UNIVERSITY of VIRGINIA

CS6501
CLOUD SYSTEM RELIABILITY

Course Introduction

Prof. Chang Lou, UVA CS, Fall 2024

AGENDA

- **What this course is about**
  - why study on cloud system reliability?
- **What you can expect from this course**
  - and what we expect from you

- **But first, tell us more about yourself!**
  - your name/background/hobby, and why you took this course

WHY YOU TOOK THIS COURSE

- **Likely answers**
  - **No exam**
  - Interested in cloud computing
  - Get experience for related jobs in industry
  - Try a small research project
  - Write more reliable codes

WHY CLOUD SYSTEM RELIABILITY MATTERS

WHY CLOUD SYSTEM RELIABILITY MATTERS

— Modern world depends on cloud systems

WHY CLOUD SYSTEM RELIABILITY MATTERS

— In 2023, 94% of enterprises use cloud services

https://map.datacente.rs/ https://www.zippia.com/advice/cloud-adoption-statistics/

WHAT IS RELIABILITY

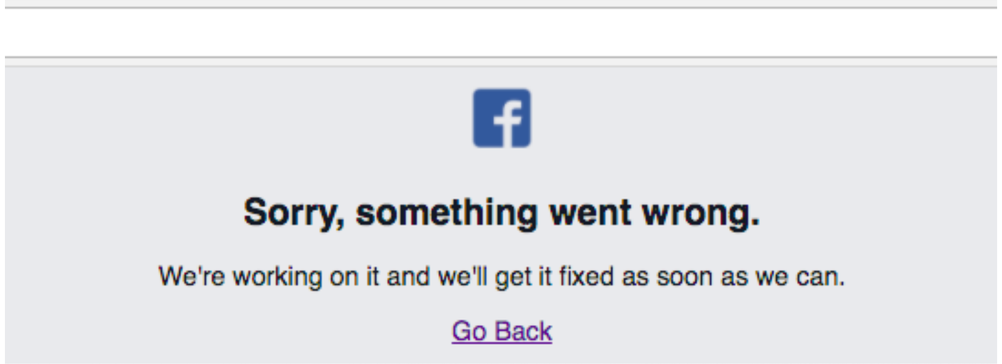
— What are some common qualities we measure on systems?

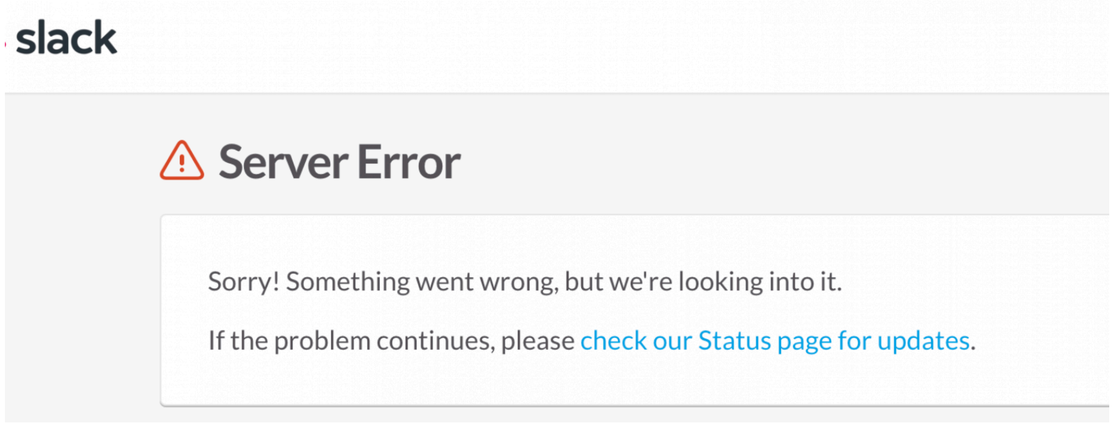
WHAT IS RELIABILITY

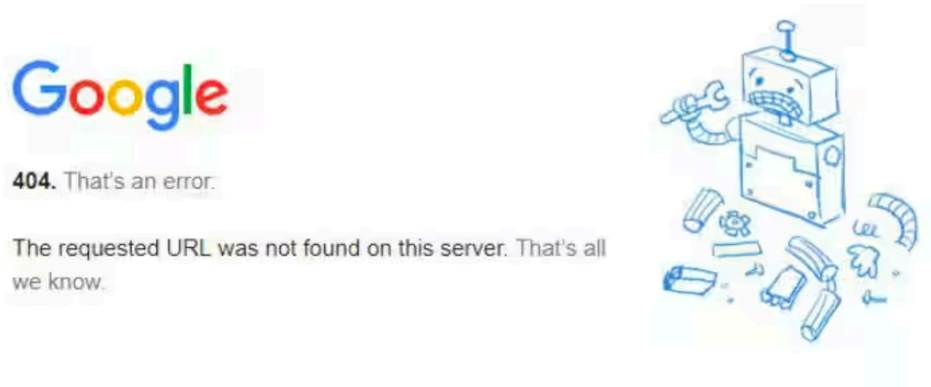
- **Reliability is not**
  - Performance: make systems faster
  - Usability: make systems more user-friendly
  - Security: make systems safer against intrusions
  - Cost-effectiveness: make systems more affordable
- **Reliability is**
  - the system's ability to consistently perform its intended function without **failure** over a given period.

CLOUD FAILURES

— Cloud failures are prevalent







CLOUD FAILURES

— Cloud failures can be really annoying

CLOUD FAILURES

— .. or much worse, huge economic loss and service unavailability

Microsoft's MFA is so strong, it locked out users for 8 hours

23 May 2013

3 difficult days for Rackspace Cloud Load Balancers

Posted by **iwgcr**

After almost 24 hours of technical difficulties, Facebook is back

Facebook blamed the issue on a "server configuration change."

Amazon 'missed out on $34m in sales during internet outage'

The e-commerce giant generates $9,615 in sales per second – but not when it's website is down

Ben Chapman • Tuesday 08 June 2021 16:54 • 1 Comments

Millions online hit by Microsoft 365 outages

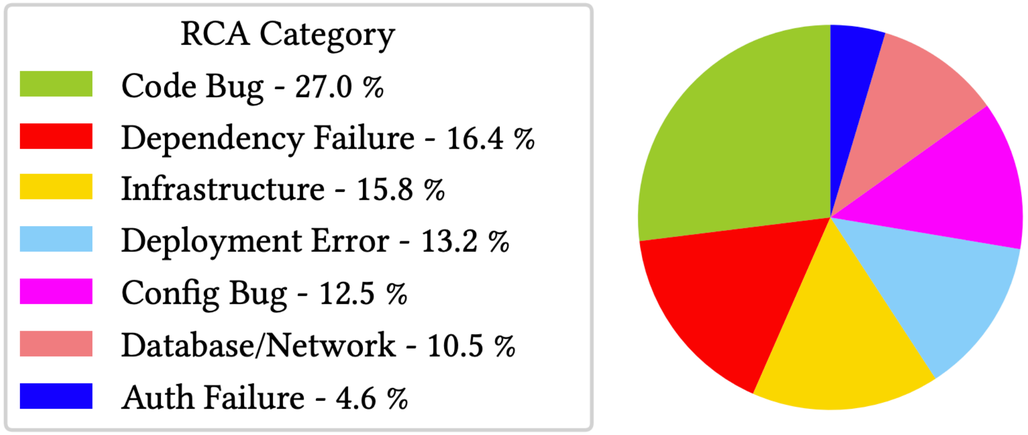
911 emergency services go down across the US after CenturyLink outage

Zack Whittaker @zackwhittaker / 4 months ago

Comment

CLOUD FAILURES

— Cloud systems fail due to different root causes



• S. Ghosh et al., How to Fight Production Incidents? An Empirical Study on a Large-scale Cloud Service, In Proceedings of the 13th Symposium on Cloud Computing (SoCC '22), November 2022.

