GMO GlobalSign Incident Report

Load-based issues in Singapore Data Centre

18th March 2020

www.globalsign.com

Background Details

At approximately 11:00 UTC on Tuesday 10th March 2020, GlobalSign's Singapore Data Centre started experiencing load-based issues on its external-facing ("DMZ") firewalls.

High traffic levels hitting the data centre overloaded the CPU of the firewalls, causing them to hang and fail to forward data. This caused intermittent failures of the following services:

- Certificate Issuance (GCC)
- Timestamping (non-DSS, and non-Japanese soil)
- CertSafe
- Certificate Validation services (OCSP and a small number of uncached CRLs)

Atlas services (including DSS and Atlas timestamping) were **not** affected.

A number of strategies were employed to distribute load around different parts of the data centre network, reducing reliance on those firewalls, and at the current time they have been nearly completely obsoleted. Unfortunately, as these processes took several days to resolve, the issue recurred a number of times between Wednesday 11th and Friday 13th March.

During the troubleshooting and recovery period, customers may also have experienced sporadically receiving errors ("HTTP 503" or "HTTP 530") when querying some OCSP responders, due to the way our CDNs handled the unavailability.

On Sunday 15th March the situation at the data centre was declared "mitigated and under monitoring". Further mitigation work is being undertaken to move some high-usage customers to dedicated equipment in order to prevent this issue happening again. Note that this is being performed in a safe, phased manner in order to prevent any impact to core GlobalSign services during the required maintenance.

Over the week commencing Monday 16th March, a small number of customers have been affected by knock-on effects, including:

1. IPv6-based availability of TSA services
2. Some HTTP 503 errors failing to be evicted from CDN cache

By the 18th March these knock-on issues have been considered resolved.

Timeline (all times in UTC)

Tuesday 10th March

11:00	Issue begins. Customers will experience occasional timeouts and HTTP 503 or 504 errors on Timestamping services, and sporadic issuance failures
12:30	Issue mitigated by moving some services to an alternative firewall and load balancer cluster. Some timeouts still remain, but issuance failures ceased.
13:00-21:25	Further troubleshooting and tuning work performed. GlobalSign OCSP failures reduced to 0.5%. Timestamping time-outs still occurring, but customers were given work-arounds.
21:25	Timestamping services using the VPN connection to Japan restored. Other timestamping services still taking 13-15 seconds to return as load drops off slowly.
23:15:	Situation declared stable.

Wednesday 11th March

03:15	Firewall sessions are seen increasing, then at 3:15 UTC reach the critical point, interrupting services again.
03:30 – 05:30	Troubleshooting and investigation of issue.
06:30	OCSP traffic moved over to DR firewalls.
07:30	The Security Team identify a source of malicious incoming traffic, this is blocked to reduce impact. Firewall vendor arrives on-site to help troubleshoot.
08:30	All inbound traffic routed through DR firewalls, improving service slightly. However, traffic between the DMZ and the App zone still travels over the affected firewalls, still impacting service.
09:30-11:30	The engineer from the vendor brings the DMZ firewalls back online and traffic is slowly ramped up through them, restoring service. Timestamping is still slow.
14:30-20:30	The network is largely stable, but with an elevated occurrence of HTTP 503 and 530 errors, and timestamps taking longer than usual to return (2-3 seconds). We see a high rate of certificate issuance failures.
20:30	Rate limits are put in place on the load balancers to prevent load causing issues as traffic ramps up in the morning.
20:30-00:30	Work with CertSafe customers on troubleshooting and resolving any outstanding issues they are still seeing.

Thursday 12th March

03:00	A burst of traffic at around 09:00 SGT (01:00 UTC) caused a build-up of sessions on the load balancers, which started to cause further issues in the data centre. The nature of the issue meant that the rate limits put in place on the Wednesday were ineffectual at resolving the problem.
03:00-06:00	Troubleshooting continues – further rate limits placed on connections.
06:00	System declared stable but with low throughput. Some certificate issuance failures showing, and performance issues for some CertSafe customers.
06:00-09:00	Monitoring and working on outstanding performance issues.
09:00-12:00	Inside leg of DMZ moved to DR firewalls to further reduce the reliance on the legacy units, which are still exhibiting high CPU at times.
12:00-20:00	Most services stable, some errors on some OCSP services. Ongoing investigation of service issues for CertSafe customers.
20:00	System stable, however a small number of OCSP services are intermittently unavailable, showing with a CDN error.
20:00-03:00	Further work with CertSafe customers resolving outstanding issues.

Friday 13th March

Most GlobalSign systems stable, however still some OCSP services erroring.

15:30-18:30	Some load-related latency and outages experience by CertSafe customers.

Saturday 14th March

07:00-12:30	Work on restoring full access to OCSP services which are still throwing errors.
12:30	All services now operational, and remain so over Sunday 15th March.

Monday 16th March

01:00-07:00	Some load-based outages on the CertSafe platform, leading to close monitoring throughout the day. CertSafe was moved to dedicated hardware at 04:00 on Tuesday to resolve this.

Impact and Risk Assessment

This problem caused widespread outages across the GlobalSign service estate, leading to a significant backlog on certificate issuance, and sporadic outages of various OCSP responders.

At times TSA services were unavailable or responding very slowly.

Root Cause Analysis

In-depth analysis of many tens of GB of firewall logs and packet captures has been performed to identify the root cause. The following factors have been identified as contributing to the outage, in a "perfect storm" for the data centre:

1. A significant amount of malicious traffic from apparent malware, attempting to perform certificate validation on non-existing certificate validation endpoints .Because the validation endpoints did not exist all these requests were forwarded by our CDNs directly to our data centre.
2. A sharp, unforeseen increase in origin traffic on the CertSafe platform, causing an increase in database contention, and leading to intermittent sharp rises in load balancer sockets in the TIME_WAIT state (waiting for a response).
3. As a result of #1 and #2, the CPU on the DMZ firewalls became saturated instantiating new sessions, which led to pauses in packet forwarding and further hanging of connections
4. This increase in established but hung connections caused database connection pools on some services to fill up, meaning useful sessions were then unable to establish connectivity to the database.

Preventative Measures

Direct mitigation of root causes:

- The malware traffic (#1) has been black-holed, and we are working to optimise our detection and prevention techniques for faster diagnosis and resolution in future
- The CertSafe platform has been moved to dedicated database and HSM hardware, meaning further changes to usage pattern cannot affect other services. (#2)
- The vendor was unable to find significant failings in the legacy DMZ firewalls (#3), however they have now been completely retired as a precaution, and more powerful alternatives are being used instead

Other mitigations and improvements underway:

- Some services (e.g. secure.globalsign.com) have been failed over to our UK data centre to share load better between the sites
- TSA services are being made available from multiple sites simultaneously, whereas previously they had been available from other data centres for DR purposes only
- Our Account Managers are working with customers to identify where moving to our higher capacity Atlas services is appropriate and recommended
- We are undertaking an internal review of Incident Communication, in order to improve proactive communication and customer notifications in the event of any future incidents.