Google Cloud Status Dashboard > Incidents > Google Cloud Networking

Google Cloud Status Dashboard

This page provides status information on the services that are part of Google Cloud Platform. Check back here to view the current status of the services listed below. If you are experiencing an issue not listed here, please contact Support. Learn more about what's posted on the dashboard in this FAQ. For additional information on these services, please visit cloud.google.com.

Google Cloud Networking Incident #19009 The network congestion issue in eastern USA, affecting Google Cloud, G Suite, and YouTube has been resolved for all affected users as of 4:00pm US/Pacific.

Incident began at **2019-06-02 11:45** and ended at **2019-06-02 15:40** (all times are **US/Pacific**).

DATE TIME

y Jun 06, 2019

09:42

ISSUE SUMMARY

DESCRIPTION

On Sunday 2 June, 2019, Google Cloud projects running services in multiple US regions experienced elevated packet loss as a result of network congestion for a duration of between 3 hours 19 minutes, and 4 hours 25 minutes. The duration and degree of packet loss varied considerably from region to region and is explained in detail below. Other Google Cloud services which depend on Google's US network were also impacted, as were several non-Cloud Google services which could not fully redirect users to unaffected regions. Customers may have experienced increased latency, intermittent errors, and connectivity loss to instances in us-central1, us-east1, us-east4, us-west2, northamerica-northeast1, and southamerica-east1. Google Cloud instances in us-west1, and all European regions and Asian regions, did not experience regional network congestion.

Google Cloud Platform services were affected until mitigation completed for each region, including: Google Compute Engine, App Engine, Cloud Endpoints, Cloud Interconnect, Cloud VPN, Cloud Console, Stackdriver Metrics, Cloud Pub/Sub, Bigquery, regional Cloud Spanner instances, and Cloud Storage regional buckets. G Suite services in these regions were also affected.

We apologize to our customers whose services or businesses were impacted during this incident, and we are taking immediate steps to improve the platform's performance and availability. A detailed assessment of impact is at the end of this report.

ROOT CAUSE AND REMEDIATION

This was a major outage, both in its scope and duration. As is always the case in such instances, multiple failures combined to

amplify the impact. Within any single physical datacenter location, Google's machines are segregated into multiple logical clusters which have their

own dedicated cluster management software, providing resilience to failure of any individual cluster manager. Google's network

control plane runs under the control of different instances of the same cluster management software; in any single location, again, multiple instances of that cluster management software are used, so that failure of any individual instance has no impact on network capacity. Google's cluster management software plays a significant role in automating datacenter maintenance events, like power infrastructure changes or network augmentation. Google's scale means that maintenance events are globally common, although

rare in any single location. Jobs run by the cluster management software are labelled with an indication of how they should

behave in the face of such an event: typically jobs are either moved to a machine which is not under maintenance, or stopped and rescheduled after the event. Two normally-benign misconfigurations, and a specific software bug, combined to initiate the outage: firstly, network control plane jobs and their supporting infrastructure in the impacted regions were configured to be stopped in the face of a maintenance event. Secondly, the multiple instances of cluster management software running the network control plane were marked as eligible for inclusion in a particular, relatively rare maintenance event type. Thirdly, the software initiating

maintenance events had a specific bug, allowing it to deschedule multiple independent software clusters at once, crucially even if those clusters were in different physical locations. The outage progressed as follows: at 11:45 US/Pacific, the previously-mentioned maintenance event started in a single physical location; the automation software created a list of jobs to deschedule in that physical location, which included the logical clusters running network control jobs. Those logical clusters also included network control jobs in other physical locations. The automation then descheduled each in-scope logical cluster, including the network control jobs and their supporting infrastructure in multiple physical locations.

Google's resilience strategy relies on the principle of defense in depth. Specifically, despite the network control infrastructure being designed to be highly resilient, the network is designed to 'fail static' and run for a period of time without the control plane being present as an additional line of defense against failure. The network ran normally for a short period - several minutes after the control plane had been descheduled. After this period, BGP routing between specific impacted physical locations was withdrawn, resulting in the significant reduction in network capacity observed by our services and users, and the inaccessibility of some Google Cloud regions. End-user impact began to be seen in the period 11:47-11:49 US/Pacific. Google engineers were alerted to the failure two minutes after it began, and rapidly engaged the incident management protocols

used for the most significant of production incidents. Debugging the problem was significantly hampered by failure of tools

handling failure of such tools, but use of these backup plans (including engineers travelling to secure facilities designed to

competing over use of the now-congested network. The defense in depth philosophy means we have robust backup plans for

