From: [redacted] </O=ITHAKA/OU=FIRST ADMINISTRATIVE GROUP/CN=RECIPIENTS/CN=[redacted]>
Sent: Wednesday, October 20, 2010 5:41 PM
To: [redacted] <[redacted]@ithaka.org>
Cc: [redacted] <[redacted]@ithaka.org>
Subject: FW: 10:00 am Update: JSTOR Abuse at MIT: All IPs Blocked

---

fyi

-----Original Message-----
From: [redacted] [mailto:[redacted]@MIT.EDU]
Sent: Wednesday, October 20, 2010 2:01 PM
To: [redacted]
Cc: [redacted]
Subject: RE: 10:00 am Update: JSTOR Abuse at MIT: All IPs Blocked

Hello [redacted], I apologize for the delay in responding. I have been waiting for more detail from our IS&T group, but I still don't have that information. In the meanwhile, I am hoping we can move forward with our discussions to make the technical changes necessary to implement our additional authorization layer, which we call "econtrol."

We would like to try an additional tweak to our normal 'econtrol' process if it is possible on the JSTOR end. If JSTOR could use an Apache mod_rewrite to redirect any activity from the MIT IP addresses (aside from those for our proxy server) to our proxy server, our patrons would not have to remember to use a special gateway to get to JSTOR. This would be a big benefit to our patrons. Would this be feasible for JSTOR?

When we are ready for the change, we will need you to reduce the authorized MIT IP ranges list you appended below to these:

18.51.1.222
18.7.29.240

Ultimately only the second address will be needed, but we are in transition from one proxy server to another. If you can implement the server-side configuration change mentioned above, we would currently be redirecting to 18.51.1.222, though we'd have to change to the other address in the next few months.

Please do not make any change yet, as we want to send information to our patrons before the switch so that they will be aware of the changes in the access model. We are preparing the communications now, but we need to know whether JSTOR can implement the apache mod_rewrite before we finalize those messages, since that information will determine what we have to tell our patrons about access.

I have copied [redacted] from our IS&T group here -- if there are technical questions about the Apache mod_rewrite you will be best served by going direct to [redacted]. I have also copied the [redacted], [redacted], as [redacted] will be the one making the changes on our end to EZproxy. Those changes will need to precede the changes on JSTOR's end.

In at least partial answer to your inquiry below, we offer guests access to the MIT network. However, once we institute our additional authorization layer for JSTOR, this route will be closed to guests. So we will have closed the pathway through which the excessive use occurred.

We look forward to moving forward with an econtrol implementation with or without the apache code --

Thanks very much,
[redacted]

[redacted]

[redacted]
MIT Libraries
P [redacted]
[redacted]@mit.edu
http://libraries.mit.edu/scholarly

-----Original Message-----
From: [redacted] [mailto:[redacted]@ithaka.org]
Sent: Monday, October 18, 2010 10:04 AM
To: [redacted]
Cc: [redacted]
Subject: RE: 10:00 am Update: JSTOR Abuse at MIT: All IPs Blocked

Thanks [redacted],

I appreciate your candor here. I have dealt with many cases over the years and understand the difficulties inherent with tracking down individual users. I am hopeful we can use this opportunity to work together towards building more stable, sustainable and secured access to JSTOR. We are meeting as a larger group to discuss this matter further this afternoon and I am wondering if you could provide one point of clarification for that conversation.

Understanding that you may not be able to identify the individual, where you able to identify the credentials used to access MIT authorization for this activity? That is, was there a shared UN/PW used for guests or an open port on a proxy used in this case. Basically, the concern is, as we sort out the IP configurations necessary with you, could this or any user use the same authorization methodology to do this again or has the pathway been identified and locked down?

As for your IP configurations and establishing an access point within your range. We have the following IP addresses currently installed for MIT. Please let me know which ones to maintain and which to remove as needed and we'll get right to it.

18.*.*.*
128.30-31.*.*
128.52.*.*
129.55.*.*
192.52.61-66.*
198.125.160-163.*
198.125.176-192.*

Best,

[redacted]
[redacted]
JSTOR | Portico

[redacted]@ithaka.org
[redacted]

-----Original Message-----
From: [redacted] [mailto:[redacted]@MIT.EDU]
Sent: Friday, October 15, 2010 5:04 PM
To: [redacted]
Cc: [redacted]
Subject: RE: 10:00 am Update: JSTOR Abuse at MIT: All IPs Blocked

Hello [redacted], I wanted to give you an update before the weekend.

At this time we have gone as far as we can in identifying the individual involved in these incidents. Our records and logs related to this activity do not allow us to definitively identify the guest. We appreciate your offer of more granular log information, but our IS&T group does not believe that such files will allow us to reach the point of identification.

We can, however, take a significant step to prevent recurrence by moving to the new protocol I'd mentioned below. Since it sounds as if that would be welcome and workable on the JSTOR end, I hope we can pursue that next week.

I would be happy to discuss any aspect of this with you further, and I'm sorry we do not have more details to report in terms of the incidents.

[redacted]

[redacted]
[redacted]
MIT Libraries
P [redacted]
[redacted]@mit.edu
http://libraries.mit.edu/scholarly

Thank You [redacted],

I appreciate your response here. It appears we still have a ways to go to reach resolution, but I am glad to assist.

