Catch Fire and Halt

Fire in the Datacenter

Jon Kuroda - UC Berkeley EECS Research
jkuroda@eecs.berkeley.edu

Follow the Bouncing Slide

Slides available to follow along at

http://tinyurl.com/firetalklisa2016

Thanks and Acknowledgments

I would like to thank the following for … Everything

UC Office of General Counsel and Outside Counsel
Manufacturer, Insurance & Forensics Investigators,
Campus Risk Services, Campus Fire Marshal,
System Integrators, Environmental Health & Safety,
CIO's Office, Datacenter Staff, EECS Department Chairs,
My Lab Director, My Faculty, My Management,

And the ~~Academy~~ … USENIX LISA16

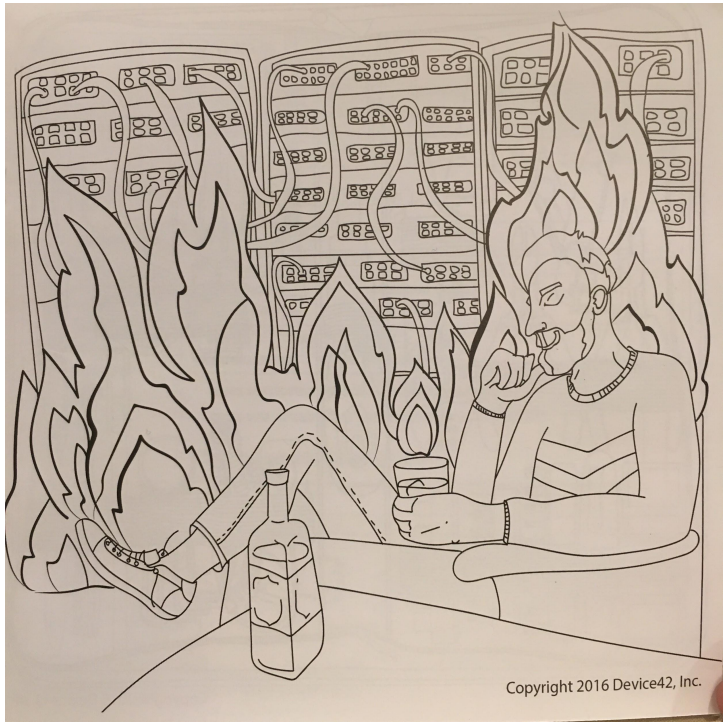
Especially LISA16 for this in the Coloring Book

Much Relevant.

Such On-Point.

Wow.

The Timely.



What Happened - TL;DR version

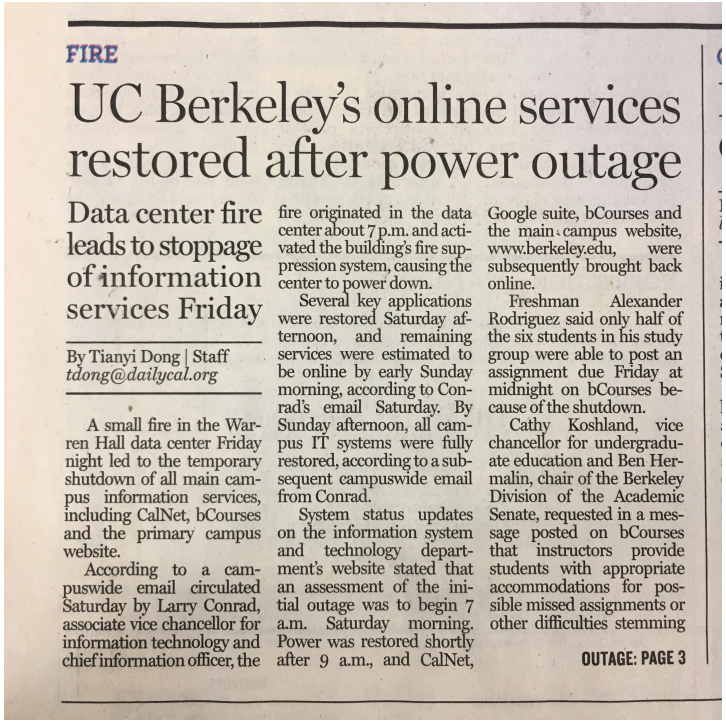
Co-located Customer System Caught Fire in Campus DC

6:48PM, 18 Sep 2015, UC Berkeley Campus Datacenter

- Fire Suppression and Emergency Power-Off (EPO) systems activated
- Almost all Campus online systems and co-located customer systems offline
- Recovery lasted through the weekend
- Forensics and Insurance Investigations through mid-2016
- Ensuing settlement talks through Fall 2016

No injuries, deaths, or anything worse than lost sleep, stress, or paperwork.

Press and Other Coverage



FIRE

UC Berkeley's online services restored after power outage

Data center fire leads to stoppage of information services Friday

By Tianyi Dong | Staff
tdong@dailycal.org

A small fire in the Warren Hall data center Friday night led to the temporary shutdown of all main campus information services, including CalNet, bCourses and the primary campus website.

