SOHOM DATTA

sohomdatta1+web(at)gmail.com

MediaWiki Gerrit • GitHub • Linkedin

EDUCATION

- **North Carolina State University, Raleigh** *August 2024 – Ongoing*
  PhD in Computer Science (Focus: Web security and privacy)
- **Manipal Institute of Technology, Manipal** *July 2019 – August 2023*
  B.Tech in Computer Science and Engineering (Coursework: Compilers, Software testing, Computer security)

PUBLICATIONS

- **[To appear] Cross-Boundary Mobile Tracking: Exploring Java-to-JavaScript Information Diffusion in WebViews**
  Sohom Datta, Michalis Diamantaris, Ahsan Zafar, Junhua Su, Anupam Das, Jason Polakis, Alexandros Kapravelos
  Proceedings of the Network and Distributed System Security Symposium (NDSS) (2026).
- **[To appear] Same Script, Different Behavior: Characterizing Platform-Specific Divergent JavaScript Execution**
  Ahsan Zafar, Junhua Su, **Sohom Datta**, Alexandros Kapravelos, Anupam Das
  Proceedings of the ACM Conference on Computer and Communications Security (CCS) (2025).

RESEARCH AND TECHNICAL EXPERIENCE

Research Assistant, Wolfpack Security and Privacy Research Lab *January 2024 – Ongoing*
North Carolina State University

- Conducted research into privacy-violating JavaScript code across different environments, like mobile browsers and WebViews.
- Maintained and extended VisibleV8, an open-source browser instrumentation framework across 25+ major Chrome releases.
- Ported VisibleV8 to Android WebViews, enabling the first dynamic analyses of privacy violations in the Android advertising ecosystem across 1K apps.
- Designed infrastructure and co-organized HackPackCTF, a competition with 1000+ participants worldwide.

Visiting researcher, Software Security Research Group *September 2023 - December 2023*
Max Planck Institute of Security and Privacy

- Worked on understanding race-condition bugs in application layer logic in across widely used web servers like Express and NextJS.
- Implemented fuzzers to automatically detect and flag security-critical race-condition issues in application logic on web servers.

Student Intern, Wolfpack Privacy and Security Lab *February 2023 - July 2023*
North Carolina State University

- Implemented large-scale architectural changes to the VisibleV8 crawling pipeline that realized a 20x improvment in crawling and post-processing logs generated by visiting websites.
- Contributed patches to fix deficiencies in VisibleV8's logging and tracing capabilities, such as it's ability to trace `eval(...)` and other code execution pathways.

Student Software Developer, Chromium *June 2022 – November 2022*
Google Summer of Code

- Worked on aligning Chromium's implementation of the Performance API with the W3C specifications.
- Discovered and fixed 3 bugs related to incorrect First Contentful Paint and Largest Contentful Paint entries while refactoring the way the PaintTiming API marked contentful images in Chromium.
- Collaborated with the W3C Web Performance Working Group to rectify an issue in the way LCP timing entries were being reported at the Painting layer.
- Discovered and patched systemic security issues (cross-site leakages) in the implementation of the ResourceTiming API in the major browser engines, Webkit, Blink and Gecko (Firefox). (`CVE-2023-1232`)

Member, Cryptonite Manipal (cybersecurity student project) *April 2021 – August 2022*
Manipal Institute of Technology

- Led a cybersecurity team of 22 engineering students that placed among the top 15 in India in CSAW CTF '21 hosted by NYU (one of the oldest capture the flag competitions for academic teams), secured 2nd place in India in ASIS CTF Finals 2021 and was ranked among the top 12 teams in India on CTFTime in 2021-2022.
- Created challenges and actively managed the security infrastructure for niteCTF, an international cybersecurity capture the flag event that attracted over 1200+ participants from 43 countries.
- Conducted workshops on control flow integrity, binary exploitation and format strings exploits detailing the state-of-the-art research in the area, including mitigation techniques such as ASLR, stack cookies, and fuzzy testing.
- Worked on building a obfuscation technique aimed at reducing the efficacy of linkage attacks against ego-centric social networks.

