Oh my god everything is on fire

A story about how we handle critical issues for maximum sanity and minimum panic.

Me

@andrewhowdencom

Anything useful will be posted to Twitter.

Software engineer @ Sitewards

--

Bored of Magento so I got into sysadmin and it's kind of fun.

You (hopefully)

I am new developer

Target Audience

I am incident god

Please Interrupt.

Please Contribute!

Thanks <3

Anton Boritskiy · Sitewards · Google SRE book authors

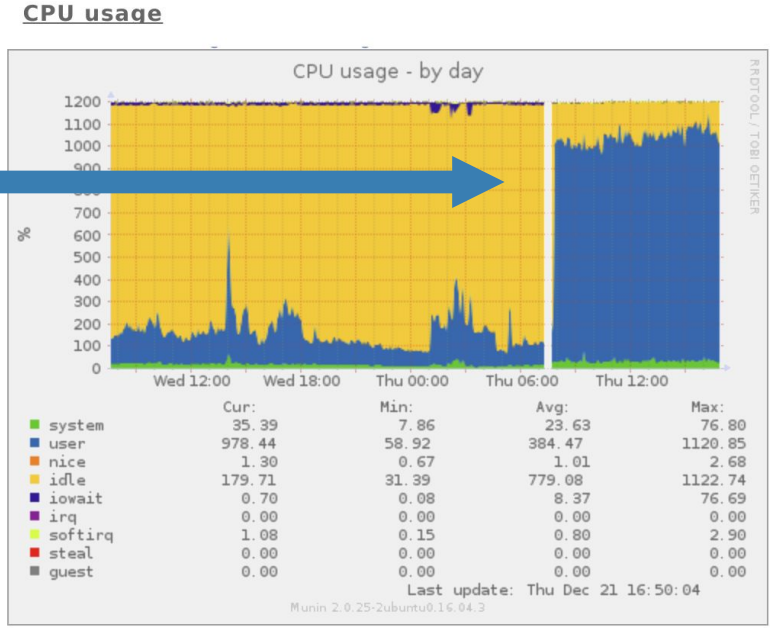
So who here's broken production?

Our Problem

(Your Example)

“Twas the night before Christmas…”

?!



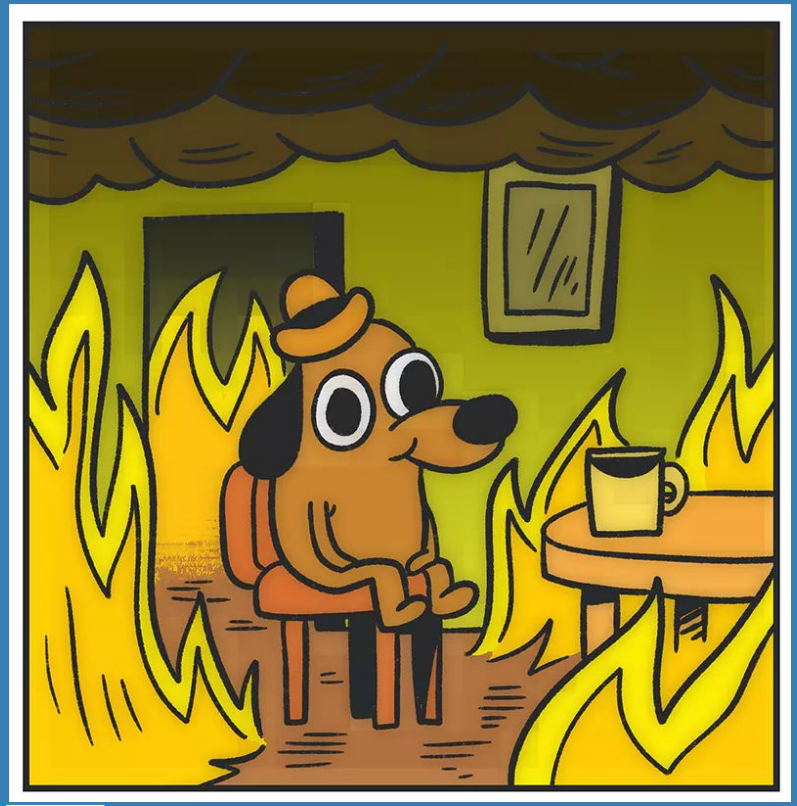
A wild cryptominer appeared

```
$ cat /etc/crontab | grep monero
*/5   *    *    *    *    root   /usr/sbin/monero
```

At least (2) problems

Problem #1

Production is borked



So what's the second problem?

