EVERY MINUTE COUNTS

COORDINATING HEROKU'S INCIDENT RESPONSE

Blake Gentry / @blakegentry

PERSONAL BACKGROUND

- Lead Engineer at Heroku since 2011
- Worked on nearly all parts of the platform
- In 2012, I led a project to overhaul Heroku's Incident Response procedures

TALK OVERVIEW

I'M NOT GOING TO TALK ABOUT HOW TO:

- Build robust systems
- Debug production issues
- Fix issues quickly
- Monitor your systems
- Set up your on-call rotations

I AM GOING TO TALK ABOUT:

- How Heroku coordinates production incident response
- How to apply it to your startup

IN PARTICULAR, HOW TO:

- Organize your company's response to incidents
- Communicate with the company about what's happening
- Communicate with your customers about the incident
- **Build customer trust**

WHAT'S THE PROBLEM?

SOFTWARE BREAKS!

- Happens to everybody
- Even if it's well-built
- Bugs, human error, power outages, security incidents, …
- Can't stop it, but you *can* control how you respond

PRODUCTION INCIDENTS ARE STRESSFUL

- A lot of stuff is happening
- Every minute counts
- High-pressure situation

EFFECTS OF POOR INCIDENT HANDLING

- Direct loss of revenue
- SLA credits
- Customers leave
- Erosion of trust

HEROKU'S INCIDENT RESPONSE IN EARLY 2012

CAMPFIRE + SKYPE

'CAN SOMEBODY FILL ME IN?'

CONTEXT-SWITCHING FOR STATUS UPDATES BREAKS FLOW

CUSTOMERS WERE KEPT IN THE DARK

ESPECIALLY AS THE INCIDENT EVOLVED

NO WAY TO IMPROVE OUTSIDE OF ACTUAL INCIDENTS

NO POST-MORTEM OWNERSHIP

MANY REASONS TO BLAME:

- Product growth
- Company growth
- Changing personnel

TL;DR: INCIDENTS WERE CHAOTIC AND DISORGANIZED.

THIS WAS AFFECTING OUR BUSINESS.

INCIDENT RESPONSE IS A SOLVED PROBLEM!

THE INCIDENT COMMAND SYSTEM

IT OPS ISN'T THE FIRST GROUP TO DEAL WITH THESE PROBLEMS

- Wildfires
- Traffic accidents
- Storms
- Earthquakes

THE INCIDENT COMMAND SYSTEM (ICS)

- Designed in the late 1960s to organize the fighting of California wildfires
- Based on the Navy's management procedures
- Has evolved into a Federal standard for emergency response

ICS: KEY CONCEPTS

- Flexible, modular, scalable org structure
- Unity of command
- Limited span of control
- Clear communications
- Common terminology
- Management by objective

OTHER GOOD RESOURCES ON ICS FOR IT

- Incident Command System for IT (Brent Chapman)
- Incident Command System in Wikipedia

APPLYING ICS TO HEROKU

THREE PRIMARY ORGANIZATIONAL UNITS

1. Incident Command
2. Operations
3. Communications

1. INCIDENT COMMANDER (IC)

A single person in charge with final decision-making authority.

By definition, the first responder is the IC until they hand over responsibilities or the incident ends.

INCIDENT COMMANDER RESPONSIBILITIES:

- Tracks incident progress
- Coordinates the response between different groups
- Decides on state changes
- Issues periodic situation reports ("sitreps")
- Handles all other unassigned responsibilities

WHAT'S A SITREP?

AWS ec2 api problems, we're keeping an eye on things. single tenant Heroku Postgres databases cannot be provisioned at this time. PX dynos are also affected, can't scale them at this time.
We have Craig from AWS in Hipchat to give us more details
We currently are OK on slack for PX dynos. If you're currently using PX dynos and don't need it right now, please switch to a non PX dyno type.
AWS Ticket: https://aws.amazon.com/support/case?caseId=175519571
Trello: https://trello.com/c/dsuS2loo
IC: Ricardo Chimal Jr.
Comms: Keiko
Noah, Greg Burek, Harold, Rodrigo
Posted 17 days ago, Mar 21, 2014 06:07:12 UTC

WHAT'S A SITREP?