According to a campuswide email circulated Saturday by Larry Conrad, associate vice chancellor for information technology and chief information officer, the fire originated in the data center about 7 p.m. and activated the building's fire suppression system, causing the center to power down.

Several key applications were restored Saturday afternoon, and remaining services were estimated to be online by early Sunday morning, according to Conrad's email Saturday. By Sunday afternoon, all campus IT systems were fully restored, according to a subsequent campuswide email from Conrad.

System status updates on the information system and technology department's website stated that an assessment of the initial outage was to begin 7 a.m. Saturday morning. Power was restored shortly after 9 a.m., and CalNet, Google suite, bCourses and the main campus website, www.berkeley.edu, were subsequently brought back online.

Freshman Alexander Rodriguez said only half of the six students in his study group were able to post an assignment due Friday at midnight on bCourses because of the shutdown.

Cathy Koshland, vice chancellor for undergraduate education and Ben Hermalin, chair of the Berkeley Division of the Academic Senate, requested in a message posted on bCourses that instructors provide students with appropriate accommodations for possible missed assignments or other difficulties stemming

OUTAGE: PAGE 3

OUTAGE: Official says department to hold review of event, response

FROM FRONT

from the outage.

The information system and technology department will conduct a review on the incident and its response procedures Monday, according to another email sent by Conrad on Sunday.

Rita Rosenthal, communications manager for campus's Office of the Chief Information Officer, said the investigation of the cause of the fire and extent of damage is still in progress, and a full report is expected to be published early next week.

On Sunday morning, bCourses was intermittently inaccessible for several hours

[NFPA 75 Whitepaper on Fire Protection and Suppression in Datacenters]

Caveats Before I Continue.

Settlement negotiations impose some restrictions.Therefore:

- **No** identification of any non-UC Berkeley parties involved
- **No** identifying photos of any involved equipment
- **No** discussion of the specifics of the system itself.
- **No** discussion of the findings of the investigation

No, really.

These decisions take place far above my “pay grade.”

Okay, it sucks, but how common is this?

Somewhat rare - but, fires and other outages do happen

- Electrical Infrastructure caught fire (Iowa State Gov't)
- Another Datacenter Fire (Samsung SDS, S. Korea)
- Outage via Auto … Intrusion (Rackspace, Dallas, TX)
- Electrical Infrastructure Fire ($10M!!) (Seattle, WA)
- Contractors … Amazon (Ashburn, VA)
- Multiple Lightning Strikes (Google, Belgium)
- … These things and more happen. Are you ready?

Extensive post-event communication with ...

System Integrator, Manufacturer, System Integrators
Campus Fire Marshal, **Campus Risk Services**,
Environmental Health and Safety,
Insurance and Forensics Investigators,
UC Legal Counsel, Outside Counsel

Many separate Post-mortem/retrospectives/investigations

So. Much. E-Mail.

A System Caught Fire? Wat?

- One of my research group's systems caught fire
  - This is our first public acknowledgement of this
  - But it's been a semi-open secret on campus a while.
- Those in the know have (mostly) been **very** good re: this.
- That said, it's been a sensitive topic the past year or so
- **Everybody** learned something from this event.

The System that caught fire

- COTS 2U x86 server, reputable vendor, delivered Q4 2012
  - Intel Sandy Bridge-EP CPUs, Samsung DDR3 RAM, Intel/Samsung SSDs, Seagate HDDs
  - 100% Commodity Components from Reputable Vendors
  - Moved from EECS Dept to Campus Datacenter Q3 2014 along with 9 others
  - Recently reconfigured (that afternoon) to a low-memory config for a researcher request
  - To be returned later to a high memory config for use in a Spark / HDFS cluster
  - 1 of 3 racks used by our research group - 2 devoted to separate Genomics Cluster
- Part of a larger order of 30 systems
  - These were our workhorse systems, ~90% of our local compute capability.

How Does a Modern Server System Catch Fire?

We cannot discuss details at this time.

Obviously, something went wrong.

Timeline for Friday, 18 Sept 2015

- ~6:20PM: Last customer (Me) signs out of facility, (tries to) get dinner.
- **[~6:45PM: System catches fire in Datacenter]**
- **[6:47PM: Smoke Sensors triggered]**
- **[6:48PM: FM-200 Clean Agent Fire Suppression / EPO systems activate]**
  - Nearly 4000 pounds / 1300kg of FM-200 Fire Suppression Agent
- ~8:30PM: Customer returns to campus office after dinner
  - Notices many Campus services offline, starting around 6:48PM
  - First thought: “Great, another weird EECS/Campus routing flap.”
- ~9:00PM: Customer sees 8PM report of “Halon System” use at campus DC
- 9:05PM: Customer sends quick FYI to his management.

Okay, maybe not a “Weird Routing Flap”

Maybe still a false alarm that triggered the fire suppression system.

Timeline for 18 Sept 2015: Oooooooh. Bad.

