

From: ██████████ </O=ITHAKA/OU=FIRST ADMINISTRATIVE GROUP/CN=RECIPIENTS/CN=██████████>
Sent: Thursday, October 14, 2010 4:55 PM
To: '██████████' <██████████@MIT.EDU>; ██████████ <██████████@ithaka.org>
Cc: '██████████' <██████████@mit.edu>
Subject: RE: 10:00 am Update: JSTOR Abuse at MIT: All IPs Blocked

Thank You ██████████,

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,

-----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]

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

From: [REDACTED] [mailto:[REDACTED]@MIT.EDU]
Sent: Wednesday, September 29, 2010 4:24 PM
To: [REDACTED]
Cc: [REDACTED]
Subject: RE: JSTOR Abuse at MIT: All IPs Blocked

[REDACTED], thanks very much. We appreciate it.

Looking to the future, would it be possible to clarify that JSTOR will follow the protocol that was ultimately used here, shutting down not the class A range but the class C range, should an excessive use case emerge again?

We have not had a history of excessive use of JSTOR content from MIT, so the problems do not seem at this time to be widespread. In addition, we are finding that the industry norm at this time is shut down of the specific offending IP.

We can see that in some cases (as here, where the initial suspension did not stop the misuse) moving to suspend the class C range is a reasonable response. But it would be very helpful on our end if we could work out an agreement that a shutdown of the class A range will not be part of the standard initial response protocol for excessive use cases.

We look forward to your thoughts --

[REDACTED]

[REDACTED]
[REDACTED]

MIT Libraries

P [REDACTED]

[REDACTED]@mit.edu

<http://libraries.mit.edu/scholarly>

Subject: JSTOR Abuse at MIT: All IPs Blocked

Dear [REDACTED],

I am writing you this afternoon to let you know that we have been forced to block access to JSTOR from MIT. Yesterday, around 6pm, we began to see hundreds of PDF downloads per minute occurring from multiple sessions at 18.55.6.215. As these requests began to affect performance of the public site, we were forced to deny access to this IP. Requests continued to pour in from this IP for some time, but were denied access. This clearly indicates robotic harvesting of PDFs which violates our Terms & Conditions of Use.

This morning, at around 8am, this activity started again from IP 18.55.6.216, forcing us to restrict access to the entire range of MIT IP addresses. We rarely take this level of response to abusive activity, but felt it necessary to maintain the stability of the web site for other institutions and users.

Once you have identified the responsible party and can assure us that this activity will not continue, we will be happy to restore access as soon as possible. That said, please note that the block had to be executed at the firewall level to prevent performance degradation (even after denying the PDF downloads, the requests themselves were so frequent it continued to be problematic) and thus will require coordination of our systems administrators to restore, which may take some time.

Please do let me know if I can be of additional assistance or if I can provide additional information. Having worked in this area for some time, I am well aware that this activity is normally a compromised username and password or a student/researcher unaware of the impact of their activities or that this method of gathering PDFs is in violation of our Terms and Conditions of Use. We routinely work with researchers through our dfr.jstor.org site or by providing data exports to accomplish the intended research aim and would be happy to do so in this case as well if that turns out to be the motivation.

Best,

[REDACTED]

[REDACTED]

JSTOR

[REDACTED]@<mailto:[REDACTED]@ithaka.org>ithaka.org

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. JSTOR is part of ITHAKA, a not-for-profit organization that helps the academic community use digital technologies to preserve the scholarly record and to advance research and teaching in sustainable ways.