withstand the most catastrophic failures, and a reduction in priority of less critical network traffic classes to reduce congestion) added to the time spent debugging. Furthermore, the scope and scale of the outage, and collateral damage to tooling as a result of network congestion, made it initially difficult to precisely identify impact and communicate accurately with customers. As of 13:01 US/Pacific, the incident had been root-caused, and engineers halted the automation software responsible for the maintenance event. We then set about re-enabling the network control plane and its supporting infrastructure. Additional problems once again extended the recovery time: with all instances of the network control plane descheduled in several locations, configuration data had been lost and needed to be rebuilt and redistributed. Doing this during such a significant

network configuration event, for multiple locations, proved to be time-consuming. The new configuration began to roll out at

In parallel with these efforts, multiple teams within Google applied mitigations specific to their services, directing traffic away

from the affected regions to allow continued serving from elsewhere. As the network control plane was rescheduled in each location, and the relevant configuration was recreated and distributed, network capacity began to come back online. Recovery of network capacity started at 15:19, and full service was resumed at

16:10 US/Pacific time. The multiple concurrent failures which contributed to the initiation of the outage, and the prolonged duration, are the focus of a significant post-mortem process at Google which is designed to eliminate not just these specific issues, but the entire class of

PREVENTION AND FOLLOW-UP We have immediately halted the datacenter automation software which deschedules jobs in the face of maintenance events. We

will re-enable this software only when we have ensured the appropriate safeguards are in place to avoid descheduling of jobs in

multiple physical locations concurrently. Further, we will harden Google's cluster management software such that it rejects such

similar problems. Full details follow in the Prevention and Follow-Up section.

14:03.

requests regardless of origin, providing an additional layer of defense in depth and eliminating other similar classes of failure. Google's network control plane software and supporting infrastructure will be reconfigured such that it handles datacenter maintenance events correctly, by rejecting maintenance requests of the type implicated in this incident. Furthermore, the network control plane in any single location will be modified to persist its configuration so that the configuration does not need

to be rebuilt and redistributed in the event of all jobs being descheduled. This will reduce recovery time by an order of

control plane, to allow an adequate window for recovery with no user impact. Google's emergency response tooling and procedures will be reviewed, updated and tested to ensure that they are robust to network failures of this kind, including our tooling for communicating with the customer base. Furthermore, we will extend our continuous disaster recovery testing regime to include this and other similarly catastrophic failures. Our post-mortem process will be thorough and broad, and remains at a relatively early stage. Further action items may be

magnitude. Finally, Google's network will be updated to continue in 'fail static' mode for a longer period in the event of loss of the

identified as this process progresses.

DETAILED DESCRIPTION OF IMPACT

Compute Engine instances in us-east4, us-west2, northamerica-northeast1 and southamerica-east1 were inaccessible for the duration of the incident, with recovery times as described above.

Instance to instance packet loss for traffic on private IPs and internet traffic:

App Engine

Compute Engine

us-east1 up to 33% packet loss from 11:38 to 12:17, up to 8% packet loss from 12:17 to 14:50. us-central1 spike of 9% packet loss immediately after 11:38 and subsiding by 12:05.

us-west1 initial spikes up to 20% and 8.6% packet loss to us-east1 and us-central1 respectively, falling below 0.1% by 12:55. us-

west1 to European regions saw an initial packet loss of up to 1.9%, with packet loss subsiding by 12:05. us-west1 to Asian

regions did not see elevated packet loss. Instances accessing Google services via Google Private Access were largely unaffected. Compute Engine admin operations returned an average of 1.2% errors.

App Engine applications hosted in us-east4, us-west2, northamerica-northeast1 and southamerica-east1 were unavailable for the duration of the disruption. The us-central region saw a 23.2% drop in requests per second (RPS). Requests that reached App

flag service_control_network_fail_open was enabled, which is disabled by default.

congestion on Google's backbone in those geographies during this same time frame.

Engine executed normally, while requests that did not returned client timeout errors. **Cloud Endpoints**

Requests to Endpoints services during the network incident experienced a spike in error rates up to 4.4% at the start of the incident, decreasing to 0.6% average error rate between 12:50 and 15:40, at 15:40 error rates decreased to less than 0.1%. A separate Endpoints incident was caused by this disruption and its impact extended beyond the resolution time above.

