White Paper

An Intelligent Services Framework

Applications running across networks encounter a wide range of performance, security, and availability challenges that can cost organizations an enormous amount in lost productivity, missed opportunities, and damage to reputations. The F5 intelligent services framework meets the challenges with granular control, scalability, and flexibility.

by Nathan Pearce
Senior Technical Marketing Manager

Contents

Introduction

The requirements for delivering employee and customer applications have never been more similar. Largely due to mobile working, the BYOD (bring your own device) movement, and the webification of the data center, organizations are forced to deliver internal employee applications to meet demands and expectations similar to those for external, customer-facing applications. With a broad range of devices to support and varying connectivity profiles, IT must deliver secure access from anywhere and on any device, quickly and with 24-hour availability, always.

The landscape for enterprises hosting customer-facing applications and services has changed, too. Competition has raised consumer expectations, and meeting those expectations demands improvement over generic, one-size-fits-all architectures and programming methodologies.

Competition, internally between IT departments and Software-as-a-Service (SaaS) providers, and externally, between competing organizations, is driving a review of not just what applications are being delivered but, more importantly, how. But the disconnect between data center networking and the ways in which applications and services are consumed is longstanding. Customers have always had the right to choose and will exert that right by taking their business elsewhere when expectations are not met. What's newer is that this right to choose now also extends to employees. The term "consumer" must now be redefined to include colleagues.

In a time of high expectations, where consumers' desire for instant gratification reigns, organizations can ill afford to allow application errors, performance related issues, and security shortfalls, however severe or temporary, to impact service delivery. Success today requires an infrastructure designed to approach application delivery from a service-to-consumer perspective—one highlighting the expectations of those consuming services. An intelligent services framework delivers on those expectations and consumer demands, whether the consumer is a customer on a mobile device in another country or a colleague working remotely via a personal smartphone or tablet.

Consumer Expectations

For many organizations, the devices used by employees to consume services, and the sites from where those services are accessed, are no longer strictly owned or controlled by the IT department. Business units have their own budgets and are able to shop around for the best provider. Furthermore, with large-scale, niche SaaS applications, implementation times are often faster than with internally provisioned and maintained solutions. The same is true of feature development and innovation, increasing the competitive pressure on IT departments to deliver on performance, availability, and functionality across a fast-changing landscape.

For customer applications, consumer choice has forced developers to progress from restrictive support of specific browser clients and operating systems, instead focusing more on open, standards-based protocols such as HTML5 over Shockwave Flash or Microsoft Silverlight—plug-in technology that's only available to some devices. This broadening of browser and operating system support has been accompanied by a growing demand for broader device support. With smartphones and tablets come mobile networks and specific screen real-estate requirements, forcing application architects to consider context in how services are consumed. That means abandoning "common" design.

When evaluating services, consumers are typically driven by three elements:

- Performance
- Availability
- Security

It is the responsibility of application, data center, and network architects to consider the context of new mobile network and device requirements against these elements. Access to increased bandwidth will continue to grow, but consumer expectations about the user experience have already outpaced throughput availability. According to Harry Shum, a computer scientist and speed specialist at Microsoft, "Two hundred fifty milliseconds, either slower or faster, is close to the magic number now for competitive advantage on the web."[1] That's one-quarter of a second: less than the blink of an eye. A delay of 250 milliseconds is easily incurred by poor mobile coverage. Conversely, 250 milliseconds readily can be shed from a transaction by optimizing both the application and communication. A high-resolution image is not necessary on the small screen of a smartphone. Consider also the amplified effect of network latency as it applies to larger data sizes.

1 "For Impatient Web Users, an Eye Blink Is Just Too Long to Wait," *The New York Times*, March 1, 2012

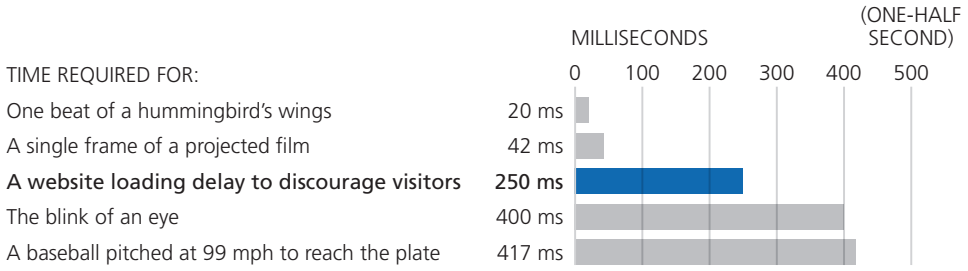

Figure 1: Consumer expectations of application delivery speed

The Internet has brought an abundance of choice regarding where and how services can be consumed. That choice has made today's consumers less forgiving and more unlikely to return to a service after a failure. The simplicity of choosing a new service greatly reduces the customer loyalty that once existed.

