I T H A K A

Root Cause Analysis (RCA)

DATE: 1/21/2011 ORIGINATOR: [redacted] BU: ITHAKA

DETAILED DESCRIPTION OF PROBLEM:

Software running on a server in the Durrance Building on MIT Campus systematically downloaded several hundred thousand PDFs from the JSTOR website.

- 9/25/2010 6:48 PM – first sign of problems was jstor.org was running slow
- 9/25/2010 6:51 PM – found that servers were bottlenecking on id generation and that there was likely scraper activity
- 9/25/2010 8:54 PM – Blocked 18.55.6.215 with literatum – application begins to return 500 errors to the client rather than pdfs
- 9/25/2010 10:00 PM – observed the rate of accesses from 18.55.6.215 had not subsided
- 9/26/2010 10:00 AM – observed the scraper moved to 18.55.6.216 at 8 am
- 9/26/2010 12:00 PM – 18.55.6.216 is blocked in the firewall – packets are dropped.
- 9/26/2010 1:13 PM – observed client changed ips again.
- 9/26/2010 5:58 PM – blocked the class c net (18.55.6.0/24) in the firewall
- Later that week, we decide to block 18.0.0.0/8 if we see the abuse change to a different class-c network since it appeared that MIT had the entire range and we were unaware of the breadth of potential addresses available to the client.
- 10/9/2010 6:50 PM – 18.0.0.0/8 blocked in the firewall due to the same abusive activity throughout the day from a different class-c range (18.55.5.100).
- 10/31/2010 -- The JSTOR website was suddenly bombarded with 40,000 requests to almost 80,000 requests over a period of 4 minutes starting at 10/31 5:20 AM. This represents 40 to 80 times more requests of this type than normal usage of the system. This degraded the system to a point where it was completely unavailable for a period of time. The requests originated from over 1500 different networks on the internet. However, they all had the same user-agent signature in the http headers: Apache Synapse -- an enterprise service bus package developed by the Apache Foundation. Someone used this package to coordinate activity of a botnet.
- After discussing with MIT, we decided that for the next incident, we need only to block the class-c.
- 12/26/2010 – the scraper returned and the class-c net of the ip he came from was directed toward an an access denied message.

The "abusive activity" had the following characteristics

- The useragent was "curl/7.19.7 (i486-pc-linux-gnu) libcurl/7.19.7 OpenSSL/0.9.8k zlib/1.2.3.3 libidn/1.15"
- Our pdf files were accessed in sequential order – generally a red flag for abusive activity.
- Over 500 journals were touched each day the client reappeared on the site.
- Millions of requests came from this client to the jstor website.

CLIENTS AFFECTED: Massachusetts Institute of Technology

DATE OCCURRED	APPLICATION / PRODUCT	DURATION
September 26, 2010	www.jstor.org	~ 3 days
October 9, 2010	www.jstor.org	~ 1 week
December 26, 2010	www.jstor.org	~ 1 month

OF ACCOUNTS OR $ AMOUNTS AFFECTED IF APPLICABLE: _____

NET IMPACT TO THE CLIENT: www.jstor.org was unavailable to a segment or all of MIT campus.

[redacted]

RESPONDING DEPARTMENT:

☒ [redacted] ☐ [redacted] ☐ [redacted] ☒ [redacted]

☐ [redacted] ☐ [redacted]

CAUSE TYPE:

☐ Hardware ☐ Software ☐ Human Error ☐ Procedures

☒ Other Abuse

Did the proper notifications/escalations occur? ☒ Yes ☐ No

RECOVERY ACTIVITIES:

[redacted] identified abusive activity before it became disruptive to other users. In each case, different parts or all of the MIT campus network was blocked.

RESOLUTION TO THE INCIDENT:

MIT's network was blocked either in the ITHAKA frontend routers or in the [redacted] load balancers. Due to the IP address of the client changing throughout the incidents, [redacted] had to widen the network range of the filter.

9/26 – A Class-C range of MIT's network was filtered in our firewalls. Packets were simply dropped.

10/9 – MIT's Class-A network range was filtered in our firewalls due to the client changing IP addresses rapidly during the previous incident.

12/26 – A Class-C of MIT's network was directed to an 'abuse farm' of web servers that display an access denied message.

In all cases above, the problem was escalated to [redacted] who reached out to [redacted] contacts at MIT to communicate what was happening. MIT would communicate with us after each incident once they felt they had resolved the problem on their side.

CLIENT IMPACT:

Parts or all of MIT campus did not have access to the JSTOR website on the days listed above.

PREVENTIVE ACTION:

MIT implemented a forward proxy on their network and JSTOR has updated its authorization for MIT to just the IP address of MIT's forward proxy. This essentially gives MIT a way to monitor and gauge their usage of JSTOR and close the loop on abusive activity on their network.

[redacted] has implemented more sophisticated monitoring to identify and escalate abnormal amounts of PDF downloads to operators to remediate. This in combination with new filtering on the load balancers allows us to block and more accurately identify abusive usage.