From Sunday 2 June, 2019 12:00 until Tuesday 4 June, 2019 11:30, 50% of service configuration push workflows failed. For the

configuration. Requests to new Endpoints services, created after the disruption start time, failed with 500 errors unless the ESP

duration of the Cloud Endpoints disruption, requests to existing Endpoints services continued to serve based on an existing

Since Tuesday 4 June, 2019 11:30, service configuration pushes have been successful, but may take up to one hour to take effect. As a result, requests to new Endpoints services may return 500 errors for up to 1 hour after the configuration push. We expect to return to the expected sub-minute configuration propagation by Friday 7 June 2019. Customers who are running on platforms other than Google App Engine Flex can work around this by setting the ESP flag service_control_network_fail_open to true. For

customers whose backend is running on Google App Engine Flex, there is no mitigation for the delayed config pushes available at

Cloud Interconnect Cloud Interconnect reported packet loss ranging from 10% to 100% in affected regions during this incident. Interconnect Attachments in us-east4, us-west2, northamerica-northeast1 and southamerica-east1 reported packet loss ranging from 50% to 100% from 11:45 to 16:10. As part of this packet loss, some BGP sessions also reported going down. During this time, monitoring statistics were inconsistent where the disruption impacted our monitoring as well as Stackdriver monitoring, noted below. As a

result we currently estimate that us-east4, us-west2, northamerica-northeast1 and southamerica-east1 sustained heavy packet loss until recovery at approximately 16:10. Further, Interconnect Attachments located in us-west1, us-east1, and us-central1 but connecting from Interconnects located on the east coast (e.g. New York, Washington DC) saw 10-50% packet loss caused by

Cloud VPN

this time.

Cloud VPN gateways in us-east4, us-west2, northamerica-northeast1 and southamerica-east1 were unreachable for the duration of the incident. us-central 1 VPN endpoints reported 25% packet loss and us-east1 endpoints reported 10% packet loss. VPN gateways in us-east4 recovered at 15:40. VPN gateways in us-west2, northamerica-northeast1 and southamerica-east1 recovered at 16:30. Additional intervention was required in us-west2, northamerica-northeast1 and southamerica-east1 to move the VPN control plane in these regions out of a fail-safe state, designed to protect existing gateways from potentially incorrect changes, caused by the disruption. **Cloud Console**

Cloud Console customers may have seen pages load more slowly, partially or not at all. Impact was more severe for customers

who were in the eastern US as the congested links were concentrated between central US and eastern US regions for the duration of the disruption. **Stackdriver Monitoring**

Stackdriver Monitoring experienced a 5-10% drop in requests per second (RPS) for the duration of the event. Login failures to the Stackdriver Monitoring Frontend averaged 8.4% over the duration of the incident. The frontend was also loading with increased latency and encountering a 3.5% error rate when loading data in UI components.

Cloud Pub/Sub experienced Publish and Subscribe unavailability in the affected regions averaged over the duration of the incident:

us-west2 did not have a statistically significant change in usage.

approximation of the error rate available at the time of publishing:

Cloud Pub/Sub

us-east4 publish requests reported 0.3% error rate and subscribe requests reported a 25% error rate. southamerica-east1 publish requests reported 11% error rate and subscribe requests reported a 36% error rate.

Additional Subscribe unavailability was experienced in other regions on requests for messages stored in the affected Cloud

regions. Analysis shows a 27% global drop in successful publish and subscribe requests during the disruption. There were two

periods of global unavailability for Cloud Pub/Sub Admin operations (create/delete topic/subscriptions). First from 11:50 to 12:05

incident, between 11:47 and 12:02 where jobs.insert API calls had an error rate of 27%. Streaming Inserts (tabledata.insertAll API

calls) had an average error rate of less than 0.01% over the duration of the incident, peaking to 24% briefly between 11:47 and

northamerica-northeast1 publish requests reported a 6% error rate and subscribe requests reported a 31% error rate.

and finally from 16:05 to 16:25. **BigQuery** BigQuery saw an average error rate of 0.7% over the duration of the incident. Impact was greatest at the beginning of the

Cloud Spanner Cloud Spanner in regions us-east4, us-west2, and northamerica-northeast1 were unavailable during the duration 11:48 to 15:44.

We are continuing to investigate reports that multi-region nam3 was affected, as it involves impacted regions. Other regions'

availability was not affected. Modest latency increases at the 50th percentile were observed in us-central1 and us-east1 regions for brief periods during the incident window; exact values were dependent on customer workload. Significant latency increases

nam-eur-asia1 had 120 ms of additional latency from 13:50 to 15:20. nam3 had greater than 1 second of additional latency from 11:50 to 13:10, from 13:10 to 16:50 latency was increased by 100 ms.

