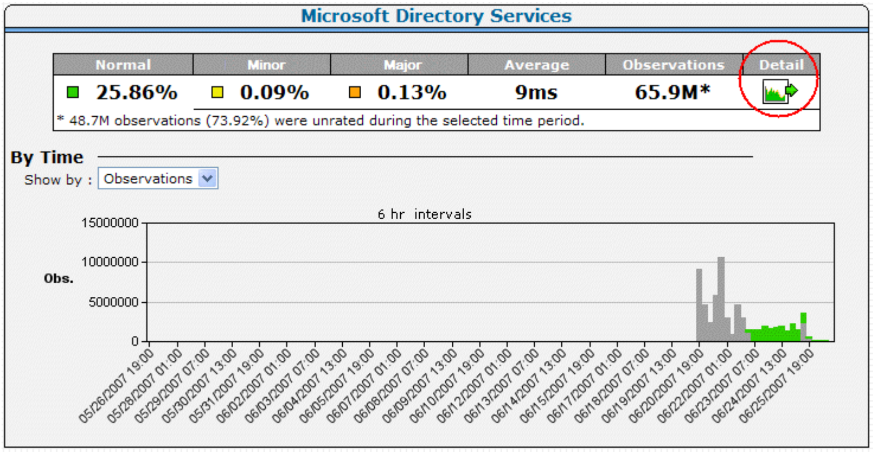
- Unrated (gray)
- Normal (green)
- Minor (yellow)
- Major (orange)

The Application List sorts the performance rating for each application by the number of observations.

Follow these steps:

1. Click the Management page.
2. Click Performance Scorecard in the Show Me menu.
3. Scroll to the Application List and click an application.
4. Click Settings to change report settings.
5. Click a color-coded performance bar or click an application name to obtain detailed information about which servers and networks are not performing in the same way as their peers.

   A detailed view of the application shows data by time intervals. Select whether to view the data by Observations or Percentage.
6. Click Performance Scorecard and Network in the Show Me menu to view the application details by network.
7. Click Performance Scorecard and Server in the Show Me menu to view the application details by server.
8. Click Detail to view a Component Report for the application.



Microsoft Directory Services

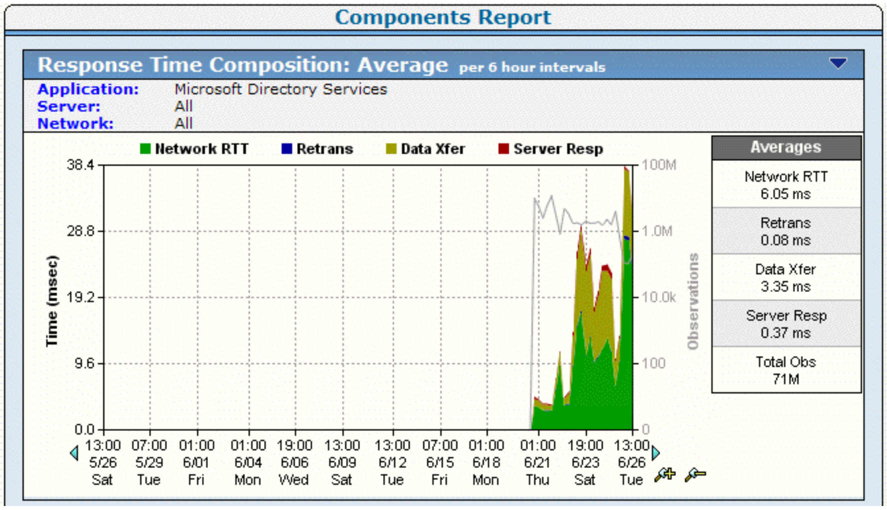
Normal	Minor	Major	Average	Observations	Detail
🟩 25.86%	🟨 0.09%	🟧 0.13%	9ms	65.9M*	

* 48.7M observations (73.92%) were unrated during the selected time period.

By Time

Show by : Observations

The Components Report page appears.



Components Report

Response Time Composition: Average per 6 hour intervals

Application: Microsoft Directory Services
Server: All
Network: All

Averages
Network RTT 6.05 ms
Retrans 0.08 ms
Data Xfer 3.35 ms
Server Resp 0.37 ms
Total Obs 71M

