Fair enough. Do we know roughly how many sessions off hand? If it takes digging, please don't, there is time to review later. My concern about sessions is, as always, related to proxy. If we set the limit to even 20 sessions in 5 minutes (a very low threshold I would guess), wouldn't these tools just stagger the session creation to adjust? Of course, it would also inadvertently catch our very larges pretty quickly I'd guess.

███

From: ███████
Sent: Saturday, September 25, 2010 10:55 PM
To: ███████
Cc: ███████
Subject: Re: website sad

We probably need a session creation rate limit to catch this one.

On Sep 25, 2010, at 10:08 PM, "███████" <███████@ithaka.org> wrote:

> The follow up question is why aren't Literatum blocking tools getting them at the 300 mark if they are indeed downloading PDFs. We aren't seeing anything on the blocking front.
>
> ███
>
> **From:** ███████
> **Sent:** Saturday, September 25, 2010 10:04 PM
> **To:** ███████
> **Cc:** ███████
> **Subject:** Re: website sad
>
> Actually, we're still getting hit heavily, but the bad guys aren't getting PDFs anymore. Just "go away" messages. So we should leave the blocking on.
>
> On Sep 25, 2010, at 10:01 PM, "███████@ithaka.org> wrote:
>
>> On second thought, screw em (very professional, I know). You mess with the bull, you get the horns. They are just as likely to start up again from an alternate IP and ruin my sleep tonight anyway. I'll reinstate em tomorrow morning if nothing else goes awry and let you know when I do.
>>
>> ██
>>
>> **From:** ███████
>> **Sent:** Saturday, September 25, 2010 8:53 PM
>> **To:** ███████
>> **Cc:** ███████
>> **Subject:** Re: website sad
>>
>> Word up. Jack em.
>>
>> On Sep 25, 2010, at 8:22 PM, "███████@ithaka.org> wrote:
>>
>>> Knock em out if you need to. If you want me to jack em, just say the word...
>>>
>>> ██
>>>
>>> **From:** ███████
>>> **Sent:** Saturday, September 25, 2010 8:15 PM
>>> **To:** ███████
>>> **Cc:** ███████
>>> **Subject:** Re: website sad

The PDF download activity has actually increased since 6pm now.
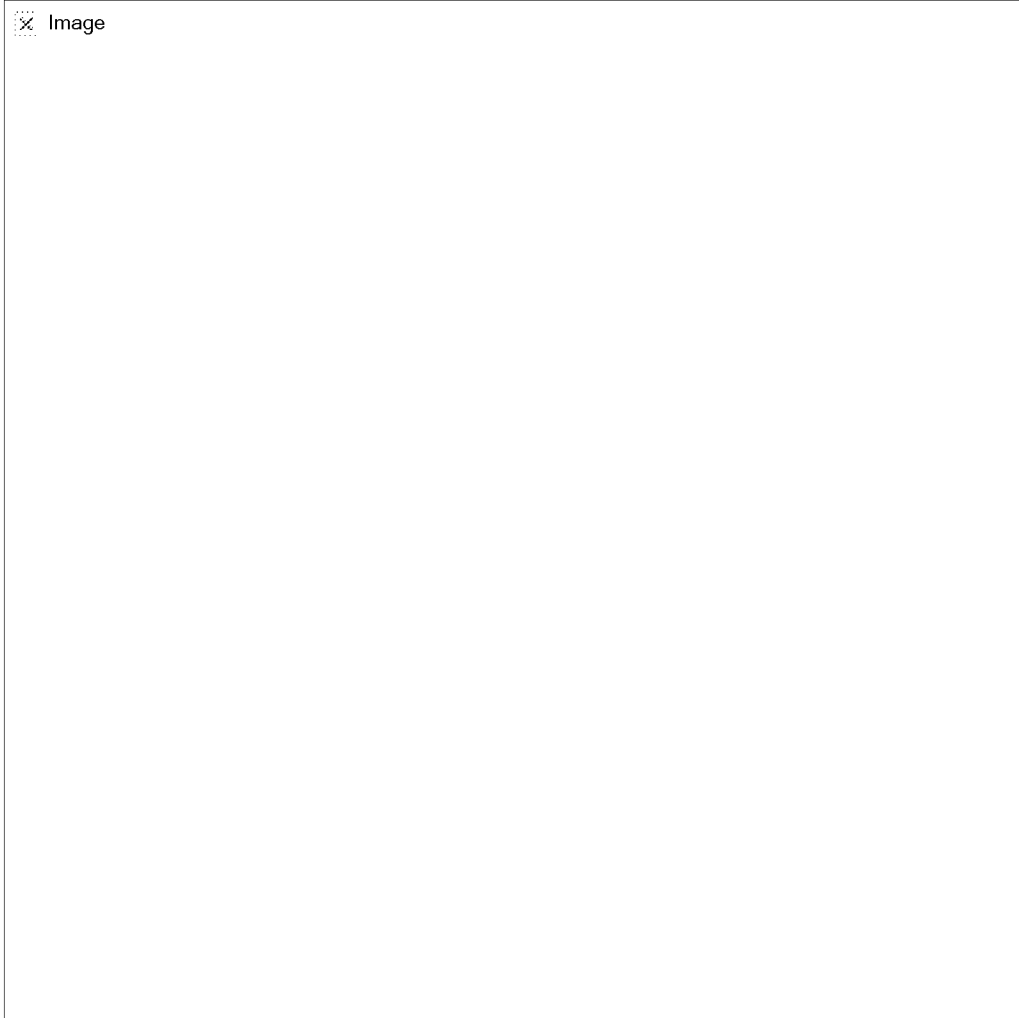
see the lighter red line.

