How NOT to do research on an open source community...

Greg Kroah-Hartman

David A. Wheeler

May 2021

THE LINUX FOUNDATION

Timeline (1 of 4)

- 2020 Aug 9..21: “Hypocrite Commits” patches sent from UMN researchers
  - Attempted to introduce vulnerabilities to see if they would be detected
  - Sent to kernel developers from false identities; without consent, notice, or ethics review
- 2020 Nov: Draft “Hypocrite Commits” paper is published
- 2020 Nov 22: Sarah Jamie Lewis calls attention to paper’s questionable ethics
- 2020 Dec 1: Lewis & others send letter to IEEE S&P, questioning ethics
- 2020 Dec [day unknown]: UMN IRB appears to give after-the-fact exemption to research on the basis that it believes the research is not human research
- 2020 Dec 15: UMN issues clarification
- 2021 Apr 6: Poor quality patches sent by UMN after ~7 months of silence
  - Raises spectre of ***continued*** attacks
- 2021 Apr 20: Greg K-H asks submitters to stop sending poor quality patches under the guise of “research on maintainers”
  - Researcher claimed new set of patches was not part of previous research
  - Greg replies, ***umn.edu submissions*** should be ***rejected*** until all figured out

UMN = University of Minnesota, IEEE = Institute of Electrical and Electronics Engineers, S&P = Security & Privacy

Source: TAB report

Timeline (2 of 4)

- 2021 Apr 21: Greg K-H requests review @umn.edu reverts, TAB begins review
- 2021 Apr 23: Linux Foundation sends letter to UMN requesting:
    - Id all proposals of known-vulnerable code from any U of MN experiment
    - Withdraw, from formal publication, research where subjects didn’t give prior consent
    - Ensure all future U of MN experiments on people *first* have review and approval
    - Ensure all future reviews of proposed experiments on people will normally ensure the consent of those being experimented on
- 2021 Apr 24: UMN publishes “An open letter to the Linux community”
- 2021 Apr 26: UMN researchers retract "Hypocrite Commits" paper from formal publication
- 2021 Apr 27: UMN published details on commits & replies to LF
    - Paper withdrawn. UMN believes it’s not “human subjects research”
    - Will do faculty ethics training in 2021-2022, explore added processes, to prevent similar situations

Source: TAB report

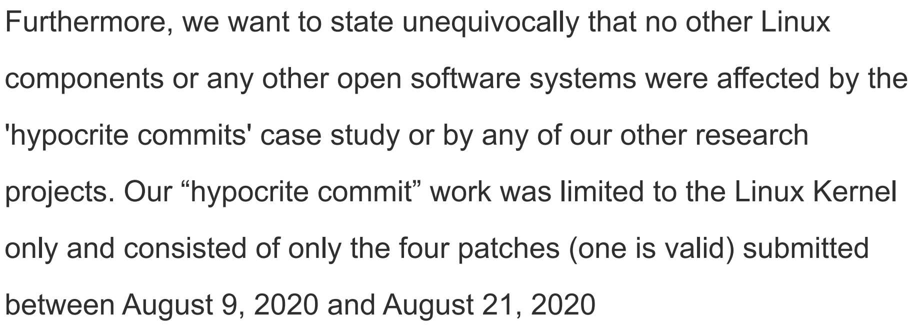
Timeline (3 of 4)

- 2021 May 3: Greg K-H posts a final set of reverts, along with correct fixes
- 2021 May 5: Linux TAB publishes detailed report, with due diligence audit results
  - 435 UMN commits were re-reviewed, thanks to 85 Linux kernel developers
  - Confirmed that all intentionally-vulnerable patches with vulnerabilities were rejected
    - One (“patch 1”) was intended to be vulnerable, but due to lack of understanding by the submitter, it was valid & was accepted
    - Yes, you read that correctly, you can’t make this stuff up :-)
    - Patch 1 was asked to be removed because submission was made under a false name (there have been exceptions, but true identities are still known to a subset)
  - Huge majority of the reviewed commits (349) were found to be correct
  - UMN overall patch quality relatively poor; 25 were fixed by later commits, 39 needed fixing

Source: TAB report

Timeline (4 of 4)

- 2021 May 6: UMN meets with Greg, Kees and LF to discuss productive ways to move forward and improve
- 2021 May 6: IEEE publishes statement about how the paper violated ethical guidelines and what would be put into place to prevent it happening again
- 2021 May 7: UMN responds to TAB report, verifying it is correct
  - Identifies one further set of patches from their team, using a private email address in February 2021. All were rejected by the community as they were invalid changes.
  - Stated that they had only done this for the Linux kernel, not for any other open source project:



Furthermore, we want to state unequivocally that no other Linux components or any other open software systems were affected by the 'hypocrite commits' case study or by any of our other research projects. Our “hypocrite commit” work was limited to the Linux Kernel only and consisted of only the four patches (one is valid) submitted between August 9, 2020 and August 21, 2020

