Accelerating Mobile Access

Mobile devices are proliferating, and their use to access applications is skyrocketing, while users are less accepting of application performance issues than ever before. Since mobile devices have limited bandwidth, IT departments need to implement services that will speed access to web applications while expanding the devices these applications support.

White Paper

by Don MacVittie

Introduction

There is no doubt that wireless devices—primarily smartphones and tablets—are proliferating. In fact, according to CTIA, an international association for the wireless telecommunications industry, in June of 2011, there were more active wireless devices in the United States than there were people.[1] These devices are being used for phone, text, and, increasingly, Internet communications. This number seems to be borne out by other research. For example, in 2010, Neilsen estimated that there were 60.7 million mobile web users.[2]

At the same time, users have become increasingly demanding about website response times. In February 2012, The New York Times reported that web users are willing to wait only about 250 milliseconds for a page to load.[3] In the ever faster and more reliable world of the desktop, this expectation might be reasonable. In the cellular networking world, it appears on its face to be unreasonable. And yet it is an expectation that enterprises must strive to fulfill.

The enterprise has very little control over the performance of the wireless network, and rapidly increasing usage conspires with the time and expense of upgrading the cellular network to keep latency high and throughput low.

Gartner predicts that through 2015, wireless traffic will increase at an average rate of 91 percent annually,[4] a rate that will challenge even fourth-generation (4G) cellular networks to provide adequate bandwidth and response times.

Thus, organizations must seek an architectural solution to wireless wide area network (WWAN) performance that can be implemented on the enterprise network and provide benefits to the wide range of devices being used on the WWAN to transfer data.

Mobile Acceleration Issues

Since the enterprise does not control the cellular network or its performance, the best solution for enterprise IT to improve application performance is to improve the way that applications communicate over that network. The two best options for improving communications are reducing connections and reducing content size.

Both options reduce the number of round trips on the cellular network, one by reducing the TCP/IP overhead of reconnecting, the other by reducing the number of packets required to complete connections. Each contributes to overall application performance, and the performance improvements can be dramatic in areas where the cellular network is weak or slow.

TCP Overhead

When TCP connects two computers, a series of TCP packets is passed between the two computers before the application sends any data at all. This process (known as handshaking) is common knowledge among IT professionals. The problem with handshaking is that as the latency in a network increases, this initial setup takes longer, making an application appear slower than it actually is. Since cellular networks are out of the control of the enterprise, reducing this handshaking exercise—a TCP optimization—and reducing the number of connections required are the most that IT staff can do to address this issue. While this optimization is not new, its benefits apply to mobile devices as well as to traditional clients.

Excess Data Transfer

The problem of excess data being sent is compounded by page load inefficiency. With mobile devices, quite often a page will fill more than a single screen, scrolling off the bottom of the screen. In those cases, to render the page, the browser only needs a subset of the total page elements—those visible when the page is first rendered. The rest of the objects need to be sent to the client, but not upfront.

Web servers send a page in the order elements are encountered, however, even if that means the entire page waits for the loading of an image the user will not see until they scroll down. Content reordering addresses this issue by changing the order in which objects are delivered to the client; as a result, page load times appear minimal because the client receives enough information to render the visible portion of the page early in the process, and objects that will not be visible at load time are downloaded later. This is referred to as progressive rendering.

BIG-IP Solution Addresses Airline Application Delivery Challenges

A Fortune 500 transportation services company addressed the following challenges by deploying F5 BIG-IP solutions:

- Slow page load and/or slow response times for end users
- Poor application performance
- Complexity managing Oracle solutions
- Bandwidth management by reducing web traffic payload sizes

Source: IT Architect, Fortune 500 Transportation Services Company
TVID: C2E-F5C-5E7

