# Post Incident Report For Microsoft 365

Report Date: February 6, 2023

Report By: ICC

The information contained in this document represents the current view of Microsoft Corporation on the issues discussed as of the date of publication. Because Microsoft must respond to changing market conditions, it should not be interpreted to be a commitment on the part of Microsoft, and Microsoft cannot guarantee the accuracy of any information presented after the date of publication.

MICROSOFT MAKES NO WARRANTIES, EXPRESS, IMPLIED OR STATUTORY, AS TO THE INFORMATION IN THIS DOCUMENT.

Microsoft may have patents, patent applications, trademarks, copyrights, or other intellectual property rights covering subject matter in this document. Except as expressly provided in any written license agreement from Microsoft, the furnishing of this document does not give you any license to these patents, trademarks, copyrights, or other intellectual property.

The descriptions of other companies' products in this document, if any, are provided only as a convenience to you. Any such references should not be considered an endorsement or support by Microsoft. Microsoft cannot guarantee their accuracy, and the products may change over time. Also, the descriptions are intended as brief highlights to aid understanding, rather than as thorough coverage. For authoritative descriptions of these products, please consult their respective manufacturers.

**Microsoft 365 Customer Ready Post Incident Report**

For further information, the Azure DNS PIR can be found at https://status.azure.com/en-us/status/history/ with Tracking ID VSG1-B90.

This PIR is specific to the impact experienced by Microsoft 365 services.

## Incident Information

| | |
|---|---|
| **Incident ID** | MO502273 |
| **Incident Title** | Users may have been unable to access multiple Microsoft 365 services |
| **Service(s) Impacted** | Multiple Microsoft 365 services |

## Summary

Starting Wednesday January 25, 2023, at 7:08 AM UTC, customers may have experienced issues with networking connectivity, manifesting as network latency and/or timeouts when attempting to connect to Azure and Microsoft 365 services.

As part of a planned change to update the IP address on a WAN router, a command given to the router caused it to send messages to all other routers within the WAN, that resulted in all of them recomputing their adjacency and forwarding tables. During this re-computation process, the routers were unable to correctly forward packets. Further, this issue impacted connectivity between clients on the internet to Azure, connectivity between services in datacenters, and ExpressRoute connections.

We identified the recent change to WAN infrastructure as the underlying cause of the issue; however, the system had already begun automatic recovery. Our telemetry indicated the service was mostly recovered by 9:05 AM UTC, and continued to saturate across all regions and services, with most of the networking equipment automatically recovering at 9:35 AM UTC.

While most regions and services had recovered by 9:05 AM UTC, intermittent packet loss issues caused some customers to continue seeing connectivity issues due to two routers not being able to recover automatically. The majority of impacted Microsoft 365 services required network connectivity to be restored prior to recovering. As such, some of Microsoft 365 services may have seen a delay in recovery as we worked to expedite the mitigation of the remaining impacted services. Our telemetry indicates that a majority of Microsoft 365 services were recovered by 12:43 PM UTC.

## User Impact

- Users may have been unable to access multiple Microsoft 365 services.
- Users who could access may have experienced degraded feature functionality within the services.

Impact was to the following services:

**SharePoint Online (SPO) and OneDrive for Business (ODB)**
The majority of impact centered around users being unable to access the service as connections were unable to reach the SPO and ODB infrastructure. Telemetry shows that during the impact window between 7:08 AM UTC and 9:50 AM UTC, there was approximately a 10% reduction in the number of requests to the SPO an ODB services worldwide in comparison to the previous week.

