How Does Prime Infrastructure Ensure Consistent Application Experiences for WAN End Users?

Ensure Consistent Application Experiences for WAN End Users

Cisco Prime Infrastructure can help ensure high-quality WAN end-user experiences across applications at multiple sites.

• View Application Key Performance Indicators for Sites
• Set Up the Application Performance Dashboard for WAN Optimization
• Identify Low-Performing WAN Applications, Clients, Servers, and Links
• View WAN Optimization Results
• View WAN Client-Server and Site-to-Site Optimized Traffic Flows

Note To use this feature, your Cisco Prime Infrastructure implementation must include Assurance licenses.

Network operations staff must share a common data resource that gives them complete visibility into network performance data throughout every stage of the optimization cycle, including:

• Identifying the sites and applications that are candidates for optimization, so that network designers can plan where application optimization is critical (see View Application Key Performance Indicators for Sites).
• Establishing site and application performance baselines (see Set Up the Application Performance Dashboard for WAN Optimization).

Cisco Prime Infrastructure performs baselining for key performance metrics and detects abnormal deviations of baselined values. The key performance metrics include:

• Server Response Time
• Client Transaction Time
• Network Round-Trip Time
• MOS score
• Jitters
• Packet loss

How Does Prime Infrastructure Ensure Consistent Application Experiences for WAN End Users?

View Application Key Performance Indicators for Sites

- Bytes sent/received
- Interface utilization
- CPU Utilization
- Memory Utilization

Cisco Prime Infrastructure determines the baseline (mean) for each metric by taking the average values of the metric during the last 30 days. Average values are computed separately for each hour of the day for each monitored entity (such as interface, host, site, or application). For example, the baseline for HTTP response time of a given server between 9AM to 10AM today will be different from the baseline of the same server between 7PM to 8PM yesterday.

Cisco Prime Infrastructure also computes the metrics' standard deviations using the last 30 days of data. Similar to averages, standard deviations are computed separately for each hour of the day for each monitored entity.

- Post-implementation validation that WAN performance and application stability have actually improved (see View WAN Optimization Results).

Because the mean and standard deviation of each metric vary over time, Cisco Prime Infrastructure continuously reevaluates the thresholds used to compute the health scores (adaptive thresholds). Cisco Prime Infrastructure computes baselines and thresholds every hour, and evaluates health scores every five minutes. In each interval:

1. Health scores are computed for every application-site combination.
2. These health scores are aggregated to derive the overall health of each business-critical application (across all sites) and overall health of each site (across all business-critical applications).

When aggregating across sites/applications, the worst scores are used. For example, if any business-critical application of a given site is rated "red," that site is also rated "red" for that interval. See Customize Service Health Rules for Application Performance for more information.

- Ongoing monitoring and troubleshooting of the optimized flows (see View WAN Client-Server and Site-to-Site Optimized Traffic Flows).

Using the baseline means and standard deviations, Cisco Prime Infrastructure can monitor application and service health issues by detecting abnormal deviations of key metrics from their baselined values and assign a health scores (red, yellow, or green) for each application and site for each monitoring interval.

- A red score indicates a highly abnormal deviation from baseline (deviations from baselines with a probability of less than 0.1%).
- A yellow score indicates a mildly abnormal deviation (deviations with a probability of less than 1%).
- A green score indicates that the metric is within its normal range.
- A gray score indicates there is insufficient data for a site/application.

Cisco Prime Infrastructure offers a consistent data resource for each of these stages in performance optimization.

View Application Key Performance Indicators for Sites

Choose **Services > Application Visibility & Control > Service Health** to view the sites and their business critical applications. Each application for a site is given a score for each of the KPIs (Key Performance Indicators) that are available in the system:

- **Traffic** (megabits per second)
- **Client Experience** (varies based on application type: average transaction time for transaction-based applications such as HTTP, or MOS code for real-time applications such as RTP)
- **Network Performance** (average network time for HTTP, jitter and Package Loss for RTP)

How Does Prime Infrastructure Ensure Consistent Application Experiences for WAN End Users?

2

How Does Prime Infrastructure Ensure Consistent Application Experiences for WAN End Users?