OPEN SOURCE EXPERIENCE

Member, Product and Technology Advisory Council *October 2024 – October 2025*
Wikimedia Foundation

- Part of a council of 8 volunteers that advised on the technological direction of the Wikimedia movement.
- Drafted recommendations to assuage community concerns surrounding deployment of artificial intelligence features
- Led working group to help prioritize work on critical community-facing unmaintained tools that were widely used by the community.

Lead Maintainer, ProofreadPage *October 2021 – Present*
Wikimedia Foundation

- Maintained the infrastructure for the ProofreadPage extension, which adds proofreading capabilities to MediaWiki and is deployed in over 70 production wikis managed by Wikimedia. (80+ major patches)
- Improved developer documentation related to the extension, providing guides and detailed walkthroughs regarding its setup and use to help ease the onboarding process for new developers.
- Introduced Selenium integration tests across the ProofreadPage codebase to better validate critical frontend changes.
- Took initiative in overhauling the preload and caching mechanism provided by ProofreadPage to decrease load times for editor-facing components.

Student Software Developer, Wikimedia *March 2020 – Sep 2020*
Google Summer of Code

- Developed a software feature that made it easier for volunteers to work with "pagelist"s, a custom syntax used to store page number information for multi-page files.
- Developed heuristics that allowed users to create and edit "pagelist" without interacting with custom XML tag-based syntax.
- Assisted in developing automated tests to detect accidental/malicious changes in minified JavaScript blobs imported as part of dependencies.

SERVICE

Artifact reviewer, IEEE Symposium for Security and Privacy 2026 Artifact Evaluation Committee *(2025 – Present)*
Member, Wikimedia Unsupported Tools Working Group *(2025 – Present)*
Member, Wikimedia Product and Technology Advisory Council *(2024 – Present)*

PROJECTS

- **VisibleV8 crawler** - a open-source tool that enables internet-scale (around 1M websites) web crawling with VisibleV8 which is used extensively internally at the Wolfpack Privacy and Security Lab.
- **WebViewTracer** - One of the first browser instrumentation frameworks for Android WebViews (an artifact of the Cross-Boundary Mobile Tracking paper which is set to appear in NDSS 2026)
- **Fuzzing sudo** - Performed a fuzzing campaign on sudo, and found latent use-after-frees, out-of-bound reads and integer overflows. (sudo-project/sudo `Issue#198`, `PR#196`, `PR#218`, `PR#227`)
- **Software sandbox using seccomp-bpf** - Developed a toy process sandbox using the kernel seccomp API that enabled users to selectively allow and deny specific sequences of syscall usages that were considered malicious according to a set of heuristics rules.

SECURITY ISSUES

- Found and fixed an authentication bypass, a cross-site scripting and a information disclosure vulnerability in English Wikipedia. (`CVE-2024-47848`, `CVE-2024-23174`, `CVE-2023-45369`)
- Awarded 5000 USD for finding a mechanism to reliably leak a user's browsing history via an experimental "origin-trial" web feature in Google Chrome 116.
- Awarded 7500 USD for finding an XSS sanitization deficiency in the html/template library in Golang. (`CVE-2023-24538`)
- Awarded 3133.7 USD for discovering authentication bypasses in the dart:core URI parsing module in Dart-lang by the Google Vulnerability Rewards Program in 2022. (`CVE-2022-3095`, `GHSA-m9pm-2598-57rj`)
- Awarded 3133.7 USD for discovering a URL validation bypass in Google's Clojure library by the Google Vulnerability Rewards Program in 2021.
- Found and reported cross-site leakages (XS-leak) issues in Google Chrome and Firefox's implementation of the ResourceTiming API. (`CVE-2022-1146`, `CVE-2022-29915`)
- Found and reported a high severity Denial-of-service attack against the popular jpeg-js JavaScript library to snyk.io. (`CVE-2022-25851`, `SNYK-JS-JPEGJS-2859218`)

TECHNICAL STRENGTHS

Programming Languages	Javascript, C, C++, Python, Golang
Domains	Browser Internals, Web Security, Dynamic Analysis, Fuzzing
Other Software/Frameworks	Chromium, Puppeteer, Ghidra, IDA, AFL++, Tensorflow, PostgreSQL, MongoDB