Reason for Outage Report

_____

Primary Outage

On Saturday april 30 at 21:43 CEST our utility provider suffered downtime on one of their transformers. Alarms were quickly raised on our end and we immediately checked to make sure that power generation was running. We were able to confirm that it was. Staff were dispatched to ensure successful failover of power generation. When arriving at the datacenter at 22:45 CEST, we found that power to the site had been lost a few minutes earlier.

Upon investigating we found that the device responsible for switching over power from utility to power generation failed even though backup generation was running. Power was restored by switching over the system from utility power to backup power manually.

Subsequent effects

After power was restored we found that network connectivity was still not working for several customers. As it turned out some network devices were having issues coming back up, and some had either lost their configuration completely or simply did not boot.

Work was immediately started to replace failed switches or restore their configuration. The majority of our time restoring services was spent in troubleshooting networking and doing manual interventions for servers that did not come back up.

Improving networking

Because we had a power outage our routing equipment went down. For customers not in our own datacenter, this meant that their network was not available even if their servers were up and running.

We will look at ways of improving our routing infrastructure so that the loss of a single site will not cause downtime for other sites.

In our data centers we make use of a common spine/leaf architecture. For many services we deploy a leaf switch that communicates with multiple redundant spines. Most of the switches we ended up having issues with were leaf switches. For customers connected to those, they were effectively down.

We found that even though we did have backups of all configuration data, resolving these issues took away a lot of engineering time that could have been spent on resolving individual customer issues. In order to improve this we will be looking at switch automation and zero-touch provisioning of switches. Meaning that if a switch were to fail we could very quickly pull it out, and simply have it boot into new configuration automatically with very little manual intervention other than replacing the device id in our provisioning system.

Reason for Outage Report

_____

Improving communications

During an event like this, communications are always difficult. Our customers are understandably worried about their services and when they can be back up and running. On our end we needed to spend as much time as possible resolving issues and communicating issues that we were made aware of for customers to onsite staff so that they could be resolved. We will work on improving our communication routines to make sure customers have as much information as we can provide while not tying up onsite staff in providing constant updates so that issues can be resolved as quickly as possible.

Improving power resiliency

In order to understand why power was lost, it is helpful to understand how backup power generation works. When utility power is lost there are microcontrollers that communicate with backup power generation and the breakers in the distribution panels.

When power is lost these controllers ask backup power to start. When the backup system has started it will turn off the utility power breakers and turn on backup power breakers.

This transitions the power feed from utility to backup power. In our case this transition did not happen. We are working with our electrical contractor to work out why this did not happen and to prevent a similar occurrence in the future.

Furthermore we will look at making breakers remotely controllable which they currently are not in our case.

Summary

Last, but certainly not least, we want to apologize. We know our services are critical to our customers' businesses. Over the coming weeks we will spend a lot of time investigating the chain of events in detail in order to improve our understanding of what happened and improve infrastructure reliability.