- 9:30PM (approx): Customer sees new post on Campus System Status site:

Fire in campus data center. Tons of stuff down. - IST Service Status

Friday 09/18/2015 21:18: There was a fire in the data center. Extent is not known at this time. Steve Aguirre ***[DC Manager]*** is on-site and actively working on the issue(s). ETA unknown at this time.
The CSS-IT Service Desk is receiving reports that internet connectivity has been interrupted for multiple applications, users, and locations across campus.
IST is working to identify the root cause and resolve the issue.
No ETA is available at this time.

Timing of my departure relative to fire ⇒ sobering observations.

How bad was it?

What was down?

“Everything.” Not really, but it sure felt like it.

- Campus SSO Authentication aka “CalNet”
  - Without this, not a lot of Campus Services work
  - WiFi. Business Services. Student Services. Storage, VMs, Oracle, Lots of stuff.
  - Many of those services not even up to complain about lack of authentication
  - Even Google Services which tie into Campus Authentication for CAS/SSO
  - Campus website down. Whoa.
- EECS Departmental Maillists, Jabber/XMPP instance
  - Hosted at or dependent upon services hosted in the Campus Datacenter
  - Heavily impacted ability to notify people within the Department

What still worked?

A few basic services: didn't seem like a lot, but it was.

- Campus Network and most basic Network Services
  - IPV4/IPV6, DNS, NTP, attwifi, but no DHCP on Campus-managed networks
- SMTP Delivery for @berkeley.edu and other hosted berkeley.edu domains
  - Offsite MX systems at SDSC for the win
- IMAP clients could still authenticate, active web client sessions too.
- Campus System Status site - hosted off campus at 3rd party provider
- Campus Emergency Website (if needed) - hosted off campus at NYU.
- Campus Emergency Notifications - "WarnMe" - use of 3rd party service
- Most Depts' "In-House" services w/o dependencies continued to work

What Else Went Down?

Our Research Group's Jenkins Instance

CI for 2 large Open Source Projects we spun out and our smaller projects.

- **Apache Spark™** and **Alluxio** (Formerly known as ***Tachyon***)

Hosted at the Campus Datacenter, exposed through our in-house hosted website

Fortunately, it was a Friday night, so few people noticed, but this is probably my research group's most notable externally facing service.

Some attempts made to notify key developers in those projects, but Friday night.

Detailed timeline: Back to Our Story

- ~10PM: Customer returns to Campus DC to assist with recovery efforts
  - Saw one developer on the way - “Don’t Commit” “No worries! Good Luck!”
  - Electricians, Other Customers, Exec Dir of Environmental Health and Safety (EHS) waiting
  - Local FD not allowing any non-essential staff into the building yet.
  - Customer’s Manager drives-by to offer moral support (He had been enjoying Friday night.)
- ~10:30PM: Non-essential staff allowed in by FD and Campus Fire Marshal
  - DC staff on duty at time of fire exits to go home having been relieved by fresh staff.
  - “Kuroda! You were the last one out! What did you do!?” Haha. LOL. If only we knew then ...

First thing we noticed upon entering the ground floor ...

Whoa, It's Warm.

Datacenter itself on the 3rd floor, but even the ground floor was warm.

~90F/30C when we arrived. Reportedly around 120F when FD arrived.

10,000 ft^2 of DC space ⇒ lots of systems ⇒ lots of residual heat

EPO ⇒ No Power ⇒ No HVAC + lots of residual heat ⇒ Warm++

Datacenters get **warm** if no HVAC even after EPO activation.

HVAC **needs** to shutdown to allow FM-200 system to work

(Reportedly not all racks powered off … Local FD used Manual EPO)

<u>This could have had major repercussions.</u>

Back to our Story: One problem to solve first

Find the ignition source of whatever triggered the smoke sensors

- Local FD and Fire Marshal cleared the scene for less essential staff
- Fire Marshal wouldn't OK power-on before ID of ignition source
- Two smoke sensors localized it to a zone of racks
  - Turned out to be hard due to FM-200/smoke sensor interaction
- Fire Marshal had detected fading odor of burnt electronics
- We had power for lights, but camera/video system offline.

It was 11PM and we needed to find what tripped the smoke sensors.

Row by Row, Rack by Rack

We had people. It was only ~90F/30C. “Rack To Rack” it was.

- Started with the rows of racks covered by the tripped smoke sensors
- I, as one of few customers there, opened up my racks in that zone
- “Let’s just get my racks out of the way and cross those off the list …”
- Yeah … Maybe not so much. Hi, Stomach of mine, meet the floor.

“I think I know what caught fire.”

That got everyone’s attention …

I considered my career options ...

OK, fine, some purposefully obfuscated pictures.

Yeah, something bad definitely happened.

First Look “Under The Hood”

Another look “under the hood”

Saving the worst for last ...

So, um, wow … like whoa.

Till I saw this, I did not once think one of my boxes had caught fire.

DETAILS OF SYSTEM DAMAGE REDACTED

Complete and Total Loss of Server System.

