Azure status history

This page contains all root cause analyses (RCAs) for incidents that occurred on November 20, 2019 or later. Each RCA will be retained on this page for 5 years. RCAs before November 20, 2019 aren't available.

Product: All

Region: All

Date: All

August 2021

8/11 RCA - Connection errors for resources leveraging Azure Front Door and Azure CDN (Tracking ID 0MQY-NPG)

Summary of Impact: Between 06:30 UTC and 09:30 UTC on 11 Aug 2021, a subset of customers leveraging Azure Front Door and Azure CDN Standard from Microsoft in Japan East, Japan West, Korea South, Korea Central and/or West US regions may have experienced intermittent HTTPS request connectivity failures when trying to reach their applications. During the incident, the average global error rate was ~2.5% and the peak global error rate was ~5%.

Root Cause: Azure Front Door and Azure CDN Standard from Microsoft serve traffic through edge locations around the world. We were in the process of rolling out a software update to prevent the use of TLS session resumption keys which were older than specific thresholds. The update followed the Azure safe deployment process and was rolling out in phases until it reached the impacted locations. A subset of edge locations in Korea, Japan, and West US were running with stale TLS resumption keys, and the rolled-out update triggered the mechanism to prevent the reuse of stale keys. However, a code defect in the rolled-out version resulted in a race condition where a few servers in the impacted locations tried to revert to a full TLS handshake. This race condition resulted in these servers dropping HTTPS requests.

Mitigation: Our monitoring detected this issue and alerted the service team. To mitigate, we removed unhealthy edge locations from serving traffic, which routed traffic to healthy edge locations. We also rolled back the update that caused the regression.

Next Steps: We apologize for the impact to affected customers. We are continuously taking steps to improve the Microsoft Azure Platform and our processes to help ensure such incidents do not occur in the future. In this case, this includes (but is not limited to):

- Deploy the fix for the regression that caused the race condition.
- Enhancements to monitoring to ensure alerting if TLS resumption is off or if the session resumption key is older than threshold.
- Enhancements in staging environments to account for additional stale TLS resumption failure modes.

Provide Feedback: Please help us improve the Azure customer communications experience by taking our survey: https://aka.ms/AzurePIRSurvey

July 2021

7/28 RCA - Azure Network Infrastructure service availability issues - Brazil Southeast (Tracking ID LNZM-TZG)

Summary of Impact: Between 13:48 UTC and 15:20 UTC on 27 July 2021, a subset of customers experienced issues connecting to their services in the Brazil Southeast region.

Root Cause: We determined that a degradation in connectivity was caused by packet loss when the metadata on one of our regional-level routers was updated incorrectly. As part of a planned network configuration refresh, an update was being performed on the regional-level routers in the Brazil Southeast region. The regional-level tier of the network is designed with redundancy to allow a subset of the routers at that network tier to be taken off-line (not serving customer traffic) for updates.

During the update, our automated network configuration system applied an incorrect IPv4 network prefix (IP Range) to a regional-level router that was taken off-line. Restoring traffic to this regional-level router resulted in packet loss for some of the Azure services in the region. The incorrect network prefix caused traffic from this region to Azure destinations in other regions and a subset of internet regions to be dropped.

Mitigation: The device with incorrect prefixes was removed from service. This mitigation took longer than expected because automated safety checks were failing for the entire region, and some human intervention was required to proceed with the traffic rollback.

Next Steps: We apologize for the impact to affected customers. We are continuously taking steps to improve the Microsoft Azure Platform and our processes to help ensure such incidents do not occur in the future. In this case, this includes (but is not limited to):

- We have audited all ongoing deployments to identify this error pattern and enhancing our validation checks to prevent such a combination of errors, including improved detection logic.
- We are continuously enhancing our alert logic to help identify issues faster and force rollbacks without human intervention.

Provide Feedback: Please help us improve the Azure customer communications experience by taking our survey: https://aka.ms/AzurePIRSurvey

7/23 RCA - Service availability issue - Brazil South (Tracking ID 5TGJ-1SZ)

Summary of impact: Between 19:00 UTC and 21:28 UTC on 23 Jul 2021, a subset of customers in Brazil South may have experienced issues connecting to a number of Azure services hosted in this region, including, but not limited to, App Service, ExpressRoute, Azure Search, and VPN Gateway.