CLOUD FAILURES

— .. sometimes very surprising root causes

TECH TECHNOLOGY GOOGLE FIBER

Google Fiber Shot Down By 'Bored' Hunters

'Bored' Hunters Shoot Down Google Fiber

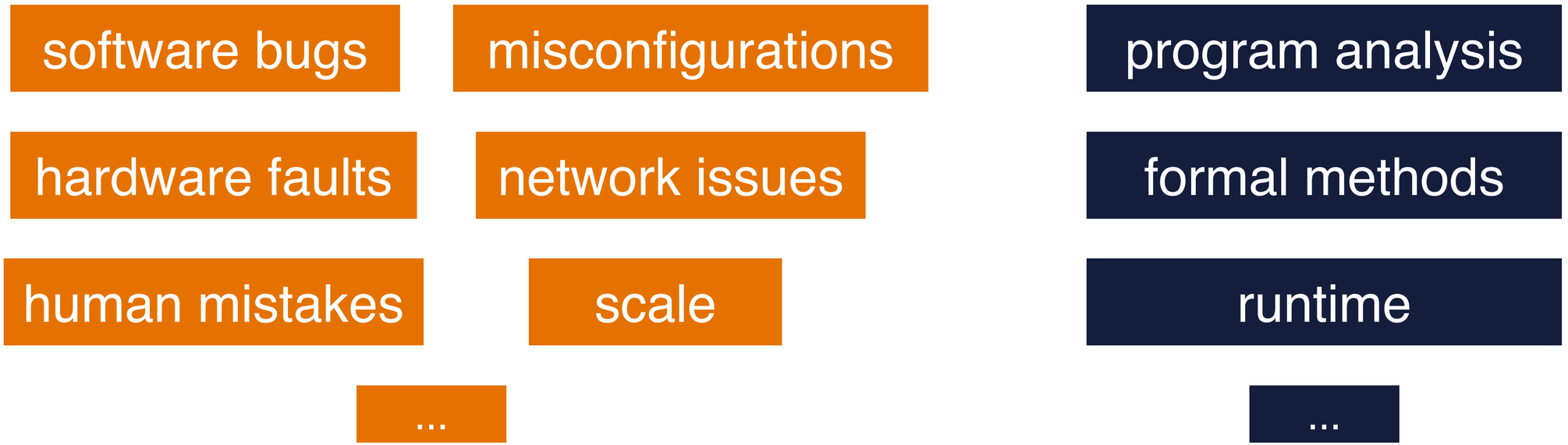
By Bianca Bosker

Nov 22, 2010, 05:12 AM EST | **Updated** May 25, 2011, 05:50 PM EDT

Google reinforces undersea cables after shark bites

Sharks have been biting down on fibre optic cables under the Pacific, possibly confused by electrical signals that resemble fish

IN THIS COURSE



GOAL OF THIS COURSE

- Introduce students to the foundational concepts of cloud computing and system reliability.
- Provide hands-on opportunities for students to engage in cutting-edge CSR research.

IN THIS COURSE

- **Review literatures about cloud reliability**
  - classic work + state-of-art
  - from top system conferences: SOSP/OSDI, NSDI, ASPLOS..
- **Learn how to present a research work**
  - ..and defend it like you were the author!
- **Explore a research topic you feel excited ★**
  - a try-out if you are considering to apply Ph.D. programs in the future

Course Info

COURSE INFO

- **Time/Location**
  - TuTh 3:30pm - 4:45pm, Rice Hall 340
- **Office Hours**
  - MoTh 5:00pm - 6:00pm, Rice Hall 304
- **Discussion Forum**
  - UVA Canvas (where you submit reviews)
- **Questions**
  - Send emails to chlou@virginia.edu or stop by at my office (open-door policy)

LECTURER INFO

- **Prof. Chang Lou**
  - Ph.D. from Johns Hopkins
  - Joined UVA CS as a faculty member in Fall 2023
- **Research Areas**
  - Distributed Systems, Operating Systems, Software Reliability
- **Research Goal**
  - Enhance cloud systems to be more resilient against arising reliability challenges

TA INFO

- **$ whoami**
  - Kahfi Soobhan Zulkifli
  - first year Ph.D. student at UVA CS
- **$ echo $OFFICE**
  - Rice Hall (TBD)
- **$ head research_interests**
  - Cloud
  - System for ML
- **$ cat ta.email**
  - kwf3wv@virginia.edu

GRADING

- Reviews: 15%
- Class Participation: 15%
- Presentation: 20%
- Project: 50%

REVIEWS

- **Each class we will discuss two papers**
  - choose one reading to write an one-page review
  - light reading on the other one
- **Submit to Canvas before 12:00 pm on class day**
  - maximum **five** reviews are allowed to miss without penalties
  - you **don't** need to submit reviews if you are the presenter for that paper (they will not be counted as missing reviews)

Violation of UVA Academic integrity: Directly copying from paper contents, peers or online resources will be considered as a violation of academic integrity and will lead to consequences.