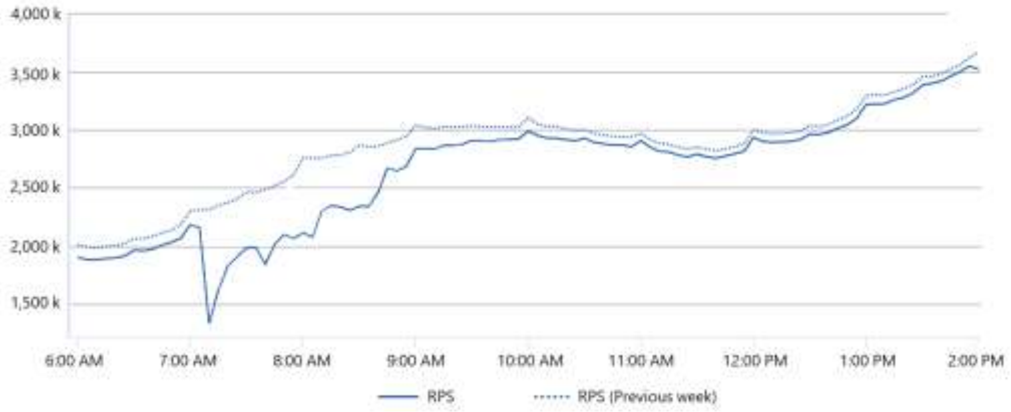

*Figure 1 – The graph shows the reduction in Requests Per Second (RPS) on Front-end components compared with the same time-period the previous week. This shows a significant drop in requests to the SPO and ODB services due to the WAN outage.*

Additionally, during the initial stages of the networking outage, users able to access the service may have experienced issues when attempting to access content or leveraging features within the service. This was seen in three waves of impact at:

- 7:09 AM to 7:22 AM UTC
- 7:42 AM to 7:47 AM UTC
- 8:27 AM to 8:29 AM UTC

This was due to internal connectivity issues within the service between User Front End (USR) and backend SQL database infrastructures caused by the wider WAN outage.

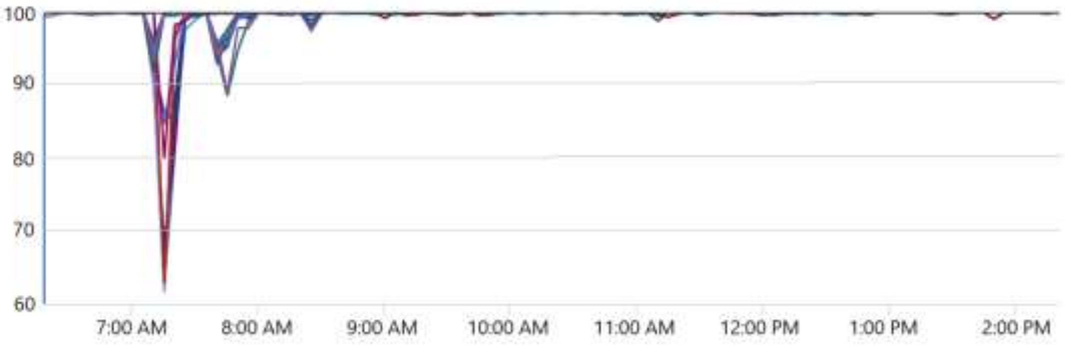

*Figure 2 – Homepage availability (%) within different SPO region infrastructures and the three separate waves of impact occurring within SPO and ODB front end components.*

**Exchange Online**

During the outage, many users were unable to connect to the Exchange Online (EXO) service as the connections failed to reach the service. Telemetry shows a significant drop in traffic in all connection methods.

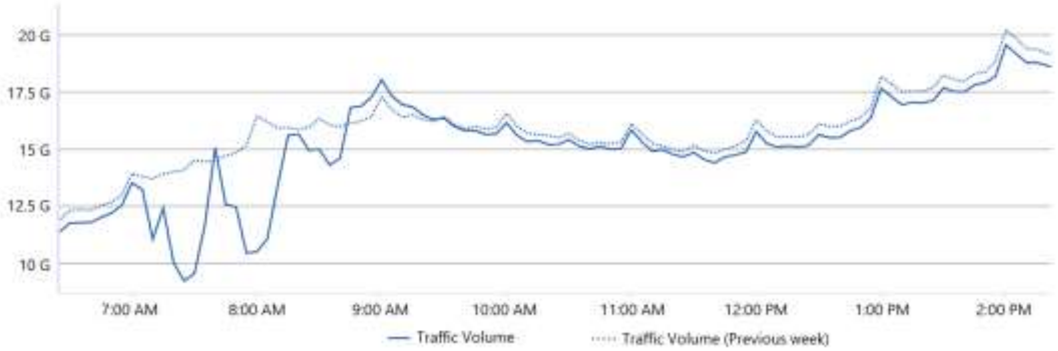

*Figure 3 – Total traffic volume to the EXO service compared with the previous week. This shows a significant drop due to the WAN outage.*

Users attempting to connect through the Outlook client recovered at approximately 9:00 AM UTC, however, other methods including Outlook on the web and Exchange ActiveSync (EAS) experienced a small amount of residual impact for an extended period of time.