In addition to the demand for performance is the necessity of resilience—the availability of a service. The cost of application downtime can be likened to that of office closure: as with productivity for employees, sales via customer-facing applications cease immediately if the application cannot be accessed. The simple fact that consumers cannot be seen turning away at the door can lull organizations into a false—and dangerous—sense of security.

Yet efforts to improve the performance and availability of a service cannot come at the risk of security. Only good security enables consumers to trust organizations with personal information, including identity, financial details, medical records, etc. And trust, as many know, takes a lifetime to build but only moments to destroy. The brand damage from data theft is simply too severe to ignore.

A Consumer-Focused Approach

An appreciation of context is the key to handling the many different requirements of delivering applications today. This context includes how greatly communication over mobile networks differs from that of fixed-line connectivity and thus how accessible the application is via a smartphone's touch screen in comparison with a high-resolution monitor and a mouse. Add to these considerations the vast differences in security vulnerabilities across a growing range of consumer platforms.

With virtualization came a great reduction in data center silos and the practice of deploying isolated infrastructure for each service or application. Virtualization

inspired shared infrastructure across tiers of functions: shared server farms, switches, and firewalls, for example. But there's still plenty of room for more shared intelligence across these tiers.

Application Delivery Networking (ADN) concerns itself with the consumer expectations of applications and the relationships between performance, availability, and security.

Fast

Anyone who's been delayed by traffic understands that the time taken by a journey from A to B is largely defined by the slowest part of that journey. Network congestion is equally as detrimental to consuming data and services over the Internet. But access to Internet-based resources often spans a number of networks, increasing the importance of evaluating performance from end-to-end—from the data all the way to the consumer device, wherever it may be—not merely from server to service provider. And mobile device proliferation demands that the focus on service performance be altered to encompass the efficiency of the last mile.

Application performance management is the translation of IT metrics into business meaning and value. Performance is often measured in two separate areas: application performance and connectivity (or network) performance.

Application performance

Application performance is typically measured by metrics that include:

- Resource utilization (CPU, RAM).
- Responsiveness.

As metrics, CPU and RAM utilization are not as effective as they once were. Modern applications consume as much RAM as possible for storing information, given how fast that information can be accessed over the disk or, worse, a call to another server. The validity of this metric is further diluted by virtualization. Hypervisors deliver a layer of abstraction between the operating system and physical hardware by means of queuing, and queuing of access to the pooled resources is fine until the hypervisor itself experiences congestion. Granted, leading hypervisors can be configured with virtual machine prioritization, but this is cumbersome to manage and not automated sufficiently to prevent the undoing of the consolidation gained from virtualization.

Responsiveness is a more intelligent means of measuring application performance. Typically this metric is the result of a bot (scripted behavior) that makes requests of the application resources, with those requests involving multiple elements: access, authentication, and database queries. Thus the metric conclusively tests the inter-server communications involved to provide a meaningful value: a time delta from request to response.

Network performance

Network performance can be broken into two segments: the data center network performance and the service provider's access performance.

Measurement of the data center network performance validates the core switching, routing, and firewalls of the data center itself. More recently, with virtualization, this could include software-based switching and routing, such as VMware's virtual switch. While measurement of these devices, virtual or physical, is important for identifying bottlenecks or periodic congestion, it does not measure the user experience, largely because of the part played by mobility and BYOD.

Monitoring of the service provider networks that connect the data center to consumers is important. However, performance degradation is equally likely, if not more likely, at the consumer end of the session. In the days of fixed-line communication, this part of the session, taking place over the final network segment connecting to the user, was referred to as "the last mile."

Taking an end-to-end user-experience approach to performance—one that encompasses everything from the data center to the last mile—requires an appreciation of the context of all elements in the communication, from data to consumer.

Available

To the consumer, how and why a service is unavailable are of little interest. They care only that the service is delivered, so availability should be the primary focus in designing a highly resilient application. Disaster recovery is what one initiates when things have gone bad. A strong disaster avoidance solution, one that makes real-time decisions that affect the user experience, is what reduces the risks to staff productivity and customer sales.

Server/application availability

Merely checking that a server is powered on provides no assurance about the user experience. To measure application availability, IT must simulate the user experience, requesting of the application all that is needed for live production use. For example, testing access to the login page of a web application only checks that it is responding to connections. It does not validate the availability of the authentication service or of the application or database accessed post-authentication. The alternative assessment of availability, a far more thorough solution, is a synthetic transaction that exercises all elements of the application.

Data center availability

The measurement of data center availability concerns both application availability and data center performance. In this regard, many organizations are already realizing the benefits of active/active data centers and hybrid private/public cloud deployments, which include the ability to distribute workloads between the two, thus enhancing availability overall. An infrastructure designed to be both dynamic and highly available ensures agility and resilience.

