

From: [REDACTED] </O=ITHAKA/OU=FIRST ADMINISTRATIVE GROUP/CN=RECIPIENTS/CN=[REDACTED]>
Sent: Tuesday, October 26, 2010 11:30 PM
To: [REDACTED] <[REDACTED]@ithaka.org>; [REDACTED] <[REDACTED]@ithaka.org>; [REDACTED] <[REDACTED]@ithaka.org>
Cc: [REDACTED] <[REDACTED]@ithaka.org>; [REDACTED] <[REDACTED]@ithaka.org>; [REDACTED] <[REDACTED]@Ithaka.org>; [REDACTED] <[REDACTED]@ithaka.org>; [REDACTED] <[REDACTED]@ithaka.org>
Subject: RE: MIT Conversation Summary & Abuse Methodology from [REDACTED]

Thanks for the report. Sounds basically encouraging. It is definitely a good thing that we have opened this channel of communication. Well done all.

[REDACTED]

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

From: [REDACTED]
Sent: Tuesday, October 26, 2010 10:27 PM
To: [REDACTED]
Cc: [REDACTED]
Subject: MIT Conversation Summary & Abuse Methodology from [REDACTED]

Good Evening,

I wanted to provide an update from the conversation we had with MIT this afternoon. [REDACTED], [REDACTED] and I spoke with their primary contact, their [REDACTED] and their [REDACTED]. It was a pleasant and engaged conversation with some meaningful outcomes.

Incident Update:

1. They confirmed that they believe the activity took place on campus. Based on their analysis of the IP addresses used, and as they explained it, they are convinced that this is the case. This is good news in that the individual might still be identified and the material recovered.
2. They are continuing their efforts to identify the individual(s) responsible. As a side note, I was paying particular attention to this point as their [REDACTED] addressed it. I get the impression that [REDACTED] means it and I also get the feeling [REDACTED] has active leads, though [REDACTED] didn't share them.

They were cognizant of the problem that the volume taken here poses and expressed understanding about our ongoing concerns. As we discussed possible next steps they hinted at infrastructure changes, already in the works prior to our issue, that this event, combined with others, would help facilitate.

Next Steps:

1. They understood the difficulties that MDC and multiple servers would pose for their proposed solution and we quickly moved on to the next solution. They did confirm that a half dozen other vendors are set up this way, but didn't mention which ones and said that it was sometimes contractually required by the vendor. I brought it up twice, asking about similar sized vendors or content delivery sites, but they weren't forthcoming about which ones and I didn't press the somewhat moot point, since we weren't going there.
2. █████ proposed a potential solution, using a visual prompt for MIT students not arriving at JSTOR through their proxy that directs them through the MIT login process. This was seen as a nice alternative from MIT's perspective and we agreed to develop a proto-type on our side. From my understanding, this functionality would work like the overlay that was used for the MyJSTOR Beta and requires a little work on █████ part, but nothing major and no web site 'development' work.

This solution doesn't really scale, but their situation is very unique. That is, there is no one of their size as an academic institution, bandwidth/IP range wise, and I don't know of any other institution configured this way. It's odd. They recognize this too and said as much, attributing it to a history of open access on campus that integrates less and less gracefully.

I will keep you updated as we learn more and make progress against the proto-type.

█████ summary of the pattern from 9/25 is as follows. Title roll-up work is continuing.

The robot was spawning sessions beginning with a particular DOI, and instructing the session to download articles sequentially. When a particular session failed, a new one was spawned at the # where the previous one died. The order in which journals were targeted was haphazard, but sessions follow the same pattern - beginning with a DOI, and incrementing it by 1, ad infinitum, until failure.

As there were 1.2MM sessions, with only ('only') 451K downloads, the session failure rate was obviously fairly high, which is likely a by-product of: 1) our abuse controls; and, 2) gaps in the number sequence in our DOI list (i.e., the robot hits the next number in a sequence, but we have no article with that DOI).

Best,

█████