# Performance vs Reliability

**Lori MacVittie, 2007-20-09**

Our own Deb Allen posted a great tech tip yesterday on conditional logic using HTTP::retry with iRules. This spawned a rather heated debate between Don and myself on the importance of performance versus reliability and application delivery, specifically with BIG-IP.

Performance is certainly one of the reasons for implementing an application delivery network with an application delivery controller like BIG-IP as its foundation. As an application server becomes burdened by increasing requests and concurrent users, it can slow down as it tries to balance connection management with execution of business logic with parsing data with executing queries against a database with making calls to external services. Application servers, I think, have got to be the hardest working software in IT land.

As capacity is reached, performance is certainly the first metric to slide down the tubes, closely followed by reliability. First the site is just slow, then it becomes unpredictable - one never knows if a response will be forthcoming or not. While slow is bad, unreliable is even worse.

At this point it's generally the case that a second server is brought in and an application delivery solution implemented in order to distribute the load and improve both performance and reliability. This is where F5 comes in, by the way.

One of the things we at F5 are really good at doing is ensuring reliability while maintaining performance. One of the ways we can do that is through the use of iRules and specifically the HTTP::retry call. Using this mechanism we can ensure that even if one of the application servers fails to respond or responds with the dreaded 500 Internal Server Error, we can retry the request to another server and obtain a response for the user.

As Deb points out (and Don pointed out offline vehemently as well) this mechanism [HTTP::retry] can lead to increased latency. In layman's terms, it slows down the response to the customer because we had to make another call to a second server to get a response on top of the first call. But assuming the second call successfully returns a valid response, we have at a minimum maintained reliability and served the user.

So the argument is: which is more important, reliability or speed?

Well, would you rather it take a bit longer to drive to your destination or run into the ditch *really fast*?

The answer is up to you, of course, and it depends entirely on your business and what it is you are trying to do. Optimally, of course, we would want the best of both worlds: we want fast ***and*** reliable. Using an iRule gives you the best of both worlds. In most cases, that second call isn't going to happen because it's likely BIG-IP has already determined the state of the servers and won't send a request to a server it knows is responding either poorly or not at all. That's why we implement advanced health checking features. But in the event that the server began to fail (in the short interval between health checks) and BIG-IP sent the request on before noticing the server was down, the iRule lets you ensure that the response can be resent to another server with minimal effort. Think of this particular iRule as a kind of disaster recovery plan - it will only be invoked if a specific sequence of events that rarely happens happens. It's reliability insurance. The mere existence of the iRule doesn't impact performance, and latency is only added if the retry is necessary, which should be rarely, if ever.

So the trade off between performance and reliability isn't quite as cut and dried as it sounds. You ***can*** have both, that's the point of including an application delivery controller as part of your infrastructure. Performance is not at all at odds with reliability, at least not in BIG-IP land, and should never be conflicting goals regardless of your chosen application delivery solution.

*Imbibing: Pink Lemonade*

Technorati tags: performance, MacVittie, F5, BIG-IP, iRule, reliability, application delivery

F5 Networks, Inc. | 401 Elliot Avenue West, Seattle, WA 98119 | 888-882-4447 | 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