Problem #2

Panic

Lots of panic panic

Developer

Oh god what did I do plz no one yell at me

Merchant

Oh god what happened to the shop I can do nothing

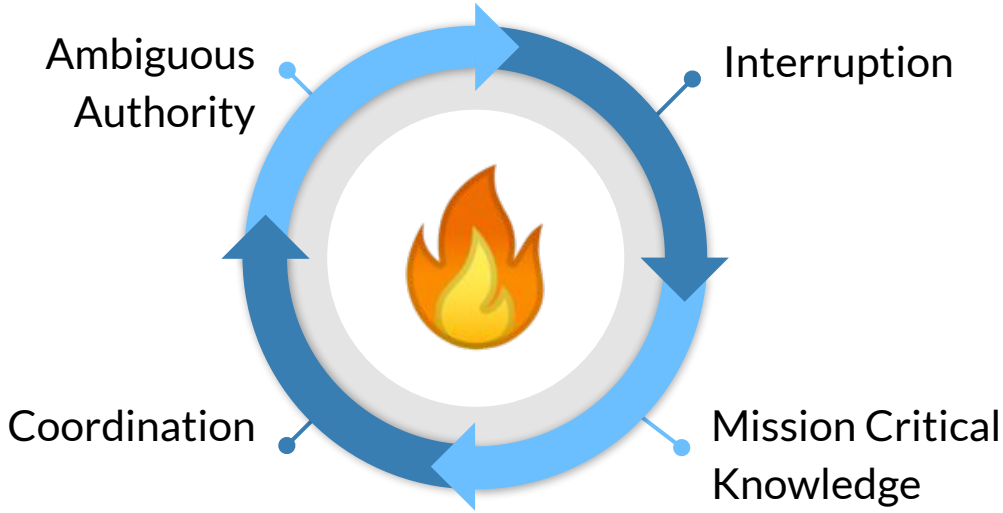
Project Manager

Oh god merchant will shortly be yelling at me

Agency

Oh god merchant will be mad and we need moneys

The many flavours of panic



Solution!

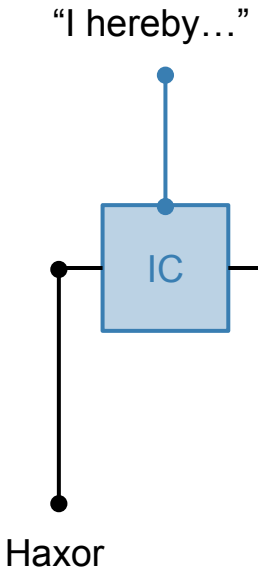
“I hereby declare this an incident”

The Goal

✗ Make production safe
✗ Find out how and why
? No one kills each other

Step #1: Declare, delegate

Timeline



Incident controller (IC)

Responsible for assigning roles, ensuring accountability.

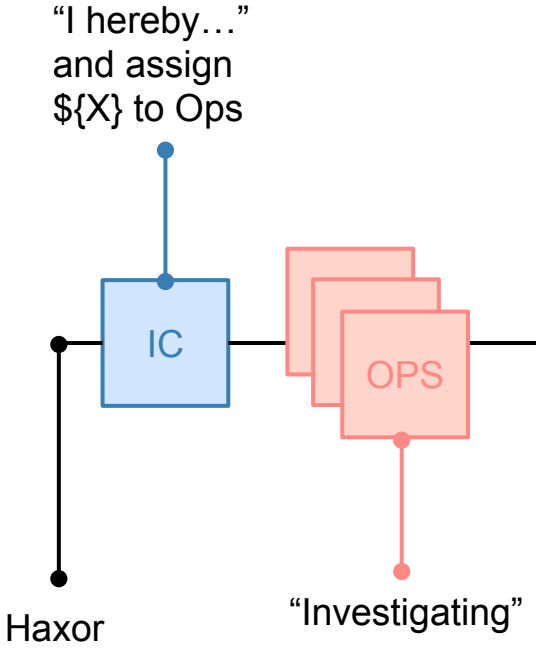
- Assigns other roles
- Creates centralised communication
- Has the authority over all
- Can add and remove team members as required
- Writes the post mortem