Create Custom Applications to Monitor Their Performance

      • **Application Response** (applicable only for transaction-based applications such as HTTP)

The KPI scores can come from multiple data sources; scores are computed across all data sources for all of the KPIs, and the overall score in the main dashboard is an aggregate of these scores. Scores are assigned as red, yellow, or green based on the warning and critical threshold values assigned in Health rules page. You can navigate to this page by clicking **Launch Health Rules** in the **Services > Application Visibility & Control > Service Health page**. You can use this option to add/modify the health rule settings at any site level by providing new standard deviation value or custom value.

For data to be displayed in Service Health, there must be at least one hour of data. After the first hour, the previous hour's data is overlaid on the data line as the historical data for the next hour. After the first day, standard deviation and mean are based on the hourly data for the previous day.

These scores are stored for seven days. When you view the data for a previous day, the maximum moving time interval is six hours (you can look at up to six hours of data at a time).

Create Custom Applications to Monitor Their Performance

Choose **Services > Application Visibility & Control > Applications and Services** to create and manage custom applications and services. Services are groups of applications. Prime Infrastructure provides a default set of applications and services consistent with the Cisco NBAR standard. (See NBAR Home page for more information.)

The **All Applications** table shows the list of all predefined and user-defined applications. You can configure some of the applications as **Business Critical applications.**

You can create custom applications that contain the definitions you require and which are not available (either from the device or from Prime Infrastructure). After you create an application, you can deploy the application to the supported devices. Deploying the application definition to the device makes Netflow exported data consistent with Prime Infrastructure and other management tools.

If you deploy a custom application to a device and later want to remove it, you must undeploy the application using the **Applications and Services** option. If you delete the custom application from Prime Infrastructure only, the custom application remains active on the device.

Applications without definitions are displayed as "unknown."

Custom applications are organized under services; services are organized by category and subcategory to align with the Cisco NBAR standard. For more information about NBAR, see NBAR Home page.

To create a custom application, follow these steps:

Step 1   Choose **Services > Application Visibility & Control > Applications and Services**.

Step 2   Click **Create**.

Step 3   Complete the required **General** and **Attributes** fields.

Step 4   Choose the traffic classification rule from the **Rule** drop-down list.

Note   Protocol is applicable for NAM and NBAR2 supported IOS devices.

      Server/DSCP is applicable for NBAR2 supported IOS devices ver 15.5(2) and above.

      URL - Applicable for NAM ver 6.0(2) above and NBAR2 supported IOS devices.

      NAM Server IP Address - Applicable for NAM 6.0(2) and above devices only.

      RTP Payload Type - Applicable only for Prime Infrastructure.

How Does Prime Infrastructure Ensure Consistent Application Experiences for WAN End Users?

3

How Does Prime Infrastructure Ensure Consistent Application Experiences for WAN End Users?

View Service Health Using the AVC Service Health Window

Step 5   Click the **Condition** drop-down list and enter the required values in the applicable fields based on the chosen rule.

Step 6   Click the Plus icon to add more traffic rules and conditions.

Step 7   Click **Create**.

The newly created application appears in the **All Applications** table.

Step 8   Choose the newly created application and click **Deploy**.

Note   You can undeploy existing application. To undeploy, select the custom application from **Application and Services** and click **Undeploy**.

Step 9   Choose the devices on which you want to deploy this application and click **Submit**.

Step 10   Click **View Jobs** to view the status of the deployment job.

Step 11   In the device selection pane of the Application Deployment dialog box, you can:

- Select a parent group or a child group to deploy all the devices listed inside the group.

Note   You must select only one parent group at a time. However, you can select one or more child groups within a parent group.

- Expand a group and select one or more devices individually.

- View the number of devices selected in the CLI Preview screen.

Note   Expanding a parent or child group and then selecting the respective check box will only select the initial set of devices listed within and not the successive devices. You must first select the check box of a parent group and then expand it to select all the underlying devices.

View Service Health Using the AVC Service Health Window

Choose **Services > Application Visibility & Control > Service Health**, then click **Health Summary**. Cisco Prime Infrastructure changes to display the heath information in a timeline.