Actually, MIT is the heaviest hitter right now for pdf hits with this ip: 18.55.6.215.

On Sep 25, 2010, at 7:29 PM, ███████████ wrote:

Any chance the offending scraper has an IP from the Portland area? We had a tool from Portland State University apologize and admit he was using 3+ PCs to mass download after they went to his house and punched him in the face (if only).

███

-----Original Message-----
From: ███████████
Sent: Saturday, September 25, 2010 7:02 PM
To: ███████████
Cc: ███████████████████████
Subject: Re: website sad

grep -hr synchronized literatum/src/java | while read line ; do FIX IT ; done :)D

On Sep 25, 2010, at 6:55 PM, ██████████ wrote:

Without a *new type of* thread pileup, I mean.

On 9/25/10 6:55 PM, ██████████ wrote:
Indeed; some champs like ████████ have over 1000 threads stuck there, while others have more like 50 or so.

Someday I'd like to have a month where we don't run into a thread pileup in this fine piece of software.

On 9/25/10 6:51 PM, ██████████ wrote:
observing lots of pdf scraping and this:

java.lang.Thread.State: BLOCKED (on object monitor)
    at
com.atypon.literatum.persistence.database.AbstractDb.generateId
(AbstractDb.java:830)
    - waiting to lock<0xffffffd7d1b810b58>   (a
com.atypon.literatum.persistence.database.████████)
    at com.atypon.literatum.acs.logger.LogIDGenerator.newID
(LogIDGenerator.java:53)
    at
com.atypon.literatum.acs.logger.LogIDGenerator.newActionID
(LogIDGenerator.java:25)
    at com.atypon.literatum.acs.logger.LogActionRow.<init>
(LogActionRow.java:99)
    at
com.atypon.literatum.acs.logger.SessionLogger.createLogAction
(SessionLogger.java:169)
    at
com.atypon.literatum.action.ActionController.dispatchRequest
(ActionController.java:213)
    at
com.atypon.literatum.action.ActionController.dispatchRequest
(ActionController.java:137)
    at
com.atypon.literatum.action.ActionController.dispatchRequest
(ActionController.java:111)
    at com.atypon.literatum.action.ControllerServlet.service
(ControllerServlet.java:85)
    at javax.servlet.http.HttpServlet.service
(HttpServlet.java:91)