



Toolforge webservices are in the final stages of [migrating to the toolforge.org domain](#).  
Please help us clean up older documentation referring to tools.wmflabs.org!

# Incident documentation/20200224-esams

[< Incident documentation](#)

**document status:** final

## Contents [\[hide\]](#)

- 1 [Summary](#)
  - 1.1 [Impact](#)
  - 1.2 [Detection](#)
- 2 [Timeline](#)
- 3 [Conclusions](#)
  - 3.1 [What went well?](#)
  - 3.2 [What went poorly?](#)
  - 3.3 [Where did we get lucky?](#)
  - 3.4 [How many people were involved in the remediation?](#)
- 4 [Links to relevant documentation](#)
- 5 [Actionables](#)

## Summary

A router hardware failure caused widespread issues. While a single hardware failure like that shouldn't cause issues, more hardware failure had preceded this one, lowering our capability to absorb failures.

## Impact

All services were affected for all types of users for large portions of Europe and everyone reaching our projects via Amsterdam. That includes parts of Asia and Africa. It is difficult to gauge the exact numbers as it was highly related to the network path from the user to our infrastructure.

Eyeballing the [traffic graph](#) for the interval, we lost about 47M queries over an interval of approx. 20 minutes.

## Detection

Humans detected this first. Icinga alerted 1 minute later with pages for cr2-esams arriving to multiple SREs

## Timeline

**All times in UTC.**

- 13:55:55 Pages for cr2-esams arrive at multiple SRE people
- 13:56:37 bast3004 is also down for multiple SRE.
- 14:00:56 XioNoX: cr2-esams: 2020-02-24 13:52:02 UTC Major FPC 0 Major Errors - Lkup Error code: 0x40038
- 14:00:04 XioNoX: looks like a linecard failure as well
- 14:03:37 \_joe\_: esams depooled
- 14:12:05 paravoid: what \*works\* now?
- 14:14:06 paravoid all cr2-esams interfaces are down
- 14:15:22 paravoid the links to asw2 are et-1/0/0 & et-1/0/1, and show up as up, but are on FPC 1
- 14:15:42 \_joe\_: now we're in the business of restoring esams
- 14:18:36 bblack: normal volume on vfe-reported GET was 86K/s before the cr2 hit, then it was around 40K just before dns depool, latest sample is ~5K and dropping
- 14:19:41 paravoid: so we have two MX480s both with major FPC errors after a JunOS upgrade
- 14:19:49 paravoid: and a site down
- 14:19:52 XioNoX: paravoid: cr3 didn't get upgraded
- 14:20:23 bblack charting codfw+eqiad+esams aggregate reqs (those involved in the shuffle), we look to be

[Main page](#)  
[Recent changes](#)  
[Server admin log \(Prod\)](#)  
[Server admin log \(RelEng\)](#)  
[Deployments](#)  
[SRE/Operations Help](#)  
[Incident status](#)

[Cloud VPS & Toolforge](#)

[Cloud VPS documentation](#)

[Toolforge documentation](#)

[Request Cloud VPS project](#)

[Server admin log \(Cloud VPS\)](#)

[Tools](#)

[What links here](#)

[Related changes](#)

[Special pages](#)

[Permanent link](#)

[Page information](#)

[Cite this page](#)

[Print/export](#)

[Create a book](#)

[Download as PDF](#)

[Printable version](#)

back at "normal" total reqs as expected, at least roughly <https://w.wiki/J65>

- 14:28:24 XioNoX: akosiaris: Service Request ID 2020-0224-0258 has been created.
- 14:35:08 marostegui: so, are we fully up? (trying to document stuff)
- 14:35:43 akosiaris: marostegui: as far as I know and as far as users are concerned, yes
- 14:36:00 Outage reported as having ended.
- 14:42:01 paravoid: esams has two MX480s, and both failed today
- 14:42:13 paravoid: eqiad and codfw also have two MX480s for core routers
- 14:48:42 mark: so for RMA
- 14:48:46 mark: we can't swap line cards at esams
- 14:48:49 mark: the PDUs are in the way :)
- 21:40 volansl bblack, XioNoX: just noticed that the CA app email still complains about ns2 IP and in effect I cannot reach ns2
- 22:20 volansl Prefix 91.198.174.0/24: NOT ADVERTISED since 2020-02-24T22:20:13.69323Z, on\_demand=True
- 22:50 XioNoX so traffic does internet -> cr3-knams -> cr3-esams -> asw2 -> server
- 22:59 XioNoX bblack, volans, ns2 is now redirected to eqiad

## Conclusions

- esams is a SPOF effectively.

### What went well?

- Automated monitoring detected the incident, outage was understood quickly enough
- esams was depooled quickly.
- The `geo-maps-esams-offline` Geographic DNS configuration, created back in January, proved itself well here: we were able to leave esams depooled for some time without overheating eqiad, which had been a problem in the past when esams was depooled during peak traffic times of day. Instead, this map shifts most of eqiad's usual North American traffic westwards to codfw (and some codfw traffic further westwards to ulsfo).

### What went poorly?

- We were unlucky enough to have 2 different hardware failures in 2 different routers coincide.

### Where did we get lucky?

- It happened during a timeframe where a lot of SREs were awake and fresh.

### How many people were involved in the remediation?

- 8 SREs, 2 SRE directors.

## Links to relevant documentation

- [https://wikitech.wikimedia.org/wiki/Network\\_monitoring#host\\_\(ipv6\)\\_down](https://wikitech.wikimedia.org/wiki/Network_monitoring#host_(ipv6)_down)

## Actionables

- Restricted phabricator tasks [T246009](#) and [T245825](#) have been filed for replacing the FPCs on the routers after RMA process was recommended by the vendor. Tasks are restricted per the usual policy for vendors

Category: [Incident documentation](#)

This page was last edited on 27 April 2020, at 14:57.

Text is available under the [Creative Commons Attribution-ShareAlike License](#); additional terms may apply. See [Terms of Use](#) for details.