- Summary of what's broken
- Describe how widespread the impact is
- Explain what's being done to fix it
- Track who's working on it
- Sent regularly (i.e. hourly or for important updates)
- Sent to the entire company

INCIDENT COMMANDER

EVENT LOOP ⟲

- Do any groups need additional support?
- Does anybody need a break or sleep?
- Are customers being kept informed?
- Do we fully understand the impact?
- Is it time for a sitrep?
- Do all groups have the info they need?
- Repeat ↻

2. OPERATIONS

- Where the actual work happens
- Mostly engineers
- Usually only a small handful of people
- Large incidents may have multiple groups w/ own supervisor

OPERATIONS RESPONSIBILITIES

- Diagnose the issue
- Fix what's broken
- Report progress

3. COMMUNICATIONS

Keeps customers informed about the status of the incident.

Typically managed by customer support personnel.

WHY USE CUSTOMER SUPPORT?

- Don't have to context switch with problem-solving
- Used to speaking customers' language
- Can report back to the IC on customer impact

CUSTOMER COMMUNICATIONS (STATUS UPDATES)

Timely public posts describing:

- What's ***broken***
- What's being done to fix it
- ***What customers can do to work around the issue***.

STATUS UPDATES

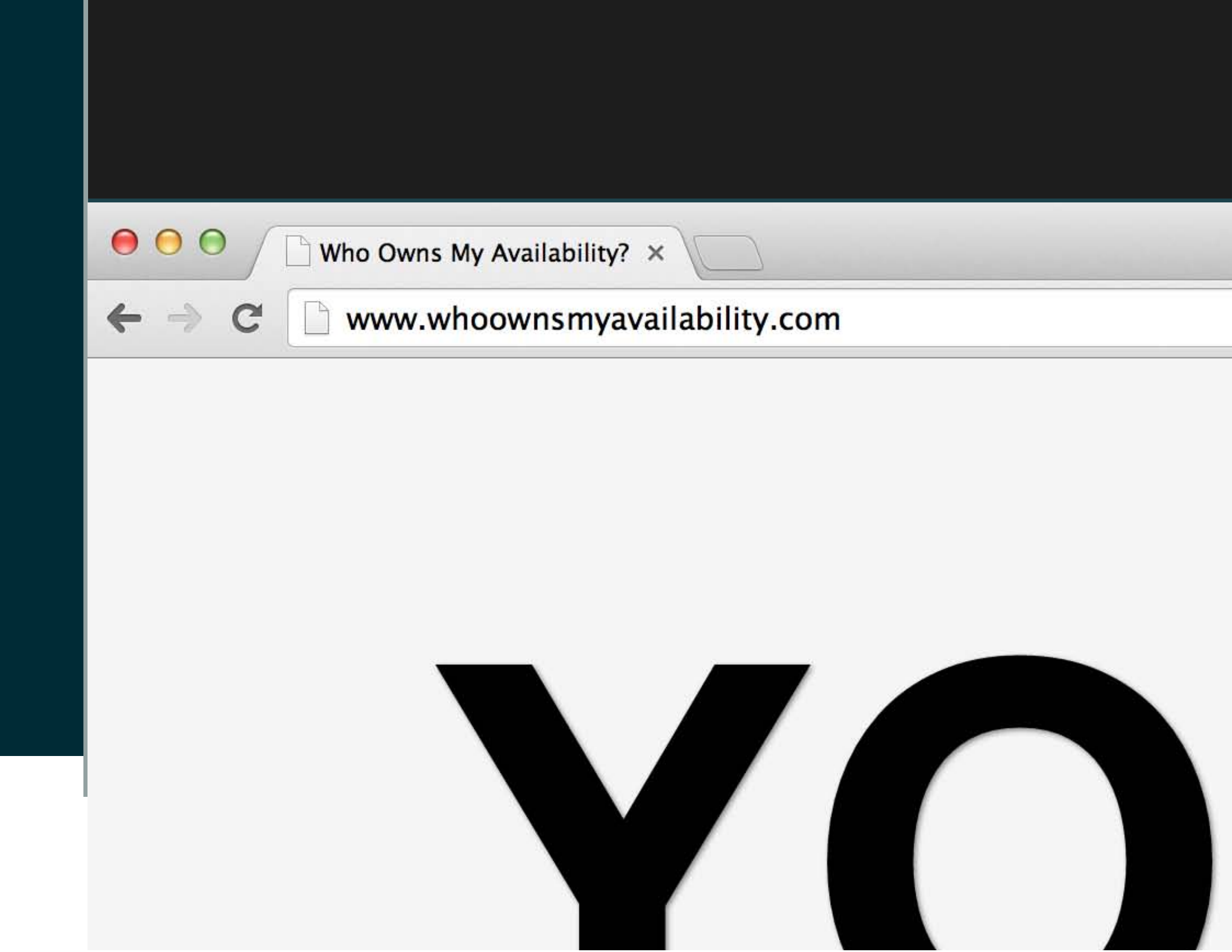
SHOULD:

- Be honest
- Be transparent and upfront
- Explain progress

STATUS UPDATES

SHOULD NOT:

- Provide an explicit ETA
- Presume to know the root cause
- Shift blame


Who Owns My Availability?
www.whoownsmyavailability.com

DON’T DO THIS:

PROACTIVE HANDLING OF TOP CUSTOMERS

HANDLING SUPPORT TICKETS DURING INCIDENTS

RECAP: ORGANIZATIONAL UNITS

1. Incident Command
2. Operations
3. Communications

COMMAND STRUCTURE ISN'T SET IN STONE.

