



GUIDE

The Five Network Metrics You Should Keep to See Into the Cloud

The whole point of moving business applications to the cloud is to take management tasks off of IT's plate and streamline processes. With those apps and workloads out of the data center, though, visibility also disappears, as the old tools don't measure the application delivery path once apps are off-premises. Cloud environments create a big challenge when it comes to tracking metrics and end-user experience.

What's changed in this modern cloud computing world isn't the actual metrics. The key performance indicators that show how users are experiencing an application are largely the same as they've always been. But IT's control over the situation is what has changed. As companies move to the cloud, whether by using SaaS apps or migrating apps to AWS, IT's ability to see typical network metrics has disappeared.

IT is still responsible for performance, but without the visibility of the old metrics, gauging performance can become near-impossible. Finding the balance of control with cloud that keeps users happy and IT mostly hands-off will be different for every organization. We recommend these five metrics as the building blocks to measuring end-user experience and network performance in the cloud.

1. Latency. This metric measures the time it takes for packets to travel from source to destination, measured asymmetrically to match the nature of the internet. Because one-way and two-way traffic are different for any given app delivery path, round-trip time is a partner metric to latency. Latency is about perception, so the perception of latency is different for different users and applications.

It might have been easy to take good latency for granted when applications were all running in an on-premises data center. In the old days, when 100 people were using a local application, 10 milliseconds of latency wasn't a problem. But now, there are thousands of people using an app in the cloud, and a tenfold increase in latency (100 milliseconds), as well as round-trip time, isn't OK anymore.

Plus, web applications are very chatty. A web app isn't just one monolithic file that's downloaded, but rather a series of requests and responses from the client to the web server. So the increase in latency will affect each of these requests and objects downloaded. When the app in question is a business-critical SaaS app that employees are using all day, the increased latency can have a big effect on productivity. This means that choosing the right cloud provider with the right hosting locations is critical to SaaS success.

Once you identify that latency is a problem, you'll need to figure out where the latency is occurring. It could be in your network, your WiFi, your WAN connection, over the open internet or even in your service provider's environment.

When there's a latency issue, try these basic troubleshooting steps:

- User/host: Are other apps on the host slow as well?

- Application: Do all other apps on that device run just fine?
- Location: Is it just at headquarters or a specific remote office?

Beyond that, you'll need better tools. Simple methods like traceroute won't help, because of how the internet is architected. A traceroute can give a rough idea of the route from the user to the app, but the internet is asymmetric, and the most efficient route from the server back to the user will be taken when that web app is actually in use. There's no guarantee that the routes will be the same every time.

2. Packet loss. This is the percentage of network packets that are lost between the source and destination. (This isn't exactly the same as data loss.) Depending on the protocol, packet loss can lead to network congestion, wasted time and frustrated users. In small bursts, networks can handle loss, but if loss compounds, then it can have severe effects on end users.

For example:

- With 1% packet loss, users will notice, as it can slow the app up to 40%
- With 2% packet loss, response time has increased eight to ten times the normal rate, due to the re-transmit rate

Back when apps were all hosted internally over the LAN, packet loss wasn't really a concern. If there was packet loss, it was pretty straightforward to find and fix. But on the open internet, it's a different story. Internet protocol TCP guarantees delivery, but if it detects packet loss and re-transmits the data, it adds latency and leads to congested networks.

Today's VoIP and video streaming applications are where packet loss can be noticeable. UDP data loss means that users will experience choppy audio and video, with dropped calls and poor quality. Compression and decompression algorithms can handle some packet loss in these apps, but anything more than 4% to 5% will be noticed by a user. It's possible to track packet loss independently on both data and voice if you're supporting those applications.

3. Capacity. This is the maximum possible transit rate between source and destination. Capacity is an end-to-end metric, limited by the most congested hop along the application delivery path. So on that entire path, speed can be limited by whatever the slowest point is. This becomes especially important when considering cloud services, since you don't have control over a provider's network, and don't know how fast the connection really is.

It's important to recognize capacity's place in a metrics report, and how it ties into user experience. Measuring capacity gauges the actual application path, including WiFi. And continuous monitoring that includes capacity is what's needed when using the dynamic internet. Monitoring the total capacity allows companies to validate service-level agreements (SLAs) with ISPs.

There are two varieties of capacity to understand: available capacity and utilized capacity. They're almost always different.

Available Capacity: Most accurate measure of network resources available to applications. This metric is used to identify or rule out the root cause of degradation.

Utilized Capacity: High utilization is a strong indicator for performance degradation. It's possible to reduce troubleshooting time when tracking capacity by isolating where the slowest hops are.

Capacity has taken over for bandwidth, which used to be a primary network metric. But path-based metrics like capacity are more indicative of user experience, which is a key goal for organizations today. Measuring capacity can reduce knee-jerk bandwidth purchases, since bandwidth sometimes gets blamed when it's a capacity problem. Improving capacity can result in a reduced MTTR for troubleshooting issues. And, historical capacity data can point to network issues that bandwidth can't, and identify future needs (read: capacity planning).

4. Jitter. This metric reflects the percentage of packets with delay variation between source and destination. This most often manifests as patchy video or choppy audio. When jitter is problematic, it's highly visible. While most networks have some amount of jitter, the quality of a call or online meeting can be affected with as low as 30 to 40ms. When infrastructure is moved to the cloud, it is more important than ever to monitor the entire delivery path of voice, video and application traffic in order to accurately identify where jitter is occurring.

5. Quality of Service. This metric ties to routing priority for traffic over specific ports or protocols. It matters when congestion hits a network, since QoS is what ensures a good experience on business-critical applications. For some apps like VoIP or video, if those routing priorities are demoted or re-marked, the network can experience jitter, data loss and latency. An organization might also need to route application priority according to business needs, such as a financial institution's high-speed trading app taking precedence over regular internet traffic.

In this world of cloud challenges and possibilities, it's easy to lose sight of the old visibility and metrics. You no longer own the application delivery infrastructure, and yet that infrastructure can radically affect performance and end-user experience. Start with these five network metrics to regain the right amount of control of your apps and users' experience.

ABOUT APPNETA

AppNeta is the leader in proactive end-user performance monitoring solutions built for the distributed digital enterprise. With AppNeta, IT and Network Ops teams can assure continual and exceptional delivery of business-critical applications. AppNeta's SaaS-based solutions give IT teams essential application and network performance data, allowing them to continuously monitor user experience across any application, network, data center or cloud. For more information, visit www.appneta.com.