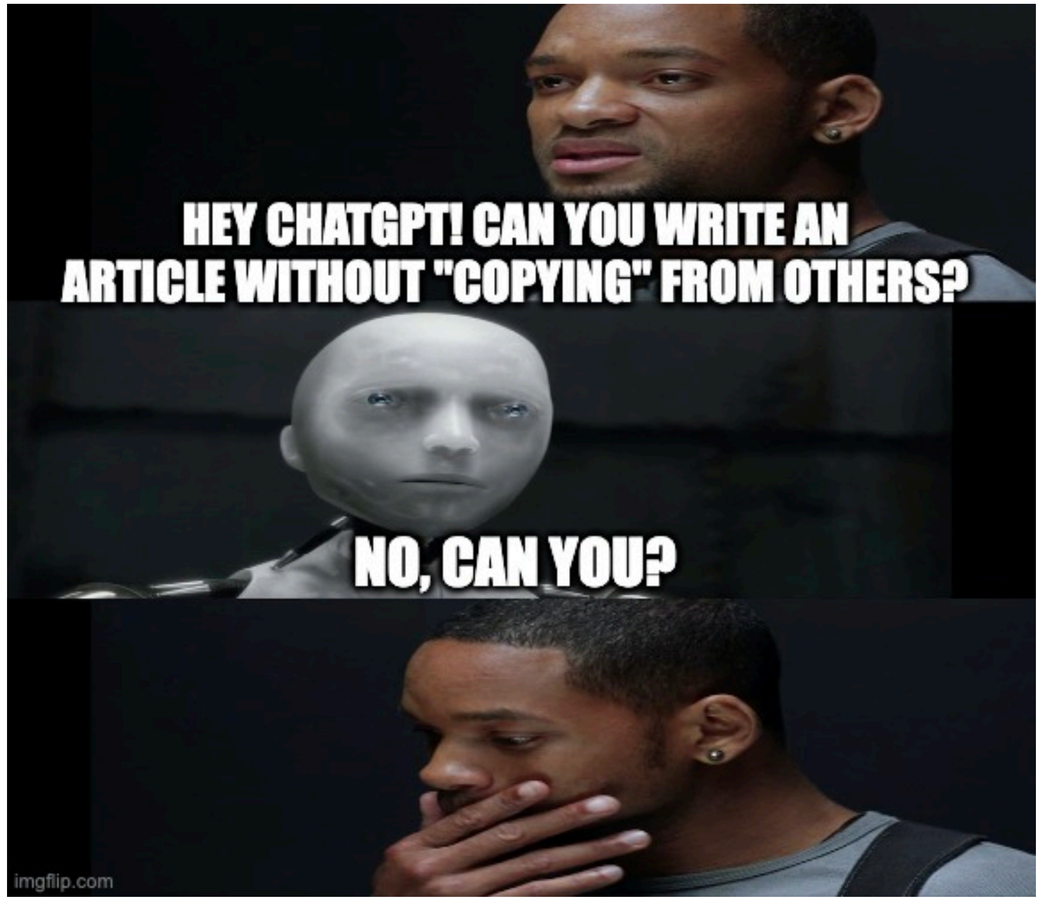
LATE POLICY

- **Everyone has 96 hour late tokens (for reviews, report ...)**
- **To use, just send an email to staff email list**
  - cs6501csrstaff@virginia.edu
  - late submissions using late tokens receive no penalty
- **What if I run out of tokens?**
  - 1 day late, 15% deduction
  - 2 days late, 30% deduction
  - 3 days late, 60% deduction
  - after 4 days, no credit.

CHATGPT POLICY

- **How to best use AI writing assistant**
  - brainstorm initial ideas
  - check grammar errors
- **You should not use AI to**
  - directly generate reviews for you

It is a violation of academic integrity as well.



But.. what if I have never read a paper before?

HOW TO READ A PAPER (3-PASS APPROACH)

by S. Keshav, University of Waterloo

- **The first pass: general idea (5-10 mins)**
  - title, abstract, and introduction
  - headings
  - conclusion
- **The second pass: content (< 1 hour)**
  - figures, diagrams and other illustrations in the paper
  - references
- **The third pass: details (may take hours)**
  - "re-implement" the paper

https://web.stanford.edu/class/ee384m/Handouts/HowtoReadPaper.pdf

HOW TO READ AN ENGINEERING RESEARCH PAPER

by William G. Griswold, CSE, UC San Diego

- What are the **motivations** for this work?
- What is the proposed **solution**?
- What is the work's **evaluation** of the proposed solution?
- What is your **analysis** of the identified problem, idea and evaluation?
- What are the **contributions**?
- What are **future directions** for this research?
- What **questions** are you left with?
- What is your **take-away message** from this paper?

CLASS PARTICIPATION

- **Attend classes**
  - discuss reviewed papers with peers
  - there might be a few random (but simple!) quizzes ☺
- **Exceptions: illness and other absences**
  - If you feel uncomfortable, it is better for you to get some rest at home
  - sometimes you need to pursue career opportunities - interviews, conferences, ..
  - **three** absences are allowed with no questions

PRESENTATION

- **You will present two papers through the semester**
  - Register at Google Sheet **TODAY**! (Link in Canvas Announcement)
  - Presenters in First Eight Slots get **bonus credits**
- **For each presentation**
  - Main body: 25-30 min
  - Q&A: 5 min
- **Lead the discussion and defend the work**
  - (it is called thesis "defense" for a reason)

QUICK TIPS FOR PRESENTATION

- **One common mistake: too many details**
  - the presenter tries to cover everything in the paper
  - however, it is impossible for anyone to learn all details in 40 min!
- **Simple tip: focus on three takeaways**
  - most people can only remember three things after the talk
  - build your slides around them to make points clear
  - key concepts + logic >> technical details