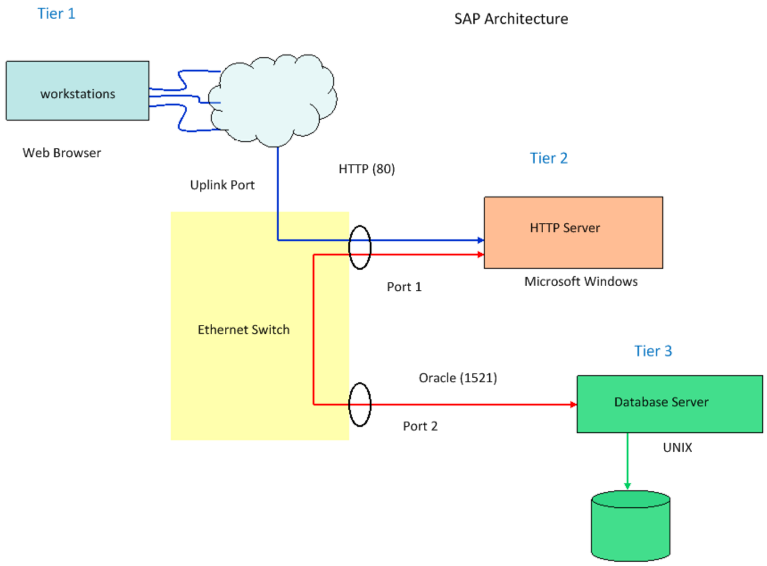
Multi-tiered Application Performance

Use the management console to gain visibility into the network, server, and application performance of each tier for a multi-tiered application. Monitor N-tier applications and obtain the necessary data to isolate an application performance problem to a specific tier and to the specific problem source of server, network, or application within each tier.

Understand Multi-Tiered Application Operation

Consider an N-Tier SAP architecture that consists of the following tiers:

- Tier 1--Internet Explorer running on a user workstation
- Tier 2--An HTTP-based application running on Windows Server 2003
- Tier 3--A database server running Oracle on UNIX



The illustration shows the following process:

1. Using Internet Explorer, a user initiates a connection to the Tier 2 HTTP server, which is illustrated by the blue line.
2. After the connection is established, the user requests application data.
3. The HTTP server forwards this request to the Tier 3 Oracle database server, depicted by the red line.
4. The Oracle server runs the user query and returns the results to the tier 2 HTTP server.
5. The HTTP server sends the data back to the Tier 1 client.

The multiple handoffs among the application tiers can make it difficult to identify the source when a performance problem occurs with an N-Tier application. Operationally, when Tier 2 waits for the Tier 3 response, its performance depends on the Tier 3 performance.

Analyze Multi-Tiered Application Performance

Use the management console to analyze multi-tiered applications the same way that you analyze single-tiered applications with two exceptions:

- A lower-tiered application's performance often depends on an upper-tiered application's performance.
- When reporting on each tier, set up the tier above as you would set up a network and set up the tier below as the server and application.

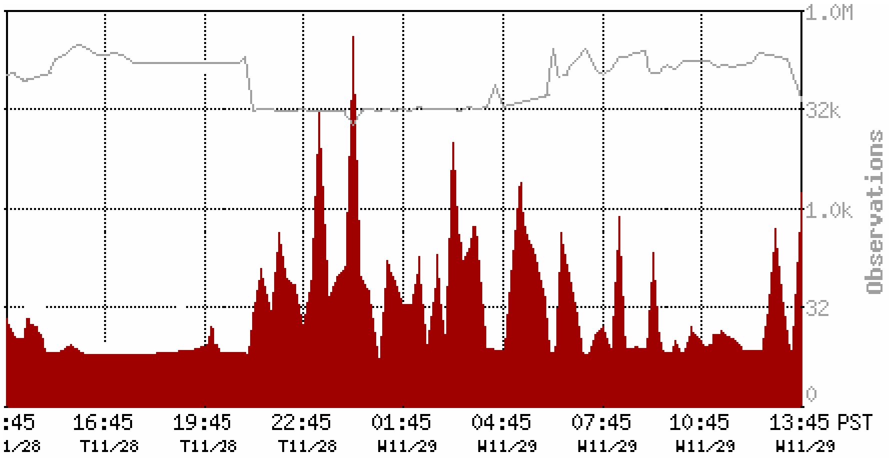
In the example in the previous section, the Tier 2 HTTP Web application's performance is a function of the Tier 3 Oracle database server performance. It follows that the Tier-N HTTP Web server appears as the client in reports while the Tier N+1 host appears as the server/application.

When you analyze N-tier applications, begin with the highest tier (the one farthest away from the end user) and work your way toward the user. Note the effects of dependent performance points on the dependent tiers. The common dependent performance points between tiers follow in the probable order of occurrence.

