Post-Mortem Analysis

Network Outage on Saturday, September 6th

Management summary

On Saturday, September 6th, a critical network outage occurred in our data center infrastructure, resulting in full loss of internet connectivity for several hours. The root cause was linked to a malfunction of the edge switch operated by our data center provider, specifically related to misbehavior of the Spanning Tree Protocol (STP). Despite redundant systems in place, packet loss and traffic disruption persisted until on-site intervention by our DevOps team resolved the issue.

We are actively working with the data center provider to clarify the root cause and ensure more robust escalation and support processes. Internal and external measures have been initiated to prevent recurrence.

Timeline of Events

All timestamps provided are CEST (UTC+2)

- 15:33 – Our DevOps team observes network issues in our monitoring systems.
- 15:45 – Incident report created with our data center operator regarding network connectivity issues.
- 15:58 – Direct contact is established between our DevOps team and the Network Operations Center (NOC) of our data center operator.
- 17:18 – Partial restoration of internet connectivity after rebooting the edge switch at the data center. We suspect Spanning Tree Protocol (STP) issues on this switch.
- 17:20 – Notable increase in packet loss observed.
- 17:20–18:15 – Direct and detailed communication ongoing between our team and the NOC at the data center.
- 18:33 – Ticket escalated to the head of the NOC. Our CEO, Thomas Eisenbarth, engages directly.
- 19:20 – Network issues persist. DevOps team members are en route to the data center to implement on-premise solutions.

- 20:30 – DevOps team arrives on-site and begins applying physical changes to the network infrastructure.
- 20:45 – Network connectivity stabilizes.
- 21:00 – DevOps team works on stabilizing infrastructure and hosted applications.
- 22:00 – Recovery of services completed. Ongoing monitoring confirms stable infrastructure and application availability.

Technical Background and Root Cause

Our network infrastructure is designed with redundancy: each server connects to two makandra-owned switches, which are redundantly connected to an edge switch provided and operated by the data center operator. This edge switch connects to the data center's core network and the internet. It also represents the demarcation point between our infrastructure and the data center operator's responsibility.

At 15:33, the edge switch at the data center began blocking traffic from our equipment. All our infrastructure remained operational, which we verified through a fallback server connected directly to the data center's switch. Our suspicion quickly focused on the data center's edge switch filtering or blocking our traffic.

The root cause involved the Spanning Tree Protocol (STP) being activated on the edge switch. STP prevents network loops by selectively disabling redundant paths. It ensures there is only one active route between two network devices, providing stability and fault tolerance. However, in this case, STP unexpectedly disabled both uplinks to our switches, isolating our infrastructure from the internet.

After requesting a reboot of the edge switch at 17:15, partial connectivity was restored at 17:18. However, high packet loss persisted. Based on our analysis, we believe the following occurred:

- STP was manually disabled before the reboot, as it had previously blocked both uplinks.
- After rebooting with STP disabled, both links to our switches remained active.
- This led to a **broadcast storm**, causing looping traffic and severe packet loss.
- At 20:45, after disconnecting one of the two uplinks, the network stabilized immediately.

Why STP was triggered initially or why it failed to manage the redundancy properly remains unclear. We are in close coordination with our data center operator to identify the underlying cause.

Lessons Learned and Preventive Measures

1. Improved Escalation Procedures:
We will work with our data center operator to ensure faster escalation paths and guaranteed access to senior-level engineers in future incidents.

2. Network Architecture Enhancements:
We are evaluating the option to connect directly to the core network at the data center to avoid the edge switch as a potential single point of failure.

3. Enhanced Incident Communication:
We will streamline internal communication and provide more real-time technical context to all stakeholders during critical incidents, e.g. by our status website https://status.makandra.de

4. Leadership-Level Involvement:
Our executive team, including the CEO, was actively involved in escalating and resolving the issue and will continue to support strategic improvements going forward.

Final Statement

We sincerely apologize for the disruption caused by this outage. We understand the severity of the impact and take full responsibility for the delay in resolution.

This was the most critical infrastructure failure we've experienced since our founding in 2009. While rare, it clearly demonstrated weaknesses in response time and coordination with our provider that we are now addressing directly and urgently.

We are committed to transparency, accountability, and investing in long-term resilience—both technically and operationally.

Thank you for your continued trust.