Security Policy Audits: Why and How

Arvind Narayanan and Kevin Lee | Princeton University

We describe a series of security policy audits that we conducted, exposing policy flaws affecting billions of users that are often exploited by even low-tech attackers. We argue that a systematic study of security policies and processes is sorely needed, and present a research agenda.

Security Policies Matter, and Researchers Should Study Them

Information security isn’t just about software flaws: It’s also about policies and processes. Whether an organization uses multifactor authentication, has an incident response plan, and trains employees to avoid phishing has as big an impact on security as whether the software it runs has exploitable bugs. Even the impact of bugs comes down to a policy question, namely the frequency of patching.

Companies use audits and penetration tests to check for vulnerabilities in their policies and processes. We argue that this isn’t enough: Researchers must also study policies from the outside. One reason is that due to misaligned incentives, companies often don’t fix policy flaws that affect their users—sometimes billions of them, as we’ll show. External audits aim to reveal these flaws, warn users of security risks, and pressure companies to change. Besides, research-oriented audits are often exploratory and uncover new policy flaws, whereas industry audits tend to be anchored in established criteria. That’s not to say that research and practice should operate on different wavelengths. On the contrary, we’ll describe how companies’ own efforts and those of external researchers can complement each other.

Security policy audits fall under the umbrella of human factors in security, and there is a strong connection to usable security. Unlike most usable security research, the objects of study aren’t end users, but companies, their policies, and processes. There have been a few studies of security practices of individuals within organizations (e.g., Muhammad Ali et al.) whereas we propose studying organizational policies. Published studies of security policies are currently rare.

Our Work: Exposing Authentication Flaws

We began this line of work in 2018 when we heard frequent reports of Subscriber Identity Module (SIM) swaps in the media. We aimed to test whether there were systemic reasons why they were extremely common. We did so by attempting SIM swap attacks on ourselves: 10 attempts at each of five prepaid wireless carriers in the United States. Thirty-nine of our 50 attempts succeeded (including all 30 at the three major carriers AT&T, Verizon, and T-Mobile) because carriers used insecure ways to authenticate the caller’s identity. For example, one carrier accepted details of recent incoming calls as evidence of identity, which can be trivially subverted by an attacker by calling the victim before attempting the SIM swap. In addition to bad policies, we stumbled upon bad processes, namely poor training of customer service representatives. Some reps simply forgot to authenticate us, others gave us hints or victims’ personal information, and yet others proceeded despite authentication failures!

During the course of this work, we happened to notice that attackers could achieve essentially the same effect by legitimately taking over someone’s number—if that person had relinquished their number. Giving up one’s phone number is extremely common: 35 million numbers are disconnected every year in the United States. Carriers recycle these numbers because the pool of 10-digit numbers is finite. Do people remember to unlink numbers from online accounts before giving them up? Or could phone number takeover lead to online account takeover?

We sampled 259 phone numbers available to new subscribers at two major carriers, Verizon and T-Mobile, and found that 171 of them were tied to existing accounts at popular websites. Worse, about 40% (100 of 259) of the numbers were linked to leaked login credentials on the web, which could enable account hijackings that defeat SMS-based multifactor authentication. We also found...
that carriers failed to include any defensive measures in their recycling policies or online interfaces, such as rate limits for acquiring and relinquishing phone numbers, making the attacker’s task trivial.

We realized that the reason that SIM swap and number recycling attacks are so problematic is that SMS as a second factor for authentication is so prevalent, so we turned to that. We reverse-engineered the authentication policies of more than 140 websites that offer phone-based authentication. Notably, we found 17 websites on which user accounts can be compromised based on a SIM swap alone, i.e., without a password compromise. Further, 83 websites defaulted to or recommended an SMS-based second factor over more secure options such as authenticator apps.\textsuperscript{13}

Finally, to get a more complete picture of online account security, we studied password policies. We examined the policies of 120 of the most popular websites for when a user sets a new password on their account. Despite well-established advice that has emerged from the research community, we found that only 13% of websites followed all relevant best practices in their password policies. For example, 75% of websites do not stop users from choosing the most common passwords—like abc123456 or P@$$w0rd.\textsuperscript{12}

**Policy Problems, Policy Solutions**

These weaknesses were caused by flawed security policies and flawed enforcement. None were the result of software flaws. Attackers could remain within the functionality of the user interface (UI) and use the system with the same privileges as any other user, albeit with malicious intent. These low-tech attackers are UI-bound adversaries. Since the adversary doesn’t need to use any tools or exploit a software vulnerability, anyone could potentially be an attacker.

