APPLICATION DELIVERY CONTROLLER

Optimize and secure Web applications to ensure successful migration to the Cloud

White Paper

by Joël Levée, CTO, ActivNetworks

ActivNetworks SAS
17 rue Dumont d'Urville
75116 Paris
France

Mail to:
1 rue de Terre Neuve
Mini Parc du Verger – BP 127
91944 Les Ulis Cedex
France

Phone: +33 (0)1 64 86 44 44
Fax: +33 (0)1 60 92 01 42
contact@activnetworks.com
www.activnetworks.com
www.boostedge.com

Table of Content

1. Opening up applications for Web technologies

Enterprise applications today are more and more often accessible via a web browser. This facilitates their use by employees and customers alike who can access remotely and even when on the move. This new approach, if it brings many benefits for businesses and for users, also introduces new constraints:

- Data rates are not always sufficient on the internet,
- Controls exercised by the enterprise stop at the boundary of its LAN,
- Application do not generate optimized data,
- Opening the network means putting security at risk.

Providing employees and customers with a web-based application that they can access all the times, from their office at corporate headquarters, from a remote site or even when on the move, produces immediate and verifiable benefits for the company. For one thing, productivity of employees increases through the sharing of information. Missions are carried out using the latest available data. Information relating to orders in progress or to projects involving partners and suppliers, is accessible instantaneously. Employees, partners and customers can be connected at absolutely any time, 24/7, to gather information, to buy and to work at the most convenient times for them.

So the advantages of a web application are manifold, providing the web application is actually available under the conditions of ‘fluidity’ and at the speeds that users demand now. If information does not arrive within a few seconds, employees will be quickly frustrated. If a picture or photo is missing from a web page, the potential buyer will likely surf to another site. It is not enough for the web application to present well and to be useful, it must be fast and fluid too.

The fact is, as perfect as it may be, the application remains dependent on the communication channels that connect it to users, whether these be employees or customers (ADSL, WiFi, 3G, satellite links etc.) –i.e. dependent, basically, on the Internet. Now, whereas the company can act on its own in-house facilities, it has no control over the Internet. The only solution is therefore to work upstream, before the data passes beyond its scope of action.

1.1. Take account of the data rates available

The web application is hosted inside the company and is supported by its IT architecture, in particular its local area network or LAN, which offers users very fast data rates.

The LAN has a very high data carrying capacity, of the order of 100 million to 1 billion characters per second. This extremely high capacity that prevails within the company is subject to an immediate and major fall-off when we move through the Internet where typical values will be divided by at least 10 to 100. And when the user is connected via satellite or via a mobile network (with a smartphone, tablet or 3G PC, for example), the transport capacities are even lower, at best 300 times less than those available on the corporate network.

The business must take this important constraint into account: it is not technically possible to make as much data transit outside the company as inside, at least not in the same time frame.

1.2. Internet, a long and complex road

The difficulty related to the data rates of the different networks is compounded by that of the distance that information must travel to get to the end user. While the company's LAN is geographically limited to the confines of its own offices, the Internet covers the entire planet and there are many segments for the data to travel before it reaches the awaiting user terminal. To direct packets of information by the most efficient route, specific equipment are located at each intersection. Their function is to calculate the best route, and this takes a little time during which data transfer is interrupted. These delays, known as "latency", accumulate and add to the total travel time.

Hence the travel time of a packet through the network, is the sum of the physical propagation times and the switching times. Of course, in the case of a satellite link, the roundtrip time required for information to reach the satellite and return to the receiving antenna is even greater and is that much more detrimental to overall transmission times.

In the same vein, when a company makes a web-based application available, it is likely to be intended for new additional users who did not previously have access to the application. Logically, this will result in additional traffic and a multiplication in queries addressing the application. This may cause the hardware to saturate. Solutions for load sharing (load balancing) across multiple machines are available on the market, mainly targeted at businesses equipped with multiple servers.

Distances and processing times are incompressible and the company itself cannot hope to reduce them. This new constraint inherent in the Internet and in the very concept of a web application open to the outside world, must also be taken into account.

