

SOHOM DATTA

sohomdatta1+web(at)gmail.com

W MediaWiki Gerrit •  GitHub •  LinkedIn

EDUCATION

- **North Carolina State University, Raleigh** *August 2024 – Ongoing*
PhD in Computer Science
- **Manipal Institute of Technology, Manipal** *July 2019 – August 2023*
B.Tech in Computer Science and Engineering

PUBLICATIONS

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, WebViews.
- Maintained and extended VisibleV8, an open-source browser instrumentation framework across 25+ major Chrome releases.
- Built the VisibleV8 Crawler, an open-source tool that enabled internet-scale web crawling with VisibleV8.
- Ported VisibleV8 to Android WebViews, enabling us to analyse privacy violations in the Android advertising ecosystem.
- 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 commonly 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

- Worked on improving VisibleV8, a research tool aimed at making it easier to perform large-scale measurements of web security and abuse patterns on the internet.
- Implemented large-scale architectural changes to the VisibleV8 crawling pipeline that improved performance in terms of crawling and post-processing the logs generated on crawling 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.
- Conducted research into measuring privacy data leakages across cross-party contexts using VisibleV8.

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 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 issues 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)

LEADERSHIP POSITIONS

Subsystem Head, 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.
- Mentored new recruits and provided resources and guidance on getting started on the basics of cybersecurity.

Web-development Head, Association of Computing Machinery Manipal *August 2021 – August 2022*
Manipal Institute of Technology

- Led and mentored a team of 10 students to develop and maintain infrastructure for Scavenger Hunt, an inter-college event with 700+ participants.
- Conducted workshops and one-on-one sessions to bring the freshers up to speed with modern secure web-development standards demonstrating web security attacks such as cross-site scripting, CSRF, DOM Clobbering etc.

OPEN SOURCE EXPERIENCE

Member, Product and Technology Advisory Council
Wikimedia Foundation

October 2024 – Present

- Part of a council of 8 volunteers that advised the CPTO’s office on the technological direction of the Wikimedia movement.
- Championed and elevated users’ concerns about Wikimedia’s handling and communication surrounding artificial intelligence and helped draft recommendations to assuage community concerns on Wikimedia’s AI strategy.
- Advised the Wikimedia Foundation to invest in a mobile-first strategy to bring in newer contributors on the mobile platform (overwhelmingly used by people from the Global South)

Lead Maintainer, ProofreadPage
Wikimedia Foundation

October 2021 – Present

- 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.
- Built EditInSequence, a community-requested feature that allows users to edit multiple pages via a fast and easy to use interface.

Student Software Developer, Wikimedia
Google Summer of Code

March 2020 – Sep 2020

- 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.

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)

PROJECTS

- **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.

TECHNICAL STRENGTHS

Programming Languages	Javascript, C, C++, Python, HTML, CSS, Golang
Other Software/Frameworks	Chromium, jQuery, NodeJS, Express, React, Make, Tensorflow, IDA, Ghidra