These unsophisticated attack vectors are considered less interesting by the security research community, yet arguably cause the most harm. SMS-based authentication is still widely used; according to a late-2021 survey, 85% of people in the United States and United Kingdom have used SMS 2FA at least once in 2021, compared to 44.4% for authenticator apps and 9.7% for security keys.\textsuperscript{4} Similarly, passwords remain the most common means of online authentication.

Call center authentication, such as for phone service, credit cards, health insurance, and utilities, is a low-tech authentication method involving no customer interaction with software. Call center fraud is reportedly the leading cause of account takeovers of financial accounts.\textsuperscript{16} Despite the insecurity, there has never been any previous academic research on call center authentication practices.

Policy problems require policy solutions. We mean policy in a broad sense: companies voluntarily changing their security policies, or industry self-regulation, or government regulation. We’d hoped that companies would fix their flawed policies when we presented our research to them. That happened in a few cases. One of the mobile carriers we studied informed us that it had partially implemented our recommendations on mitigating SIM swaps. Four of the 17 websites that were especially vulnerable to SIM swap attacks informed us that they had fixed the flaws in their authentication policy after our outreach. In light of our research on recycled numbers, the two carriers in our study have made changes to better inform their subscribers about reassigning phone numbers. Our work also received a modicum of media attention. Journalists have an important role in security: Their work can help companies to improve their security.

Yet, more often than not, we found that the market didn’t correct itself, or took too long. In fact, policy audits are so unusual that most companies have no idea how to deal with them. In our disclosure to the 17 websites in the SIM swap study, five did not understand our vulnerability report despite our attempts to make it as clear as possible; three websites acknowledged SIM swap attacks but failed to realize that their authentication policy was allowing accounts to be vulnerable, and the other two websites misinterpreted our disclosure as feature requests. Moreover, four websites relied solely on third-party bug bounty platforms to be notified about vulnerabilities. Since members of the platform were only focused on reviewing software bugs, three of our four reports were dismissed. All in all, only four websites made changes in response to our disclosure. As for the wireless carriers, while we didn’t have trouble reporting our findings, only one of our five reports to the major carriers resulted in (partial) fixes.

If responsible disclosure and pressure from users do not lead to improvements to security practices, policymakers should require companies to make the improvements or to bear the consequences of noncompliance. We have had some success in influencing policy improvements that address the flaws we found. Most notably, in September 2021, the Federal Communications Commission launched a formal rulemaking process to protect consumers from SIM swap and number portability attacks, citing our research on SIM swaps as justification.\textsuperscript{7}

**Methods and Challenges**

We encountered a few major challenges in this research. First, there is no avoiding the need for manual work. In some cases, such as testing SIM swap vulnerabilities, automation was impossible. In other cases, such
as reverse engineering password policies, automation yielded poor results despite our best efforts. Crowd work wasn’t effective either because understanding and executing the tasks required some security expertise.

While repetitive, the work was not mindless. Tasks such as adapting to unexpected scenarios during customer service calls or reverse engineering a password policy require some creativity and ability to think on one’s feet. Besides, we found that doing it manually left us with a level of insight into the data that would otherwise have been hard to obtain and led to new research insights.

The second challenge is ensuring reproducibility and consistency in manual data collection. In the SIM swap study, researchers who called the companies followed a script which was developed with various contingencies in mind. In the password policies study, a second researcher studied a sample of websites to ensure that the results were consistent. Manual data collection also means that if reviewers identify issues with the methods and request modifications, the study would essentially have to be redone, a daunting prospect. Although this is unfamiliar territory for most computer scientists, it is the norm in experimental fields; time and care invested in methodological rigor up front can help avoid wasted effort.

