After Action Review: InCommon Federation MDQ Service Degradation 2021-10-07

Date: 2021-11-02
Prepared by: Nicole Roy, Internet2 Trust and Identity Services

Summary

Early on the morning of October 7, 2021, Amazon Web Services (AWS) started rolling out a regular update to systems that support the AWS Lambda@Edge service, which is a fundamental component of InCommon's per-entity metadata query (MDQ) service. This update was unannounced, as are all such routine updates of underlying infrastructure. It was not intended to change the way the service behaved. However, this change resulted in a double-URL-encoding of MDQ request parameters (URL-encoded SAML entityIDs). This, in turn, resulted in degraded fulfillment of MDQ service requests for the period 5:37 a.m. US EDT on Thursday, October 7, 2021, to 5:00 p.m. US EDT on Friday, October 8, 2021.

Timeline

```
2021-10-07 5:37 a.m. US EDT
First known failure of the service to respond to a metadata query request

2021-10-07 10:16 a.m. US EDT
Voicemail message is left on the InCommon helpdesk line. This voicemail
contains the first report of this issue by a participant to InCommon.
```

2021-10-07 10:30 a.m. US EDT
TI Operations made aware of the issue via escalation from first-line support
TI Operations begins to investigate

2021-10-07 11:06 a.m. US EDT
TI Operations sends a notice to the inc-ops-notifications mailing list to acknowledge that we have received reports of an MDQ service disruption, and are investigating.

2021-10-07 11:11 a.m. US EDT
TI Operations determines a way to query individual Lambda@Edge/CloudFront edge servers and start to measure the scope of the impact. It is determined that some edge servers are responding correctly, and some are failing with an HTTP 404, indicating an issue with resolution of the individual metadata entity descriptors based on the query string supplied to the service. Failures appear to be most prevalent in the US-East-1 / “IAD” CloudFront location, but are spreading to other locations over time.

2021-10-07 11:12 a.m. US EDT
More helpdesk tickets arrive with reports of InCommon participants affected by the problem.

2021-10-07 11:28 US EDT
TI Operations posts an outage notice at https://status.incommon.org, which will be continually updated throughout the outage.

2021-10-07 12:30 US EDT
InCommon Operations determines that our Lambda@Edge code appears to be affected by an unannounced change made by AWS, and escalates the issue to our managed services partner, DLT.

TI Operations starts to analyze CloudFront logs to determine if there is a pattern to which sites may be affected.

2021-10-07 12:30 p.m. US EDT
The status page has a length limitation to its notes section, so a wiki page is created where future updates are posted: https://spaces.at.internet2.edu/display/federationops/2021-10-07+MDQ+intermittent+outage

2021-10-07 01:10 p.m. US EDT

TI Operations analyzes logs associated with our production MDQ CloudFront deployment, and determines that the affected edge locations are increasing over time. We let DLT/AWS know about this via our open DLT service issue.

2021-10-07 01:42 p.m. US EDT
DLT starts a screen sharing session with TI Operations, in order to collaboratively investigate the problem. At the conclusion of the call, TI Operations requests hourly updates from DLT/AWS. DLT agrees to do this.

2021-10-07 4:21 p.m. US EDT
InCommon Operations identifies a number of edge locations which are still working correctly, and updates our service outage wiki page with information about how participants may configure their metadata clients in order to temporarily pin MDQ requests to a known good location. A method for determining if an edge location is good or bad is also supplied in this update.

2021-10-07 4:36 p.m. US EDT
DLT reports that AWS is still investigating the issue and that there is no update or estimated time to resolution. InCommon continues to update our inc-ops-notifications mailing list and the outage wiki page.

InCommon continues to request updates from DLT throughout the night of 2021-10-07 through 10-08. TI Operations staff are assigned to handle the case in shifts, continuing to prompt DLT for follow-ups throughout the night. No response or update from AWS is given until 9:30 p.m. US EDT

2021-10-07 9:30 p.m. US EDT

AWS requests relevant logs from TI Operations, via DLT. The requested logs are supplied.

2021-10-07 11:30 p.m. US EDT

AWS requests further specific logs, which TI Operations supplies via DLT.

2021-10-08 10:01 a.m. US EDT

TI Operations follows up with DLT and supplies additional logs.

2021-10-08 11:15 a.m. US EDT

DLT/AWS request information on what TI Operations' Lambda@Edge code does. TI Operations provides a description.

We continue to receive none of the requested status updates from AWS, and continue to prompt them via DLT.

2021-10-08 2:25 p.m. US EDT

DLT starts another screensharing session with TI Operations, DLT, and AWS' engineering case manager. TI Operations is informed that there appears to be an issue with a routine update to edge location infrastructure which caused a double-encoding of URL-encoded query parameters. A fix is in development but no ETA for resolution can be supplied yet.

2021-10-08 3:41 p.m. US EDT

TI Operations requests another status update

2021-10-08 3:51 p.m. US EDT

AWS informs TI Operations, via DLT, that a fix is starting to roll out to edge locations.

TI Operations tests some formerly failing edge locations, and notes that MDQ queries which had been failing, are now succeeding. We report this to AWS via DLT.

2021-10-08 4:51 p.m. US EDT

DLT informs TI Operations that AWS has fully resolved the problem with the edge servers.

2021-10-08 5:30 p.m. US EDT

After further testing, TI Operations determines that all testable edge locations appear to be responding correctly again.

TI Operations sends an issue resolution message to the inc-ops-notifications mailing list and updates the outage wiki with issue resolution status info. We also resolve the issue note on https://status.incommon.org.

Scope/impact of the outage

MDQ service requests were degraded (failing for certain requests) at the service locations of a number of InCommon participants, depending on the location and DNS resolution parameters of the metadata clients. This service degradation lasted from approximately 5:37 a.m. US EDT on Thursday, October 7, 2021, until approximately 4:51 p.m. US EDT on Friday, October 8, 2021.

Results of the investigation

TI Operations and AWS determined that the service disruption was the result of an unannounced, normal service change by AWS, which should not have affected MDQ, but some part of the AWS testing procedures did not reveal the double-encoding issue. This caused a partial MDQ service outage across a number of AWS edge locations as this update rolled out. AWS confirms that there was nothing that TI Operations could have done to prevent the issue.

Recommendations

1) The InCommon TAC should convene a working group to look into the feasibility of additional measures within the MDQ service offering portfolio which may allow different types of proactive local caching of metadata, different ways that this metadata could be served, and fallback options for future possible CloudFront/Lambda@Edge service disruptions.
2) Separately, InCommon Operations should undertake an architecture review of the MDQ service and determine if there are changes that could feasibly be made to ensure greater availability and/or independence from reliance on a sole solution provider.