From: [redacted] </O=ITHAKA/OU=FIRST ADMINISTRATIVE GROUP/CN=RECIPIENTS/CN=[redacted]>
Sent: Sunday, October 10, 2010 8:49 PM
To: [redacted] <[redacted]@ithaka.org>; [redacted] <[redacted]@ithaka.org>
Cc: [redacted]@ithaka.org>
Subject: RE: MIT is currently denied JSTOR Access

Hi [redacted]

I have spoken with [redacted] and [redacted] has started to gather the relevant data. To track progress against this request, I filed a JIRA ticket **OPS-1843** at [redacted] request.

Let me try and address each of these as best I can at present…

It is a good point to loop in [redacted] and [redacted] we usually do. I will do that shortly. The journals being downloaded will have to come from the [redacted] investigation.

1. Ramping up enforcement. This is certainly possible, but aside from the considerations about the PR of it all, there are a few road blocks. The primary one, as I see it, is that these users are almost always behind a proxy. This means that the Institution would have to dig through their logs to tell who the actual user was, requiring their cooperation. I don't know this area well, but my instinct is that institutions may not be inclined to do this willingly if prosecution of one of their students is on the other side. A second problematic piece is how to verify where the data went… cloud storage? portable storage devices? Again, I don't know this well, but I am guessing it would take a skilled techie to uncover the trail of the data, if the device(s) in use is/are locatable.

2. We can support an alternate access method at their request, but typically the blocking of their IPs is a strong signal and they could see the offering of a UN/PW as unusable on their end. That said, we can do it. I'll also have a look to see what we might do via referring URL or a more scaled solution if we decide to go that route. I must say I am little reluctant to offer an olive branch at this point. I am slightly concerned that their initial report feels misleading and that, even after specifically discussing that we would prefer not to block all of their IPs, but might have to, it recurred. It has been an effective mechanism in the past to get the issue the attention it needs and MIT is not the first to tell us they took care of it and have it recur. Nor are they the first one we have denied for multiple days. An event of this variety is a very rare event indeed and this one is not see as more nefarious than others at this moment, just more successful due to the pattern change mentioned earlier today.

3. Posting to their web site. We certainly can ask them to do so. Note that these hackers can be as much of a pain for MIT and their institutions as they are for us, especially with all of their IPs denied. I'll just weigh in plainly here, forgive the candor, I don't think librarians would be likely to take this suggestion in the spirit it is intended and often don't have control over their library pages.

We'll have to circle in o this one this week, once the data is presented, to make recommendations for next steps. MIT has not responded as yet.

Thanks,

[redacted]

From: [redacted]
Sent: Sunday, October 10, 2010 6:35 PM
To: [redacted]
Cc: [redacted]
Subject: RE: MIT is currently denied JSTOR Access

p.s. In addition to [redacted], [redacted] also should be contacted in order to be prepared for any inquiries from publishers re weird usage stats.

Do we know what principally was down loaded?

From: [redacted]
Sent: Sunday, October 10, 2010 1:37 PM
To: [redacted]
Cc: [redacted]
Subject: RE: MIT is currently denied JSTOR Access

Thanks for this background, [redacted] A few thoughts for all of us:

1. I'd like us to think about ramping up how we deal with piracy-as-industrial-theft on a case-by-case basis. [redacted] is doing its job in monitoring and modifying thresholds as needed, and [redacted] is reaching out to librarians (fyi, [redacted] et al. should be alerted as well). But, it some instances, it may not be not enough to learn that "the activity has stopped". Where is the content that was downloaded living? Has the institution ensured that it's not on someone's hard drive that will go back to, say, China? In our agreement with institutions, they commit to being cooperative with us in dealing with matters of abuse, and we may want to pursue this more aggressively. If someone is downloading 1000s of articles (what seems like a reasonable threshold for us to take action), what's wrong with us – or the university in collaboration with us -- alerting the cyber-crimes division of law enforcement and initiating an investigation, having a cop search a dorm room and try to retrieve any hard drive that contains our content, etc. Our content is extraordinarily valuable and hard to replicate by the sweat of one's brow, but can be duplicated by savvy hackers, and who knows what they want to do with the content (we've already witnessed an interested in pirating our content from senior university officials in China….). So, beginning with this instance, I'd luck us to think about this. One step at a time, I realize, but I would like to be in touch with university officials and with this group or others as needed think about initiating some form of law enforcement investigation.
2. Are we prepared to create an alt method for MIT users to make use of the database while this is being resolved? It's a big deal as we know to shut down a whole university (and do we know why they'd have only one OP address vs. a range). Can we offer limited password access (heavy duty admin work for us) or some other mechanisms so that we don't lose good will here.
3. Is there a way for a notice to go out or be posted on the MIT library site making it clear that we are down as a result of extreme hacking by someone in the MIT community. Otherwise users won't know this and will think we're unreliable…..

[redacted]

From: [redacted]
Sent: Sunday, October 10, 2010 10:52 AM
To: [redacted]
Cc: [redacted]
Subject: RE: MIT is currently denied JSTOR Access

Thanks [redacted]

Just to clarify at the outset, the 1st incident that happened 2 weeks ago (9/25 – 9/26) was the same incident they

reported a visiting scholar responsible for, the 2nd incident started yesterday. I was mistaken, It was not 3 weeks ago, it was 2 weeks.