Another challenge is that of generating quantitatively reliable measurements. Unlike software vulnerabilities, it is not enough to show that policy flaws exist. Fixing these flaws is rarely easy, and there are always tradeoffs. So, for policy makers to act, it is important to quantify the scale of the problem in some way: e.g., the likelihood of the adversary succeeding or the number of users affected. Otherwise, the affected companies could claim that they are one-off occurrences. Careful evidence gathering and statistical analysis are necessary.

A final set of challenges is around ethics. We took many steps to minimize risk. Usually this was straightforward, but in some cases, we had to come up with alternative research methods, and even abandoned some research questions because we couldn’t find a way to answer it using ethical methods. In the United States, human-subjects research is overseen by Institutional Review Boards (IRBs). IRBs have a narrow definition of human-subjects research, and their understanding of this type of research is still evolving, which can lead to frustration. In our experience, they sometimes overemphasized minor risks while ignoring others. For example, they ruled that if we obtained people’s recycled phone numbers and looked through their texts to see if they are receiving sensitive messages, that’s not human-subjects research! (We decided that this crossed an ethical line, so we did not do this.)

**A Research Agenda**

In this section, we’ll describe five major interrelated directions for future work that we think are promising. The first topic is authentication. Our work only scratches the surface of what is possible. Password recovery procedures are often the Achilles’ heel of online authentication, and yet have never been audited at scale, to our knowledge. Besides, we did not study authentication policies at some of the most important websites, such as banks because those require accounts with those institutions. Studying them would probably require crowdsourcing. Similarly, we only studied call center authentication for wireless carrier accounts. There is a vast set of other organizations that rely on call center authentication for sensitive business, including banks, credit card companies, government services, and hospitals. Anecdotally, offline authentication—such as when requesting a replacement key at a hotel—seems even more badly broken than online authentication. Studying these weaknesses would help convince companies and, when necessary, policy makers to act.

Next, we turn to the integrity of physical systems. Voting infrastructure, the power grid, transportation systems, and many other physical systems cannot rely primarily on software to ensure integrity in the face of attacks, so policies and processes are paramount. Among these,

---

**Four of the 17 websites that were especially vulnerable to SIM swap attacks informed us that they had fixed the flaws in their authentication policy after our outreach.**

---

www.computer.org/security

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
untapped questions. For instance, survivors of intimate partner violence can regain some privacy and safety by leaving family phone plans, but do carriers’ policies make it possible for them to do so without the abuser’s permission? Another research question: On gig economy platforms such as Uber, Lyft, and Airbnb, how effective are platforms’ efforts to ensure physical safety? There have been many reports of inadequate (or, sometimes, overly harsh) preventive measures taken by the companies, but not few systematic studies. Finally, what happens to people’s online accounts when they die? If service providers don’t have measures to hand over control to the next of kin, it can lead to a loss of treasured memories, or finances, on top of the tragedy of death.

The fourth and related research direction is about privacy. Privacy auditing is an extremely active area of research, but most of it consists of technical audits, such as uncovering what data apps are sending to third parties. On the other hand, privacy policies are a rich source of information about apps’ privacy practices, but they are not directly useful to end users. Privacy policy audits neatly bridge these two types of studies: They enable uncovering a company’s actual, operative privacy policy, and see whether it matches its expressed policy. To some extent, aspects of privacy practices can be uncovered by technical analysis of apps, but in other cases it requires unusual methods such as posing as a potential commercial partner and requesting to buy data from the company of interest. Such methods have been used by journalists and privacy advocates but not yet by researchers.

Privacy policy audits have uncovered some distressing findings. For example, Facebook was caught soliciting phone numbers from users for two-factor authentication but then misusing them for targeted advertising, resulting in a US$5 billion settlement with the Federal Trade Commission.

The final research direction is about platforms—search engines, social media, e-commerce, gig economy marketplaces, and so on. Platforms have outsized power in our society and are inadequately regulated, especially in the United States. There is a growing line of work on auditing platforms for algorithmic bias, pricing policies, and so on. We think their security and safety related policies should also be audited.