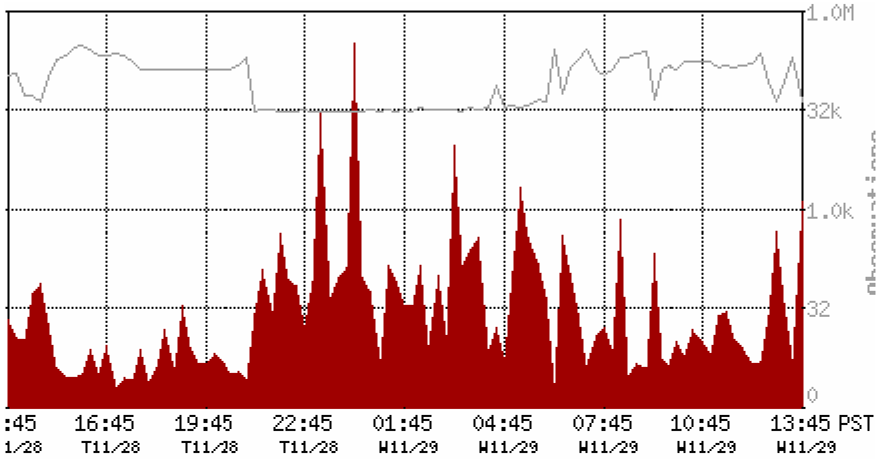
1. Server Response Time (SRT) of Tier N+1 impacts SRT of Tier N.
    - SRT is a function of server resource utilization. High values indicate the server might not have enough memory, CPU, or disk I/O resources to service the application for a given load.
    - Review the Sessions and QoS views to determine what session or user is placing too much load on the server during instances of slow SRT.
2. Network Round Trip Time (NRTT) between Tier N and Tier N+1 is high and affects data throughput.
    - NRTT within a LAN environment is a function of the switch backplane speed and contention for bandwidth on shared uplinks between switches. Best practice is to always put the primary NICs of N-Tier applications on the same switch and within the same VLAN to remove contention for uplink bandwidth issues and to remove two extra switch hops between servers. This can dramatically increase N-Tier application performance.
    - Review the Traffic and Data Volume views to determine if NRTT increases with an increase in volume. If so, there might not be sufficient bandwidth between the two servers for the application.
3. Data Transfer Time (DTT) between Tier N and Tier N+1 is exactly zero or converges to zero. Review the Response Size and Data Transfer Time by Response Size views to see if the application uses diverse response sizes greater than 1.45 KB, which is the amount of data that fits into one packet. A DTT of zero or near zero typically indicates that a single packet is being sent by the back office server for each user request. For database servers, this is typically in the form of a query requesting one row of a data many times instead of asking for all the rows in a single query. Rewrite the queries for optimal performance.

4. Retransmission Time between Tier N and Tier N+1 indicates significant packet loss.
   - Review the Traffic and Data Volume views to determine if large volumes of data are being transferred during times of high retransmission times.
   - Review the Sessions and Network Connection Time views to determine what the latency is for TCP session startup. High values indicate congestion between the two tiers.

The following charts illustrate a Server Response Time dependency between Tier 2 and Tier 3 of an application architecture. The SRT of the Tier 2 server follows the SRT of the Tier 3 server during periods of high response times.



Server Response Time - Tier 2



Server Response Time - Tier 3

Trailing data points between the two tiers indicate that the Tier 3 server performance affects the Tier 2 server performance. The Tier 3 server is a bottleneck in this application architecture.

When the Tier N application latency has the same general curve as the Tier N+1 application latency during peak latency spikes, the Tier N application is probably being adversely affected by the Tier N+1 application.

After you identify the performance bottleneck in the Tier N+1 application (high SRT indicating server issues, high NRTT and retransmission times indicating network issues, zero or high data transfer times generally indicating application issues when SRT and NRTT are low) and correct it, repeat the analytical process to identify secondary bottlenecks.

Glossary

5-minute summary file

A *5-minute summary file*, created by a CA Application Delivery Analysis Standard Monitor, Multi-Port Monitor, Virtual Systems Monitor, CA GigaStor, or a Cisco NAM, consists of 5-minute averages for each performance metric, application, server, and network combination (see page 197).

access layer

In a typical three-tier LAN network (access, distribution, core), the *access layer* is the closest layer to the server and connects servers to the network. Switches and hubs usually fall into the access layer. Typically, all server traffic can be seen at this layer, but this requires the most monitoring points.

ACK packet

During the TCP connection setup, an *ACK packet* is sent by the client to the server to acknowledge receipt of a SYN-ACK packet (see page 207) from the server.

action

See responsive action (see page 204).

application

An *application* specifies a TCP port or port range to monitor across a range of server IP addresses, such as TCP-80 traffic across a /29 server subnet.

Application Connection Time investigation (term)

An *application connection time investigation* is an application incident response (see page 195) that provides IT staff with information about how long it takes to connect to a TCP/IP application port. This includes time for the server to respond with a connection acknowledgment. A CA ADA administrator can also launch or schedule this investigation.

application incident

An application incident occurs when a Network Incident or a Server Incident impacts the performance of an application.