Cloud Storage

us-west2 96.2%

us-east1 1.7%

us-west1 1.2%

G Suite

Jun 04, 2019

Jun 02, 2019

Jun 02, 2019

Jun 02, 2019

Jun 02, 2019

All times are US/Pacific

07:09

16:56

14:08

14:08

US/Pacific

us-central 10.7%

southamerica-east1 79.3%

at the 99th percentile were observed:

12:02.

us-central1 had an additional 80 ms of latency between 11:50 to 13:10, from 13:10 to 16:50 latency was increased by 10 ms. us-east1 had an additional 2 seconds of latency between 11:50 to 13:10, from 13:10 to 15:50 latency was increased by 250 ms. us-west1 had an additional 20 ms of latency between 11:50 to 14:10.

nam6 had an additional 320 ms of latency between 11:50 to 13:10, from 13:10 to 16:50 latency was increased by 130 ms.

Cloud Storage average error rates for bucket locations during the incident are as follows. This data is the best available

us-east4 62.4% northamerica-northeast1 43.4% us 3.5%

The impact on G Suite users was different from and generally lower than the impact on Google Cloud users due to differences in architecture and provisioning of these services. Please see the G Suite Status Dashboard (https://www.google.com/appsstatus) for details on affected G Suite services. **SLA CREDITS** If you believe your paid application experienced an SLA violation as a result of this incident, please populate the SLA credit request: https://support.google.com/cloud/contact/cloud_platform_sla A full list of all Google Cloud Platform Service Level Agreements can be found at https://cloud.google.com/terms/sla/.

A formal incident report is still forthcoming. The network congestion issue in eastern USA, affecting Google Cloud, G Suite, and YouTube has been resolved for all affected users as of 4:00pm US/Pacific.

We will conduct an internal investigation of this issue and make appropriate improvements to our systems to help prevent or

We continue to experience high levels of network congestion in the eastern USA, affecting multiple services in Google Cloud, G

Suite and YouTube. Users may see slow performance or intermittent errors. Our engineering teams have completed the first phase of their mitigation work and are currently implementing the second phase, after which we expect to return to normal

We are experiencing high levels of network congestion in the eastern USA, affecting multiple services in Google Cloud, G Suite

and YouTube. Users may see slow performance or intermittent errors. We believe we have identified the root cause of the

congestion and expect to return to normal service shortly. We will provide more information by Sunday, 2019-06-02 15:00

We are experiencing high levels of network congestion in the eastern USA, affecting multiple services in Google Cloud, G Suite

and YouTube. Users may see slow performance or intermittent errors. We believe we have identified the root cause of the

and YouTube. Users may see slow performance or intermittent errors. We believe we have identified the root cause of the

About • Send Feedback

minimize future recurrence. We will provide a detailed report of this incident once we have completed our internal investigation.

For G Suite, please request an SLA credit through one of the Support channels: https://support.google.com/a/answer/104721

Jun 02, 2019 16:04 The network congestion issue affecting Google Cloud, G Suite, and YouTube is resolved for the vast majority of users, and we expect a full resolution in the near future. We will provide another status update by Sunday, 2019-06-02 17:00 US/Pacific with current details.

This detailed report will contain information regarding SLA credits.

service. We will provide an update by Sunday, 2019-06-02 16:00 US/Pacific.

G Suite Service Level Agreement can be found at https://gsuite.google.com/intl/en/terms/sla.html

Additional information on this service disruption has been published in the Google Cloud Blog:

https://cloud.google.com/blog/topics/inside-google-cloud/an-update-on-sundays-service-disruption

Jun 02, 2019 We continue to experience high levels of network congestion in the eastern USA, affecting multiple services in Google Cloud, G 14:59 Suite and YouTube. Users may see slow performance or intermittent errors. Our engineering teams have completed the first phase of their mitigation work and are currently implementing the second phase, after which we expect to return to normal service. We will provide an update at 16:00 US/Pacific.

and YouTube. Users may see slow performance or intermittent errors. We believe we have identified the root cause of the congestion and expect to return to normal service shortly. We will provide more information by Sunday, 2019-06-02 15:00 US/Pacific Jun 02, 2019 We are experiencing high levels of network congestion in the eastern USA, affecting multiple services in Google Cloud, G Suite 14:08

congestion and expect to return to normal service shortly. We will provide more information by Sunday, 2019-06-02 15:00 US/Pacific We are experiencing high levels of network congestion in the eastern USA, affecting multiple services in Google Cloud, G Suite Jun 02, 2019 13:45

congestion and expect to return to normal service shortly.

Jun 02, 2019 We are investigating an issue with Google Cloud Networking. We will provide more information by Sunday, 2019-06-02 13:30 12:53 US/Pacific.