First, this activity is not continuing at the moment. Given that we saw it twice in two weeks, starting on a Saturday, I will hazard a guess that if this does recur, it will begin again on a Saturday. That said, if an when it does recur, we will be denying IP ranges significant enough to prevent it from continuing, while hopefully avoiding the need to block the entire range again. Internally, we are agreed on this point.

Second, we typically follow each case of excessive downloading with a three step process for considering the incident resolved...

1. Is it continuing? Not at the moment, but the jury is still out and will be for a few weeks.

2. Did the institution take the necessary steps to prevent recurrence? I see your suggestions here and have some thoughts on it as a follow on conversation. At present however, it is very important for us to understand if the users password has been changed and if the user has been contacted directly to address this issue. As a guest user, and likely the same user involved previously, using an efficient robot to grab lots of content, this is paramount to solve at the individual user level. If it is a shared account or used by multiple users, this is even more critical.

3. Was the content acquired deleted? This can be tricky, we understand, but if you can identify the user, in combination with adjusting their credentials, we must request that the best effort be made to insure that the content acquired is deleted from the storage device or web space in which they are storing it.

We can give you very granular log files from our end if identifying the user is problematic, but not identifying the user and assuring that the content is deleted, especially on a incident of this size, is a sizeable barrier to bringing this incident to a close.

As for your suggestion, we would gladly adjust the IPs that have access to JSTOR at your request. Note that some of our very large institutions do authenticate in this way. Also note that most very large institutions that do use proxy servers, use 2 or 3 to meet their bandwidth and access control needs. That said, I want to make sure we are on the same page here. Adjusting your configurations to prevent future occurrences is separate from bringing resolution to this incident.

If your IS&T group need additional information for activities between the time frames already provided, please do let me know what kind of information they are looking for and how much. Like, logs for at least 30 consecutive actions from an MIT IP between the times of 16:00 and 16:30 on Saturday, and we'll be happy to provide them.

Thanks,

[redacted]

-----Original Message-----
From: [redacted] [mailto:[redacted]@ithaka.org]
Sent: Friday, October 15, 2010 9:37 AM
To: [redacted]
Cc: [redacted]
Subject: RE: 10:00 am Update: JSTOR Abuse at MIT: All IPs Blocked

Dear [redacted],

I am wanting to keep you up to date on our findings here. I am copying an incident summary below for the two occurrences. I am hopeful that you understand our urgency as not intended to be confrontational at all, but as rooted in our deep investment in our ability to be good stewards of the content in our care. Noting that, the 9/25 and 9/26 incidents yielded more than 400K PDFs from JSTOR to this user. This is both unprecedented and very concerning. Any progress you could provide on identifying the user responsible and the steps I mentioned yesterday would be very helpful.

We are also mindful that the weekend is upon us and are looking for collaboration on your end and acknowledgement that MIT staff will be monitoring your systems closely for any recurrence, as we are, until we can reach resolution.

Incident on 9/25 & 9/26

IP = 18.55.6.215
Start = 25-SEP-10 05.06.49.109524 PM
End = 26-SEP-10 04.24.54.297995 AM
Total Sessions = 1,256,249
Total Articles Downloaded = 453,570
Total Journals Affected = 562

Incident on 10/9

IP = 018.055.005.100
Start = 2010-10-09 14:53:18 from
End = 2010-10-09 19:08:01
Total Sessions = 8,515
Total Articles Downloaded = 8,422
Total Journal Affected = 714

Best,

[redacted]

-----Original Message-----
From: [redacted] [mailto[redacted]@MIT.EDU]
Sent: Thursday, October 14, 2010 12:44 PM
To: [redacted]
Cc: [redacted]
Subject: RE: 10:00 am Update: JSTOR Abuse at MIT: All IPs Blocked

Hello [redacted] and [redacted],
Our investigations here point to the same guest that was involved in the 9/27 incident. We don't have enough information to follow the trail completely, but the signs suggest that the same guest user was responsible for this latest activity. To pursue this further, our IS&T group would need more information. Specifically, they are wondering if you are seeing any robotic activity from MIT currently and if so, whether you have any information about the IP addresses involved.

Given that it appears all of this excessive use was caused by a guest visitor at MIT, we have been considering next steps, and would like to suggest that we move to a new access model that will eliminate use by guests. We have recently developed an additional authorization layer that we can apply to particular products to prevent access by guests/walkins. We've tried this approach with one or two publishers where we had seen repeated excessive use, and it has stemmed the problem in those cases.

We would orchestrate this change by changing the proxy configuration on this end, and then we'd ask you to change the list of acceptable MIT IPs to only our proxy server's address -- a single IP.

If this sounds like an acceptable approach, let's discuss the next steps. To carry out the change, I'd have JSTOR work with [redacted], copied here.

Best,

[redacted]

[redacted]
[redacted]
MIT Libraries
P [redacted]
[redacted]@mit.edu
http://libraries.mit.edu/scholarly

-----Original Message-----
From: [redacted] [mailto:[redacted]@ithaka.org]
Sent: Tuesday, October 12, 2010 10:09 AM
To: [redacted]
Cc: [redacted]
Subject: 10:00 am Update: JSTOR Abuse at MIT: All IPs Blocked

Hello Again,

We have requested that the IP range be unblocked at the firewall and that process is currently underway. I will confirm when that is accomplished and report the IPs and timestamps surrounding the event shortly.

[redacted]