PROJECT

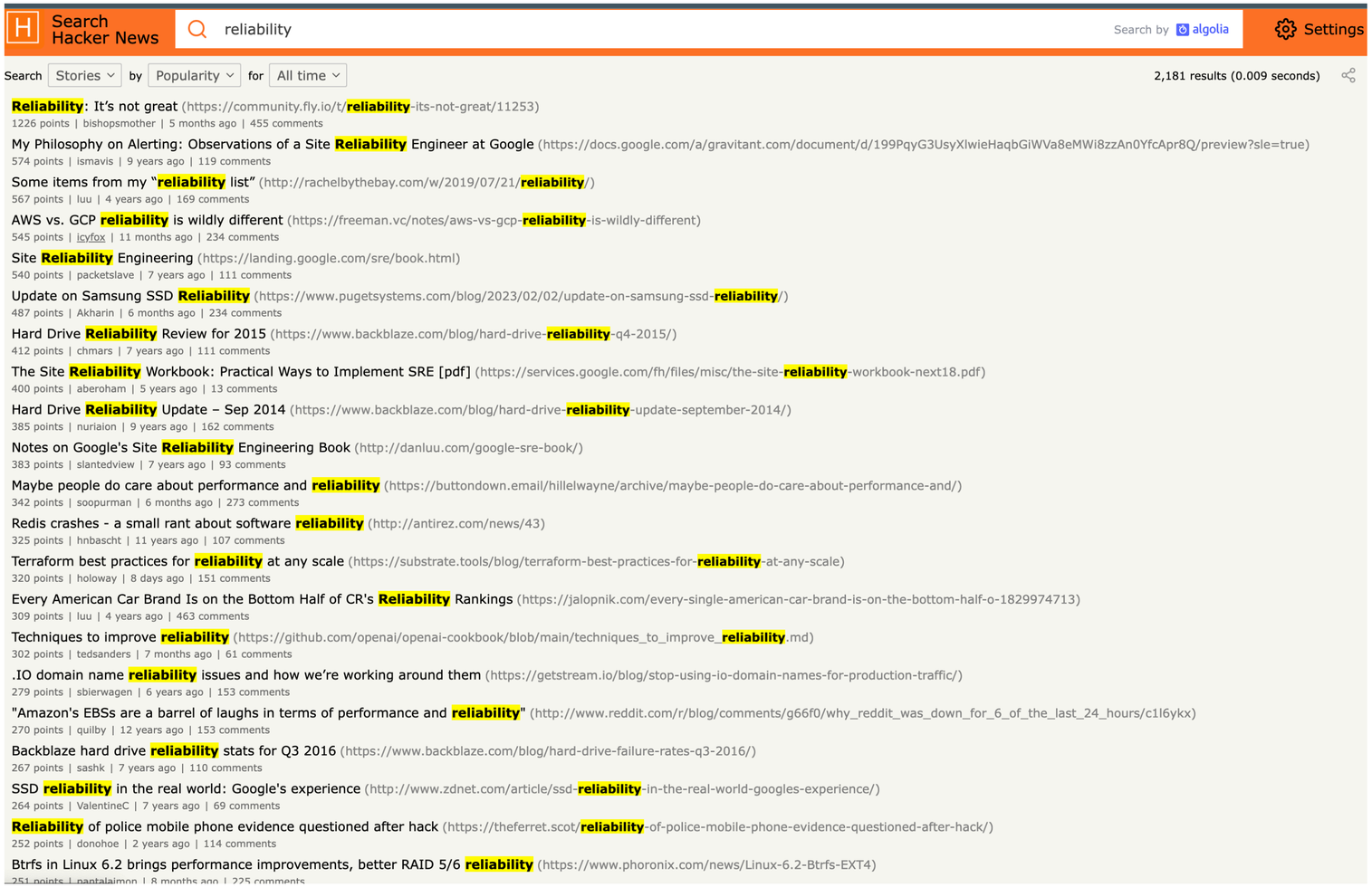
STEPS OF A RESEARCH PROJECT

- 1. Problem Identification
- 2. Proposing Hypothesis
- 3. Review of Related Literature
- 4. Preparation of Design
- 5. Experimentation
- 6. Results and Discussion

FIND YOUR TOPIC

- **Approach 1. Address the limitation of an existing paper**
  - "in this work we made the assumption that ..."
- **Approach 2. Revisit classic problems in new scenarios**
  - e.g., reliability problems in serverless/micro-services
- **Approach 3. Find inspirations from external sources**
  - what is the problem people complain about but no good solution yet?

INSPIRATIONS: HACKERNEWS



https://hn.algolia.com/?q=reliability

INSPIRATIONS: TECH BLOGS

One of my favorite pastimes is lunchtime with software engineers where I like to pose the following question: on average, how many software defects exist per 1,000 lines of delivered code? I've gotten answers across the board, and research backs up similar results – answers vary widely. But no one says "none." Results are subject to many factors, from programming language to developer skill, but some seem to suggest it is in the range of 10-20 per 1,000 lines.

For comparison, I just finished a small piece of C-code totaling 780 lines. Once all the obvious errors were dealt with and it compiled, the memory checker found four more serious issues. That's roughly five in 1,000 lines – but those are just memory issues like uninitialized memory reads and array overruns by one, etc. It does not indicate completeness, correctness, or even testing for infrequent code paths not included in the run. 10-20 per 1,000 seems right on the money to me.

The problem is cryptography may be mathematical algorithms, but the math is implemented in code. Bugs in cryptographic code happen. Recall the Java 15+ certificate validation bug? The bug was failing to check that the integers used in the algorithm were of sufficient size. In fact, integers of 0 qualified just fine, and so a certificate of all zeroes would be accepted as valid for whatever identity you tried to assume. Big bug, big consequences.

Share

Trending Blogs

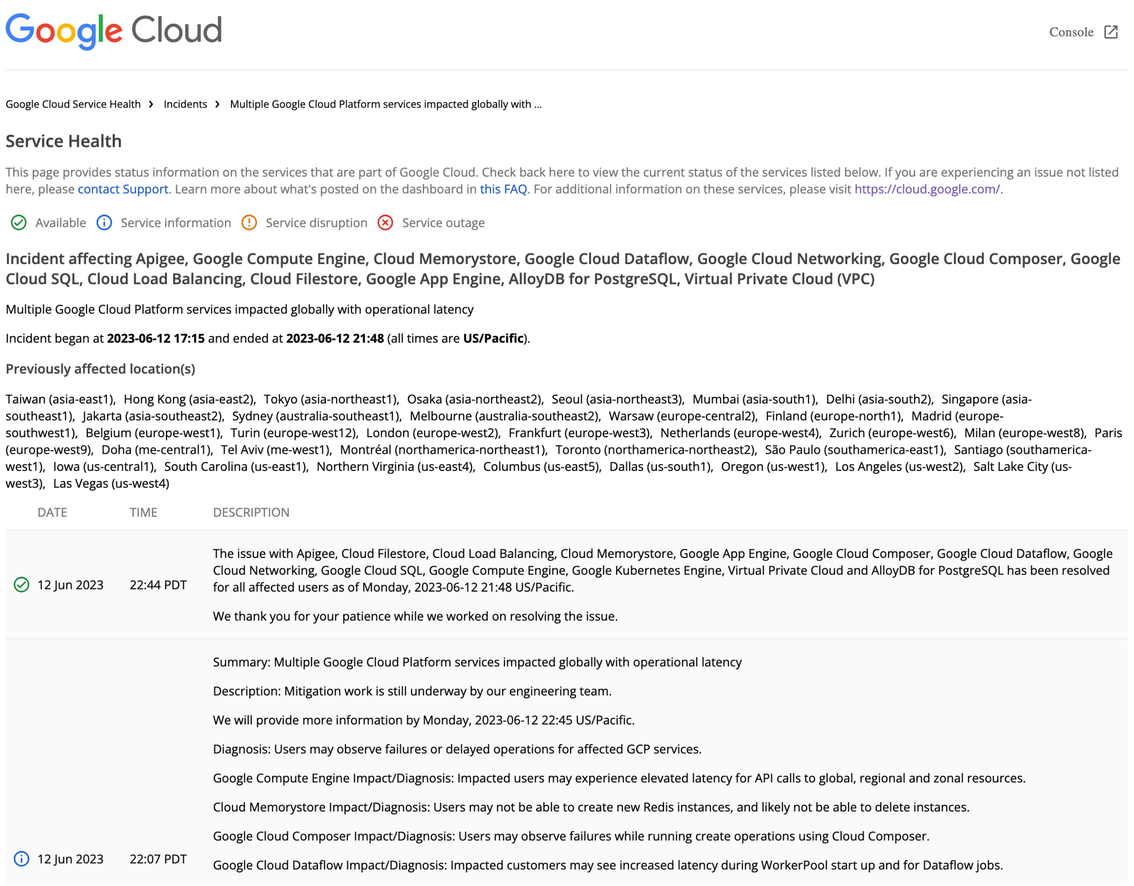
Crypto Convos Episode 1 Part 1 with Guest Adam Gordon

Crypto Convos Episode 2 with Guest Retired Admiral Mike Rogers

Crypto Convos Episode 3 with Guest Roger Grimes

https://quantumxc.com/blog/single-points-of-failure-in-cryptography-post-6-bugs-in-software/

INSPIRATIONS: POST MOMENTUM



Google Cloud

Console

Google Cloud Service Health › Incidents › Multiple Google Cloud Platform services impacted globally with ...