Contact Heat Damage to adjacent Server System.

SLIDE REDACTED

SLIDE REDACTED

SLIDE REDACTED

The Flow of Information to Campus Leadership

“Which department did you say you were from?” I was asked.

“EECS,” I replied

“Thanks *taptaptap* *send*”

“Who is that going to?” as I saw the e-mail go out from his phone.

“Lyle. [Campus Deputy CIO]”

“Ah, okay.”

I mean, as long as it wasn’t the Daily Cal or something.

What now?

I was now the man on the scene.

I hadn't intended it when I came down to lend a hand … but there I was. No running away - here's my info, my management's, my faculty's …

Upsides:

- Ignition source Found. Scope and extent of fire defined by midnight
- Recovery operations could begin - starting with HVAC and Power.
- But these things are never easy.

Which came 1st, ~~Chicken~~ Power or ~~Egg~~ HVAC?

How often do you do cold-start fire drills of your facilities?

- HVAC control systems needed power.
- Can’t turn on power until the HVAC is running …
- But that UPS room is still running ...

No restoration of power until early AM.

While waiting on status update about power ...

- [REDACTED]
  - [REDACTED]
  - [REDACTED]
  - [REDACTED]
- Customer notified his own management
- Customer unplugged all systems of the same type
  - There were a total of 30 from that order
  - There goes 90% of my group’s local compute.

Our story so far

- One of my research group's servers caught fire in the Campus DC
- Campus DC shutdown, crippling Campus computing services
- Recovery stalled till HVAC/Power circular dependency resolved
- Everyone to reconvene at 7AM to begin organized recovery.
- Rest of order of 30 servers taken offline
  - Total of 10 at Campus DC or already on RMA
  - 20 more at Departmental in-house facilities
  - Major conference deadline for our researchers in a week.

Not much local general purpose compute for my people ☹

Morning Recovery (Sat)

- Power and HVAC was up by 7AM
- We (EECS) showed up - **we had to** (I got there at 8AM)
- Everyone - DC Staff, other Customers - was cordial to us
- Had to let people see the system-most-at-fault - no hiding
- Brought up our Jenkins instance - first cold boot in ages
- Broke for food
- I started drafting initial communication to our vendor
- Most Critical Systems up Sat, rest on Sun, Cleanup Mon

It’s Monday AM. What worked?

- **No injuries or deaths**
- Fire Suppression System worked, EPO (mostly) worked
- Backup Comms: Systems remained up, were used well.
- Save for WiFi/DHCP, Campus Network continued working
- Very few systems lost as result of power cutoff
  - They were old and tasted like sour grapes anyway
- **We got lucky - this happened early on a Friday evening.**
- Recovery began Sat AM, everything back on or by Mon AM.
- We all kept our jobs

Why was Friday evening such lucky timing?

- Reduced Impact
  - 6:48PM ⇒ After “Business Hours”
  - Negligible weekend “Business” usage.
  - Still had Student usage (Homework, etc) though.
- Staff availability
  - 6:48PM ⇒ around typical “Shift Change” time, Staff++
  - Early enough to realistically reconsider weekend plans.

It’s Monday. What didn’t work so well?

- Smoke Detectors: didn’t detect early enough for staff response
  - VESDA planned (Installed Q3 2016),
  - Even then, DC not staffed Late PM - Early AM
- Automatic EPO didn’t get everything. FD hit manual EPO for the rest.
- UPS Room still ran, but without HVAC due to EPO
- HVAC/Power Circular Dependency - Delayed restoration of service
- CAS/Auth, Monitoring, Notification systems unavailable
  - Full active/active offsite DR had been/always in progress
- Facility Camera System wouldn’t export video of Hot Aisle from fire
  - DC Manager screen-scraped with his phone camera.

The 2 months After

- Sep: Notified Vendor of event (That Monday)
  - Immediate response from Vendor with Manufacturer included
- Sep: Engaged with Campus Risk Services
  - Connected us with Insurance and UC Legal Counsel
- Oct: Dept Chair contact re status of investigation
- [REDACTED]
- [REDACTED]
- [REDACTED]
- Nov: Affected Systems Replaced

Costs and Impact of the Fire itself:

Campus/Department (All depts, not just EECS) Level:

- FM-200 system replenishment/inspection: ~$100K
- Overtime for Data Center Staff, Physical Plant
- Loss of most campus online services for the weekend
- Minor loss of customer HW due to EPO power cutoff

Hard to put a cost on moving homework/academic deadlines

(It would be a few weeks before the FM-200 was replenished ☹)

Costs and Impact

To our research group:

- Loss of use of 480 Cores, 8TB of RAM, 300TB HDD/SSD
- Already Expensive AWS $$/month spend doubled
- Borrowing systems from other research groups
- Engineering Staff time devoted to managing all this

Potential Server Replacement Cost: $300K

Pace of Research Marched on.