The threshold for a combined metric is crossed when an underlying Network Incident or Server Incident causes a combined metric for an application to cross a performance threshold.

When a combined metric crosses a threshold, the management console rates the performance impact to the application as Major (orange) or Minor (yellow), but does not create an application incident response. You must define the application incident response to launch when an underlying Network or Server Incident occurs for the application.

application incident response

An *application incident response* is an application response to a Network incident (see page 201) or a Server incident (see page 205). For example, if you configure an application incident response for the Exchange application, the management console launches the incident response when a Network incident is created by clients accessing the Exchange application, or a Server incident is created by a server that hosts the application. The management console does not launch an application incident response when the threshold for a Combined metric (see page 197), such as Data Transfer Time, is crossed. The management console lets you assign the following responses to a application: Email notification (see page 199), SNMP Trap notification (see page 206), and Application Connection Time investigation (see page 194).

availability operational level agreement (availability OLA)

An *availability operational level agreement (availability OLA)* reports the percentage of time that an application is available. For example, an application must be available on a server 99% of the time over a one month period.

baseline

A *baseline* enables you to see what is historically normal performance in your network. The management console automatically reports baselines for all TCP sessions between an application port on a server, and a client network. Use baselines to compare the current performance of an application to a historical average of past performance. A crossed baseline does not necessarily indicate a problem. Baselines are calculated hourly, and take into account hour of the day, day of the week and day of the month.

CA ADA Availability Poller service

The *CA ADA Availability Poller service* checks the availability of an application. If the server that hosts an application is monitored by a CA Standard Monitor, the check is performed by the monitoring device. Otherwise, the CA ADA Availability Poller service on the CA ADA Manager checks the availability of the application.

CA ADA Batch service

The *CA ADA Batch service* stages .dat data files for processing by the CA ADA Master Batch service on the CA ADA Manager. This service runs on the CA Standard Monitor.

CA ADA Data Pump service

The *CA ADA Data Pump service* performs weekly database maintenance on the CA ADA Manager.

CA ADA Data Transfer Manager service

The *CA ADA Data Transfer Manager service* synchronizes Cisco WAE device monitoring based on the applications, servers, and client networks defined on the CA ADA Manager. This service runs on the CA ADA Manager.

CA ADA Inspector Agent service

The *CA ADA Inspector Agent service* launches an investigation on an application, server, and its related networks. If the server that hosts an application is monitored by a CA Standard Monitor, the investigation is launched from the monitoring device. Otherwise, the CA ADA Inspector Agent service on the CA ADA Manager launches the investigation.

CA ADA Inspector service

The *CA ADA Inspector service* loads five minute .dat files processed by the CA ADA Master Batch service into the CA ADA Manager database and communicates with the CA ADA Inspector Agent service to launch investigations. This service runs on the CA ADA Manager.

CA ADA Master Batch service

The *CA ADA Master Batch service* runs on the management console to receive data files from the CA ADA Batch service on the CA Standard Monitor for processing into 5-minute .dat files. This service runs on the CA ADA Manager.

CA ADA Messenger service

The *CA ADA Messenger service* synchronizes monitoring on any assigned CA Standard Monitor, CA Multi-Port Monitor, and CA GigaStor monitoring devices, based on the applications, servers, and client networks defined on the CA ADA Manager. This service runs on the CA ADA Manager.

CA ADA Monitor Management service

The *CA ADA Monitor Management service* responds to requests from the CA ADA Manager to transfer .dat files. This service runs on the CA Standard Monitor.

CA ADA Monitor service

The *CA ADA Monitor service* receives mirrored TCP packets and packet digest files from a CA ADA monitoring device. This service runs on the CA Standard Monitor and the CA ADA Manager.

CA ADA Reader service

The *CA ADA Reader* service runs on a CA GigaStor to send packet digest files, which consist of TCP headers, to the assigned CA ADA Standard Monitor or Multi-Port Monitor for metric calculation.

CA Application Delivery Analysis Manager (CA ADA Manager)

The *CA Application Delivery Analysis Manager (CA ADA Manager)* is a component of the CA ADA architecture that provides central configuration, analysis, management, and reporting across multiple monitoring devices. The CA ADA Manager receives response time metrics from any assigned monitoring devices, including a CA ADA Standard Monitor, Multi-Port Monitor, Virtual Systems Monitor, CA GigaStor, or a Cisco NAM.

CA Observer Expert

CA Observer Expert is bundled with CA GigaStor. It combines application response time monitoring from CA ADA, with the ability to drill down into packet level data for root cause analysis.

combination

A *combination* identifies the time frame, application port, server, network, and performance metric where CA ADA computes response time metrics. For example, the management console can report on the average Network Connection Time for all applications and servers that communicated with the Development client network in the last 24 hours.

