

Root Cause Analysis Report

Incident Number

INC0047261

Webex Services Incident

Incident Duration

July 28, 2022, 16:55 – 19:55 UTC

Incident Details

At 16:55 UTC on July 28th, a service provider hosting some of the Webex services experienced a significant outage. Services for messaging, devices, authentication, and analytics were hosted in the affected service provider data center, which caused multiple microservices to fail. Users were unable to authenticate to Webex, register or connect to meetings consistently from Webex devices, use messaging, or log into the Control Hub administration console. The Webex status page was also hosted in the same service provider data center, which caused delayed access to the status page.

Root Cause

The Webex engineering team attempted to redirect services outside of the affected environment; however, the redirects were not successful due to the outage affecting multiple redundant service zones. Due to a core component of the service architecture becoming unavailable, the redirects to alternative zones were ineffective. Engineering was unable to successfully stop the services in the hosted datacenter due to connectivity failures. This caused services to remain unstable as device and software clients continued to send connection retries. The combination of incomplete service termination, the unhealthy data center, and the multiple client retries caused the connections to overload the available service capacity, and the capacity of the edge and microservices in the redundant environment was unable to process the traffic.

Corrective Actions

Engineering worked with the service provider to restore the core service architecture and scaled up the edge and service capacity within the environment which allowed the clients to successfully connect, and the additional traffic generated by the retries to stop. This caused services to stabilize, which allowed engineering to take additional remediation steps, including restarting unhealthy instances and additional load balancing, leading to full service recovery. Engineering completed remaining service clean-up and load balancing, and services were fully recovered at 19:55 UTC.

The service team has identified areas of improvement for our incident response time, including changes to the service metrics for faster root cause identification, improved runbooks to speed up scale-up and deployment of additional micro-services, and revisiting client retry values which will improve service restoration should a similar incident occur in the future. The messaging architecture team is also engaged to identify service architecture improvements which will allow Webex services to better withstand a multi-zone failure.

Timeline

16:55: Webex alerts indicate multiple service failures; incident process began
17:15: Engineering began service redeploys
17:33: Services fully stopped in the affected DC
17:45: Service provider reported partial recovery
18:00: Metrics indicate high traffic volume due to client retries
18:30: Multiple scale-up efforts completed
19:10: Capacity increase across both pools completed; 80% increase in traffic observed. Service redeploys began
19:45: Vendor confirms services fully restored. Additional capacity deploys completed in all three zones
19:55: Final redeploys completed; services restored.