One set of platform policies that deserve scrutiny is around content moderation: removal or flagging of content that violates policies. A recent report by the U.S. Federal Trade Commission lists more than a dozen categories: scammy content, deepfakes, fake reviews, sale of illegal substances, child sexual abuse, revenge pornography, cyberstalking, hate crimes, glorification of violence, incitement to violence, terrorists’ use of platforms, disinformation campaigns, and sale of counterfeit products. When these policies have loopholes or are unevenly enforced, it harms legitimate users while allowing harmful content through.

Another set of policies relates to advertising: Ads can be discriminatory, promote misleading or deceptive content, and enforcement can be poor. One important hurdle is that companies, notably Facebook, have been using legal and technical restrictions to actively inhibit researchers’ efforts to study them.

Final Thoughts
We think that security policy auditing could be an entire subfield of research, rather than the occasional one-off studies we’ve seen so far. The kind of technology-enabled harm that happens in reality mostly involves attacks that exploit current or older technologies, with attackers using the products as intended, albeit with malicious intent. But keep in mind that it’s currently hard to publish security policy audits at top security venues. Despite acknowledging its real-world impact, reviewers at top conferences disproportionately merit work for publication based on factors like emerging technologies and sophisticated attack vectors. We hope that this article serves as a small first step toward changing this bias. We urge you to consider societal impact as a criterion in your own reviewing.

We end with suggestions for practitioners. Just as companies have embraced software vulnerability research—rather than ignoring or threatening researchers—they should do the same for security policy flaws. Our experiences reporting these flaws suggest that there is a long way to go. Companies should also consider publishing redacted audits and, more generally, sharing their internal expertise regarding security policies and processes. This would allow companies to learn from each other and invite productive dialogue with academia. Collaboration with researchers on security policy audits is key to improving user safety.

Acknowledgment
We conducted the research on which this paper is based in collaboration with Ben Kaiser, Jonathan Mayer, and Sten Sjöberg. We are grateful to Joseph A. Calandrino, Mihir Kshirsagar, Heather Lipford, Andrés Monroy-Hernández, and Jessica Staddon for useful feedback. The genesis of this paper was a talk given by Narayanan at the Cambridge security seminar.

References
1. M. Ali et al., "Discrimination through optimization: How Facebook’s ad delivery can lead to biased outcomes," Proc. ACM Human-Comput. Interact., vol. 3, Nov. 2019, Art. no. 199, doi: 10.1145/3359301.
2. A. W. Appel, "Security seals on voting machines: A case study," ACM Trans. Inf. Syst. Secur. vol. 14, no. 2, p. 18, Sep. 2011, doi: 10.1145/2015999.2019603.

3. R. Chatterjee et al., "The spyware used in intimate partner violence," in Proc. IEEE Symp. Secur. Privacy (SP), 2018, pp. 441–458, doi: 10.1109/SP.2018.00061.

4. D. Childers. "State of the auth 2021: Experiences and perceptions of multi-factor authentication." Duo Labs. Accessed: Jan. 9, 2023. [Online]. Available: https://duo.com/assets/ebooks/state-of-the-auth-2021.pdf

5. J. Cox. "I gave a bounty hunter $300. Then he located our phone." Motherboard Tech by Vice. Accessed: Jul. 20, 2022. [Online]. Available: https://www.vice.com/en/article/nepbxz-i-gave-a-bounty-hunter-300-dollars-located-phone-microbilt-zumigo-tnmobil

6. L. Edelson et al., "An analysis of United States online political advertising transparency," 2019, arXiv:1902.04385.

7. "In the matter of protecting consumers from SIM swap and port-out fraud. Notice of proposed rulemaking," Federal Communications Commission, Washington, DC, USA, 2021. Accessed: May 4, 2022. [Online]. Available: https://docs.fcc.gov/public/attachments/FCC-21-102A1.pdf

8. "Combating online harms through innovation," Federal Trade Commission, Washington, DC, USA, Tech. Rep., Jun. 16, 2022. Accessed: Jul. 20, 2022. [Online]. Available: https://www.ftc.gov/reports/combating-online-harms-through-innovation

