Re: http pipelining

Source: http://coding.derkeiler.com/Archive/Python/comp.lang.python/2007−04/msg04489.html

- *From*: jjl@xxxxxxxxx (John J. Lee)
- *Date*: Sun, 29 Apr 2007 23:38:18 GMT

Steve Holden <steve@xxxxxxxxxxxxx> writes:

> swq22@xxxxxxxxx wrote:
>
> > Which python module is capable of pipelining http requests?
> > (I know httplib can send mulitple requests per tcp connection, but in
> > a strictly serial way.)
>
> Oops, sorry, you meant sending requests in parallel, right?
>
> You'll need to use either urllib or urllib2 for the web, and the
> threading module is one way to run parallel requests. It's fairly easy
> to use as long as you keep your tasks properly isolated form each
> other.

No, he means "HTTP pipelining", which means sending multiple requests down a single TCP connection, without waiting for the first response.

httplib's module−level docstring says (reformatted here):

"""
.... The HTTPResponse class does not enforce this state machine, which implies sophisticated clients may accelerate the request/response pipeline. Caution should be taken, though: accelerating the states beyond the above pattern may imply knowledge of the server's connection−close behavior for certain requests. For example, it is impossible to tell whether the server will close the connection UNTIL the response headers have been read; this means that further requests cannot be placed into the pipeline until it is known that the server will NOT be closing the connection.
"""

So, sort−of−yes, if you know what you're doing and you're lucky.

Certainly urllib and urllib2 don't support pipelining. There were plans for a new HTTP client in Twisted with pipelining support, but I

don't know if that ever came about. AFAIK not many libraries (in any language) support it −− e.g. none of "Jakarta commons HTTPClient", libwww−perl, and libcurl currently support it. libwww (without the "−perl") does claim to support it (I say "claim" merely because I haven't used it or read the source −− no FUD intended).

Side note: As the OP mentions in a followup, by default firefox does NOT do pipelining (to the disbelief of the people I told about this when it came up in a previous job −− but I just tried running tcpdump and indeed about:config seems to be telling the truth; fortunately, in response to the limitations imposed by RFC 2616, somebody has thoughtfully arranged for the speed of light to be fast enough for this not to be a major problem when using websites on the other side of the planet). The firefox people say:

http://www.mozilla.org/support/firefox/tips#oth_pipelining

"""
Pipelining is an experimental feature, designed to improve page−load performance, that is unfortunately not well supported by some web servers and proxies.
"""

Instead of pipelining, it uses multiple connections (2 when I tried it, which is the maximum RFC 2616 says SHOULD be used by a "single−user client"). I didn't try IE, but apparently it has the same behaviour (2 connections, no pipelining):

http://blogs.msdn.com/ie/archive/2005/04/11/407189.aspx

I wonder if the right economic pressures are there for SCTP ever to get used for everyday web HTTP stuff...

John
.