The Service Health window allows you to view the information shown in the figure below:

How Does Prime Infrastructure Ensure Consistent Application Experiences for WAN End Users?

4

How Does Prime Infrastructure Ensure Consistent Application Experiences for WAN End Users?

Customize Service Health Rules for Application Performance

The information displayed in the Service Application Visibility and Control Service Health Window is described below:

Table 1: Services Application Visibility and Control Service Health Window Descriptions

1	Lists the location groups for the filter you selected.
2	Click to toggle between the Health Summary and the Health Timeline.
3	Provides quick links to: • **Health Rules page**, where you can modify the health rule settings as necessary for your network. • View and modify the currently defined business critical applications.
4	Displays the filter you're currently viewing. You can click any filter to remove it and refresh the window.
5	Lists the business critical applications.
6	Colored symbols indicate good, warning, and critical threshold values based on the health rule setting specified in **Health Rules page**.
7	Move the slider to specify the time range in which you want to view service health information.

Customize Service Health Rules for Application Performance

The data displayed in **Services > Application Visibility & Control > Service Health** is computed using health rules. You can customize the health rules by clicking the desired row and editing the Critical and Warning values.

How Does Prime Infrastructure Ensure Consistent Application Experiences for WAN End Users?

5

> • Critical—turns red when the data value exceeds the specified Critical value.
> • Warning—turns yellow when the data value exceeds the Warning value.

If the health rule does not exceed the specified Critical or Warning values, it is green.

For example, for Traffic Rate, you might specify the T1 the baseline value of 100 Mbps for a given site, application, and datasource, and the standard deviation value of 20 Mbps.
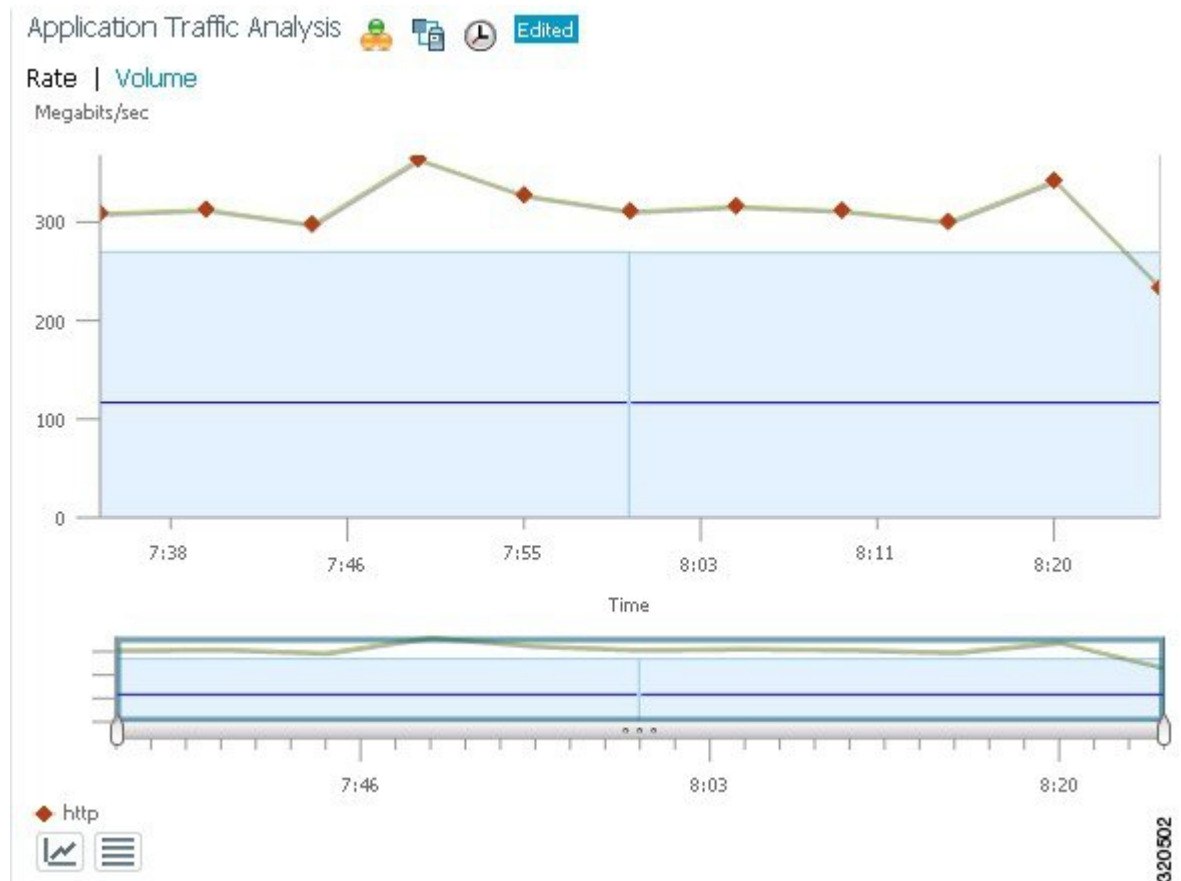
If the Traffic Rate exceeds 161.8 Mbps, which is 100+(3.09 x 20), you see a red bar indicating a critical warning.