The Monday (and 2-3 weeks) After

- We notified System Integrator Vendor of event
  - Expressed need to inspect or replace all systems from order
  - Immediate response with Manufacturer included
- Worked with Campus Risk Services
  - Connected us with Insurance and UC Legal Counsel
- Gaps in experience/expertise levels between disciplines
- Experienced means/methods “impedance mismatches”
- We all wanted “the same thing” but how to get there ...

Gaps in Expertise

No one party had broad level of understanding of everything.

- Risk Services/Legal were not engineers
- Engineers were not attorneys or investigative specialists
- Investigators were not datacenter specialists or attorneys
- Few of us were Fire Suppression knowledgeable

I answered quite a few questions re: “The Computer Industry”

Creative solutions for equiv of “Discovery Process” needs

Creative Solutions for Discovery needs

- How to provide 2 years worth of email threads re servers
  - Need to preserve order, threading, headers
  - Create read-only email lists w/ no subscribers
  - Inject all messages into those lists
  - Let list archiver create easy to read email threads
  - Provide read-only access to necessary parties.

Can’t send a gzip’d mbox or or maildir
Better than 10s-100s of print-msg-to-PDFs?

Impedance Mismatches: Assets vs Relationships

We all wanted the same thing. … Really!

But different underlying background priorities.

- Maintaining relationships with Companies, rest of Campus
- Defending University assets, funding, risk management
- Protecting Company names and reputation
- Reducing Liability

This is natural, but could have potentially derailed progress.

We all wanted to “make things right”.

- Replace all suspect components or replace all systems
- Conduct full forensics/FA on the system that caught fire
- Enact changes to prevent similar occurrences
- Cover UC Berkeley costs from the event
- Still work with and trust each other in the future

This was an unfortunate incident - not malice.
But this is not the the Airline Industry.
There is no NTSB/CPSC/FAA for the Datacenter.

Your Forensics Lab or Mine?

DETAILS REDACTED

Obviously, there was an investigation

This is a good outcome

Equipment Replacement?

DETAILS REDACTED

All Equipment Replaced

Opportunities for Improvement:

- Better Smoke Detection (VESDA): Done
- Reduce HVAC/Power Circular Dependency: In Progress
- Improve Automatic EPO effectiveness: In Progress
- Better Isolation between Enterprise/Customer Loads: ...
- Better Campus Auth/Monitoring/Notification: In Progress
- **[REDACTED]**
- More diligent Customer IPMI monitoring: neverending
- Examine countless smaller facilities across Campus

(My) Thoughts and Observations

About the investigation itself:

- “There is no NTSB/CPSC/FAA for the Datacenter.”
- No reasonably neutral external party “in charge”
- Unlike Plane Crashes, no lives lost or injuries incurred
  - Or Flaming Exploding Batteries ...
  - Otherwise, investigation would’ve been very different
  - I, for one, am very thankful - for many reasons
- We worked together pretty well given we met that week.

(My) Thoughts and Observations

"No NTSB/CPSC/FAA for Datacenter Incidents"

- No single investigative thread tying it all together
- Multiple Investigations / Post-Mortems / Reviews
  - Insurance/Legal/Manufacturer (The most extensive)
  - Datacenter, Campus IT (Quasi-separate)
  - Individual Departments w/ co-located equipment
- Less Transparency in investigation/reporting

(My) Thoughts and Observations

About Responsibility and Blame

- I'd like to be able to tell you specifics (Part Numbers, Pics)
- But (understandable) concerns by … everybody re blame
- Manufacturer exceeds "Industry Standards" so many ways
  - There is no "Industry Standard" … "Industry Typical"?
  - They're just so very far ahead of the pack in this area
- **[REDACTED]**
- Blame or Fear of Blame is a powerful disincentive. Fight it.

Actionable Suggestions

Ask your vendors (Manufacturers, Integrators, etc) about:

- Advisories - do they have/get them?
- How do they distribute them? How do you get them?
- RMA Chain Transparency - why did failed server fail?

Basically, if there's a problem, how do you find out about it?

Compare with Security Vulnerability Disclosures.

Actionable Suggestions

Ask your Datacenter/Colo/etc:

- What sort of Fire Detection/Suppression system?
- Enterprise/Less-Enterprise/Non-Enterprise Isolation?
- EPO, UPS, HVAC, Video, etc - Tested how often? How?
- Fault dependencies between those?

Ask yourself:

- Do you have any in-house/ad hoc/unofficial facilities?

Actionable Suggestions

Ask yourself / your whole Organization:

- How resilient are all of your basic Dialtone Services?
- How do you communicate when things are down?
- Do you have a holistic post-event retrospective template?
- How often do Eng/Legal/Facilities/etc work together?
  - I met most of these for the first time due to the fire.
  - Organized “Go Teams” to promote collaboration?
- How thoroughly do you track down every IPMI message?

A personal comment

This could have gone **FAR** worse but for some:

- Some good choices made about a decade ago
- Some very fortunate timing

Recent tragedy in Oakland CA has brought home just how important all of this this, no matter the venue.