Combined metric

A *Combined metric* indicates that an application performance problem is caused by either a server that hosts the application or a network that is communicating with the application, or both. The CA ADA management console sets a performance threshold for each of the following Combined metrics: Data Transfer Time (see page 198) and Transaction Time (see page 208). Note that the management console does not create application incidents. However, because Combined metrics include both Network and Server metrics, the management console can rate a server or network as Minor (yellow) or Major (orange), and rate the corresponding performance impact on the application itself. For example, if a Server metric is rated as Minor, the management console can also rate a Combined metric for the application as Minor.

control port application

A control port application uses two TCP ports. The control port sends and receives the request information, and the data port sends and receives the actual data. The same monitoring device must monitor both the control port and the data port traffic to determine the transaction response time. Any type of monitoring device can monitor a control port application.

core layer

In a typical three tier LAN network (access, distribution, core), the *core layer* enables high speed interconnection of distribution layer devices. the core layer usually has the highest speed interconnections and the highest power routers and switches in the network. Typically only client to server transactions are seen at this layer.

Data Transfer Time

Data Transfer Time is a Combined metric (see page 197) that measures the time it takes to transmit a complete application response from the first response (the end of the Server Response Time (see page 206)) to the last packet sent in the request.

Data Transfer Time excludes the initial server response time and includes Network Round Trip Time if there is no more data to send that fits in the TCP window. The response time can be impacted by the design of the application, or the performance of the server or network.

The management console does not open an incident when the Data Transfer Time threshold is crossed.

device

A *device* can be any TCP/IP system connected to the monitored network.

discarded packet

A *discarded packet* is a packet that was intentionally discarded by a monitoring device because the packet did not match the list of applications, servers, and client networks that are specified on the management console.

distribution layer

In a typical three tier LAN network (access, distribution, core), the *distribution layer* is where routing, filtering, and policy management take place. This layer usually includes routers and layer three switches. Data is gathered when access switches send data to the distribution layer. Some server-to-server traffic can be seen at this layer, as long as the servers are on different switches.

domain

A *domain* separates client IP traffic for reporting purposes and identifies the DNS server that the management console uses to resolve the host name of a server.

dropped packet

A *dropped packet* is a packet that is not analyzed by the packet capture driver on a CA Standard Monitor or by the GigaStor Connector on a CA GigaStor, because the monitoring device is too busy to process all of the packets it receives. If the monitoring device drops too many packets, the management console creates a Major monitoring device incident. The management console does not monitor packet loss at the server switch port or the monitor NIC on the monitoring device.

Effective Network Round Trip Time

Effective Network Round Trip Time is a Network metric (see page 202) that consists of Network Round Trip Time (see page 202) plus Retransmission Delay (see page 204). Note that Retransmission Delay is not the delay due to retransmissions; it is the average amount of retransmission delay per round trip. It is important to note that the management console is adding two averages, and is actually combining two metrics.

Email notification

An *Email notification* is an application incident response (see page 195), server incident response (see page 205), or network incident response (see page 202) that notifies the recipients about a threshold violation for the affected applications, servers, or networks.

Estimated Hop Delay

Estimated Hop Delay is the estimate of how much delay was encountered between two nodes. The management console determines this estimate by using the average of all the samples taken, for example, during a Trace Route investigation (see page 207).

Expired Sessions

Expired Sessions measures the number of TCP sessions where the CA ADA Monitor service did not see the TCP session tear down (FIN or RST packet). Sessions which are inactive for a period of time are cleared out of memory and marked as Expired. The management console classifies a session as Expired if it does not observe any packets in a 15 minute period. Too many expired sessions left open can cause servers to become unresponsive.

FIN packet

In the TCP protocol, a SYN packet is used by the client when establishing a TCP connection to a server. Likewise, the *FIN packet* is used to begin the tear down or termination of the TCP connection. The monitoring device determines that a TCP conversation is being terminated when it receives a FIN or RST packet.

fragmented packet

A *fragmented packet* is a packet that has been split into multiple packets as it traverses the network.

hop

A *hop* is the logical link between two gateways in a network. When a data packet traverses a network, it will usually pass through one or more routers or gateways. The path between any two logically adjacent gateways is considered a *hop*.

incident

An *incident* is opened by the management console to raise awareness about a period of anomalous behavior on an application, server, or client network. See responsive action (see page 204).

incident response

An *incident response* helps you troubleshoot a problem at the time it occurs, and reduce the mean time to repair. Assign incident responses to your business-critical applications, servers, and networks. Incident responses notify your team about a performance degradation and actively investigate a problem to gather additional information that can help you identify the root cause of a performance degradation.

investigation

An *investigation* is an active inquiry into specific performance data on applications, networks, and servers. The management console can automatically launch an investigation in response to an incident. A CA ADA administrator can also launch or schedule an investigation.

keep-alive messages

A method to keep a TCP connection active and established persistently rather than establishing a new TCP connection for each request/response. TCP *keep-alive messages* follow a known format and do not skew response time metrics. Application keep-alives, which increment sequence numbers and contain payload, can skew measurements of some server metrics, such as Server Response Time (SRT).

