GMO GlobalSign Incident Report

Outage of Japan data centre load balancing cluster

13-14th August 2019

www.globalsign.com

Background Details

At 21:15 UTC on Monday 12th August (06:15 JST on Tuesday 13th August), the redundant Load Balancer cluster in GlobalSign's Japan data centre catastrophically failed, preventing access to services presented from this data centre. These services comprise:

- GCC API
- GCC Web UI
- SiteSeal
- Timestamping (rfc3161timestamp.globalsign.com and rfc3161-timestamp.globalsign.com)

A number of internal services used for supporting customers were also affected.

Note: Timestamping services provided from Japanese soil but presented through the Singapore data centre, e.g. timestamp.globalsign.com, aatl-timestamp.globalsign.com, were not affected.

The issue was found to be a previously unknown bug in the load balancer firmware which caused high CPU and system crashes under certain edge conditions. GlobalSign's Japan Infrastructure team worked closely with engineers from the manufacturer throughout Tuesday night to apply workarounds.

The problem was resolved by a new firmware patch released by the manufacturer.

Follow-up issues

The following, more minor, problems were detected following resolution of the main outage:

Intermittent timestamping problem: A combination of intermittent load balancer issues and caching logic at one of our CDNs caused intermittent issues with timestamps served from rfc3161timestamp.globalsign.com, causing an empty response to be returned in up to 25% of cases. With CDN caching disabled for this domain, the error rate fell to 0.3%, and then was resolved by the firmware upgrade on 19th August.

Timeouts when connecting to system.globalsign.com: Some customers who had white-listed a specific IP for system.globalsign.com on their firewall received timeouts when connecting. As the workaround for the main issue involved enabling CDN caching, DNS started presenting our CDN's IP addresses instead of the origin IP, preventing these customers from connecting. GlobalSign's staff worked with affected customers to find a suitable workaround in each case.

Session expiry errors: For a brief period on 15th August, some customers received errors relating to expired sessions, through the web interface and API. This was also related to a workaround put in place to protect the service until the new firmware version was available. In this case, investigations by the Infrastructure team deemed this to be reversible without unacceptable increase to risk, so the workaround was removed.

Timeline (UTC)

13th August

21:15: Load balancer issue begins

21:54: Service outage, telephone alerts sent to Infrastructure Engineers

22:10 – 23:00: Investigation identifies the cause of the problems to be the load balancers. Remotely rebooting both primary and secondary does not resolve the issue, the Engineer investigating leaves for the data centre.

14th August

00:00 – 02:00: Infrastructure Engineer arrives at the data centre, local troubleshooting begins, looking at both software and hardware including around the load balancers (switches, cabling etc). Issue is confirmed to be related to the load balancers, and the vendor is contacted for support.

03:00: The Vendor's remote investigation shows no known issues, vendor dispatches engineers to investigate hardware replacement

05:30: Vendor's engineers arrive on-site with replacement hardware.

06:30: Vendor's engineers perform analysis of logs from active units, recommend hardware replacement.

06:30-12:00: Replacement of the two load balancers.

12:00 – 14:00: Monitoring of new hardware under load, smoke and unit tests of issue.

14:00: As problem isn't fully resolved, vendor is instructed to perform further investigations from logs and core dump.

16:00 – 18:00: "Plan B" software load balancer is built in a virtual machine and cabled to re-enable basic functionality if the hardware appliances cannot be easily recovered.

18:00 – 20:30: Investigations and unit tests demonstrate that the issue relates to a previously unknown bug in the load balancer firmware. Various workarounds are put in place to alleviate the issue and the load balancers are brought back online.

21:05: Service recovery (single load balancer) successfully brought online.

21:30: Redundant service restored.

15th August

15:00: IP-based session tracking is utilised following a number of further crashes/hangs on the load balancers, in order to prevent further outages. However, this is later rolled back due to problems reported by some enterprise customers who use pooled outbound (NAT) IP addresses.

16th August

New firmware patch received from vendor. As system is stable, decision is taken to schedule the upgrade for Monday, when the vendor's engineers can be on-site to support. An emergency plan is formulated for GlobalSign engineers to install the new firmware should any further outages or problems be experienced over the weekend.

19th August

01:00 – 02:30: Firmware replacement – under 1 minute's disruption is experienced.

Record of Actions

In order to mitigate the outage, the following actions were taken:

1) CDN caching was enabled on all endpoints (domains) being served from these load balancers.
2) The session tracking ("stickiness") mechanism was changed on the load balancers to work around the bug once discovered.
3) A firmware update was prepared by the hardware vendor and installed, mitigating the issue.

Due to the "follow-in" issues mentioned above, both mitigation #1 (for certain endpoints) and #2 were rolled back as soon as it was considered safe to do so. However, the availability of the platform as a whole was prioritised over resolution of these individual issues.

Impact and Risk Assessment

This issue caused a lengthy global outage across one of GlobalSign's primary issuance platforms, as well as timestamping services for a small number of customers, and unavailability of internal tools used to support customers.

As a large number of partners and resellers use the GCC API, the impact was felt by their customers as well.

Although this was identified to be due to a firmware bug on the hardware load balancers, the incident is not related to any issues related to the security of the load balancers or our platform.

Preventative Measures

The following areas have been identified for analysis and/or improvement:

Issue prevention

Analysis from the manufacturer indicates that this was a new, previously unknown bug, and a remediation patch was promptly issued. The Infrastructure department are constantly working to remove any single points of failure in our platforms and will be looking at how we can mitigate any future issues on this load balancer cluster.

Issue communication

Infrastructure Management is working with stakeholders across the business to ensure that internal communications always reach the relevant managers and staff, ensuring that Support staff are armed with all the relevant information in the event of a major incident such as this.

They will also be working with these managers to ensure that all staff have access to all internal status information and known workarounds (including for any related or knock-on effects) that have been provided, and have these to hand as customers call in, rather than needing to escalate and delay the response to the customer.

There is already a project in progress to make the status.globalsign.com page more dynamic, allowing Infrastructure engineers and management direct access to immediately communicate updates as they happen.