- Cannot investigate or resolve

Incident Controller

- Opens a new slack channel #incident--2018-12-23
- Summarises what's known about the problem
- Tasks ${DEVELOPER_A}, ${DEVELOPER_B} as **operations** to start investigating
- Tasks ${PROJECT_MANAGER} as **communications** to send an email to client
- Opens a Google doc to start writing the post mortem, shares that in slack

Timeline



Operations (Ops)

Responsible for investigating and resolving the issue.

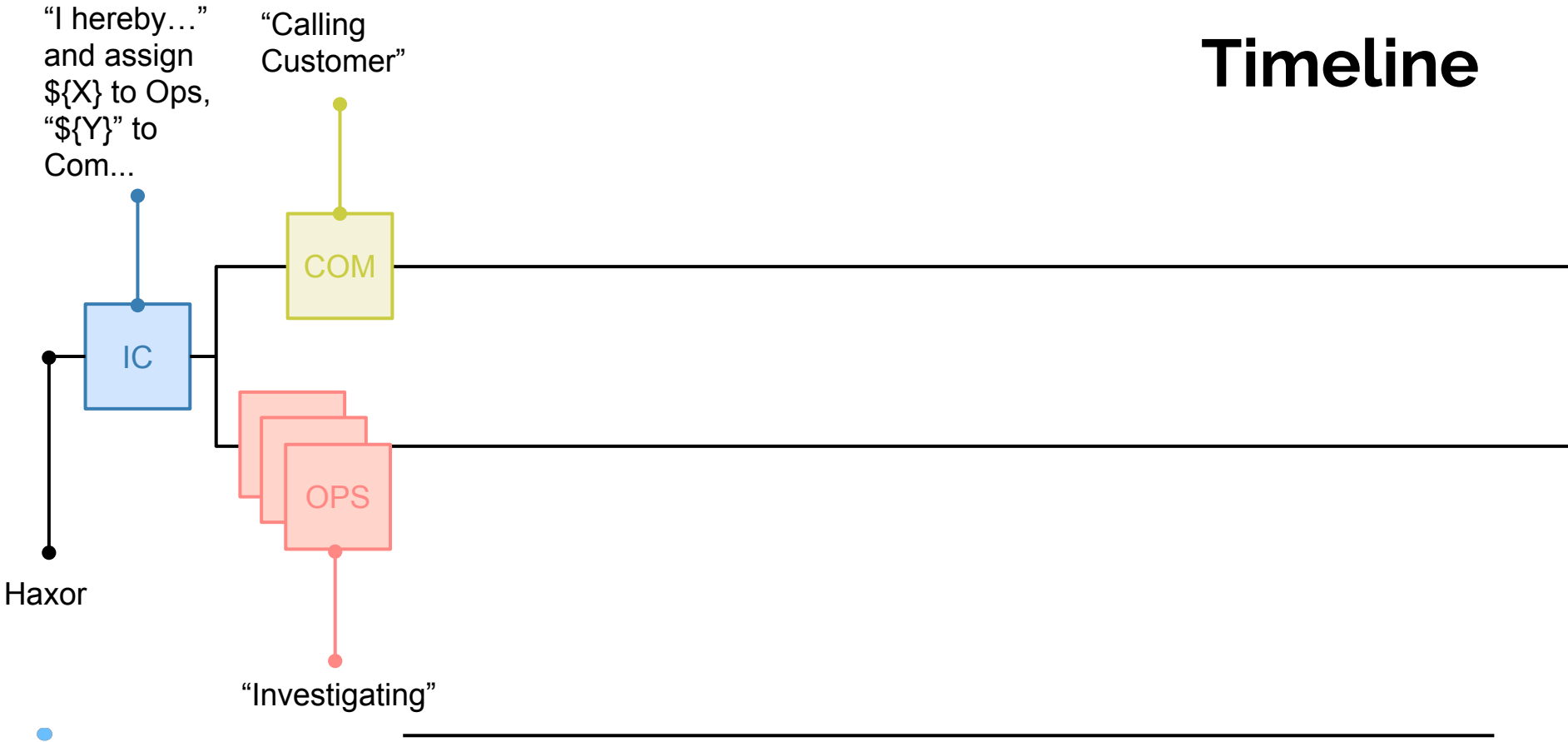
- Investigates
- Resolves
- Nothing else. Especially no communication!