1.3. The data is not optimized

In addition to all these issues of data flow rates (network capacity), latency (traveling time) and server load, there is the issue of actual volumes of data. Insofar as an application was hitherto only used on the local corporate network, and was therefore not subject to the difficulties already mentioned, it could afford to be "verbose".

It wasn't particularly penalizing if there was more data than necessary, so images and graphics could afford to be burdensome, with lots of colors for instance. Layout and content of documents didn't need to be optimized, because the company's local network compensated with its very high speeds. Also, the terminals used, the workstations installed inside the enterprise, were powerful enough and equipped with software capable of interpreting all the content produced.

But when data is delivered over the Internet, and with the limits and constraints set out above, the overall volume of information sent and the terminal's capacity to interpret that data, become important criteria. Because these are essentially the only parameters over which the company has any real control.

Since it isn't possible to control data flow rates or reduce transmission leadtimes, and without having control over the Internet or remote user terminals, the only recourse is to work on reducing the volumes of data, taking into account the different capacities of the browsers on the market, all to some degree capable of decompressing the data received.

1.4. Opening up to the Web... and to potential attacks!

We have to be aware that opening up our applications to the web and making them accessible via the Internet means taking our data and software outside of a secure environment. With "conventional" applications, sheltered behind the company's firewall, small failings (minor functional defects, coexistence of standard data and sensitive information, failure to apply updates or patches, etc.) could be overlooked without posing serious risks. In this context, by and large, only well-intentioned users can access the application and manipulate the data. But once it becomes accessible from any terminal connected to the Internet, the application is immediately a target for hackers, who have long since perfected techniques for running automated searches for new targets and pinpointing vulnerabilities.

Clearly, migrating an application to a web model and offering web access to remote users/customer is not as simple as rolling out a web server or extending access to an existing application. There are many instances of negative impact on the smooth running of companies and the risks should not be ignored or underestimated. To address these difficulties, the company can only act on parameters within its scope of action, i.e. its own network infrastructure.

2. The Application Delivery Controller

In recent years, network technologies have evolved to provide coherent solutions for companies wishing to deploy a web application while countering the difficulties set out above. A number of manufacturers now offer a device known as an Application Delivery Controller (or ADC) whose goal is to optimize the connection of web applications onto the networks in order to enhance performance and in some cases the security of applications.

The ADC is therefore a critical component in the application architecture as it has a direct impact on user satisfaction and the effectiveness of the company's web strategy. Before making a choice, it is essential to study the different offers on the market to identify the one that provides concrete and effective solutions to the problems cited, without disturbing the architecture or running of existing systems, without calling for numerous and therefore expensive interventions or being burdensome for the business' employees and customers.

2.1. Accelerate applications without changing terminals

It shouldn't be forgotten that web applications are rarely designed at the outset to maximize speed of access. In some cases, they have been ported from non-web technologies. Often the code hasn't been optimized and runs slowly. Applications sometimes generate content that is overly verbose (e.g. PDF files), heavy (graphics, etc.) or simply not suitable for the new user terminals (files that can't be read on a smartphone, for instance). These shortcomings are particularly problematic when users are no longer on a company network but on the Internet.

To reduce the amount of traffic and thereby speed up response times, many equipment manufacturers offer compression/decompression devices installed symmetrically at the point of entry to each of the company's LANs. But whereas this solution can accelerate data transfer between company locations, it is not suitable for web applications. By definition, a web application has to be supported for use from anywhere on the Internet, in other words from connection points that are not equipped with special decompression equipment.

An ADC will therefore need to implement standard mechanisms and innovative methods to achieve similar performance gains without having to install proprietary equipment at the point where users connect to the internet, and without requiring them to reconfigure their workstations or install special software. The ADC has to accelerate web applications for all users.

2.2. Compress the data

Given that it’s not possible to reduce transmission times, and in order not to have to modify the application, the simplest solution to speed response times for applications is to compress the data using standard algorithms such as gzip. Indeed, almost all currently available web browsers natively support display of compressed files in this way.

However, users will connect to the web application with a range of terminals running different browsers with varying decompression capabilities. The compression device will therefore need to be able to identify the user’s terminal and adapt to the actual capabilities of the browser.