Root cause: On 23 Jul 2021, at 19:00 UTC, we received an alert from our health monitoring that there were network connectivity issues in the Brazil South region. On initial investigation, we identified a problem in the physical network infrastructure. Upon further investigation, we found that the physical network links to some of the devices within a single cluster were disrupted during a planned maintenance activity. During this maintenance activity, a technician was executing a change that impacted the incorrect set of cables. The work that was executed on site led to the disruption of services.

Mitigation: The Data Center Operations team worked at restoring some of the physical links which restored network connectivity to operational threshold by 21:20 UTC. By 21:28 UTC most of the impacted services were recovered.

Data Center Operations team then continued to restore the remaining subset of cables and completed this activity at 00:52 UTC on 24 Jul 2021, to bring the services back through normal connectivity.

Next steps: We apologize for the impact to affected customers. We are continuously taking steps to improve the Microsoft Azure Platform and our processes to help ensure such incidents do not occur in the future. In this case, this includes (but is not limited to):

- Reviewing existing network datacenter operations cabling processes and standards.
- Revisiting and retraining our data center staff on processes.

Provide feedback: Please help us improve the Azure customer communications experience by taking our survey: https://aka.ms/AzurePIRSurvey

7/21 RCA - Azure Cognitive Services (Tracking ID 4LRL-V8G)

Summary of impact: Between approximately 07:00 UTC and 13:55 UTC on 21 Jul 2021, customers using Cognitive Services may have experienced difficulties connecting to resources. Impacted customers may have experienced timeouts or 401 authorization errors.

Root Cause: We determined that a subset of cognitive services backend compute clusters were impacted during this incident, causing them to be unable to service customer requests. Our investigation revealed that this issue was triggered by a pending (deployed, but awaiting reboot) OS security update which had introduced a networking change on the updated clusters. The networking update unintentionally caused a networking state which was not supported, and thus when the update was staged, the machines lost network connectivity and could no longer handle requests.

This issue had two separate impact streams as there was a loss of available compute clusters due to the update, and also there was impact to the control plane functionality, thus preventing the dynamic addition of resources to meet with demand. In addition, the standard auto-mitigation workstream of rebooting impacted nodes could not execute, as the nodes could not be contacted due to the networking connectivity issues on the individual nodes. The ultimate outcome was a significant impact to our ability to service Cognitive Services requests for some customers due to loss of compute nodes, and impairment of our ability to auto-mitigate using the standard frameworks.

Mitigation: Our internal monitoring detected the problem and alerted the on-call engineers. We initially mitigated customer impact by recycling the impacted nodes of the control plane which restored the self-healing capabilities of the system. This mitigation had to be applied on multiple parts of the service as the issue affected several internal subsystems (e.g., authentication, speech-to-text, post-processing, text-to-speech). We then validated that we were able to service customer requests.

Separately, an updated security patch was produced that reverted the unintentional change in behavior such that future critical security patches for this component will not trigger this failure pattern again.

Next Steps: We apologize for the impact to affected customers. We are continuously taking steps to improve the Microsoft Azure Platform and our processes to help ensure such incidents do not occur in the future. In this case, this includes (but is not limited to):

- Improvement of the validation mechanisms of update testing to ensure similar issues are detected pre-deployment
- Improvement of the patch staging/pre-reboot process to reduce the footprint of potential impact during updates
- Improvement of the cluster network connectivity detection and automatic remediation/reset processes

Provide Feedback: Please help us improve the Azure customer communications experience by taking our survey: https://aka.ms/AzurePIRSurvey

7/21 RCA - Azure Cognitive Services - Azure Government (Tracking ID 4VWK-LPZ)

Summary of impact: Between approximately 03:00 EDT and 09:55 EDT on 21 Jul 2021, customers using Cognitive Services may have experienced difficulties connecting to resources. Impacted customers may have experienced timeouts or 401 authorization errors.

Root Cause: We determined that a subset of cognitive services backend compute clusters were impacted during this incident, causing them to be unable to service customer requests. Our investigation revealed that this issue was triggered by a pending (deployed, but awaiting reboot) OS security update which had introduced a networking change on the updated clusters. The networking update unintentionally caused a networking state which was not supported, and thus when the update was staged, the machines lost network connectivity and could no longer handle requests.