- There may be more than one

Operations

- Removes the bitcoin miner
- Starts looking through the system for other indicators of compromise
- Starts looking through the logs to determine how system was compromised
- Posts various graphs, logs, and other status updates as attachment sin slack

Timeline


“I hereby…” and assign ${X} to Ops, “${Y}” to Com...
“Calling Customer”
COM
IC
OPS
Haxor
“Investigating”

Communication

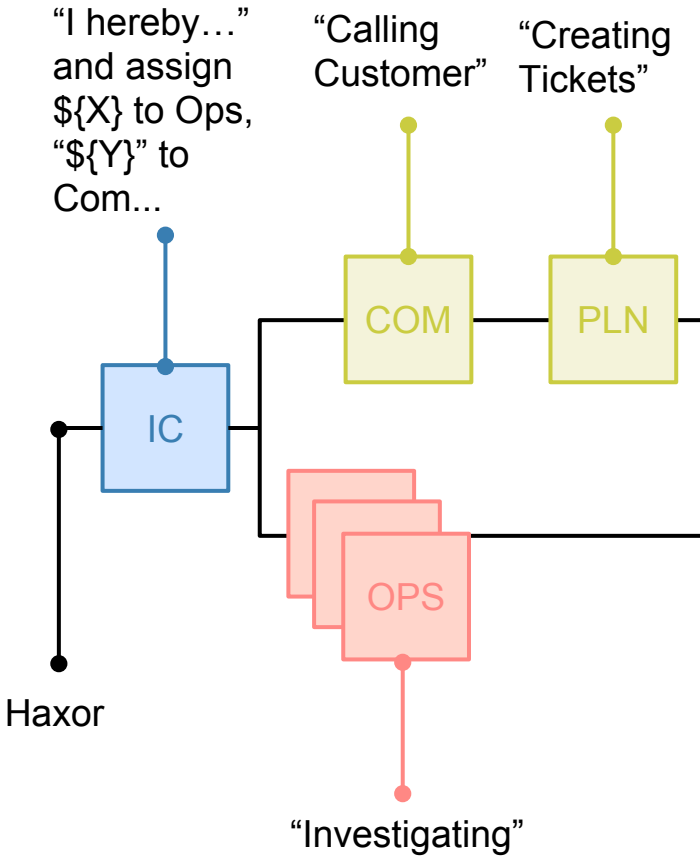
Talks to people

- Proactively informs client, business, other stakeholders
- Sets up points of contact with the stakeholders
- Acquires additional knowledge “did X change Y”

Communication

- Emails the client indicating system has been compromised
- Summaries the detail
- Answers common questions (“how is this possible” “what is next” “what do I do”)
- Requests further information as requested by ops

Timeline


“I hereby…” and assign ${X} to Ops, “${Y}” to Com...
“Calling Customer”
“Creating Tickets”
COM
PLN
IC
OPS
Haxor
“Investigating”