Source: TAB report

Issues Created

- Submitting patches using false identities intending to deceive a community
- Submitting patches with known vulnerabilities (versus innocently submitting poor quality code)
- Researching on a community without notice or consent
- Every security conscious community scrambled to identify “Did UMN contribute known-vulnerable code to our project?”
    - UMN has assured us that only the Linux kernel was targeted
- Pattern of poor quality proposals (even when not intentional)
    - Asked UMN to designate a set of experienced developers to review and provide feedback on proposed kernel changes before those changes are submitted publicly; UMN agreed
    - Identical to what has been put into place for other companies

The BAD News (1 of 2)

- It’s unclear other communities without Linux kernel-level review practices would have caught these issues
- Researchers created massive amount of extra work for developer community
- IRB & other ethics process scope/definitions may not clearly cover research on community processes, even if humans are involved in those processes
  - UMN says this was a mistake & apologizes, but claims it’s not “Human Subjects Research” per US federal regulations (e.g., 45 CFR 46.102)
  - Yet US Belmont Report, Common Rule, & Menlo Report emphasize *consent*
  - Human Subject Regulations Decision Charts suggest to us it *was* (for US)
  - IEEE responded, “paper does not follow [ethical] guidelines”
  - James Davis argues researchers don’t grok sociotechnical systems
  - NSF has been notified, need to watch this carefully

The BAD News (2 of 2)

- Researchers sometimes do not interact with production development environments appropriately
  - Due to incentive misalignment & lack of guidance for researchers
  - TAB is working to develop guidance specifically for researchers

- Issues apply far more broadly than UMN, or US, or the Linux kernel
  - UMN promises to add ethics training & code review
  - How can we scale beyond UMN to all research?
  - How can we ensure OSS community issues are *included* in ethics decisions?

The GOOD News

- The Linux kernel code review process ***worked***
  - ***All*** UMN intentionally-vulnerable buggy patches were ***caught*** and not accepted
    - Note: One patch was accepted because it was unintentionally correct
- The Linux kernel developers rapidly reviewed all UMN contributions
  - Double-check of code should increase confidence by users & potential users
- Strong public support for the Linux kernel developers response and position, including from researchers who have been working with the kernel community for decades
- UMN apologized & actively working to prevent recurrence
- Many organizations around the world have seen the fallout from submitting intentionally weak patches and are on notice they might be banned by the community

Where do we go from here?

- The LF and U of MN had productive discussions, including a call with key deans and leadership
- The TAB will facilitate identifying a technical mentor for U of MN similar to what we do for member companies in need of help
- U of MN will be reviewing and revising its ethical research policy, outside of IRB purview
    - U of MN promised to give the LF a heads up on a future draft
- Greg K-H and the TAB will be working with research institutions that have worked well with the kernel community to publish best practices for community research that can be a future guide for U of MN and others

Contact Us

Learn more at linuxfoundation.org.

For general inquiries, questions related to membership, or about our events or training offerings, please visit linuxfoundation.org/about/contact/

Mailing Address

548 Market St
PMB 57274
San Francisco, California
94104-5401 US
Phone/Fax: +1 415 723 9709

Legal Notice

The Linux Foundation, The Linux Foundation logos, and other marks that may be used herein are owned by The Linux Foundation or its affiliated entities, and are subject to The Linux Foundation's Trademark Usage Policy at https://www.linuxfoundation.org/trademark-usage, as may be modified from time to time.

Linux is a registered trademark of Linus Torvalds. Please see the Linux Mark Institute's trademark usage page at https://lmi.linuxfoundation.org for details regarding use of this trademark.

Some marks that may be used herein are owned by projects operating as separately incorporated entities managed by The Linux Foundation, and have their own trademarks, policies and usage guidelines.

TWITTER, TWEET, RETWEET and the Twitter logo are trademarks of Twitter, Inc. or its affiliates.

Facebook and the "f" logo are trademarks of Facebook or its affiliates.

LinkedIn, the LinkedIn logo, the IN logo and InMail are registered trademarks or trademarks of LinkedIn Corporation and its affiliates in the United States and/or other countries.

YouTube and the YouTube icon are trademarks of YouTube or its affiliates.

All other trademarks are the property of their respective owners. Use of such marks herein does not represent affiliation with or authorization, sponsorship or approval by such owners unless otherwise expressly specified.

The Linux Foundation is subject to other policies, including without limitation its Privacy Policy at https://www.linuxfoundation.org/privacy and its Antitrust Policy at https://www.linuxfoundation.org/antitrust-policy. each as may be modified from time to time. More information about The Linux Foundation's policies is available at https://www.linuxfoundation.org.

Please email legal@linuxfoundation.org with any questions about The Linux Foundation's policies or the notices set forth on this slide.