You can click any of the colored bars to get further details.

Enable Baselines for Computing Application Performance

Standard deviation and mean values are used to compute the scores in Service Health. Baselining is not enabled by default. When baselining is enabled:

> • The blue box indicates the standard deviation.
> • The blue line indicates the mean value for that hour.

To enable baselining, follow these steps:

Step 1   Choose **Dashboard > Performance > Application**.

How Does Prime Infrastructure Ensure Consistent Application Experiences for WAN End Users?

Set Up the Application Performance Dashboard for WAN Optimization

Baselining is supported by these dashlets:

a. Application Traffic Analysis—Shows the aggregate bandwidth rate/volume for a site/enterprise one application, service, or set of applications.

b. Application ART Analysis—Shows the response times for a transaction.

Step 2 To enable application traffic analysis baselining:

a) Open the **Application Traffic Analysis** dashlet, hover your cursor over the dashlet icons and click **Dashlet Options**.

b) Select the **Baseline** check box and save your changes.

Step 3 To enable application response time analysis baselining:

a) Open the **Application ART Analysis** dashlet, hover your cursor over the dashlet icons and click **Dashlet Options**.

b) Choose a metric from the **Metric Type** drop-down list.

If you choose the **Server Response Time** metric, you can select an individual Application Server to see what the response time of that server has been in the past.

c) Select the **Baseline** check box and save your changes.

Set Up the Application Performance Dashboard for WAN Optimization

Follow these steps to establish the standard performance characteristics of your candidate applications and sites before implementing WAN optimizations.

Step 1 Choose **Dashboard > Performance > Application**.

Step 2 Add the following dashlets (see Add Dashlets to Dashboards) to this page:

- Worst N Clients by ART Metrics

- Worst N Sites by ART Metrics

- Application Server Performance

- Application Traffic Analysis

Step 3 Use these dashlets to establish the performance characteristics of your optimization candidates as currently configured.

- **Worst N Clients by ART Metrics**: For the worst-performing clients and applications: Maximum and average transaction times, and 24-hour performance trend.

- **Worst N Sites by ART Metrics**: The same information for the worst-performing sites and applications.

- **Application Server Performance**: For all application servers: the maximum and average server response time, and a 24-hour performance trend.

- **Application Traffic Analysis**: Gives 24-hour application traffic metrics in bytes per second and packets per second. Calculates statistical mean, minimum, maximum, median, and first and second standard deviation for the period.

You can sort by any column in any dashlet by clicking the column heading. You can also filter the data in the dashlets by **Time Frame**, **Site**, and **Application**.

How Does Prime Infrastructure Ensure Consistent Application Experiences for WAN End Users?

7

How Does Prime Infrastructure Ensure Consistent Application Experiences for WAN End Users?

Identify Low-Performing WAN Applications, Clients, Servers, and Links

Step 4 Click the **Site** tab and use **Top N Applications, Top N Devices with Most Alarms**, **Top N Clients** and **Worst N Clients by ART Metrics** as you did in Step 3.

Identify Low-Performing WAN Applications, Clients, Servers, and Links

Follow these steps to identify your network's lowest performing applications, clients, servers, and network links.