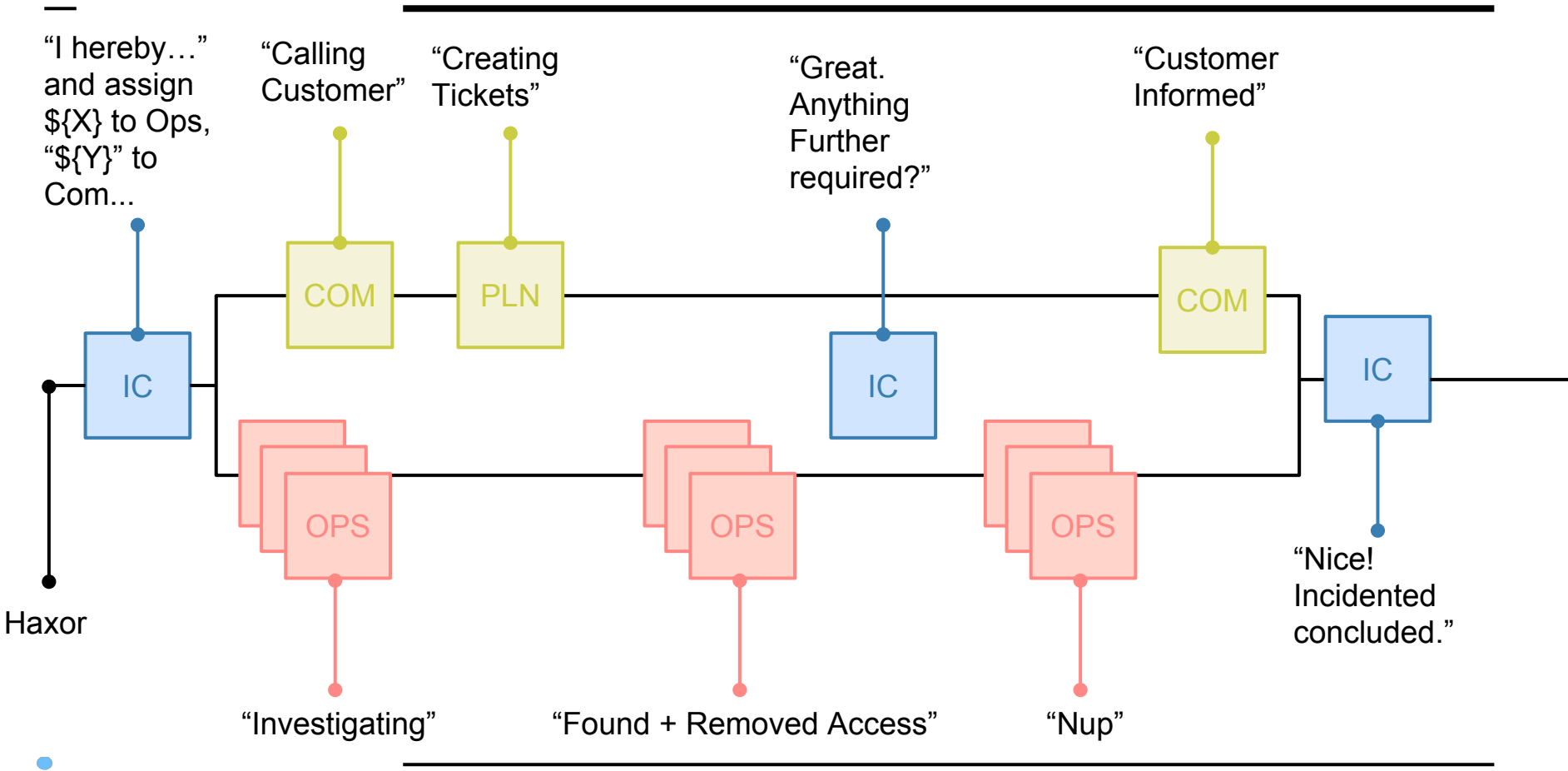
Planning (Pln)

Ongoing issue follow up

- Plans work to permanently fix an issue into sprints
- Plans follow up for post mortems