Service Health

This page provides status information on the services that are part of Google Cloud. Check back here to view the current status of the services listed below. If you are experiencing an issue not listed here, please contact Support. Learn more about what's posted on the dashboard in this FAQ. For additional information on these services, please visit https://cloud.google.com/.

Available Service information Service disruption Service outage

Incident affecting Apigee, Google Compute Engine, Cloud Memorystore, Google Cloud Dataflow, Google Cloud Networking, Google Cloud Composer, Google Cloud SQL, Cloud Load Balancing, Cloud Filestore, Google App Engine, AlloyDB for PostgreSQL, Virtual Private Cloud (VPC)

Multiple Google Cloud Platform services impacted globally with operational latency

Incident began at **2023-06-12 17:15** and ended at **2023-06-12 21:48** (all times are **US/Pacific**).

Previously affected location(s)

Taiwan (asia-east1), Hong Kong (asia-east2), Tokyo (asia-northeast1), Osaka (asia-northeast2), Seoul (asia-northeast3), Mumbai (asia-south1), Delhi (asia-south2), Singapore (asia-southeast1), Jakarta (asia-southeast2), Sydney (australia-southeast1), Melbourne (australia-southeast2), Warsaw (europe-central2), Finland (europe-north1), Madrid (europe-southwest1), Belgium (europe-west1), Turin (europe-west12), London (europe-west2), Frankfurt (europe-west3), Netherlands (europe-west4), Zurich (europe-west6), Milan (europe-west8), Paris (europe-west9), Doha (me-central1), Tel Aviv (me-west1), Montréal (northamerica-northeast1), Toronto (northamerica-northeast2), São Paulo (southamerica-east1), Santiago (southamerica-west1), Iowa (us-central1), South Carolina (us-east1), Northern Virginia (us-east4), Columbus (us-east5), Dallas (us-south1), Oregon (us-west1), Los Angeles (us-west2), Salt Lake City (us-west3), Las Vegas (us-west4)

DATE	TIME	DESCRIPTION
12 Jun 2023	22:44 PDT	The issue with Apigee, Cloud Filestore, Cloud Load Balancing, Cloud Memorystore, Google App Engine, Google Cloud Composer, Google Cloud Dataflow, Google Cloud Networking, Google Cloud SQL, Google Compute Engine, Google Kubernetes Engine, Virtual Private Cloud and AlloyDB for PostgreSQL has been resolved for all affected users as of Monday, 2023-06-12 21:48 US/Pacific. We thank you for your patience while we worked on resolving the issue.
12 Jun 2023	22:07 PDT	Summary: Multiple Google Cloud Platform services impacted globally with operational latency Description: Mitigation work is still underway by our engineering team. We will provide more information by Monday, 2023-06-12 22:45 US/Pacific. Diagnosis: Users may observe failures or delayed operations for affected GCP services. Google Compute Engine Impact/Diagnosis: Impacted users may experience elevated latency for API calls to global, regional and zonal resources. Cloud Memorystore Impact/Diagnosis: Users may not be able to create new Redis instances, and likely not be able to delete instances. Google Cloud Composer Impact/Diagnosis: Users may observe failures while running create operations using Cloud Composer. Google Cloud Dataflow Impact/Diagnosis: Impacted customers may see increased latency during WorkerPool start up and for Dataflow jobs.

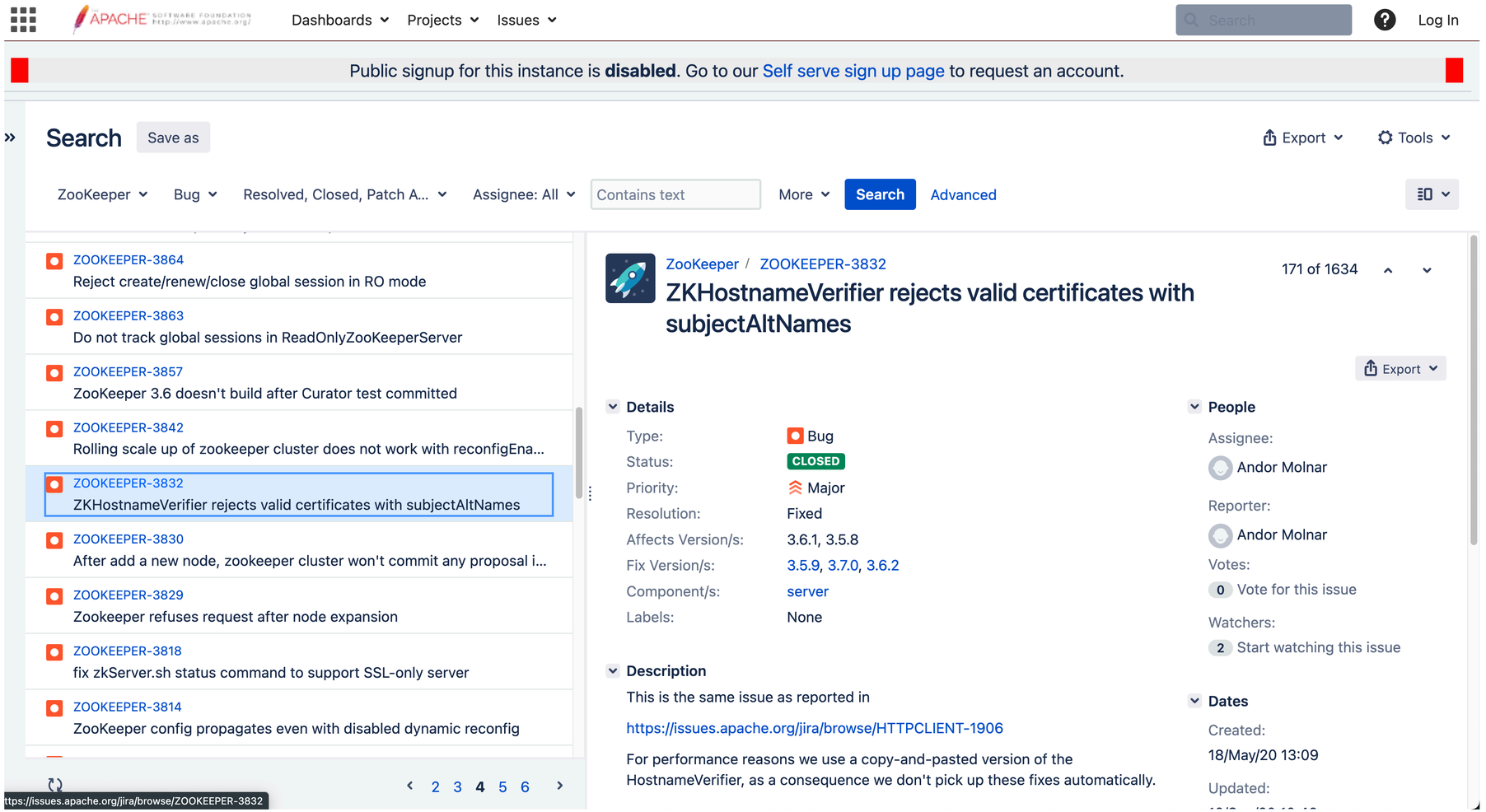
https://status.cloud.google.com/incidents/VuCtCwkRXueAyusvrXfG