“When we tech an event, we have people's lives in our hands.”

<u>This Is Important Stuff</u>

Questions. Comments. “Better you than me.”

I also have 20+ slides of detailed timeline, some random notes about FM-200, Halon, and other Gaseous Fire Suppression systems, and background info on the Campus Datacenter.

If you have more questions, we can meet up back here @830PM for a “Fires and Other Disasters at Work” BOF

No Cute Kitten Pictures though - I went through my entire stash to prepare for giving this talk.

Timeline for Friday, 18 Sept 2015

- ~6:20PM: Last customer (Me) signs out of facility, (tries to) get dinner.
- **[~6:45PM: System catches fire in Datacenter]**
- **[6:47PM: Smoke Sensors triggered]**
- **[6:48PM: FM-200 Clean Agent Fire Suppression / EPO systems activate]**
  - Nearly 4000 pounds / 1300kg of FM-200 Fire Suppression Agent
- ~8:30PM: Customer returns to campus office after dinner
  - Notices many Campus services offline, starting around 6:48PM
  - First thought: “Great, another weird EECS/Campus routing flap.”
- ~9:00PM: Customer sees 8PM report of “Halon System” use at campus DC
- 9:05PM: Customer sends quick FYI to his management.

Okay, maybe not a “Weird Routing Flap”

Maybe still a false alarm that triggered the fire suppression system.

Timeline for 18 Sept 2015: Oooooooh. Bad.

- 9:30PM (approx): Customer sees new post on Campus System Status site:

Fire in campus data center. Tons of stuff down. - IST Service Status

Friday 09/18/2015 21:18: There was a fire in the data center. Extent is not known at this time. Steve Aguirre ***[DC Manager]*** is on-site and actively working on the issue(s). ETA unknown at this time.
The CSS-IT Service Desk is receiving reports that internet connectivity has been interrupted for multiple applications, users, and locations across campus.
IST is working to identify the root cause and resolve the issue.
No ETA is available at this time.

Timing of my departure relative to fire ⇒ sobering observations.

How bad was it?

What was down?

“Everything.” Not really, but it sure felt like it.

- Campus SSO Authentication aka “CalNet”
  - Without this, not a lot of Campus Services work
  - WiFi. Business Services. Student Services. Storage, VMs, Oracle, Lots of stuff.
  - Many of those services not even up to complain about lack of authentication
  - Even Google Services which tie into Campus Authentication for CAS/SSO
  - Campus website down. Whoa.
- EECS Departmental Maillists, Jabber/XMPP instance
  - Hosted at or dependent upon services hosted in the Campus Datacenter
  - Heavily impacted ability to notify people within the Department

What still worked?

A few basic services: didn't seem like a lot, but it was.

- Campus Network and most basic Network Services
  - IPV4/IPV6, DNS, NTP, attwifi, but no DHCP on Campus-managed networks
- SMTP Delivery for @berkeley.edu and other hosted berkeley.edu domains
  - Offsite MX systems at SDSC for the win
- IMAP clients could still authenticate, active web client sessions too.
- Campus System Status site - hosted off campus at 3rd party provider
- Campus Emergency Website (if needed) - hosted off campus at NYU.
- Campus Emergency Notifications - "WarnMe" - use of 3rd party service
- Most Depts' "In-House" services w/o dependencies continued to work

What Else Went Down?

Our Research Group's Jenkins Instance

CI for 2 large Open Source Projects we spun out and our smaller projects.

- **Apache Spark™** and **Alluxio** (Formerly known as ***Tachyon***)

Hosted at the Campus Datacenter, exposed through our in-house hosted website

Fortunately, it was a Friday night, so few people noticed, but this is probably my research group's most notable externally facing service.

Some attempts made to notify key developers in those projects, but Friday night.

Detailed timeline: Back to Our Story

- ~10PM: Customer returns to Campus DC to assist with recovery efforts
  - Saw one developer on the way - “Don’t Commit” “No worries! Good Luck!”
  - Electricians, Other Customers, Exec Dir of Environmental Health and Safety (EHS) waiting
  - Local FD not allowing any non-essential staff into the building yet.
  - Customer’s Manager drives-by to offer moral support (He had been enjoying Friday night.)
- ~10:30PM: Non-essential staff allowed in by FD and Campus Fire Marshal
  - DC staff on duty at time of fire exits to go home having been relieved by fresh staff.
  - “Kuroda! You were the last one out! What did you do!?” Haha. LOL. If only we knew then ...

First thing we noticed upon entering the ground floor ...

Whoa, It's Warm.

Datacenter itself on the 3rd floor, but even the ground floor was warm.

~90F/30C when we arrived. Reportedly around 120F when FD arrived.

10,000 ft^2 of DC space ⇒ lots of systems ⇒ lots of residual heat

EPO ⇒ No Power ⇒ No HVAC + lots of residual heat ⇒ Warm++

Datacenters get **warm** if no HVAC even after EPO activation.