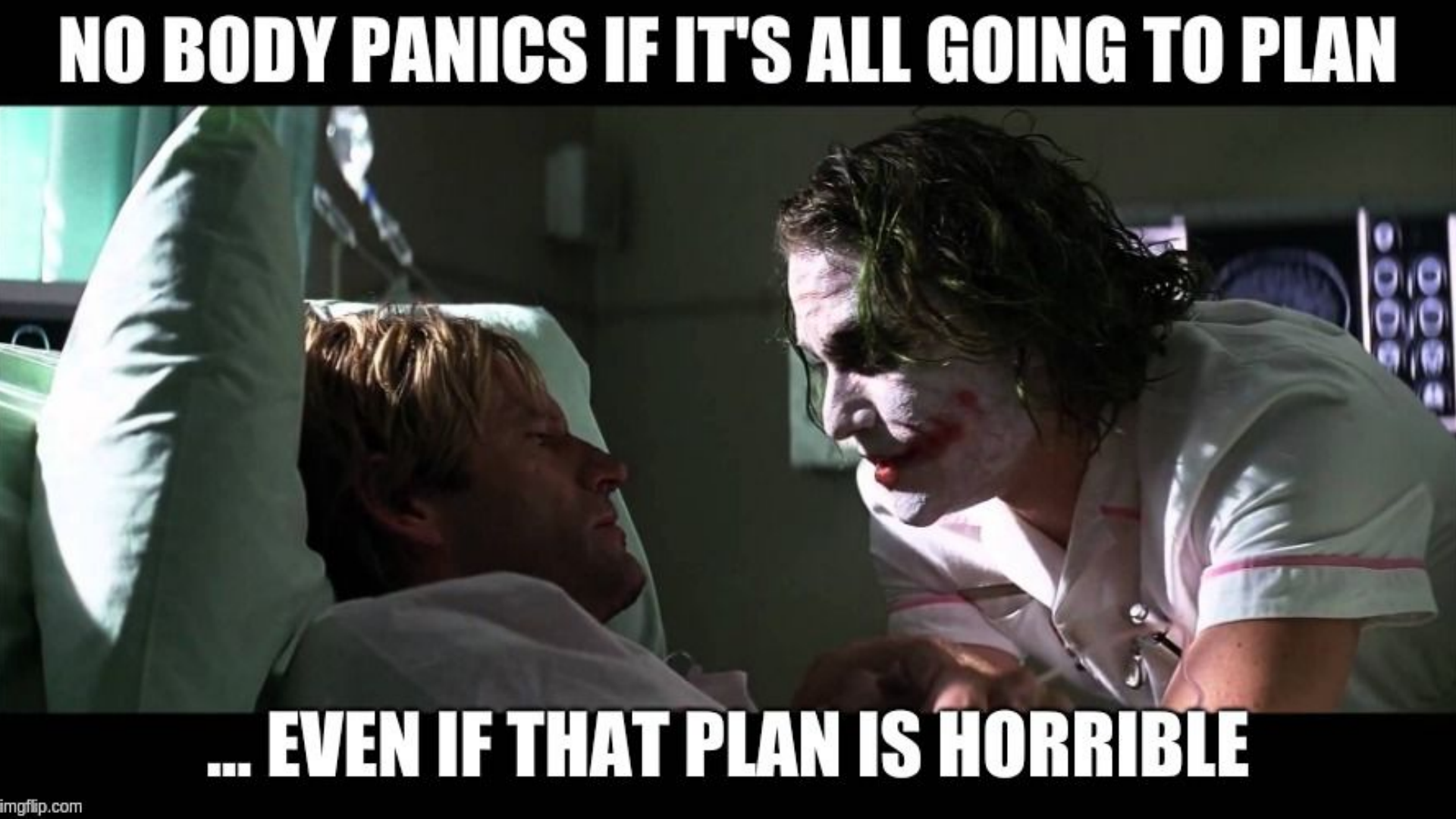
Planning

- Watches the slack channel
- Creates ongoing jira tasks for “cleanup” and permanent resolution


"I hereby…" and assign ${X} to Ops, "${Y}" to Com...
"Calling Customer"
"Creating Tickets"
"Great. Anything Further required?"
"Customer Informed"
COM
PLN
COM
IC
IC
IC
OPS
OPS
OPS
Haxor
"Nice! Incidented concluded."
"Investigating"
"Found + Removed Access"
"Nup"

In Summary

- All work is parallelised
- Each person only has to focus on one task
- All people work together
- All communication is shared


NO BODY PANICS IF IT'S ALL GOING TO PLAN
... EVEN IF THAT PLAN IS HORRIBLE
imgflip.com

Resolution?