This issue had two separate impact streams as there was a loss of available compute clusters due to the update, and also there was impact to the control plane functionality, thus preventing the dynamic addition of resources to meet with demand. In addition, the standard auto-mitigation workstream of rebooting impacted nodes could not execute, as the nodes could not be contacted due to the networking connectivity issues on the individual nodes. The ultimate outcome was a significant impact to our ability to service Cognitive Services requests for some customers due to loss of compute nodes, and impairment of our ability to auto-mitigate using the standard frameworks.

Mitigation: Our internal monitoring detected the problem and alerted the on-call engineers. We initially mitigated customer impact by recycling the impacted nodes of the control plane which restored the self-healing capabilities of the system. This mitigation had to be applied on multiple parts of the service as the issue affected several internal subsystems (e.g., authentication, speech-to-text, post-processing, text-to-speech). We then validated that we were able to service customer requests.

Separately, an updated security patch was produced that reverted the unintentional change in behavior such that future critical security patches for this component will not trigger this failure pattern again.

Next Steps: We apologize for the impact to affected customers. We are continuously taking steps to improve the Microsoft Azure Platform and our processes to help ensure such incidents do not occur in the future. In this case, this includes (but is not limited to):

- Improvement of the validation mechanisms of update testing to ensure similar issues are detected pre-deployment
- Improvement of the patch staging/pre-reboot process to reduce the footprint of potential impact during updates
- Improvement of the cluster network connectivity detection and automatic remediation/reset processes

Provide Feedback: Please help us improve the Azure customer communications experience by taking our survey: https://aka.ms/AzurePIRSurvey

7/12 RCA - Azure Front Door and Azure CDN - Connectivity issues and increased latency (Tracking ID 0TYG-DPG)

Summary of impact: Between approximately 14:55 UTC and 19:50 UTC on July 12th, 2021, a subset of customers primarily in the US and Canada experienced increased latency and connection timeouts when connecting to Azure Front Door and Azure CDN Standard from Microsoft. Retries may have been successful on the resources/environments that experienced issues. The peak failure rate was approximately 70% in some East US metro locations. Aggregate availability drop average across the US and Canada was around 10%. A subset of customers using Windows Virtual Desktop may have experienced connection failures when attempting to access remote applications and resources.

Root Cause: Azure Front Door and Azure CDN Standard from Microsoft service run a periodic background task to process customer configuration updates. We made a recent software update in the background process following our safe deployment guidelines. The release started on July 2nd, 2021 and was completed on Sunday morning July 11th, 2021. A software regression in this release caused intermittent CPU consumption to spike to 40-60% of CPU capacity, approximately 4 times an hour, lasting less than a minute each. This reduced the overall available capacity for processing incoming requests. During the rollout, we did not observe customer traffic availability degradation which allowed the rollouts to proceed. This was because the initial rollout was in edge locations with low traffic and overall reduced traffic volume during the July 4th week in the USA. Peak traffic increase in the busiest metropolitan areas on Monday morning in the USA initiated the incident due to the reduced capacity caused by regression. This resulted in intermittent increased latency, timeouts, and connection failures observed by customers. The Windows Virtual Desktop infrastructure services makes use of Azure Front Door. The increased latency and timeouts mentioned above caused some user requests for these services to fail.

Mitigation: Our internal monitoring detected the drop in availability and issued service alerts. Resource Health Check (RHC) system also detected the drop in availability and issued automated alerts to customers. Our first mitigation was to re-balance incoming traffic to other nearby edge locations to alleviate the issue. However, we observed that even though re-balancing traffic reduced the incident's severity, the overall CPU consumption on some edge sites remained high. We then further diagnosed the issue to be related to the recent background service update. A second mitigation was then applied to limit CPU utilization on the background service to ensure that it does not go over a threshold. The combination of re-balancing traffic and reducing CPU consumption limit for background service successfully mitigated the incident. We temporarily reconfigured the Windows Virtual Desktop infrastructure services to bypass the use of Azure Front Door during a portion of the outage period.