PROJECT TOPIC

- **What if I cannot think out of anything?**
- **It's fine, we have a list of prepared topics**
  - but we suggest coming up with your own idea, something you feel passionate about
  - a fun and novel idea with okay execution is better than a boring idea executed perfectly

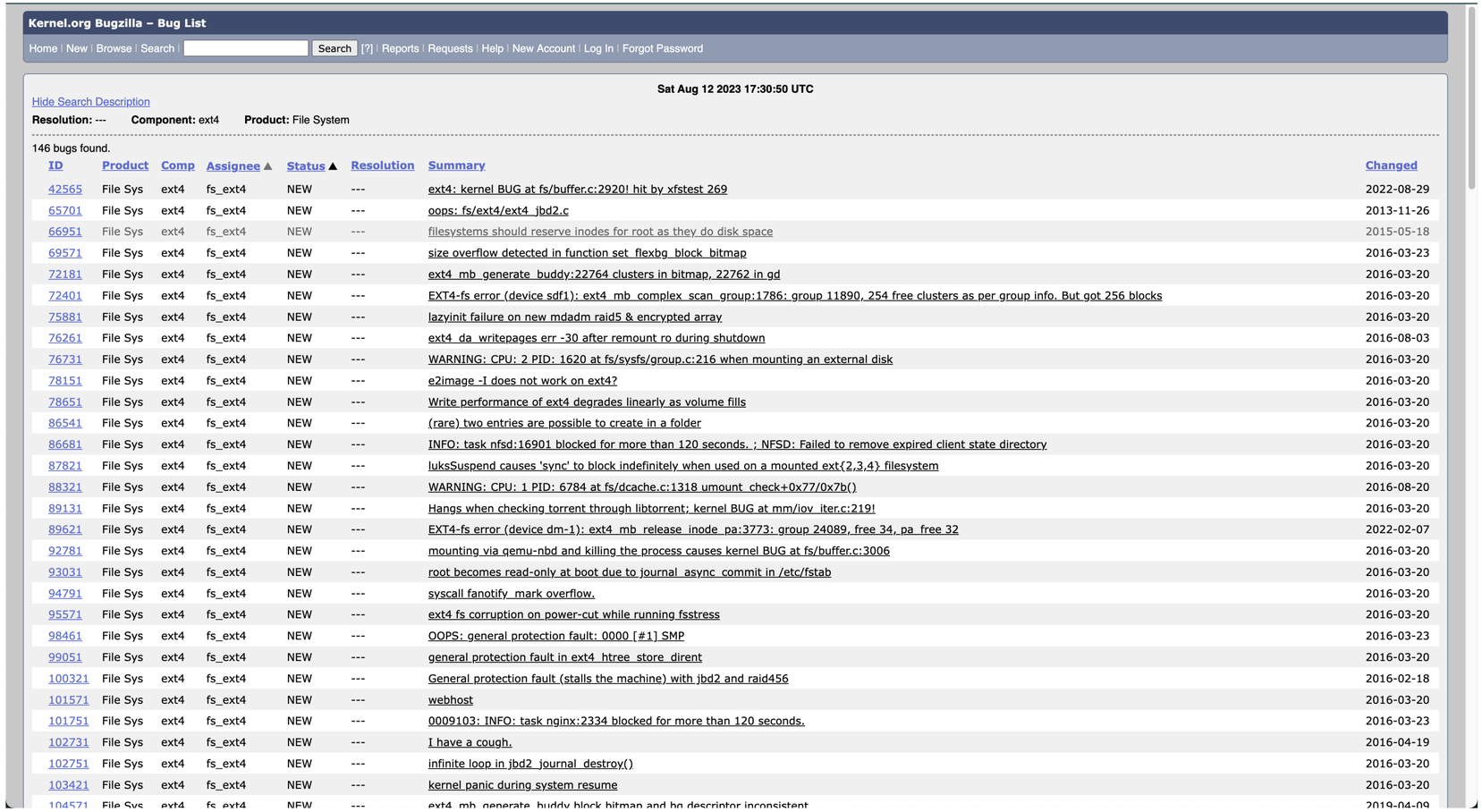
Exercise: bug detection tool for ChatGPT generated codes in distributed systems

SOURCES: JIRA



https://issues.apache.org/jira/browse/ZOOKEEPER-3832

SOURCES: BUGZILLA



Kernel.org Bugzilla – Bug List

Home | New | Browse | Search | Search [?] | Reports | Requests | Help | New Account | Log In | Forgot Password