The residual impact was communicated separately under EX502694 and was fully resolved by Thursday, January 26, 2023 at 3:50 AM UTC. The cause of the residual impact was due to a race condition between BitLocker and the Shared Cache service on Client Access Front End (CAFE) components after machine restarts, which caused the Shared Cache service on the component to enter a degraded state, causing increased request latency.

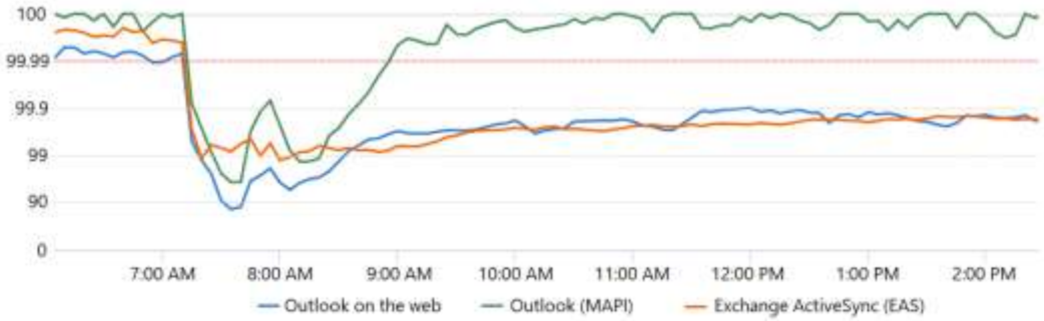

*Figure 4 – Availability (%) for different protocols connecting to the Exchange Online service. The graph shows Outlook recovering and the residual impact for Outlook on the web and EAS.*

**Microsoft Teams**
Many Teams users were unable to access or connect to the Microsoft Teams service as requests failed before they reached the Teams service.

Users able to access the Microsoft Teams may have experienced issues with:
- Chat functionality – Chat messaging, Group messaging and Channel posting
- Calling functionality – Joining meetings, failed Audio/Video calls
- Other features such as Presence, Calendar and Search

The Microsoft Teams service recovered at approximately 9:40 AM UTC.

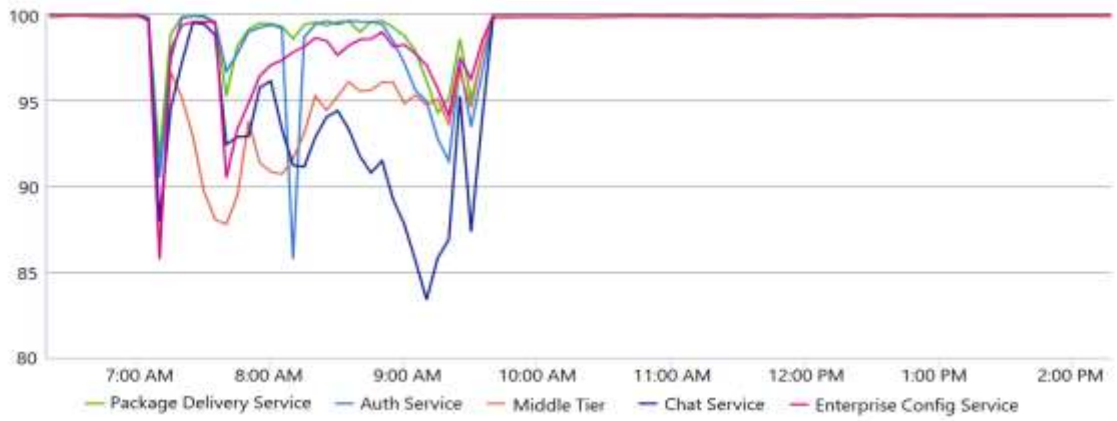

*Figure 5 – Availability (%) of core Microsoft Teams internal services.*

**Other Microsoft 365 services** which were impacted with access issues included, but was not limited to:
- Microsoft Graph
- Power BI and Power Platform
- Microsoft 365 admin portal
- Microsoft Intune
- Microsoft Defender for Cloud Apps, Identity and Endpoint.

## Scope of Impact

Any user serviced by the affected infrastructure may have been unable to access multiple Microsoft 365 services. The majority of impact occurred to users located in the Asia Pacific and Europe, Middle East, and Africa (EMEA) regions as the issue coincided with core business hours, however, users in other regions were also affected.