Secure

It is paramount that security provisions and measurements have a data-to-consumer focus. Only this approach will mitigate against the growing trend toward diverse, distributed, denial of service (3DoS) attacks. There are many elements in data-to-consumer security, however, and good security requires them all to work together in harmony.

Security starts with context: understanding the location, device, device integrity, and identity involved. An appreciation of delivery location empowers businesses to control from where sensitive and confidential information can be accessed—an issue of growing importance with the increasingly mobile-enabled workforce. Device identity and integrity open up options for organizational confidence in and management of how applications and services are consumed. If the source device is unknown, then a virtual desktop infrastructure (VDI) alternative can be provided, protecting the service from direct access to a potentially compromised machine. Only when these context-related security requirements have been met should IT be concerned with identity and authentication challenges.

An Intelligent Services Framework

Enterprise IT teams are concerned with maintaining control over applications as they move in and out of the data center, making sure those applications always remain fast, available, and secure. Although all three are equally important, since they all impact the applications and the infrastructure, each can require unique solutions. The integrated F5 intelligent services framework delivers a single platform that nonetheless enables and optimizes unique solutions targeted toward the specific application needs of the enterprise.

The F5 intelligent services framework is a set of extensible and programmable ADN capabilities based on F5® BIG-IP® Application Delivery Controllers (ADCs). To address the application delivery challenges affecting enterprises today, a new, more sophisticated model needs to replace simple ADCs. An intelligent services framework is the key. Only a completely integrated intelligent services framework from F5 provides discreet services throughout the Application Delivery Network—including mobile optimization and application access management, total DNS delivery services, and an application delivery firewall—in one unified platform. An F5 intelligent services framework manages users from any location and on any device, applies application delivery policies to all application requests in both directions, and connects those users to applications regardless of where the applications live.

Strategic, Multi-Faceted Intelligence

In the new world of on- and off-premises applications accessed by users with multiple devices, there is no way to manage mobile device and application growth simultaneously without an intelligent services framework acting as a full proxy. As a broker between the users and applications, this F5 solution provides:

- **Application awareness:** Total insight into how the application is supposed to look on the wire.
- **User awareness:** The ability to see which users are trying to access what application, from which devices.
- **Resource awareness:** The intelligence to tie all the pieces of the application delivery infrastructure together to provide real-time visibility into the entire Application Delivery Network.

Application, user, and resource awareness come together to provide strategic awareness: application delivery awareness at a strategic point of control for

applications that are both on and off the premises. This strategic awareness is made possible by the completely integrated platform supporting F5's powerful solutions across the Application Delivery Network—a platform designed from the ground up solely to make applications fast, available, and secure.

Application-Centric Services

The F5 intelligent services framework starts with hardware and software designed to deliver applications. The application-fluent hardware and software components work together to provide complete control and can readily be scaled up in the data center and out to off-premises locations around the world or in the cloud.

Many point products on the market try to solve a singular problem at one location in the data center. But point products aren't capable of addressing all application delivery services at once, nor do they have deep visibility or an appreciation of context across the entire infrastructure. The unified services of the F5 intelligent services framework deliver the insight and control to manage application delivery comprehensively, from the app all the way to the user.

Finally, although the F5 intelligent services framework can readily manage application delivery in thousands of scenarios, every enterprise environment is unique. F5 devices provide unmatched programmability and can be configured to match the individual challenges faced by any organization.

Conclusion

Changes in access devices and connectivity profiles are shifting the technology landscape. This change is being reflected across both enterprise and customer-facing applications, forcing IT departments into a service provider model in which they are often being held to service-level agreements by internal business units. Increasingly, those business units have other choices, including SaaS choices, if the IT department can't deliver.

In addition to the increased pressures of competition, consumers are more informed, have access to all the information they need to find the best solution, and are becoming increasingly aware of their ability to choose. With time being the most precious commodity, the failure of any application or service to meet consumers' performance, availability, and security expectations will drive these consumers to exercise their right to choose an alternative path—a path of least resistance.

Fortunately, consumers' performance and availability expectations can be met—without compromising on security—by architecting a consumer-focused data center network. This feat is made possible only by injecting programmable intelligence into the application delivery path.

The F5 intelligent services framework provides that programmable intelligence, bringing context to every aspect of application delivery in what has been previously an unintelligent architecture. That context is the key to addressing device- and connectivity-specific challenges to provide a faster, more secure, and more reliable user experience.

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

F5 Networks, Inc.
Corporate Headquarters
info@f5.com

F5 Networks
Asia-Pacific
apacinfo@f5.com

F5 Networks Ltd.
Europe/Middle-East/Africa
emeainfo@f5.com

F5 Networks
Japan K.K.
f5j-info@f5.com