Cloud-IAM's Scaleway infrastructure outage

Summary

Today, a major outage on Scaleway private network and internal DNS has impacted several deployments and Cloud-IAM internal infrastructure.

From a customer perspective, the consequence was the unavailability of certain deployment.

Severity: Major, customer were not able to connect to their deployment

Time to identification: 6 minutes

Time to first corrective action: 7 minutes

Time to resolution: ~2h30

Recovery & Mitigation

- Hard reboot of the nodes to completely reset the network stack
- Disaster recovery for deployment that were too long to recover

Timeline

- 10:07 CET The first deployment became unavailable. The deployment is not able to access its database, causing LOGIN error and triggering the on-call team.
- 10:13 CET More deployments become unavailable. They are all running on Scaleway.
- 10:13 CET: The on-call team makes the link between the incidents and the status of Scaleway https://status.scaleway.com/incidents/2qn5s5sq74m5
- 10:13 CET: The on-call team restarts various deployment nodes to mitigate the impact for the customer
- 10:41 CET: Scaleway announces a fix has been implemented, but deployments still have the network issue
- 11:37 CET: Scaleway announces a second fix has been deployed
- 12:35 CET: The on-call team was able to restart all the deployments, the situation is back to normal

What happened

There was an outage on the private network and internal DNS from Scaleway. This directly impacted the deployment which were not able to renew their DHCP lease, leading to the loss of network connectivity in the private network and also losing access to the database and finally to the internet.

This prevents the cluster from being able to:

- read data from database
- writing events to the database
- custom extensions to access an external service
- properly replicating the sessions across the nodes

All these issues led to cluster instability or unavailability.

During this period, the internal Cloud-IAM infrastructure was also impacted, making the observability and the pipelines used to apply corrective actions slow and unstable.

External messaging

• 11:38 CET The incident is reported on Cloud-IAM status page

Short term planned actions

- 1. Expose on the Cloud-IAM status page (https://status.cloud-iam.com) a section with the cloud provider overall status.
 - This will ease the announcement of outage that might impact deployments running on this cloud provider
- 2. Run additional resources (observability, runners) on a different cloud provider to avoid being blocked because of an incident on Scaleway