The initial wave of network connectivity issues occurred at 7:08 AM UTC when the operation was first performed, with auto-recovery features starting to take affect soon after. However, there was an additional wave of network connectivity impact which occurred approximately 33 minutes later, when the same operation was performed again.

## Incident Start Date and Time

Wednesday, January 25, 2023, at 7:08 AM UTC

## Incident End Date and Time

Wednesday, January 25, 2023, at 12:43 PM UTC

## Root Cause

At 7:08 AM UTC a network engineer was performing an operational task to add network capacity to the global Wide Area Network (WAN) in Madrid. The task included steps to modify the IP address for each new router, and integration into the IGP (Interior Gateway Protocol, a protocol used for connecting all the routers within Microsoft's WAN) and BGP (Border Gateway Protocol, a protocol used for distributing Internet routing information into Microsoft's WAN) routing domains.

Microsoft's standard operating procedure (SOP) for this type of operation follows a 4-step process that involves: [1] testing in our Open Network Emulator (ONE) environment for change validation; [2] testing in the lab environment; [3] a Safe-Fly Review documenting steps 1 and 2, as well as a roll-out and roll-back plans; and [4] Safe-Deployment which allows access to only one device at a time, to limit impact. In this instance, the SOP was changed prior to the scheduled event, to address issues experienced in previous executions of the SOP. Critically, our process was not followed as the change was not re-tested and did not include proper post-checks per steps 1-4 above. This unqualified change led to a chain of events which culminated in the widespread impact of this incident. This change added a command to purge the IGP database – however, the command operates differently based on router manufacturer. Routers from two of our manufacturers limit execution to the local router, while those from a third manufacturer execute across all IGP joined routers, ordering them all to recompute their IGP topology databases. While Microsoft has a real-time Authentication, Authorization, and Accounting (AAA) system that must approve each command run on each router, including a list of blocked commands that have global impact, the command's different, global, default action on the router platform being changed was not discovered during the high-impact commands evaluation for this router model and, therefore, had not been added to the block list.

Azure Networking implements a defense-in-depth approach to maintenance operations which allows access to only one device at a time to ensure that any change has limited impact. In this instance, even though the engineer only had access to a single router, it was still connected to the rest of the Microsoft WAN via the IGP protocol. Therefore, the change resulted in two cascading events. First, routers within the Microsoft global network started recomputing IP connectivity throughout the entire internal network. Second, because of the first event, BGP routers started to readvertise and validate prefixes that we receive from the Internet. Due to the scale of the network, it took approximately 1 hour and 40 minutes for the network to restore connectivity to every prefix.

Issues in the WAN were detected by monitoring and alerts to the on-call engineers were generated within 5 minutes of the command being run. However, the engineer making changes was not informed due to the unqualified changes to the SOP. Due to this, the same operation was performed again on the second Madrid router 33 mins after the first change, thus creating two waves of connectivity issues throughout the network impacting Microsoft customers.

## Actions Taken (All times UTC)

January 25, 2023

7:08 AM – Telemetry indicated that this is when impact started.

7:11 AM – Our monitoring alerted of us of a potential networking issue, and we started to review Domain Name System (DNS) configurations.

7:27 AM – We published MO502273 on the Service Health Dashboard (SHD). As there was impact to the Microsoft 365 admin center, we also published communications to https://status.office.com.

7:38 AM – The SPO and ODB failed their services off of Azure Front Door (AFD) routing to see if it would provide relief, however, this did not correct the problem.