Haha no we got pwnwed remember

The Goal

✓ Make production safe

✗ **Find out how and why**

? No one kills each other

Finding out how and why

Post Mortem	Tickets	Follow Up Meetings
“An analysis or discussion of an event held soon after it has occurred, especially in order to determine why it was a failure.”	“a request logged on a work tracking system detailing an issue that needs to be addressed or task that must be performed.”	“a meeting designed to share insights into the issue and ensure all follow up tasks have been completed”

Step #2: Post Mortem

Has anyone done a post mortem?

People

Who was involved. The only place names are used.

People

- Incident Controller: Andrew Howden
- Operations A: __REDACTED__
- Operations B: __REDACTED__
- Communications A: __REDACTED__
- Communications B: __REDACTED__
- Project Owner A: __REDACTED__
- Project Owner B: __REDACTED_

Problem

A concise summary of the issue that presented

Problem

At 7:30 AM on Thursday, 21 of December an alert was triggered that indicated servers were unavailable for a period of a period of 5 minutes. Initial investigation into this outage noted that the machine was rebooted at this time, and CPU usage was several orders of magnitude higher on average than prior to the reboot. A cryptocurrency miner was discovered running and consuming large amount of CPU time. A short investigation into this showed that in order to create this minor “root” or administrative privileges were required, and the machine was denoted compromised to the highest possible degree.

Timeline

A verbatim timeline of the issue.
Often simply chat logs.

Timeline

This one is long.

TLDR we caught it within about half an hour (thanks to the reboot), and removed it. We found persistence in a common pattern shortly thereafter.

The machines were torn down and rebuilt in the following day.

Impact

A description of the (especially financial) impact of the incident on service owner, business, teams etc.]

What are some impacts?

Impact

- The data breach was valued at __REDACTED__ based on the IBM cost of data breach calculator
- The machines needed to be rebuilt
- The response team worked exhaustive hours

Contributing Factors

An analysis of all the things that went wrong for this thing to go wrong.

Designed to avoid a “root cause”.

What are some contributing factors?

Contributing Factors

- Unknown IPMI interface to manage the machine
- Firewall failed-open allowing access to that IMPI interface to the public
- Alerting was not specific to this problem
- No tools were able to pick up the unauthorised user behaviour
- Unclear mechanisms to establish access to the machine
- Unclear procedure for incident management
- Rescue software not compatible with modern operating systems
- Ansible configuration contained hidden bugs

Abstract

The abstract is a short summary of the article, including:

- Problem
- Impact
- Resolution
- Summary of causal factors

Abstract

Early on Thursday, 21st of December a system compromise was discovered. Investigation showed it to be a total machine compromise, likely by an automated tool against a previously unknown management interface. This machine was rooted, persistence established by way of adding an extra user and a cryptominer installed. At this stage, it appears no data was stolen. The impact of this breach appears to be extremely limited, though this is due to fortuitous luck rather than anything implemented by the project teams. However, presuming the theft of customer data and required follow up actions this breach has an estimated potential impact of __REDACTED__ number of issues were highlighted as a result of the breach investigation, but broadly include:

- A previously unknown management interface that was compromised
- Limitations in the tooling designed to pick up these breaches
- Limitations in process around this tooling

Jira tickets were created to address each of these issues, and this will be followed up by a meeting on Monday, 29th of January.