Major performance rating

A *Major performance rating* is a severity state (orange) that is represented in the management console to indicate that the metric value exceeds the Major threshold. The management console sets thresholds for both Minor and Major performance degradations.

metric digest file

A *metric digest file* contains the pre-calculated response time metrics from a Cisco NAM. The CA ADA Manager receives metric digest files from a Cisco NAM.

Minor performance rating

A *Minor performance rating* is a severity state (yellow) that is represented in the management console to indicate that the metric value exceeds the Minor threshold. The management console sets thresholds for both Minor and Major performance degradations.

monitor feed

A *monitor feed* is a source of response time information, such as a CA Standard Monitor.

monitoring device

A *monitoring device* monitors TCP transactions and calculates application, server, and network response time metrics.

monitoring device incident

A *monitoring device incident* is created by the management console if the threshold for the performance and availability of a monitoring device is violated. For example, if a device becomes unreachable, no data is seen by the device, or a device discards packets.

monitoring unit

A *monitoring unit* is the processing load that is created on the CA ADA Manager by adding a monitoring device. For example, a CA Standard Monitor utilizes one monitoring unit. The CA ADA Manager supports up to 15 monitoring units.

multi-tier application

A *multi-tier application* is an application with more than one server, and communication between servers is performed by at least one server that acts as both a server to client requests, and a client to another server.

NetQoS MySql51 service

Starts and stops the MySql server which hosts the CA ADA Manager database.

Network Connection Time

Network Connection Time (NCT) is a Network metric (see page 202) that measures the amount of time between the Syn-Ack sent by the server and the Ack received back from the client. When a network is uncongested, it is a measurement of network latency that represents the minimum latency due to distance and serialization, and is the best possible round trip time for your network architecture. Sudden spikes in this value are commonly attributed to congestion, while a plateau (which goes up and stays up) typically indicates a path change.

Network incident

The management console creates a *Network incident* when the threshold for a network metric, such as Network Round Trip Time, Network Connection Time, Effective Round Trip Time, or Retransmission Delay, is exceeded during a 5-minute interval for a particular application, server, and network combination.

network incident response

An *Network incident response* is a CA ADA response to a Network incident (see page 201). The CA ADA administrator can assign an Email notification (see page 199), SNMP Trap notification (see page 206), and Trace Route investigation (see page 207) to a Network incident.

Network metric

A *Network metric* indicates an application performance problem is caused by a network that is communicating with the application. Use the CA Application Delivery Analysis management console to customize the performance thresholds for each of the following Network metrics: Network Round Trip Time (see page 202), Network Connection Time (see page 201), Effective Network Round Trip Time (see page 199), and Retransmission Delay (see page 204).

network region

A *network region* is a management console tool used to automatically expand a broad subnet definition into narrower subnets. You can define a network with up to 256 regions, such as a /16 network with 256 regions, to define 256 /24 networks. If you use network regions, the management console reports on the narrower network region subnet definitions rather than the broader network definition.

Network Round Trip Time

Network Round Trip Time is a Network metric (see page 202) that measures the time that a packet takes to travel across the network in both directions between the server and clients on a network, excluding lost packets. Application, server, and client processing time are excluded.

network tap

A *network tap* is a hardware device that lets you access the data flowing across a computer network. After a tap is in place, you can connect a monitoring device to it without impacting the monitored network. Taps enable you to view traffic occurring in both directions (upstream and downstream), but only in one broadcast domain in a switched network.

network type

A group of networks that share the same physical access to the application. For example, all of the subnets at a remote site would share the same WAN link to the data center.

observation count

The observation count measures the number of times during a five minute interval that a monitoring device calculates a performance metric for a particular application, server, and network combination. Within a TCP transaction there can be different numbers of observations for different metrics. For example, there may be more observations of Network Round Trip Time than Server Response Time. Other metrics are links and will always have the same number of observations. For example, each TCP transaction has one Server Response Time observation and one Data Transfer Time observation. To rate a metric as Normal, Minor (yellow), or Major (orange), the metric must have a minimum number of observations.

OLA

See performance operational level agreement (performance OLA) (see page 203) and availability operational level agreement (availability OLA) (see page 195).