Step 1 Choose **Dashboard > Performance > WAN Optimization**.

Step 2 Add the following dashlets (see Adding Dashlets) to this dashboard.

- Application Traffic
- Server Traffic
- Client Traffic
- Network Links

Step 3 Using these dashlets, identify the optimization candidates.

- All of the dashlets show the current traffic rate (in bytes per second), average number of concurrent connections, and average transaction time in milliseconds, for every application, client, server, or network link.
- **Network Links** also shows the sites for that client and server endpoints of each link, and the average length of time that the link exists.
- **Server Traffic** shows both the server IP address and the application that it serves.

Step 4 Sort and filter the performance data as needed.

- To sort on any column in any dashlet, click the column heading.
- To filter the data displayed in all of the dashlets by **Time Frame**, **Site**, or **Application**, enter or select the filter criteria that you want on the **Filters** line and click **Go**.
- To filter within a dashlet, click its Filter icon and specify a Quick or Advanced Filter, or use a Preset Filter.

Step 5 For a quick report of the same data:
a) Choose **Report > Report Launch Pad**.
b) Specify filter and other criteria for the report, then click **Run**.

View WAN Optimization Results

After you have deployed changes at candidate sites, follow these steps to validate the return on your optimization investment.

Step 1 Choose **Dashboard > Performance > WAN Optimization.**

The dashlets on this page show:

- **Transaction Time (Client Experience)**—Graphs average client transaction times (in milliseconds) for the past 24 hours, with separate lines for optimized traffic and pass-through traffic (in which optimization is turned off). With optimization enabled, you should see a drop in the optimized traffic time when compared to the pass-through time.

- **Average Concurrent Connections (Optimized vs Passthru)**—Graphs the average number of concurrent client and pass through connections over a specified time period.

- **Traffic Volume and Compression Ratio**—Graphs the bandwidth reduction ratio between the number of bytes before compression and the number of bytes after compression.

- **Multi-Segment Network Time (Client LAN-WAN - Server LAN)**—Graphs the network time between the multiple segments.

Step 2 You can filter the data in the dashlets by **Time Frame**, **Client Site**, **Server Site**, and **Application**.

Step 3 To generate a report:

a) Choose **Tools > Reports > Report Launch Pad**, then choose **Performance > WAN Application Performance Analysis Summary**.

b) Specify the filter and other settings for the report, then click **Run**.

View WAN Client-Server and Site-to-Site Optimized Traffic Flows

Follow these steps to monitor optimized WAN traffic.

Step 1 Choose **Dashboard > Performance > WAN Optimization.**

Step 2 In the **Multi-Segment Analysis** dashlet, click **View Multi-Segment Analysis**.

Step 3 Click the **Conversations** tab to see individual client/server sessions, or the **Site to Site** tab to see aggregated site traffic. For each client (or client site) and server (or server site) pair and application in use, these pages show:

- Average and Max Transaction Time—The time between the client request and the final response packet from the server. Transaction time will vary with client uses and application types, as well as with network latency. Transaction Time is a key indicator in monitoring client experiences and detecting application performance problems.

- Average Client Network Time—The network time between a client and the local switch or router. In Wide Area Application Services (WAAS) monitoring, Client Network Time from a WAE client data source represents the network RTT between the client and its edge WAE, while Client Network Time from the WAE server data source represents the WAN RTT (between the edge and core WAEs).

- Average WAN Network Time—The time across the WAN segment (between the edge routers at the client and server locations).

- Average Server Network Time—The network time between a server and NAM probing point. In WAAS monitoring, Server Network Time from a server data source represents the network time between the server and its core WAE.

- Average Server Response Time—The average time it takes an application server to respond to a request. This is the time between the client request arriving at the server and the first response packet being returned by the server. Increases in the server response time usually indicate problems with application server resources, such as the CPU, Memory, Disk, or I/O.

- Traffic Volume—The volume of bytes per second in each of the Client, WAN, and Server segments.

Step 4   Sort and filter the performance data as needed.

- To sort any column, click the column heading.

- You can filter the data displayed by **Time Frame**, or click the Filter icon and specify a Quick or Advanced Filter, or use a Preset Filter.