

**From:** [REDACTED] </O=ITHAKA/OU=FIRST ADMINISTRATIVE GROUP/CN=RECIPIENTS/CN=[REDACTED]>  
**Sent:** Sunday, September 26, 2010 9:53 PM  
**To:** [REDACTED] <[REDACTED]@ithaka.org>  
**Cc:** [REDACTED]@ithaka.org>; [REDACTED] <[REDACTED]@ithaka.org>  
**Subject:** RE: website sad

---

Alright. Let's chat tomorrow whence we get settled...

-----Original Message-----

**From:** [REDACTED]  
**Sent:** Sunday, September 26, 2010 9:51 PM  
**To:** [REDACTED]  
**Cc:** [REDACTED]  
**Subject:** Re: website sad

So far blocking the class C net has kept the scraper at bay. MIT has an entire class A so the whole range for MIT is actually quite large :): 18.0.0.0/8.

Hopefully everything you need to communicate to MIT is in this thread.

Also, I bet this may not be the last script kiddie to figure out that we only limit the pdfs downloaded per session and that we don't limit sessions this fall. We may want to plan out what we need to enable the existing session limit controls literatum provides.

On Sep 26, 2010, at 9:05 PM, [REDACTED] wrote:

> Sorry Fellas, otherwise occupied today. Silly me for thinking that blocking them outright would allow a breath of relief. Anything I can do? Need the whole range from MIT?

>

> [REDACTED]

>

> -----Original Message-----

> **From:** [REDACTED] [mailto:[REDACTED]@ithaka.org]  
> **Sent:** Sunday, September 26, 2010 8:27 PM  
> **To:** [REDACTED]  
> **Cc:** [REDACTED]  
> **Subject:** Re: website sad

>

> And just because [REDACTED] taunted it, a few minutes later the servers got  
> all complainy for 20-30 minutes.

>

> Maybe they'll do it yet again tomorrow at the same time,  
> just to prevent us from being able to leave work for dinner.

>

> By the way, I have a class on Wednesday evenings for the next 6 weeks,  
> and I \*will\* be leaving work on time to go to it. You hear that,  
> Literatum? Get your worst stuff ready!

>  
>

> On 9/26/10 5:57 PM, [REDACTED]:

>> [REDACTED] added the class c net below to the block and we've been in the clear for two hours now.

>>

>> On Sep 26, 2010, at 1:15 PM, "[REDACTED]" <[REDACTED]@ithaka.org> wrote:

>>

>>> Unfortunately, it didn't take long for them to switch to another address this time. Please now block 18.55.6.0/24.

>>>

>>>

>>> On Sep 26, 2010, at 12:31 PM, [REDACTED] wrote:

>>>

>>>> MIT has been contacted about this activity and our resulting actions.

>>>>

>>>> [REDACTED]

>>>>

>>>> -----Original Message-----

>>>> From: [REDACTED]

>>>> Sent: Sunday, September 26, 2010 10:47 AM

>>>> To: [REDACTED]

>>>> Cc: [REDACTED]

>>>> Subject: Re: website sad

>>>>

>>>> scraper is back on 18.55.6.216 since 8am. Still MIT's network range -- which is 18.0.0.0/8 I guess.

>>>>

>>>>

>>>> On Sep 25, 2010, at 10:54 PM, [REDACTED] wrote:

>>>>

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

>>>>>

>>>>> On Sep 25, 2010, at 10:08 PM, "[REDACTED]" <[REDACTED]@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.

>>>>>>

>>>>>>

>>>>>>

>>>>>> [REDACTED]

>>>>>>

>>>>>>

>>>>>>

>>>>>> From: [REDACTED]

>>>>>> Sent: Saturday, September 25, 2010 10:04 PM

>>>>>> To: [REDACTED]

>>>>> Cc: [REDACTED]

>>>>> 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, "[REDACTED]" <[REDACTED]@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.

>>>>>

>>>>>

>>>>>

>>>>> [REDACTED]

>>>>>

>>>>>

>>>>>

>>>>> From: [REDACTED]

>>>>> Sent: Saturday, September 25, 2010 8:53 PM

>>>>> To: [REDACTED]

>>>>> Cc: [REDACTED]

>>>>> Subject: Re: website sad

>>>>>

>>>>>

>>>>>

>>>>> Word up. Jack em.

>>>>>

>>>>> On Sep 25, 2010, at 8:22 PM, "[REDACTED]" <[REDACTED]@ithaka.org> wrote:

>>>>>

>>>>> Knock em out if you need to. If you want me to jack em, just say the word...

>>>>>

>>>>>

>>>>>

>>>>> [REDACTED]

>>>>>

>>>>>

>>>>>

>>>>> From: [REDACTED]

>>>>> Sent: Saturday, September 25, 2010 8:15 PM

>>>>> To: [REDACTED]

>>>>> Cc: [REDACTED]

>>>>> 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, [REDACTED] 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).

>>>>>  
>>>>> [REDACTED]  
>>>>>

>>>>> -----Original Message-----

>>>>> From: [REDACTED]  
>>>>> Sent: Saturday, September 25, 2010 7:02 PM  
>>>>> To: [REDACTED]  
>>>>> Cc: [REDACTED]  
>>>>> 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, [REDACTED] wrote:

>>>>>  
>>>>>  
>>>>>  
>>>>>  
>>>>>

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

>>>>>  
>>>>>  
>>>>>  
>>>>>  
>>>>>

>>>>> On 9/25/10 6:55 PM, [REDACTED] wrote:

>>>>>

>>>>> Indeed; some champs like [REDACTED] 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, [REDACTED] 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<0xffffd7d1b810b58> (a com.atypon.literatum.persistence.database.[REDACTED])

>>>>>

>>>>> 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)

>>>>>