

From: [REDACTED] </O=ITHAKA/OU=FIRST ADMINISTRATIVE GROUP/CN=RECIPIENTS/CN=[REDACTED]>
Sent: Wednesday, February 2, 2011 9:59 AM
To: 'Heymann, Stephen (USAMA)' <Stephen.Heymann@usdoj.gov>
Subject: RE: MIT Update: It's worse than we know

The counts are for PDF downloads, so the 1,385,569 figure represents full articles.

From: Heymann, Stephen (USAMA) [mailto:Stephen.Heymann@usdoj.gov]
Sent: Wednesday, February 02, 2011 9:06 AM
To: [REDACTED]
Subject: RE: MIT Update: It's worse than we know

[REDACTED]

In [REDACTED] numbers below, does 1,385,569 represent separate research articles, or parts (like some title pages, some parts of multi-part articles, etc) or is it impossible to tell for the moment? If you don't know for the moment, is it possible to estimate the number of separate articles?

Steve

From: [REDACTED]@ithaka.org]
Sent: Friday, January 28, 2011 3:05 PM
To: Heymann, Stephen (USAMA)
Subject: FW: MIT Update: It's worse than we know

... and this too.

From: [REDACTED]
Sent: Friday, January 28, 2011 3:02 PM
To: [REDACTED]
Cc: [REDACTED]
Subject: RE: MIT Update: It's worse than we know

I do know from [REDACTED] initial analysis that the downloading was done systematically using sequential increases in our stable URLs. That is, get stable URL 12345, get stable URL 12346, 12347, 12348 and so on.

This tells us a few things. One, that the previous activity was similar to the pattern [REDACTED] is seeing. That is, not targeted towards types. Two, it lends credence to an entire corpus grab approach. Don't care what it is I am getting, just get me the next one.

[REDACTED]

From: [REDACTED]
Sent: Friday, January 28, 2011 2:50 PM
To: [REDACTED]
Cc: [REDACTED]
Subject: Re: MIT Update: It's worse than we know

This doesn't appear to be targeted towards research articles or any particular titles, collections, or disciplines. For the 2.8 million in Nov and Dec, the breakdown by article type is:
Research articles – 1,385,569

Reviews – 938,063
Misc – 459,457
News – 62,127
Editorial – 9,472

Those numbers more or less correlate to the corpus as a whole. I'd say the he was going after everything.

From: [REDACTED]@ithaka.org>
Date: Fri, 28 Jan 2011 14:28:43 -0500
To: [REDACTED]@ithaka.org>, [REDACTED]@ithaka.org>
Cc: [REDACTED]@ithaka.org>
Subject: Re: MIT Update: It's worse than we know

Attached are 2 screen shots depicting PDF download activity from MIT for November and December. One show's all downloads and totals 2,854,824 for the 2 months. The other filters out downloads from the 3 IP's that look to be associated with the download abuse (18.55.6.240, 18.55.7.240, 18.55.5.237) and totals 17,865 for the 2 month period. Recognizing that some legitimate downloads may have occurred from the 3 filtered IP's, it would still be safe to say that about 2.8 million illegal downloads occurred during November and December. We know that some illegal downloading occurred prior to November and into January. I don't have those numbers yet. But looking at the graph you can see that some pretty aggressive downloading was taking place the last week of Dec (over 100k.day). It seems likely this extended into January for some period of time. It wouldn't be much of a stretch to say that as much of a million or more additional downloads may have occurred that are not reflected on this chart. I expect to have January data available for review by Monday. I'll also start loading Oct and Sept numbers as well to complete the picture.

[REDACTED]

From: [REDACTED]@ithaka.org>
Date: Fri, 28 Jan 2011 14:13:57 -0500
To: [REDACTED]@ithaka.org>
Cc: [REDACTED]@ithaka.org>, [REDACTED]@ithaka.org>
Subject: RE: MIT Update: It's worse than we know

So, with September and October, what does the number look like? Still looking like the entire corpus?

From: [REDACTED]
Sent: Friday, January 28, 2011 2:10 PM
To: [REDACTED]
Cc: [REDACTED]
Subject: RE: MIT Update: It's worse than we know

[REDACTED]

Still digging and [REDACTED] is going to pass along a screen shot of what he is seeing for November and December. It appears as though the activity was less impactful in November, but just these two months ballpark ~2 million + PDFs over their normal usage. [REDACTED] is also seeking out data prior and since for review.

[REDACTED]

From: [REDACTED]
Sent: Friday, January 28, 2011 1:53 PM
To: [REDACTED]
Cc: [REDACTED]
Subject: RE: MIT Update: It's worse than we know

I will call in a moment.

From: [REDACTED]
Sent: Friday, January 28, 2011 1:43 PM
To: [REDACTED]
Cc: [REDACTED]
Subject: MIT Update: It's worse than we know
Importance: High

Hi [REDACTED]

Speaking with [REDACTED] just now about making sure we have [REDACTED] time as needed for the MIT evaluation currently underway and discovering that the IP addresses associated with these specific incidents have numerous additional days of mass downloading.

It would take some time to normalize against usual MIT usage, but at first glance, it is reasonably safe to assume from what [REDACTED] and I covered that the individual responsible has already acquired the entire JSTOR corpus. Glad to have a call ASAP if you think it useful and I let [REDACTED] know that I thought you might be calling him shortly after receiving this message for clarification. In light of this information, it would seem that we need to try and understand the full picture outside of the identified incidents going forward. Certainly our tack here merits some re-evaluation, both concerning this case and the potential for additional measures of prevention as we move forward. Also copying in [REDACTED] at this juncture.

[REDACTED]

Thanks,

[REDACTED]

JSTOR | Portico

[REDACTED]@ithaka.org
[REDACTED]