9. D. Kerr, "More than 50 U.S. gig workers murdered on the job in five years," Markup, Apr. 6, 2022. Accessed: Jul. 20, 2022. [Online]. Available: https://themarkup.org/working-for-an-algorithm/2022/04/06/more-than-50-u-s-gig-workers-murdered-on-the-job-in-five-years

10. V. L. Pochat et al., "An audit of Facebook’s political Ad policy enforcement," in Proc. 31st USENIX Secur. Symp. (USENIX Security), Boston, MA, USA, Aug. 2022. [Online]. Available: https://www.usenix.org/conference/usenixsecurity22/presentation/lepochat

11. K. Lee and A. Narayanan, "Security and privacy risks of number recycling at mobile carriers in the United States," in Proc. APWG Symp. Electron. Crime Res. (eCrime), 2021, pp. 1–17, doi: 10.1109/eCrime54498.2021.9738792.

12. K. Lee, S. Sjöberg, and A. Narayanan, "Password policies of most top websites fail to follow best practices," in Proc. 18th Symp. Usable Privacy Secur. (SOUPS), Aug. 2022. [Online]. Available: https://www.usenix.org/conference/soups2022/presentation/lee

13. K. Lee et al., "An empirical study of wireless carrier authentication for SIM swaps," in Proc. 16th Symp. Usable Privacy Secur. (SOUPS), Aug. 2020. [Online]. Available: https://www.usenix.org/conference/soups2020/presentation/lee

14. F. Li et al., "Keepers of the machines: Examining how system administrators manage software updates for multiple machines," in Proc. 15th Symp. Usable Privacy Secur. (SOUPS), 2019. [Online]. Available: https://www.usenix.org/conference/soups2019/presentation/li

15. J. N. Matias, A. Hounsell, and N. Feamster, "Software-supported audits of decision-making systems: Testing Google and Facebook’s political advertising policies," Proc. ACM Human-Comput. Interact., vol. 6, no. CSCW, pp. 1–19, Apr. 2022, doi: 10.1145/3512965.

16. "2021 state of call center authentication," Neustar, Reston, VA, USA, 2021. Accessed: May 13, 2022. [Online]. Available: https://www.cdn.neustar/resources/whitepapers/risk/neustar-state-of-call-center-authentication-2021.pdf

17. "NYU ad observatory," New York Univ., New York, NY, USA, 2022. Accessed: Jul. 20, 2022. [Online]. Available: https://adobservatory.org/

18. R. Slavin et al., "Toward a framework for detecting privacy policy violations in android application code," in Proc. 38th Int. Conf. Softw. Eng., New York, NY, USA, 2016, pp. 25–36, doi: 10.1145/2884781.2884855.

19. D. Springall et al., "Security analysis of the Estonian internet voting system," in Proc. ACM SIGSAC Conf. Comput. Commun. Secur., New York, NY, USA, 2014, pp. 703–715, doi: 10.1145/2660267.2660315.

20. G. Venkatadri et al., "Investigating sources of PII used in Facebook’s targeted advertising," in Proc. Privacy Enhancing Technol., 2019, vol. 2019, no. 1, pp. 227–244, doi: 10.2478/popes-2019-0013.

21. S. Zimmek et al., "MAPS: Scaling privacy compliance analysis to a million apps," in Proc. Privacy Enhancing Technol., Jan. 2019, vol. 2019, no. 3, pp. 66–86, doi: 10.2478/popes-2019-0037.

Arvind Narayanan is a professor of computer science at Princeton University, Princeton, NJ 08540 USA. His research focuses on the societal harms of digital technologies. Narayanan received a Ph.D. in computer science from the University of Texas at Austin. He is a recipient of the Presidential Early Career Award for Scientists and Engineers (PECASE), twice a recipient of the Privacy Enhancing Technologies Award, and thrice a recipient of the Privacy Papers for Policy Makers Award. Contact him at arvindn@cs.princeton.edu.

Kevin Lee is a research scientist at JPMorgan Chase, New York, NY 10179 USA, where he works on bettering user safety practices. His research interests include usable security and technology policy. Lee received a Ph.D. in computer science from Princeton University. Contact him at kvnl@cs.princeton.edu.