Sat Aug 12 2023 17:30:50 UTC

Hide Search Description

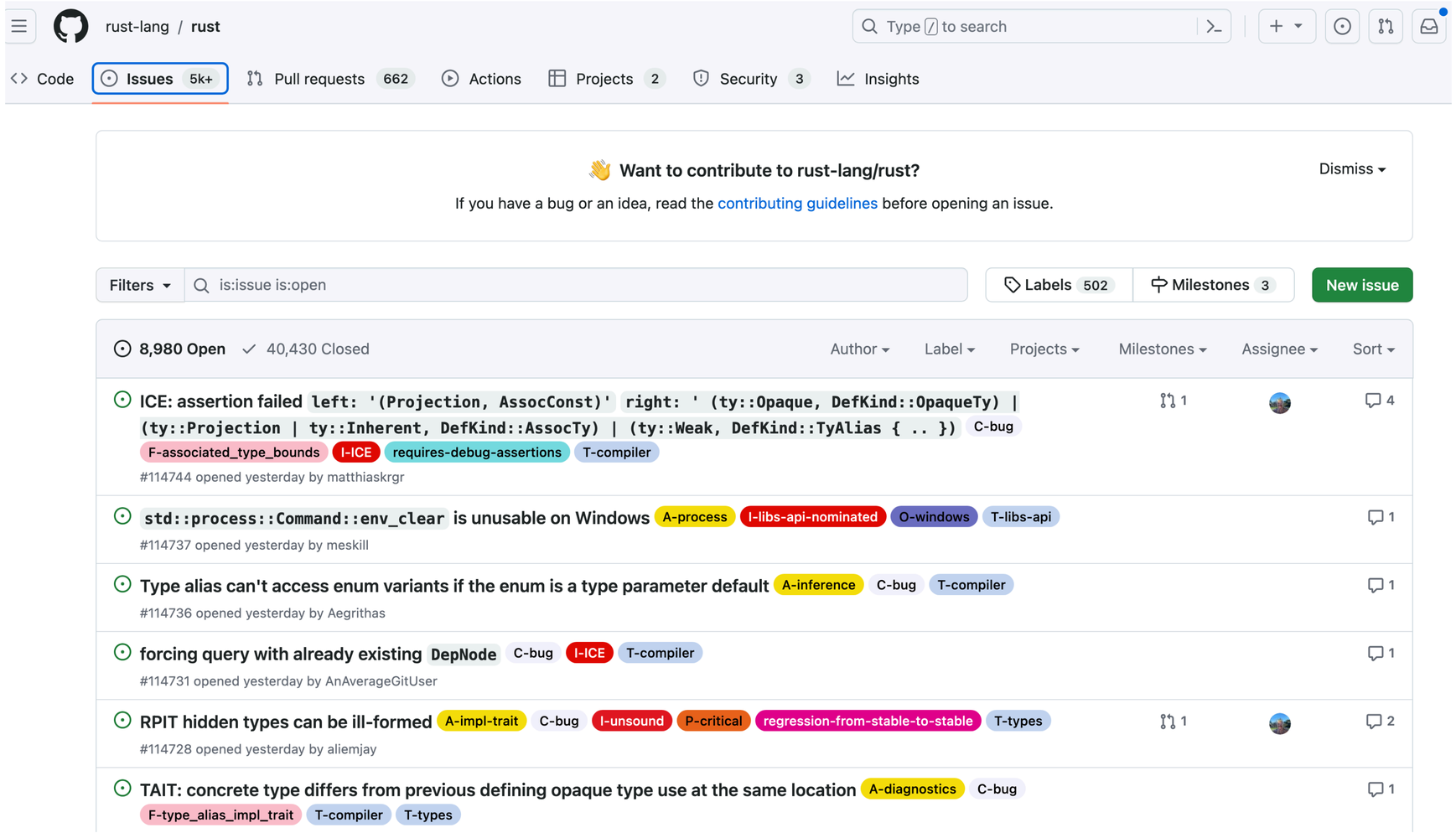
Resolution: --- **Component:** ext4 **Product:** File System

146 bugs found.

ID	Product	Comp	Assignee ▲	Status ▲	Resolution	Summary	Changed
42565	File Sys	ext4	fs_ext4	NEW	---	ext4: kernel BUG at fs/buffer.c:2920! hit by xfstest 269	2022-08-29
65701	File Sys	ext4	fs_ext4	NEW	---	oops: fs/ext4/ext4_jbd2.c	2013-11-26
66951	File Sys	ext4	fs_ext4	NEW	---	filesystems should reserve inodes for root as they do disk space	2015-05-18
69571	File Sys	ext4	fs_ext4	NEW	---	size overflow detected in function set_flexbg_block_bitmap	2016-03-23
72181	File Sys	ext4	fs_ext4	NEW	---	ext4_mb_generate_buddy:22764 clusters in bitmap, 22762 in gd	2016-03-20
72401	File Sys	ext4	fs_ext4	NEW	---	EXT4-fs error (device sdf1): ext4_mb_complex_scan_group:1786: group 11890, 254 free clusters as per group info. But got 256 blocks	2016-03-20
75881	File Sys	ext4	fs_ext4	NEW	---	lazyinit failure on new mdadm raid5 & encrypted array	2016-03-20
76261	File Sys	ext4	fs_ext4	NEW	---	ext4_da_writepages err -30 after remount ro during shutdown	2016-08-03
76731	File Sys	ext4	fs_ext4	NEW	---	WARNING: CPU: 2 PID: 1620 at fs/sysfs/group.c:216 when mounting an external disk	2016-03-20
78151	File Sys	ext4	fs_ext4	NEW	---	e2image -I does not work on ext4?	2016-03-20
78651	File Sys	ext4	fs_ext4	NEW	---	Write performance of ext4 degrades linearly as volume fills	2016-03-20
86541	File Sys	ext4	fs_ext4	NEW	---	(rare) two entries are possible to create in a folder	2016-03-20
86681	File Sys	ext4	fs_ext4	NEW	---	INFO: task nfsd:16901 blocked for more than 120 seconds. ; NFSD: Failed to remove expired client state directory	2016-03-20
87821	File Sys	ext4	fs_ext4	NEW	---	luksSuspend causes 'sync' to block indefinitely when used on a mounted ext{2,3,4} filesystem	2016-03-20
88321	File Sys	ext4	fs_ext4	NEW	---	WARNING: CPU: 1 PID: 6784 at fs/dcache.c:1318 umount_check+0x77/0x7b()	2016-08-20
89131	File Sys	ext4	fs_ext4	NEW	---	Hangs when checking torrent through libtorrent; kernel BUG at mm/iov_iter.c:219!	2016-03-20
89621	File Sys	ext4	fs_ext4	NEW	---	EXT4-fs error (device dm-1): ext4_mb_release_inode_pa:3773: group 24089, free 34, pa_free 32	2022-02-07
92781	File Sys	ext4	fs_ext4	NEW	---	mounting via qemu-nbd and killing the process causes kernel BUG at fs/buffer.c:3006	2016-03-20
93031	File Sys	ext4	fs_ext4	NEW	---	root becomes read-only at boot due to journal_async_commit in /etc/fstab	2016-03-20
94791	File Sys	ext4	fs_ext4	NEW	---	syscall fanotify_mark overflow.	2016-03-20
95571	File Sys	ext4	fs_ext4	NEW	---	ext4 fs corruption on power-cut while running fsstress	2016-03-20
98461	File Sys	ext4	fs_ext4	NEW	---	OOPS: general protection fault: 0000 [#1] SMP	2016-03-23
99051	File Sys	ext4	fs_ext4	NEW	---	general protection fault in ext4_htree_store_dirent	2016-03-20
100321	File Sys	ext4	fs_ext4	NEW	---	General protection fault (stalls the machine) with jbd2 and raid456	2016-02-18
101571	File Sys	ext4	fs_ext4	NEW	---	webhost	2016-03-20
101751	File Sys	ext4	fs_ext4	NEW	---	0009103: INFO: task nginx:2334 blocked for more than 120 seconds.	2016-03-23
102731	File Sys	ext4	fs_ext4	NEW	---	I have a cough.	2016-04-19
102751	File Sys	ext4	fs_ext4	NEW	---	infinite loop in jbd2_journal_destroy()	2016-03-20
103421	File Sys	ext4	fs_ext4	NEW	---	kernel panic during system resume	2016-03-20
104571	File Sys	ext4	fs_ext4	NEW	---	ext4_mb_generate_buddy block bitmap and bg descriptor inconsistent	2019-04-09

https://bugzilla.kernel.org/buglist.cgi?component=ext4

SOURCES: GITHUB ISSUES



https://github.com/rust-lang/rust/issues

FIND LITERATURES

- **From top system conferences**
  - SOSP/OSDI/Eurosys/ATC (system)
  - HotOS (workshop)
  - NSDI/SIGCOMM (network)
  - ASPLOS (programming language and arch)
  - FAST (storage)
  - SOCC (cloud)