7:40 AM – Our monitoring alerts confirmed that this issue was occurring within Microsoft WAN and not DNS.
7:44 AM – We received reports that this issue was impacting multiple services.
**FOR AZURE** – Out of an abundance of caution, we reverted a recently deployed IPv6 change which we thought may have been the cause of the issue. However, we later confirmed that this was unrelated to the issue.
8:20 AM – We identified the source of the issue and confirmed that automated recovery features had already started to recover the network, starting the recovery actions just after 8:10 AM UTC.
9:00 AM – Telemetry indicated that the majority of network devices had recovered.
9:35 AM – Our telemetry indicated that the WAN had self-recovered, and all networking equipment stabilized. After this completed, downstream Microsoft 365 services started recovering.
9:40 AM – Telemetry indicated that Microsoft Teams had fully recovered.
9:50 AM – Telemetry indicated that the SPO and ODB services had recovered.
9:51 AM – Some of the customers who had previously reported impact began reporting recovery.
10:11 AM – We identified some additional instances of region-specific packet loss in the India region.
10:27 AM – Some users may have still been experiencing residual downstream impact to Microsoft 365 services.
11:04 AM – We began reviewing options to restart affected application pools to mitigate downstream impact.
11:30 AM – Our telemetry indicated that the impact was no longer occurring for most customers. We continued to take mitigation actions to ensure full recovery.
11:58 AM – We monitored telemetry which continued to show that the service was stable, and the majority of users were able to access the service successfully.
12:25 PM – After confirming access to the admin center had been fully recovered, we closed the communication on https://status.office.com and directed affected organizations to view further details on the issue through MO502273 in the Microsoft 365 admin center.
12:43 PM – The majority of services had recovered at this time, the overall Microsoft 365 service was stable, and packet loss had returned to normal levels. Telemetry indicated that there was a small amount of residual impact for Outlook on the web connectivity and engineers continued to take actions to investigate and mitigate impact.
1:30 PM – We continued to investigate the residual impact for Outlook on the web and in parallel, restarted application pools on mailbox components to provide relief.
2:10 PM – After an extended period of monitoring, the majority of the Microsoft 365 services remained stable. We made the decision to close MO502273 as resolved and opened a new post under EX502694 to provide updates on the ongoing work to recover the residual impact for Outlook on the web connectivity.

## Next Steps

| Findings | Action | Completion Date |
|---|---|---|
| There were two main factors contributed to the incident:<br>1. A change was made to a standard operating procedure that was not properly revalidated and left the procedure containing an error and without proper pre- and post- checks.<br>2. A standard command that has different behaviors on different router models was issued outside of standard procedure that caused all WAN routers in the IGP domain to recompute reachability. | Audit and block similar commands that can have widespread impact across all three vendors for all WAN router roles. | February 2023 |
| | Publish real-time visibility of approved-automated and approved-break glass, as well as unqualified device activity, to enable on-call engineers to see who is making what changes on network devices. | February 2023 |
| | Continued process improvement by implementing regular, ongoing mandatory operational training and attestation of following all SOPs. | February 2023 |
| | Audit of all SOPs still pending qualification will immediately be prioritized for a Change Advisory Board (CAB) review within 30 days, including engineer feedback to the viability and usability of the SOP. | April 2023 |

Outside of the specific Azure networking repair items and Next Steps, Microsoft 365 services are also making improvements to the resiliency of the service. Some of them are outlined below:

| Findings | Action | Completion Date |
|---|---|---|
| A small amount of residual impact occurred within the EXO service due to a race condition impacting the Shared Cache service on routing components. | We're deploying a fix to address the race condition between BitLocker and the Shared Cache service after a restart, to prevent this type of issue reoccurring. | February 2023<br>(99.48% as of February 6, 2023) |
| | We're deploying multiple changes in detection and routing logic to prevent components in a known unhealthy state, such as with a degraded Shared Cache service, from receiving production traffic to stop this type of impact happening again. | March 2023 |

| Findings | Action | Completion Date |
|---|---|---|
| | We're fixing an additional bug that we identified that caused unexpected machine reboots when a machine experiences low memory conditions, preventing the scenario where the race condition could have occurred. | February 2023 |
| SharePoint Online and OneDrive for Business monitoring triggered a large number of alerts for individual farm availability issues, rather than detect a wider regional issue quickly. | We're improving existing logic within our SPO and ODB monitoring systems and updating alerting thresholds targeted at unexpected regional changes in Request Per Second. | March 2023 |
| Microsoft Teams monitoring triggered a large amount of alerts for both different regions and specific features, rather than detect a wider service-wide issue. This made internal Incident-Management processes not as efficient as designed. | We're reviewing options on how to cut down overall alerting noise during wide outages and improve the Incident Management process. | May 2023 |
| We identified that some of the anomaly detection systems for specific Microsoft Teams features were slightly delayed. | We're improving the existing monitoring logic so that future incidents impacting those specific areas would be detected more quickly. | March 2023 |
| During the WAN outage, some Microsoft Teams service-side management operations could not be performed due to access issue into Azure-based portals. This would not have mitigated impact for Microsoft Teams sooner in this scenario, however, it is a key learning for improving our ability to mitigate future incidents earlier. | We're reviewing options for improved automation and resiliency options to perform management operations during this type of issues. | May 2023 |