Depending on browser and build, decompression capabilities differ. For instance, sending a compressed JavaScript to a browser that cannot interpret a compressed script may make the application completely unusable. In addition, some versions of browsers have bugs that prevent them from handling compressed data properly.

Compressing data using standard algorithms therefore requires real intelligence and substantial processing capacity. Some companies may choose to enable the compression function on their web server despite the risk of not differentiating processing per user, and in particular running the risk of penalizing the performance of the server itself. The ideal option is actually to have the Application Delivery Controller perform this function, if it has the ability to do so.

For this, the ADC will need a real-time knowledge base, with a record of all browsers and their exact (and not supposed) capabilities. An update feature attached to this database will ensure that the latest developments of all browsers are known. The updates must of course be driven remotely to avoid the need for calling out technicians.

2.3. Adapt the data to each user

Beyond the decompression capabilities, it is essential the Application Delivery Controller be able to detect the user's communication media and equipment type so that it can adapt the content accordingly. In this way the amount of data transmitted will not exceed the transmission capacity of the media to which the user is connected.

For example, there is no point passing on information that cannot be displayed (due to limitations of the available display screen) or cannot be properly exploited (because the target terminal does not run the necessary software). The ADC must take into account the specific technical features of each of the terminals in order to transmit only targeted and usable content. This means that some of this information has to be adapted on the fly.

Similarly, not only will the ADC automatically adapt the content based on the user's terminal (browser version, etc.), but also according to the access network and the location from which the user connects. The administrator may for instance decide to greatly reduce the size of the content for users connected in 3G in order to privilege browsing speed, or not to reduce the size of content for users connecting from a location near a fast network, but to reduce the size for those connecting from another continent (where networks are considered slower), etc.

The ADC will have to offer flexible configuration to allow the administrator to easily build different rules for each application, each type of file, device, user, access network, etc. Ideally, these optimization rules should be configurable via a user friendly web interface, but should also be scriptable for automated operation of the ADC depending on the rules of operation of the business.

2.4. Optimize non-compressible files

Beyond simple compression, the ADC will also need to be capable of reducing the size of non-compressible files such as images (JPG, PNG, etc.), PDFs and video.

These non-compressible files are increasingly prevalent on websites and represent much greater data volumes than simple text files. The capability to reduce their size will have an immediate and marked impact on display speeds.

PDF files are sometimes generated on the fly by web applications without it being possible to act on the optimization settings. Images, PDFs and videos are often pre-existing objects and it would be too resource-consuming to recreate them one by one as optimized files. Having an ADC that can handle optimization of these non-compressible files automatically and intelligently (according to the characteristics of each user) provides both real flexibility and major operating efficiencies.

Optimizing non-compressible files requires rework of their content, and thus reordering and optimizing of the information they contain. This requires the ADC to have the power to do so in real-time and above all the intelligence to do so without visibly degrading the quality. This may involve, for example, restricting the color palette of an image, removing redundant components in a PDF, reducing the bitrate of a video but not deleting frames, etc.

2.5. Secure data and servers

Opening applications up for web-based access poses security problems for both servers that can be the victim of attacks, and data that can be intercepted during transmission.

Because any Internet user can connect to the web application, a hacker can simply send malicious queries to break into the databases, execute their own code, saturate the bandwidth and cause servers to crash or prevent applications from running. The economic consequences can be dramatic: theft of data, inability for customers to place orders, information lost or compromised, etc.

Detecting these attacks and countering them requires an understanding of the requests issued by the user and in particular of the content of exchanges with the server. As the Application Delivery Controller's mission is to analyze each request to the server in order to optimize it and speed up traffic, it is ideally placed to filter the queries. The ADC will therefore need to offer an application firewall (also known as a WAF or Web Application Firewall) preconfigured to block attacks.

It is also advisable to encrypt the data to prevent it from being exploited by someone having intercepting it during transfer to the user. The technology universally used for this encryption is SSL. Some companies will choose to activate SSL on their servers or to completely reconfigure their applications. It is far preferable however to delegate this role to the ADC.