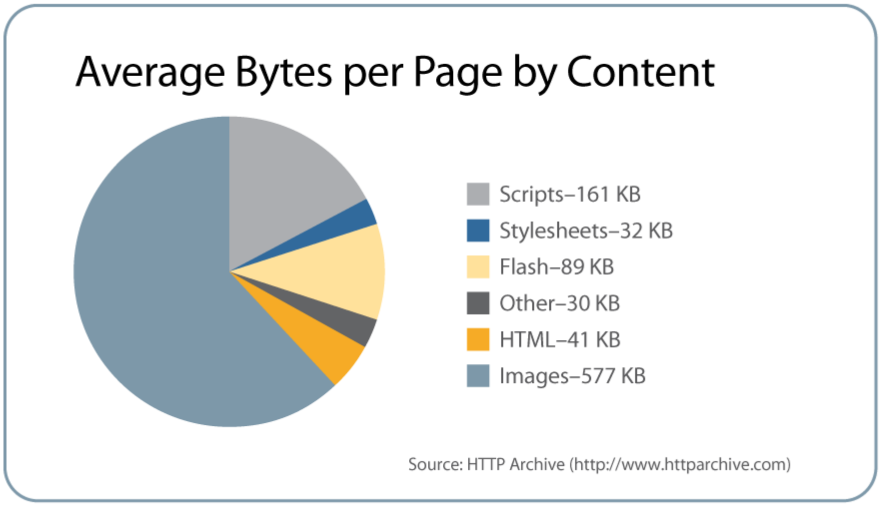

Figure 1: Images account for the majority of content, by type, in average web page composition

Images, a large part of modern web applications, must be downloaded to a client before display. The issues with this process in a mobile environment are image size and quality. Images originally developed for the desktop are very large and of unnecessarily high quality for display on the typical small mobile screen. Again mobile access involves the transfer of data that is effectively not used, since the mobile client must resize the image to fit onto the smaller screen. It is better to reduce the image as it is sent, reducing the total number of bytes that must be communicated and increasing the perceived performance of the application by reducing dependence upon network speeds. Automation of this process assumes there is a way to determine properties of the client, of course, or there would be no way to know how far to reduce the image.

Too Many Round Trips

When web applications are served to clients, the browser creates a base connection and then asks separately for each individual piece of data from the base HTML—or, at most, requests a few objects at a time. Notably, CSS and images are requested individually. The result is a tumultuous back-and-forth process that works, but has more protocol overhead than is strictly necessary.

Google has identified this problem and developed the SPDY protocol to reduce—sometimes drastically—the time required to render a web page in a browser. The only problem is that both sides must support the SPDY protocol or the connection defaults to HTTP, and while the penetration of SPDY on the client is rapid, uptake on the server side is slower due in large part to required changes to web server configurations and the maintenance of those servers over time. Most operational staff groups avoid changing production servers until they must.

Since Android devices based upon Honeycomb, Firefox, Opera for desktop, Opera Mobile, and Chrome all support SPDY and, combined, comprise a large portion of the market, exploring SPDY support is worthwhile. To avoid changing production servers and maintaining those changes regularly on many servers, IT departments need a form of gateway that can turn existing HTML applications into SPDY applications without the requiring changes to servers.

Faster and More Secure

One problem with mobile connections is the distance over the public Internet that data must travel. Security, reliability, and performance across the Internet portion of the connection can be questionable. Knowing who is connecting, where they're connecting from, and what client they are using helps with security issues, and encrypting data over a secure tunnel relieves concerns. Geographic proximity speeds connections, and compression reduces the size of the data on the wire.

For mobile customers, geographic location moves. This means that even if users are generally local, they will need to connect from remote locations at times. A system for determining which connection point is closest, as well as if the user is valid, improves security and performance more generally than if the location is assumed based upon some external factor.

Such a system, combined with a tool set that allows an enterprise to set up multiple locations quickly and easily, vastly improves performance speeds by bringing access to the data center closer to the user. When web application acceleration controls how much data is sent and how it is formatted—from scaling images sent to the device to stripping out comments and whitespace from text-based objects—performance gets another boost.

The F5 ADC System: An End-to-End Solution

Slow loading times and the challenges of mobile acceleration can mean lost revenue for the business, but there is a solution, completely within the IT sphere of control, that addresses all of these problems and more. That solution, based upon the award-winning F5 Application Delivery Controller (ADC), brings enterprises the tools needed to speed mobile applications. An F5 infrastructure also brings benefits to the network and applications above and beyond those related to mobile acceleration.

TCP Optimizations

Any F5 BIG-IP ADC product can perform TCP optimizations to reduce the amount of protocol-induced overhead by reducing the number of administrative packets sent back and forth. In high-latency situations like a WWAN, TCP optimization can have a drastic impact on the performance of web-based applications. TCP optimizations are a core part of the F5 TMOS operating system at the heart of all F5 ADC products. These optimizations are easy to activate and configure, reducing the chances of operator error.

Data Transfer Optimizations

The amount of data being sent over a given connection can be reduced in a number of ways.

All F5 ADC products support NTLM connection pooling, a technology that reduces the number of roundtrips spent authenticating to Microsoft applications. This reduces setup and teardown times for connections to the server, thus improving back-end performance.

F5 BIG-IP Application Acceleration Manager (AAM), which takes advantage of the level of compression supported by the connecting browser, allows the page to transfer less data to the client, so low bandwidth connections such as cellular networks have a smaller impact on page load times.