Step #3: Tickets

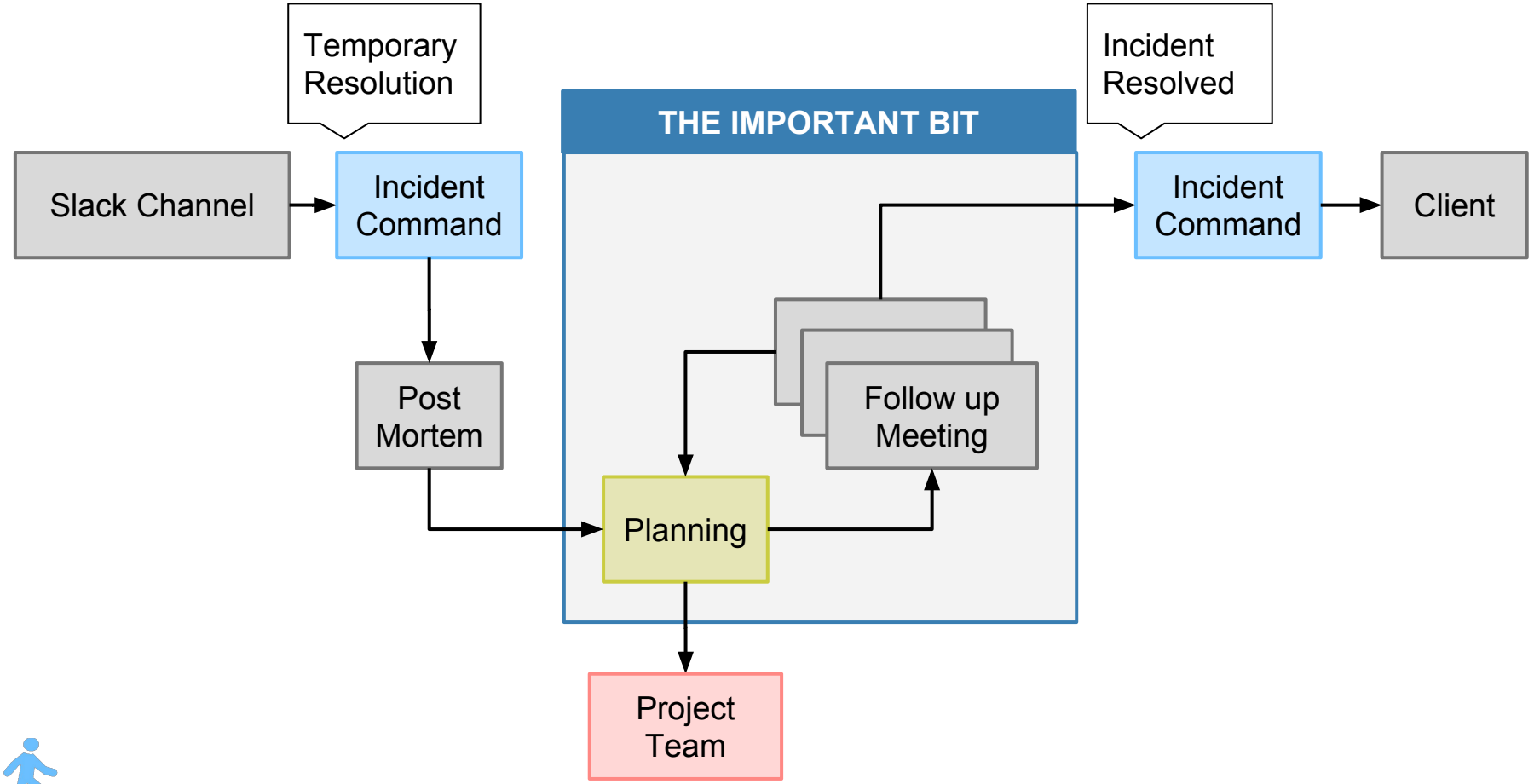
Tickets

Bug tickets, user stories and whatever else that falls into the normal teams sprints

Step #4: Meetings

Follow up meetings

- Budget goes away when pain goes away
- Problems do not go away mysteriously
- Meetings keep the pain fresh


Temporary Resolution
Incident Resolved
THE IMPORTANT BIT
Slack Channel
Incident Command
Incident Command
Client
Post Mortem
Follow up Meeting
Planning
Project Team

The Goal

✓ Make production safe
✓ Find out how and why
✓ No one kills each other

Resolved!

I didn’t get what you just said

People who know more

Talesh Seeparsan

Magento consultant

John Allspaw

Co-founder of
@AdaptiveCLabs

Riccardo Tempesta

CTO at MageSpecialist

Willem de Groot

Security consultant &
malware hunter

See slides for additional notes

What questions do you have?

Find all information and give feedback at:

https://git.io/fxkKi