Next Steps: We sincerely apologize for this incident and any inconvenience it may have caused to our customers. In our constant effort to improve the stability of the Azure platform, we have undertaken the following repair items:

• Completed changes to enforce resource utilization limits to help prevent impact to capacity.

• Apply fix to regression that caused high CPU utilization.

• Improve monitoring/alerting to help improve detection of resource usage anomalies.

• Apply additional resource limits on applicable background services to help prevent capacity impact to other services.

• Fine tune resource health alerts to include a list of affected edge locations.

• Enhancements to traffic management to better redirect incoming requests from edge locations that have high CPU.

Provide Feedback: Please help us improve the Azure customer communications experience by taking our survey: https://aka.ms/AzurePIRSurvey

7/7 RCA - Service management operation failures - North Europe (Tracking ID 0_JL-9SG)

Summary of impact: Between 21:19 UTC on 07 Jul 2021 and 12:10 UTC on 08 Jul 2021, a subset of customers in North Europe may have intermittently received errors when performing service management operations for services that rely on compute resources in this region.

Root Cause:

The request queue limit on some frontends for the disk management that is responsible for service management operations in North Europe reached their limits and were rejecting a subset of requests to it.

This resulted in intermittent failures for service management requests for virtual machines and disks. The failed requests generally succeeded on retries. The issue was triggered by a platform update on the servers running the disk management service. The platform update caused expected batched failovers between the replicas of the disk management service which unexpectedly led to high latency for some calls and build-up of queues on the service front-ends. This in turn resulted in higher latency and failures for subsequent calls. The detailed sequence of events which led to this situation is under investigation.

Mitigation:

Automated service monitoring raised immediate alerts for customer impacting failures and engineers for the impacted services were engaged immediately. This was a complex situation where a number of events and changes needed to be investigated by engineers from multiple Azure components. Throughout the incident, multiple mitigations strategies were deployed with partial success. Finally, the mitigation to perform an upgrade of service frontend components fully mitigated the issue. It was once completed the frontends were re-initialized, that the issue was fully mitigated.

Next Steps

- We apologize for the impact to affected customers. We are continuously taking steps to improve the Microsoft Azure Platform and our processes to help ensure such incidents do not occur in the future. In this case we are adding instrumentation to our service code to allow quicker diagnosis and mitigation for similar issues.

Provide Feedback: Please help us improve the Azure customer communications experience by taking our survey: https://aka.ms/AzurePIRSurvey

June 2021

6/29 RCA - Azure Resource Manager - Degraded Performance managing resources (Tracking ID 1V9K-PSZ)

Summary of Impact: Between as early as 22:24 UTC on 29 Jun 2021 and 14:30 UTC on 30 Jun 2021, a subset of Azure customers may have experienced intermittent errors when attempting to access the Azure portal and other Microsoft and Azure services. Impact was observed across multiple services and regions to varying degrees. Recovery time varied by service, and most services fully recovered by 14:30 UTC. During this time, retries may have succeeded.

Root Cause: Azure Resource Manager service has a configuration to connect to a required backend storage. The connection to some of these backend storage endpoints started failing after a maintenance to update their configuration was deployed. Over time, these machines naturally restarted as needed, picking up the new configuration - which contained an issue. Once restarted, these machines would fail to connect to the storage endpoints. Over time, this led to degraded service performance in the regions where this configuration was rolled out to, as a subset of machines would become affected.

The configuration change was designed to be non-impacting and thus was rolled out in parallel, leading to impact across more than one region.

Mitigation: Once we detected the issue, we stopped the rollout of the new configuration. In order to mitigate the issue, we took unhealthy nodes out of rotation, and patched and rolled out the correct configuration in a staged manner. After mitigation at 14:30 UTC, we continued to monitor the platform to ensure stability of the service after the rollout.