With the image reduction provided by BIG-IP AAM, not only is the file size of the image being transferred reduced by 20 to 40 percent, but unused image header information—such as the location where a photograph was taken—is stripped out. The goal is to reduce the amount of data transferred, but an added benefit is increased security about image source. Since these optimizations are configurable based upon the connecting client, they can be customized and modified to best serve the organization.

BIG-IP AAM also supports content reordering, which is useful for downloading style sheets to progressive rendering clients first while holding objects that often are not downloaded in parallel—such as scripts—until the rest of the page has been downloaded. Simple content reordering shortens the time from the user requesting the page to the browser rendering it, and even though it does not change total page load time, it does improve the user experience by showing them the page long before the original content would have been downloaded and displayed.

SPDY Support

Most organizations would gladly support the SPDY protocol if it were simple to deliver while still supporting native HTTP for those browsers that do not yet understand SPDY. Today, though, the cost of making all of an organization's applications compatible with SPDY is high. Each web server that will support SPDY must be modified to understand the protocol and determine whether a given client can support SPDY. In addition, every device—firewalls, intrusion detection systems, anything that sits between the client and the server and needs access to the packets—must support SPDY. F5 BIG-IP products solve these problems by providing a centralized location on the network that can support SPDY and translate between SPDY and HTTP when necessary. When a client that supports SPDY connects and the back-end application server has not been modified to recognize SPDY, the BIG-IP platform can accept the SPDY protocol and translate between SPDY and HTTP. The net effect is a SPDY gateway that allows clients to determine which protocols are used, while saving IT staff the time and risk of modifying each and every web server. Since the slowest part of the communication—over the public Internet—utilizes SPDY, the performance benefits SPDY provides are achieved while servers remain stable.

In short, SPDY is an HTTP improvement that allows multiple connections over a single stream and compresses the streams. This reduces the number of connections from a client to a server and improves overall page load times significantly. An F5 ADC can act as a SPDY proxy, translating between SPDY and HTTP so that each of the application servers and devices in the network do not need upgrades to support the SPDY protocol.

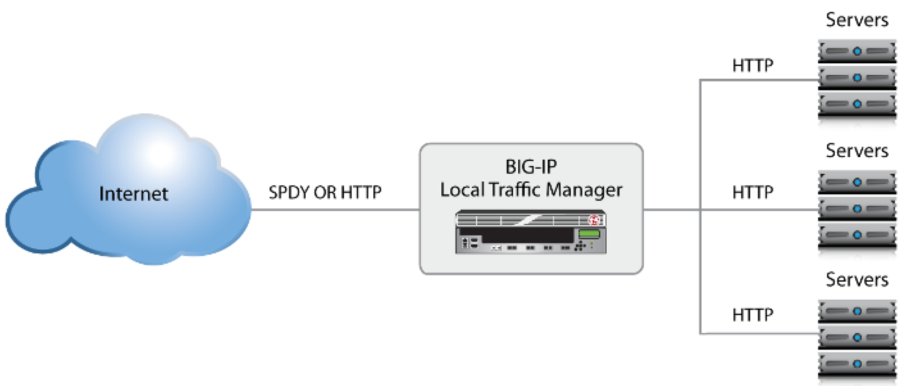

Figure 2: SPDY translation with F5 products

Conclusion

Wireless communications involve both latency and overall performance challenges. While the speed of wireless communications is growing with improved technology, the rollout of a new wireless technology takes an extended time, and wireless clients are being added at a faster pace than the implementation of bandwidth improvements.

In this environment, the application delivery optimizations delivered by an F5 architecture make wireless web applications secure, fast, and available. Advanced features like SPDY translation and image optimizations give IT departments control over the optimizations applied to each application, while F5 ADC features such as F5 iApps Templates, which allow for rapid and portable configuration, help ensure easy and error-free deployment.

---

[1] http://www.ctia.org/media/industry_info/index.cfm/AID/10323

[2] http://blog.nielsen.com/nielsenwire/press/nielsen-fact-sheet-2010.pdf

[3] For Impatient Web Users, an Eye Blink Is Just Too Long to Wait

[4] Wireless Performance Issues and Solutions for Mobile Users, Gartner Analyst Eric Siegel, January 20, 2012

F5 Networks, Inc.
401 Elliott Avenue West, Seattle, WA 98119
888-882-4447 www.f5.com

Americas
info@f5.com

Asia-Pacific
apacinfo@f5.com

Europe/Middle-East/Africa
emeainfo@f5.com

Japan
f5j-info@f5.com