The 1st incident was excessive, but seen as one off and part of the normal abuse protocols. Robot being used, IP access being denied, institution being responsive and asserting they had taken care of it. We have any number of these on a weekly basis, but to the tune of a few hundred PDFs per case. This was to the tune of several thousand from what I understand.

I recognize that this is not an ordinary case and I should have looped you in on it at the time, that was poor judgment on my part. I'll be sure to do so going forward. At the time, [redacted] was satisfied with the response, as was I, and the activity did not recur. It was seen as an efficient robot, but not special, just using a new patter and with access to fantastic bandwidth through MIT. Definitely not any kind of denial of service attack.

Basically, when you apply limits based on session as we have, the natural progression is to adjust the pattern, one of which is to simply start new sessions more frequently.

[redacted] and I discussed this very pattern last year when abuse protocols were first put in place. The only weapon we have against it, based on the pattern, is some version of CAPTCHA, requiring entry for the 1st download of every session and once again, randomly, every N downloads. This would prevent this particular kind of scripting. My opinion on that is that if these are isolated incidents, that barrier is too high for the regular user. [redacted] may not feel that way. From my perspective MIT is a special case and the first of this kind of scripting we have seen on any scale and the first to threaten the live site with its efficiency. They will likely be embarrassed that this has recurred.

It is also not clear to me if they were scraping the site heavily for metadata, stable URLs, downloading PDFs or all of the above. We are adjusting the abuse tools down to try and auto-prevent, but we need precise data in order to both prevent this kind of abuse pattern and not disrupt normal use elsewhere. To that end I will ask [redacted] and [redacted] to provide a recap of the incident and the numbers surrounding it as they never hit our abuse monitoring rules the 1st or 2nd time, I have no window into the numbers themselves.

The second occurrence, as it stands, started at some point yesterday, but I am unclear as to when that was. The [redacted] summary will give us that.

As for the timeline to this point, [redacted] began communicating broadly about it shortly before 6 pm yesterday and the IP range was denied around 6:45 pm. I first saw the email chain from [redacted] around 8:30 pm and began communicating to MIT, [redacted] and [redacted] around 9 pm. MIT has not responded and the IPs will remain blocked until they do.

I will be sure to keep you both looped in as progress is made against both resolving this issue and providing a summary report.

[redacted]

From: [redacted]
Sent: Sunday, October 10, 2010 9:33 AM
To: [redacted]
Subject: Re: MIT is currently denied JSTOR Access

Thanks, [redacted] Very troubling.

Also, this is the first I've heard of this. [redacted] needs to be involved earlier in these types of matters to ensure we adhere to contract terms and pursue appropriate remedies. This is part of the protocol we established. The activity noted is outright theft and may merit a call with university counsel, and even the local police, to ensure not only that the activity has stopped but that -- e.g. the visiting scholar who left -- isn't leaving with a hard drive containing our database.

[redacted] please clarify the history of this activity. What happened 3 weeks ago? How did we respond? What has happened in the intervening time?

And what is the "last time" incident with the visiting scholar? Is this the one from three weeks ago? Again, this is the first I've heard of this. Going forward, we need to make sure [redacted] is aware of these cases as they are emerging. The [redacted] component is critical here, but there are other avenues that often need to be pursued concurrently.

[redacted]

From: [redacted]
Sent: Saturday, October 09, 2010 11:18 PM
To: [redacted]
Subject: Fw: MIT is currently denied JSTOR Access

Just want to be sure [redacted] is aware of the situation.

[redacted]

From: [redacted]
Sent: Saturday, October 09, 2010 10:30 PM
To: [redacted]
Cc: [redacted]
Subject: MIT is currently denied JSTOR Access

Good Evening,

I want to make you aware that the MIT abuse case that showed up 3 weeks ago came back today, forcing [redacted] to deny 18.0.0.0/8 at the firewall. Half of Manchester needed restarting this afternoon to address servers that got jammed up due to this activity. I have emailed our contacts at MIT and informed them of the situation.

Just for clarity, that's the whole range, the largest we have ever denied. Last time they reported that a visiting scholar was responsible and that it should not recur as the scholar had left.

I am in conversation with [redacted] and [redacted] is reporting that this scraping is very intensive and threatening the website when unblocked. The block of their range has brought the incident under control and they are currently getting deny pages and not threatening the website.

The pattern is simple… they start a session, download 1 pdf, start a new session, download 1 pdf on and on. [redacted] can comment on the specific volume and duration.

Also, after the last incident at MIT, we implemented Literatum's # of sessions per hour IP blocking rule to 5000 sessions in 60 minutes. It did not fire. We are digging deeper, but the earliest speculation is that this rule is applied 'per server', which we did not anticipate, meaning we'll need to adjust the number down based on data gathered from this incident and elsewhere.

I don't know if this will cause any negative reaction in the public and haven't heard anything through our feedback

channels as yet, but wanted to make sure we were all on the same page and that there are no surprises here. This is an extreme block to combat an extreme attack.

More as the situation gets resolved.

Best,

[redacted]
[redacted]
JSTOR | Portico

[redacted]@ithaka.org
[redacted]