Packet Capture investigation

A *Packet Capture investigation* is an application incident response (see page 195) or a server incident response (see page 205) that performs a filtered capture of a particular server, application port, and network experiencing a problem. A CA ADA administrator can also launch or schedule this investigation.

packet digest file

A *packet digest file* contains the TCP headers from a Cisco WAE device, or CA GigaStor Connector.

Packet Loss Percentage

Packet Loss Percentage is a Network metric (see page 202) that measures the ratio of retransmitted data to total data within the network from the vantage point of the monitoring device adjacent to the server. The monitoring device can see packets retransmitted by the server due to data loss in the server-to-client direction along the network path. When data loss occurs in the client-to-server direction before reaching the server, the monitoring device cannot observe the packet loss, and that loss is not included in the Packet Loss Percentage. On the Engineering page of the management console, Packet Loss Percentage is part of the QoS report.

performance operational level agreement (performance OLA)

A *performance operational level agreement* (performance OLA) lets you evaluate compliance with application performance goals for a remote site. By default, the management console does not define operational levels for application performance.

performance thresholds

A performance threshold is a boundary of acceptable performance behavior that exists by default for each application. Thresholds enable the management console to rate data. They contribute to incident creation, incident responses, and investigations.

Performance via SNMP investigation

A *Performance via SNMP investigation* is a server incident response (see page 205) that uses SNMP to poll a server for performance information, such as CPU and memory utilization. A CA ADA administrator can also launch or schedule this investigation.

permission set

A defined list of application, server, and network aggregations that a user has permission to view. An aggregation can be a member of one or more permission sets.

Ping Response Time investigation

A *Ping Response Time investigation* is a server incident response (see page 205) that measures the time it takes to receive a ping reply after sending a ping request, and report on the packet round trip time. A CA Application Delivery Analysis administrator can also launch or schedule this investigation.

Ping Response Time vs. Packet Size investigation

A *Ping Response Time vs. Packet Size* investigation measures the time it takes to receive a ping reply for ping requests (data packets) of various sizes. This investigation helps track excessive delays and lack of connectivity at various packet sizes. A CA ADA administrator can manually launch or schedule this investigation.

Refused Session Percentage

Refused Session Percentage is a Server metric (see page 206) that measures the percentage of connection requests the server explicitly rejects during the reporting interval. This metric is part of the Unfulfilled TCP/IP Session Requests report in the CA ADA management console.

report page

The management console organizes report data under standard *report pages* designed for particular types of users such as operations personnel, executives, and engineers.

responsive action

A *responsive action*, such as sending a notification or starting an investigation, is a response to a performance threshold violation.

Retransmission Delay

Retransmission Delay is a Network metric (see page 202) that measures the elapsed time between the original packet send and the last duplicate packet send. The management console reports Retransmission Delay as an average across observations and not just for the retransmitted packets. For example, if one packet in a set of 10 requires 300 ms of retransmission time, the Retransmission Delay is reported as 30 ms (300 ms/10 packets).

role

A *role* specifies the pages of the CA ADA management console that are displayed to a CA ADA user.

RST packet

A RST packet is a normal way to end a TCP session. A web browser typically ends a session with RST rather than FIN. The management console counts a RST packet during the connection handshake as an Unfulfilled Session Request. If the monitoring device sees a RST before the TCP three way handshake has completed, the management console considers the session to be rejected.

sensitivity level

The *sensitivity level* is a unitless measure on the scale of 0-200 that is applied to a proprietary formula which calculates a new threshold for each client, server and application combination, based on historical data. The management console automatically generates a new threshold value for a metric each night at midnight GMT, using percentile statistics from the last 30 days. The management console automatically generates a separate set of threshold values for the users who access an application from each client network.

Server Connection Time

Server Connection Time (SCT) is a Server metric (see page 206) that measures the amount of time that a server takes to acknowledge the initial client connection request by sending a Syn-Ack in response to the client's SYN packet.

Server incident

A Server incident is created by the management console when the threshold for a Server metric, such as Server Response Time, Server Connection Time, Refused Session Percentage, or Unresponsive Session Percentage, is exceeded during a five minute interval for a particular application, server, and network combination.

server incident response

A *Server incident response* is a response by the management console to a Server incident (see page 205). The management console lets you assign the following responses to a Server incident: Email notification (see page 199), SNMP Trap notification (see page 206), Ping Response Time investigation (see page 204), Performance via SNMP investigation (see page 204), and Packet Capture investigation (see page 203).

Server metric

A *Server metric* indicates that an application performance problem is caused by a server that hosts the application. Use the CA ADA management console to customize the performance thresholds for each of the following Server metrics: Server Response Time (see page 206), Server Connection Time (see page 205), Refused Session Percentage (see page 204), and Unresponsive Session Percentage (see page 208).