OTHER IDEAS FROM THE ICS

TRAINING AND SIMULATIONS

INCIDENTS ARE STRESSFUL.

REALISTIC TRAINING IS ESSENTIAL.

TO RESPOND QUICKLY AND EFFECTIVELY, THE PROCESS MUST BE SECOND-NATURE.

TRAINING AND SIMULATIONS

Mimic production env as much as possible
Should happen regularly
Focused on procedures, not technical resolution

CLEAR COMMUNICATIONS

EXPLICIT STATE CHANGES AND HAND-OFFS

Use clear messaging when responsibilities transfer or state changes.

EXAMPLES:

```
@all: IC -> Ricardo
@all: Comms -> Chris Stolt
@all: Incident Confirmed
@all: Incident Resolved
```

DEDICATED COMMUNICATIONS CHANNEL

Must be defined in advance.

For us, this is a single-purpose HipChat room.

DEFINE TERMINOLOGY, PROCESS, AND GOALS UPFRONT

PRODUCT HEALTH METRICS

No more than 2-3 high-level metrics to determine whether your product is healthy.

Harder than it sounds.

PRODUCT HEALTH METRICS

OUR METRICS:

Continuous platform integration tests
HTTP availability numbers
# of apps/customers impacted

TOOLS AND CHAT OPS

TOOLS AND CHAT OPS

Greg Burek	@Sentinel page ic: trouble adding pg dbs impacting deploys
Sentinel Bot	@gregburek An incident was opened on your behalf: http://heroku.pagerduty.com/incidents/PW90UAF.

TOOLS AND CHAT OPS

Only helpful if everyone knows how to use them!

INCIDENT STATE MACHINE

0. Everything is normal
1. Investigating an incident
2. Confirmed incident underway
3. Major incident underway

FOLLOW-UPS AND POST-MORTEMS

MAKE SURE SOMEBODY OWNS THIS

HOW TO WRITE A GOOD POST-MORTEM?

1. Apologize
2. Demonstrate understanding of events
3. Explain remediation

The Mark Imbriaco formula.

HOW HAS THIS WORKED FOR US?

Mark Imbriaco
@markimbriaco

Following

Seriously good work by the @heroku team on the status.heroku.com site keeping customers up to date re: #heartbleed

Reply Retweet Favorite ••• More

RETWEETS	FAVORITES
4	3

7:31 AM - 8 Apr 2014

Reply to @markimbriaco @heroku

Jason Dunne @jbdunne · Apr 8
@markimbriaco @heroku I agree, I'm extremely impressed with the site.
Details Reply Retweet Favorite ••• More

Kevin Stone @kjstone00 · Apr 8
@markimbriaco @heroku they were taught well. One of my status sites to aspire to, both format and detail.
Details Reply Retweet Favorite ••• More

Andromeda Yelton
@ThatAndromeda

Follow

@jacobian speaking of which, Heroku wins for best communication I've gotten from any of my accounts re heartbleed. Not even a close contest.

3:24 PM - 9 Apr 2014

1 FAVORITE

Wade Wegner
@WadeWegner

Follow

I'm impressed with the @heroku team's quick actions and response to #heartbleed. bit.ly/1eeCXMp

9:26 AM - 8 Apr 2014

1 RETWEET 1 FAVORITE

WE ARE FAR FROM PERFECT, THOUGH.

RECAP: APPLYING TO YOUR COMPANY

1. DEFINE ORG STRUCTURE

2. STANDARDIZE TOOLING AND PROCESS (NOT AD-HOC)

3. PICK PRODUCT HEALTH METRICS & THRESHOLDS

4. ESTABLISH GOALS FOR CUSTOMER COMMS

5. EXPLICIT HAND-OFFS
6. EMBRACE THE SITREP
7. OWN THE POST-MORTEM

8. REALISTIC TRAINING

THANKS!

BY BLAKE GENTRY / @BLAKEGENTRY