CS290F - Paper Reviews for 2010.01.19

Author's note: for this first round of reviews, I have chosen to write a single in-depth review of one of the required papers, as I felt that I had quite a bit to say about it.

In-depth review for: F. Leighton and D. Lewin, "Global Hosting System," U.S. Patent No, 6,108,703, August 22, 2000

This "paper" is actually a patent filed with the USPTO in August, 2000, claiming to cover a network architecture to support global scale content distribution. The patent makes grandiose claims of an invention that will provide "unlimited cost effective scalability", with little effort from the content provider, at significantly decreased operations cost.

The authors of this patent, and the patent officer that granted this patent, should go read Fred Brooks' No Silver Bullet. The idea that you can get something that's all good and completely transform the status quo is one that should always be taken with a (large) grain of salt. In this case, the claims made of the invention in this patent are at best ridiculously overstated, and at worst completely unsustainable.

In summary, the invention described in the patent consists of three major components:
1. "Ghost" content-servers
2. A multi-tiered DNS system
3. A content re-writer service, function, or library

The "Ghost" content servers are globally spread out among existing ISPs, such that there are enough servers near any connected region to meet content demand. These servers are essentially dynamic mirror sites where content is pushed to or requested from when a request for content originates from a nearby region.

The location awareness portion of the invention is provided by the multi-tiered DNS system. This system ensures that requests for content indicated by specialized URLs is always delivered from a ghost content server located near the originating request.

Finally, the content re-writer is a service (or library, or function...) that is run jointly with the content provider's web services. When an embedded object is included in a page, the service re-writes the URL reference to that object to be prefixed with a well-known URL that directs the multi-tiered DNS service to direct requests to the ghost content servers near the originating request.

This patent embodies nearly everything wrong with the American patent system, at least with respect to software-based inventions. The patent is extremely verbose, without saying very much (though it is quick to point out wherever possible that the invention is "novel" and "inventive"). The patent is full of typos and grammatical awkwardness and mistakes (a handful of these were addressed later however). The figures included to illustrate the patent do very little to help the reader understand anything about the invention (in fact, figures 1 and 2 are completely useless with respect to the invention). A great deal of time is spent discussing the expected computer architectures of a client or server machine or the typical

format of an HTTP request for a web page, all without ever being referred to again in the invention.

But then, all of that is merely status quo for a U.S. patent. Even if we ignore all of those problems, there is still a great deal we can say about the claims made about the invention itself.

Regarding claims about related art: the patent does a decent job explaining why other content distribution mechanism are expensive -- the need for a content provider to host, co-locate, or lease server time is prohibitively expensive for large-scale deployments. However, the claim that these techniques cannot scale to more than a few ("i.e., less than 10") web sites is unsupported, and unwarranted. Further, the claim that existing content providers have problems with ISPs caching results, hindering or limiting dynamic content, is misguided: the HTTP 1.1 protocol, released in RFC 2616 in 1999, a year before this patent was granted (though, admittedly, after the patent was originally filed in 1998), explicitly defines caching behavior by intermediaries. Even ignoring the caching issue, however, there is no guarantee that this invention will not fall prey to the same problems, albeit at the regional level instead of the global level.

While summarizing the "specific advantages afforded by the inventive global hosting scheme", several claims are made that are untenable: First, the idea that content providers will have decreased operational expenses is based on the claim that global ISPs are responsible for replicating content and maintaining servers. That might work... if ISPs were charities that existed to give away service resources to companies. What incentive would an ISP have to implement such a scheme, even if they were in a globally-wide enough position to do so? Even if the ISP wanted to implement this scheme to reduce utilized bandwidth, there is still a real cost involved in hosting the servers and replicating the data that someone must pay for, and ISPs aren't likely to give away that cost for free.

The discussion about the advantage of "intelligent and efficient data replication" is similarly flawed. The data must still be sent to the ghost servers at some point. This requires less data transfer than traditional mirroring, but only if we can assume that a region never needs all of the content from a web site (which seems unlikely, unless a region only encompasses a very small). Furthermore, given the proposed saturation of ghost servers, it is likely that more data is required than mirroring in order to get enough content to any given region. Further more, unlike simple mirroring of data, in cases where data hasn't yet been pushed to a nearby ghost node, the data will take longer to be returned to the end client as the request is redirected to a server that has the content available.

As a final criticism, the claims of advantages of "compatibility with content provider software" are completely unsupported -- the patent enumerates a handful of problems in existing solutions (using specific servers or databases), but does nothing to explain how the content re-writing service is implemented. How do content providers access or utilize the service that re-writes content URLs in their pages? What about cases of dynamically generated pages? What about different operating systems? All of these questions are left completely unanswered by the patent.

Now, it may be the case that my personal view is biased by years of working with computers and specifically internet services. It is possible that in 2000 the invention described by this patent really was state-of-the-art (after all, I hadn't even graduated high school at the time). It is even possible that I have mis-understood the entire application of this invention, and that some or all of my criticisms are misguided. However, from what is laid out before me in the patent itself, I can't make a case to support any of these

possibilities. Instead, I'm left reading another patent that really shouldn't have been granted by the USPTO.