Next Steps: We sincerely apologize for the impact to affected customers. We are continuously taking steps to improve the Microsoft Azure Platform and our processes to help ensure such incidents do not occur in the future. In this case, this includes (but is not limited to):

- Review of requirements regarding safe deployment for configuration changes
- Updates to prevent this component from failing the rest of the service in the event of an issue
- Adding this update scenario to automation to make configuration changes when required

Provide Feedback: Please help us improve the Azure customer communications experience by taking our survey: https://aka.ms/AzurePIRSurvey

6/14 RCA - Error notifications for service management operations - West US 2 (Tracking ID LL1H-9CZ)

Summary of Impact: Between 22:00 UTC on 14 Jun 2021 and 11:15 UTC on 15 June 2021, a subset of customers with resources hosted in the West US 2 region may have received errors while performing service management operations - such as create, update, delete - for multiple services. On 15 Jun 2021 09:20 UTC mitigation was applied, and services gradually began to recover as load from queued service management requests reduced. Full recovery was confirmed for all impacted services at 11:15 UTC

Root Cause: We established that there were several factors that contributed to this customer impact:

- One of the backend access control services specifically serving service management requests in West US 2 experienced a period of unexpected high CPU consumption, because of an anomalous spike in internal traffic. This resulted in requests to the service timing out.
- Additionally, resources hosting this service became unavailable due to a code defect in a driver. This defect manifested itself under this specific load, which both exacerbated the issue and lengthened mitigation efforts.

Due to the nature of this backend service, automatic scaling is not possible. We rely on stress-testing to predict capacity needs. We have identified that our stress tests did not account for the configuration present in this West US 2.

Mitigation: Mitigation workstreams continued over an extended period due to complications in recovering a low-level internal service under high load with the crashing driver. At 6:39 UTC on June 15, targeted network rules were introduced to block specific internal traffic on a subset of underlying backend service instances to reduce the load. At 6:51 UTC, we applied a configuration change to the infrastructure and removed the impacted driver. At 9:20 UTC, additional capacity was applied to the internal infrastructure. This allowed impacted customer-facing services to stabilize, thus mitigating the issue.

Next Steps: We sincerely apologize for the impact to affected customers. We are continuously taking steps to improve the Microsoft Azure Platform and our processes to help ensure such incidents do not occur in the future. In this case, this includes (but is not limited to):

- We will continue to investigate the cause for the anomalous traffic spike.
- Re-running stress tests across all regions to account for configuration differences across regions.
- Scaling capacity to what is predicted by the stress-tests above.
- Removing the driver causing failures during the incident.

Provide Feedback: Please help us improve the Azure customer communications experience by taking our survey: https://aka.ms/AzurePIRSurvey

May 2021

5/20 RCA - Issues accessing the Azure portal and other Microsoft services (Tracking ID KN22-39Z)

Summary of Impact: Between 06:52 UTC and 16:20 UTC on 20 May 2021, a subset of Azure customers may have experienced intermittent errors when attempting to access the Azure portal and other Microsoft and Azure services. Impact was observed across multiple services and regions to varying degree. Recovery time varied by service, and most services fully recovered by 16:20 UTC.

Root Cause: We identified a series of transient name resolution issues that impacted a subset of Azure regions. The impact was seen as follows:

- 06:52 UTC to 07:10 UTC - regions in Europe
- 09:00 UTC to 09:30 UTC - regions in India
- 15:53 UTC to 16:20 UTC - regions in Europe (primarily UK)

The name resolution issues were caused by a code regression in a recent deployment to our edge DNS servers. The regression introduced lock contention issues which, when triggered, caused some processes on our edge servers to go into a paused state and stop serving traffic for some time. The paused processes auto recovered and started service traffic again. This led to intermittent query drops and degraded service performance. During this time retries may have been successful. The issue had a low probability of being triggered and it only started manifesting itself several days after the gradual deployment completed.

Mitigation: To resolve the issue, we have rolled back the recent deployment using our safe deployment practices (SDP) to a previously known healthy state, first to the impacted regions and then globally. After mitigation at 16:20 UTC, we continued to monitor the platform to ensure stability of the service both prior and during the roll back of the deployment.

Next steps: We sincerely apologize for the impact to affected customers. We are continuously taking steps to improve the Microsoft Azure Platform and our processes to help ensure such incidents do not occur in the future. In this case, this includes (but is not limited to):

- Introducing software and hardware diversity on our DNS stack to ensure that a code regression does not impact the service resiliency.
- Improving the stress/non-functional test coverage to handle additional fault injection scenarios.

Provide Feedback: Please help us improve the Azure customer communications experience by taking our survey: https://aka.ms/AzurePIRSurvey .

◂ 1 2 3 4 ... 9 ▸