PROJECT

- **Week 1: Form a team (2-3 students)**
  - brainstorm on your project topic (related to cloud system reliability)
- **Week 2-3: Schedule a meeting with me to discuss**
- **Week 4-5: Write and submit proposal (1-2 pages)**
- **Week 6-14: Do research**
  - Week 11: Checkpoint report due (~3 pages)
- **Week 15-16: Presentation**
- **Week 17: Final report (including codes)**

IMPORTANT TIMELINE

Concepts

RELIABILITY VS AVAILABILITY

- **Reliability**
  - the probability that a system operates without failure in a given period of time.
  - how to compute probability: Mean Time Between Failures (MTBF)

$$Reliability = 1 - \frac{1}{MTBF} = 1 - \frac{NumofBreakdowns}{E[uptime]}$$

- **Availability**
  - the percentage of time that the system operates satisfactorily.

$$Avalability = \frac{E[uptime]}{E[uptime] + E[downtime]}$$

RELIABILITY VS AVAILABILITY

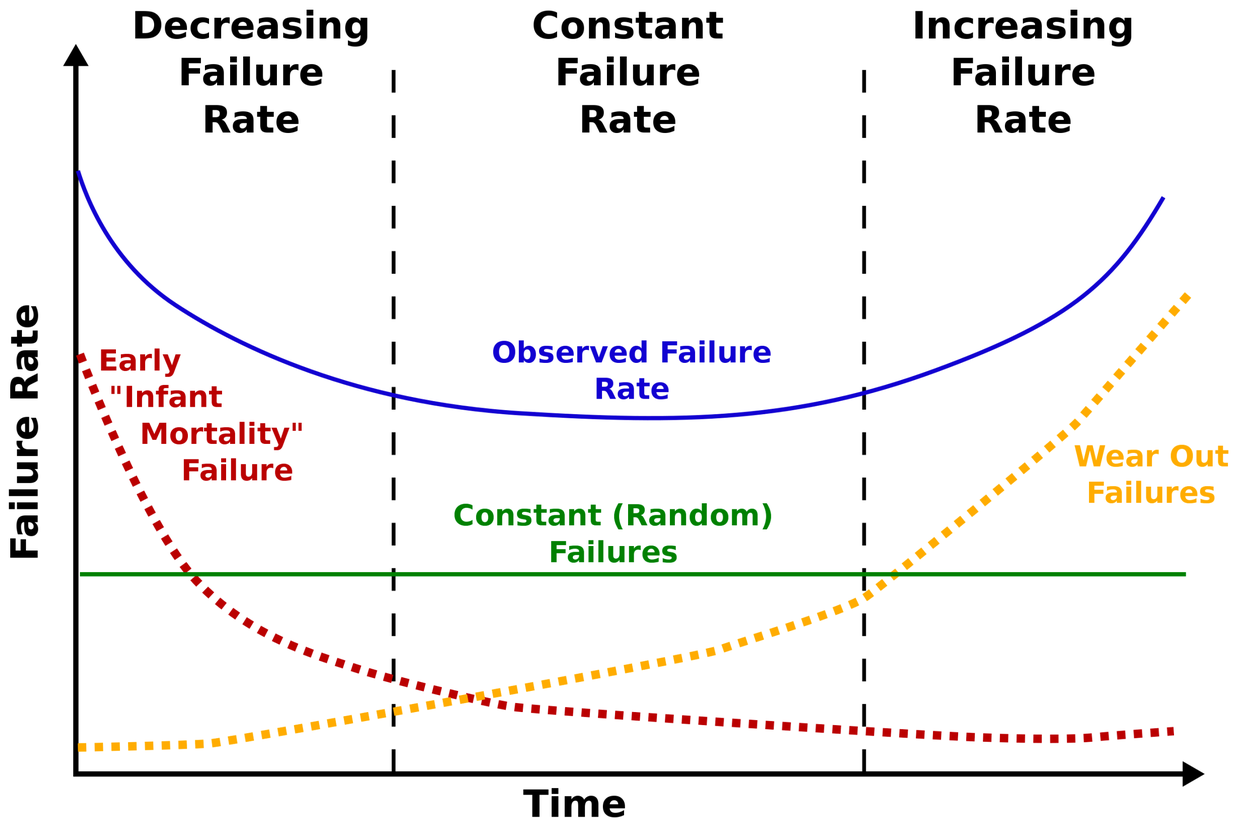
- **Reliability and availability are related concepts**
  - but not same! Imagine there are two bad laundry machines:

Stop working every 6 months,
takes 1 week to repair

Stop working every 3 weeks,
takes 1 hour to repair

- **Modern distributed systems: unreliable but high available**

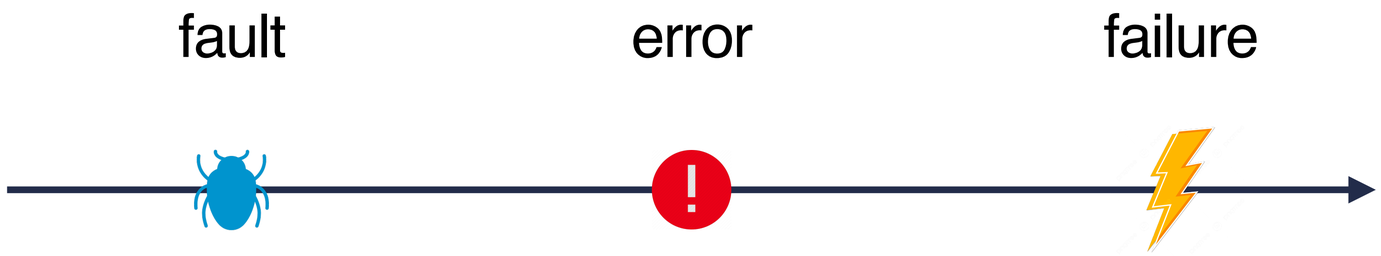
THE BATHTUB CURVE



RELIABILITY VS SECURITY

- **Both are essential aspects of system design**
  - reliability and security research have a lot of similarities
  - many bugs can both hurt system reliability and security
- **Key difference: adversary**
  - reliability: bad designs and mistakes are their own worst enemies
  - security: threats with bad intentions (malware, hackers, etc.)

FAULT, ERROR, FAILURE



root cause: a bug, misconfiguration, hardware issue, etc.

fault triggered, sth goes wrong (e.g., I/O error), could be handled

visible impact to users

UNIVERSITY of VIRGINIA

Q & A

- **ToDo 1) Team up**
  - find your teammates and discuss potential ideas
  - team leader emails me name list and schedules meeting **(ddl: 9/3)**
- **ToDo 2) login Canvas**
  - sign up in the presentation schedule for two lectures
  - submit review for Thursday class

chlou@virginia.edu

ACKNOWLEDGEMENT

PROF. RYAN HUANG (UMICH) AND PROF. TIANYIN XU (UIUC) FOR SHARING THEIR COURSE MATERIALS AND TEACHING INSIGHTS