HVAC **needs** to shutdown to allow FM-200 system to work

(Reportedly not all racks powered off … Local FD used Manual EPO)

<u>This could have had major repercussions.</u>

Wait. What do you mean “Repercussions”?

- FM-200 system was depleted
- Sprinkler system only backup to prevent Loss of Building
- If temperature rose enough at the sensors, it rains inside.
- Sprinklers are zoned to limit impact to affected area ...
- **Sprinklers activate at 165F - ceiling temp of 180F-200F**

This would protect the building, but at the risk of servers/etc.

But wait! There’s more!

But wait! There's more!

- UPS room on the roof with the generator
- Remained online despite DC EPO - separate "fire zone"
- HVAC controller system shutdown due to EPO - no HVAC

Even if low system loads, UPS batteries charging ⇒ **Heat**

Sprinklers activate at 165F - ceiling temp of 180F-200F

Back to our Story: One problem to solve first

Find the ignition source of whatever triggered the smoke sensors

- Local FD and Fire Marshal cleared the scene for less essential staff
- Fire Marshal wouldn't OK power-on before ID of ignition source
- Two smoke sensors localized it to a zone of racks
  - Turned out to be hard due to FM-200/smoke sensor interaction
- Fire Marshal had detected fading odor of burnt electronics
- We had power for lights, but camera/video system offline.

It was 11PM and we needed to find what tripped the smoke sensors.

Row by Row, Rack by Rack

We had people. It was only ~90F/30C. “Rack To Rack” it was.

- Started with the rows of racks covered by the tripped smoke sensors
- I, as one of few customers there, opened up my racks in that zone
- “Let’s just get my racks out of the way and cross those off the list …”
- Yeah … Maybe not so much. Hi, Stomach of mine, meet the floor.

“I think I know what caught fire.”

That got everyone’s attention …

I considered my career options ...

OK, fine, some purposefully obfuscated pictures.

Yeah, something bad definitely happened.

First Look “Under The Hood”

Another look “under the hood”

Saving the worst for last ...

The Flow of Information to Campus Leadership

“Which department did you say you were from?” I was asked.

“EECS,” I replied

“Thanks *taptaptap* *send*”

“Who is that going to?” as I saw the e-mail go out from his phone.

“Lyle. [Campus Deputy CIO]”

“Ah, okay.”

I mean, as long as it wasn’t the Daily Cal or something.

What now?

I was now the man on the scene.

I hadn't intended it when I came down to lend a hand … but there I was.
No running away - here's my info, my management's, my faculty's …

Upsides:

- Ignition source Found. Scope and extent of fire defined by midnight
- Recovery operations could begin - starting with HVAC and Power.
- But these things are never easy.

Which came 1st, ~~Chicken~~ Power or ~~Egg~~ HVAC?

How often do you do cold-start fire drills of your facilities?

- HVAC control systems needed power.
- Can’t turn on power until the HVAC is running …
- But that UPS room is still running ...

No restoration of power until early AM.

While waiting on status update about power ...

- [REDACTED]
  - [REDACTED]
  - [REDACTED]
  - [REDACTED]
- Customer notified his own management
- Customer unplugged all systems of the same type
  - There were a total of 30 from that order
  - There goes 90% of my group’s local compute.

Our story so far

- One of my research group's servers caught fire in the Campus DC
- Campus DC shutdown, crippling Campus computing services
- Recovery stalled till HVAC/Power circular dependency resolved
- Everyone to reconvene at 7AM to begin organized recovery.
- Rest of order of 30 servers taken offline
  - Total of 10 at Campus DC or already on RMA
  - 20 more at Departmental in-house facilities
  - Major conference deadline for our researchers in a week.

Not much local general purpose compute for my people ☹

Morning Recovery (Sat)

- Power and HVAC was up by 7AM
- We (EECS) showed up - **we had to** (I got there at 8AM)
- Everyone - DC Staff, other Customers - was cordial to us
- Had to let people see the system-most-at-fault - no hiding
- Brought up our Jenkins instance - first cold boot in ages
- Broke for food
- I started drafting initial communication to our vendor
- Most Critical Systems up Sat, rest on Sun, Cleanup Mon

The Monday (and 2-3 months) After

- Sep: Notified Vendor of event (That Monday)
  - Immediate response from Vendor with Manufacturer included
- Sep: Engaged with Campus Risk Services
  - Connected us with Insurance and UC Legal Counsel
- Oct: Dept Chair contact re status of investigation
- [REDACTED]
- [REDACTED]
- [REDACTED]
- Nov: Affected Systems Replaced

Datacenter backgrounder

- Repurposed / Retrofitted office building
- 10,000 ft^2 of datacenter space on 3rd floor
- Commissioned 2004
- Allowed 2005 DC move out of a flood-prone basement
- Also: CIO, IST, Controller’s office, EOC, Seismo Lab, etc
- UPS / Generator. 2 unscheduled outages in 12 yrs
- Staffed 6A - 9P M-Friday, 8A - 5P weekends / holidays
- Offsite DR/BC facilities at SDSC on UCSD Campus

Fire! How Does It Work?