Server Response Time

Server Response Time is a Server metric (see page 206) that measures the time it takes for a server to send an initial response to a client request, or the initial server think time. Increases in the Server Response Time generally indicate a lack of server resources such as CPU, memory, disk I/O, a poorly written application, or a poorly performing tier in a multi-tier application.

server subnet

A server subnet identifies a contiguous range of server IP addresses that are monitored by each monitoring device. When defining an application, you can assign a particular server subnet to an application, to enable the management console to automatically monitor application performance across a contiguous range of server IP addresses.

severity

Severity classifies performance data as None, Unrated, Minor, Major, and Unavailable, by established thresholds over a time period.

SNMP profile

A *SNMP profile* is used by the management console to manage SNMPv3 user credentials and SNMPv1 and SNMPv2 community names. A SNMP profile maintains the SNMP user credentials required by the management console to query the SNMP agent on a server or network device and to send a SNMP trap message.

SNMP Trap notification

A *SNMP Trap notification* is an application incident response (see page 195), network incident response (see page 202), or server incident response (see page 205) that notifies a SNMP Manager about the Open or Closed incident status of the affected applications, servers, or networks.

SPAN

Switched Port Analyzer (*SPAN)*, also known as port mirroring, is used on a Cisco network switch to send a copy of all network packets seen on one switch port to a network monitoring connection on another switch port. This is commonly used by network appliances to monitor network traffic. SPAN enables monitoring devices to view traffic occurring on multiple broadcast domains on one or more switch ports. SPAN capabilities vary by chassis.

SYN packet

In the TCP protocol, conversations (connections) between clients and servers are established by means of a three-way handshake. The *SYN packet* is sent by the client to a server to initiate the connection setup. A monitoring device uses the SYN packet in the timing and analysis of monitored connections on the network.

SYN-ACK packet

During the TCP connection setup, a *SYN-ACK packet* is sent by the server to the client to acknowledge receipt of a SYN packet (see page 207) from the client. A monitoring device uses the SYN-ACK packet in the timing and analysis of monitored connections on the network.

synchronize monitoring devices

Synchronize monitoring devices to monitor TCP sessions based on the current client network, server subnet, and application definitions on the management console. To minimize temporary interruptions to monitoring during synchronization, complete all changes before synchronizing monitoring devices.

tap

See network tap (see page 202).

three-way handshake

A *three-way handshake*, in the TCP protocol, is used to establish a connection between a client and a server. The SYN packet (see page 207) is sent by the client to a server to initiate the connection setup. A SYN-ACK packet (see page 207) is sent by the server to the client to acknowledge receipt of a SYN from the client. Finally, an ACK packet *(see page 194)* is sent by the client to the server to acknowledge receipt of a SYN-ACK from the server and to establish the TCP connection. A monitoring device uses the three-way handshake in the timing and analysis of monitored connections on the network.

thresholds

See performance thresholds (see page 203).

Trace Route investigation

A *Trace Route investigation* is a network incident response (see page 202) that records the path and each hop between a monitoring device and end-points to monitor latency and routing issues, and optionally, SNMP poll each router for its performance information. A CA ADA administrator can also launch or schedule this investigation.

traceroute

Traceroute refers to either of two types of diagnostic tools used in incident analysis: ICMP or TCP.

transaction

A *transaction* is a TCP request and all the subsequent responses. A single application transaction, such as loading a web page, can consist of multiple TCP transactions.

Transaction Time

Transaction Time is a Combined metric (see page 197) that measures the amount of time elapsed from when the client sends the request to when it receives the last packet in the response. Transaction Time is the sum of Server Response Time (see page 206), Network Round Trip Time (see page 202), Retransmission Delay (see page 204), and Data Transfer Time (see page 198). The management console does not open an incident when the Transaction Time threshold is crossed.

Unrated performance rating

An *Unrated performance rating*, indicated by a gray severity state on the Operations page of the management console, means that either there is insufficient past data to establish a threshold (two full business days of data are needed), or there were not enough observations to exceed the minimum observations threshold.

Unresponsive Session Percentage

An *Unresponsive Session Percentage* is a Server metric (see page 206) that measures the percentage of sessions where a connection request was sent, but the server did not respond. This metric is part of the Unfulfilled TCP/IP Session Requests view.

WAN

A *wide area network (WAN)* is a network that typically covers a large, diverse area comprised of multiple Local Area Networks or LANs. WANs can be private, used by different offices of a single enterprise, or public, such as the Internet.

WAN optimization device

A *WAN optimization device* reduces the volume of traffic that is transferred between the data center and a remote office through compression or other algorithms. Cisco WAE devices and Riverbed Steelhead appliances are examples of WAN optimization devices.

Index