SSL is in fact relatively demanding in terms of computing power and can penalize the performance of the server. In addition, SSL encryption generates exponential flow volumes, so it is clearly preferable to optimize the data and reduce its size before encrypting. In this way, even if the web user accesses the application from a slow network (such as a smartphone on a 3G network), the data can be encrypted and the improved response times will offer the user a more comfortable browsing experience.

Moreover, encrypting the data streams in the ADC at the output of the network will ensure that all data is encrypted even if it originates from several different servers. The ADC will also need to be able to communicate with servers in encrypted mode in the event that a server is remote or simply if a server has been configured in this way and there is no intention of changing its configuration. The ADC must therefore have the power necessary to decrypt the data sent by servers, optimize then re-encrypt to send to the user and speed up time-to-display for each web page.

2.6. Simple, seamless and invisible for security

Ideally, it should be possible to insert and implement the chosen solution in the datacenter without reworking network architecture or modifying applications. It must therefore be as simple as possible to connect up with a minimum of procedure. With current web technology supporting brief interruptions in communications, connection time for the Application Delivery Controller shall lie within that interrupt constraint and be limited to plugging in a network bypass.

Of course, the ADC must not disrupt traffic until its logic is operational and it must require only minimal configuration. The implementation of the ADC must be seamless and invisible for the local network, for the application and for users who should simply notice an improvement in performance.

This invisibility on the network must be exploited to enhance the security of the application and of the company's LAN. As we have seen, the embedded intelligence in the ADC should allow it to intercept and reject attacks and suspicious queries. If the presence of the ADC is undetectable due to its mode of insertion in the network, hackers will not be able to detect that their requests are filtered, will not be able to reach the ADC to attack it and will not be able to bypass it to reach the servers themselves.

The Application Delivery Controller is therefore an essential component for the company, placed at the junction between its own local network and the Internet. Installed quickly, without disruption, it will enhance fluidity of traffic flow by optimizing and compressing data before sending it to colleagues or customers. Its ability to intercept incoming data and user queries will fulfill a security role and also enable information previously requested and stored in its memory to be delivered faster, without loading the web application and without waiting time for users.

Its embedded intelligence ensures that information will be optimized to match both the capabilities of the network that the user is connected to and the technical capabilities of the hardware and software that he/she is using. Moreover, it must be built on a standard platform that meets market standards and incorporates the latest technological advances. This last point offers the company the guarantee that its investments are optimal and that the equipment can be upgraded as and when necessary.

2.7. In conclusion...

The first Application Delivery Controllers and Load Balancers emerged to address the need for increased traffic with web applications. Today a new generation offering greater performance and modularity is available under the name Advanced Platforms (AP-ADC).

Few in number, these solutions provide -at one and the same time- multiple functions including acceleration, fluidity, security for applications and data and, of course, load balancing.

Alongside historical players such as F5 or Citrix whose solutions are based on proprietary hardware, ActivNetworks technologies are totally hardware independent. This allows the company to exploit the latest advances in technology in order to deliver more power at the lowest cost, and also importantly to offer its technology in virtualized architectures with the same level of functionality.

With its range of BoostEdge ADCs, ActivNetworks is the only provider of complete solutions that are quick and easy to deploy, offering seamless rollout (patented technology) over a network in operation. This technological advance is also reflected in its unique ability to compress and optimize PDF files, images and videos, to adjust the flow for each user, and to significantly reduce the effects of latency on remote connections.

3. Glossary

Hosting: The application can be installed on a machine that is not physically on the company's own premises but located with a specialized service provider.

LAN: Local Area Network, restricted to a limited geographic area, such as a building, and offering high data rates.

WAN: Wide Area Network, extending the company's network across a very large geographical area, often with a role of connecting multiple LANs together. The available speeds are significantly smaller.

AP-ADC (Advanced Platform Application Delivery Controller): ADC designed and built to be enriched with new software or hardware features.

Leadtime: Amount of time taken to do something, here for data to go from point A to point B.

Latency: Period of time during which a main operation is suspended while a secondary operation takes place.

Compression: Operation whose purpose is the reduction of volume. The compressed object must be decompressed to recover its intelligibility.

Optimization: Operation which aims to change something for the purpose of improving or easing its use.