Heat + O_2 + Fuel

=

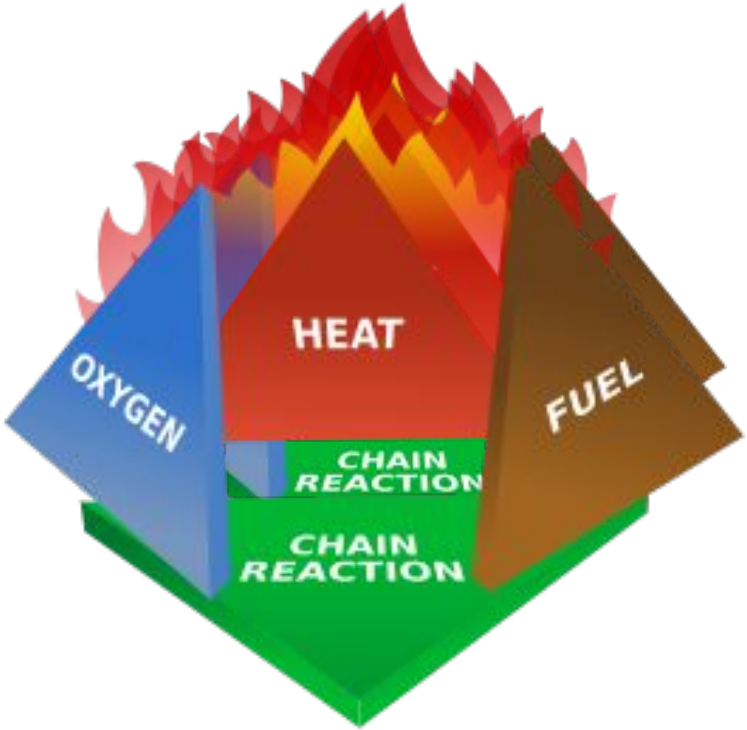
Rapid Self Sustaining

Exothermic Oxidizing

Chemical Reaction

AKA - FIRE

How Do We Put It Out?



In the Beginning, There Was ….

- Water (or Dirt), Cut off O_2, Lower Temperature.
- 🌊 + 🔥 = !(🔥) unless ...
- 🌊 + (⚡⚡⚡🔥)= ⚡⚡⚡☠⚡⚡⚡
- 🌊 + (🔥) = 🔥🔥🔥☠🔥🔥🔥
- 🌊 + (🔥) = 💥💥💥☠💥💥💥

We don’t like putting Water or Dirt on our computers.

What about my Fire Extinguisher at home?

ABC Dry Chemical Fire Extinguisher = Also Bad

- $NH_4H_2PO_4$ - Monoammonium Phosphate - main agent
- Coats fuel (computer parts), barrier to Oxygen
- Also somewhat corrosive. Baking Soda isn't really much better

Won't conduct electricity, so safe(r) for humans to use on live gear.

But still bad for computers.

Just Get Rid of the Oxygen?

- Flood the facility with CO_2 / N_2 / Inert Gas / Hard Vacuum
- Puts out fire, doesn't harm gear … but …
- O_2 level req'd to sustain life **vs** O_2 level that puts out fire
- This can kill (**and has killed**) people
- But such systems have been common off and on …
- They are effective, cheap, and reliable …

But deaths are bad ☹☹☹☹

Ooooh, “Halon”

HALON!!!

- Very Broad Class of Chemicals, dates back to 1839 (CCl_4)
- Some Refrigerants, some Fire Suppressants, Some both
- Halon 1301 and Halon 1211 most typical
- Montreal Protocol 1987 - Ozone Depletion Potential
- Equipment-safe, effective, $$ vs H_2O/Inert Gas Flooding
- Halon 1301/1211 - **Mostly** human-safe @ useful levels

No new Halon installs in the US anymore

So, a nozzle and two FM-200 tanks walk into a bar …

Halon is Dead, Long Live Halon Replacements

- **FM-200**, Fluoroform - Ozone OK, high Global Warming
- NOVEC, Halotron - Ozone OK, low Global Warming

Levels for effective fire suppression, little effect on humans.

FM-200 is also known as **heptafluoropropane,** a typical propellant for asthma inhalers.

Higher levels affect CNS. Very heavy inhalation … ☠

Your Fire Detection and Sprinkler System is ...

A
Series
Of
Tubes
Like
The
Internet

Fire Detection, Sprinkler Systems

- Typical Passive Smoke Detection is relatively slow
- VESDA - Very Early-warning Aspirating Smoke Detection
  - Much faster at detecting smoke, at loss of location-accuracy
- VESDA allows for staff intervention before problem escalates
  - However, sometimes it’s so early, nobody finds anything ☹
  - This can lead to belief system prone to “false alarms” ☹
- Double Interlocked Pre-Action Sprinkler System
  - Requires smoke/heat sensor activation
  - Requires individual sprinkler head activation
  - Then Water can flow to activated sprinklers