Where Are the Red Lines?
Towards Ethical Server-Side Scans in Security and Privacy Research

Florian Hantke*, Sebastian Roth†, Rafael Mrowczynski*, Christine Utz*, Ben Stock*
*CISPA Helmholtz Center for Information Security †TU Wien
{florian.hantke, mrowczynski, christine.utz, stock}@cispa.de sebastian.roth@tuwien.ac.at

***Abstract*—Comprehensive and representative measurements are crucial to understand security and privacy risks on the Web. However, researchers have long been reluctant to investigate server-side vulnerabilities at scale, as this could harm servers, disrupt service, and cause financial damage. This can lead to operator backlash and problems in peer review, as the boundaries posed by the law, ethics, and operators' stance towards security research are largely unclear.**

In this paper, we address this research gap and investigate the boundaries of server-side scanning (3S) on the Web. To that end, we devise five typical scenarios for 3S on the Web to obtain concrete practical guidance. We analyze qualitative data from 23 interviews with legal experts, using German law as a case study, members of Research Ethics Committees, and website and server operators to learn what types of 3S are considered acceptable and which behavior would cross a red line. To verify our findings, we further conduct an online survey with 119 operators.

Our analysis of these different perspectives shows that the absence of judicial decisions and clear ethical guidelines poses challenges in overcoming the risks associated with 3S, despite a slight majority (57%) of operators having a positive stance towards such academic research throughout the interviews and the survey. As a first step to mitigate these challenges, we suggest best practices for future 3S research and a pre-registration process to provide a reliable and transparent environment for 3S-based research that reduces uncertainty for researchers and operators alike.

1. Introduction

As digital systems keep evolving rapidly, it is crucial for security and privacy research to continuously study the prevalence and trends of security vulnerabilities, as this can help prioritize future research and guide the development of robust solutions. On the Web, security research has mainly focused on client-side risks, which can be explored in a researcher-controlled environment. By contrast, server-side risks still remain largely unexplored due to the very real risk of harming remote systems, which could lead to operator backlash, legal repercussions, and problems in publishing the research results. This leaves a crucial research gap, as data breaches frequently originate on the server side and are often attributed to unpatched known vulnerabilities [58].

A realistic and holistic picture of server-side security risks and vulnerabilities requires large-scale scans of the Internet. The boundaries set to this by the law are often unclear and subject to interpretation [8]. This has led researchers to discontinue their research [28] or only analyze self-hosted open-source projects [22, 25, 49, 56], which cannot provide a comprehensive view of the Web. Reporting vulnerabilities in (open-source) frameworks also does not necessarily fix issues at scale, as operators are often slow to apply patches [23] and should rather be addressed directly.

Still, recent real-world studies evaluated, e. g., SQL injections on orphaned web pages [57], HTTP desync issues [40], or ReDOS vulnerabilities [63]. According to anecdotal evidence, some of them had caused discussions in the research community, as unpredictable server reactions make it difficult to estimate potential harm to operators or users. In general, our community is still in the discussion of what is ethically acceptable [21, 45, 55, 73].

This work aims to advance this mostly underexplored topic and investigates the feasibility of server-side research from both ethical and legal standpoints. Our goal is to get a holistic understanding of the problem space and devise best practices for future research based on server-side scanning (3S). To gain broad perspectives on the risks and feasibility of 3S, we conducted 23 semi-structured interviews with German legal experts, members of Research Ethics Committees, and website and server operators to identify the boundaries within which such research might be permissible. To evaluate our findings at scale, we conducted an online survey with another 119 operators. Drawing on these insights, we suggest best practices and propose a pre-registration procedure to make 3S-based research more reliable and transparent.

In summary, this paper investigates the question, *How can we enable server-side scanning research within a framework that prevents harm for both researchers and server operators?* We investigate this via three sub-questions:

1) RQ1: *How can server-side scanning-based research meet legal standards?* Researchers should not fear and risk legal repercussions. Yet, understanding the law, with its intricate nuances, can be challenging for non-experts. We therefore aim to understand the boundaries within which 3S research can be conducted legally.
2) RQ2: *How can server-side scanning-based research meet ethical standards?* Research should follow ethical

standards, which can be hard in the complex Web ecosystem. Hence, this work investigates concrete ethical challenges that researchers face.

3) RQ3: *What problems do website and server operators see and how can they be minimized?* As operators are directly impacted by web research, their views are indispensable. We investigate what they consider harmful and permissible, aiming to address their concerns.

Our contribution comprises the following key points:

- We present the first qualitative and quantitative study with 23 interviews and 119 survey responses that investigates diverse perspectives about 3S, focusing on ethics committees of major security conferences and German jurists.
- We explore five concrete scenarios of how 3S is typically applied in Web security research, along with expert and operator assessments. These can guide future work and help the community identify the general boundaries of 3S research.
- We propose best practices on how to conduct 3S in security and privacy research and suggest a preregistration board for future work in the area.

2. Background & Related Work

In this section, we introduce server-side scanning and typical vulnerabilities that researchers would investigate. We describe potential legal pitfalls using the example of the German legal system and provide an introduction to research ethics. Along the way, we outline previous work related to our research.

2.1. Server-Side Scanning

Web technology can be divided into two components: the *server side*, where most of a website's logic runs, and the *client side*, which comprises the code executed on the user agent, i. e., the browser. For decades, web security researchers have been analyzing client-side vulnerabilities.

Vulnerabilities are flaws in the design and implementation of web applications and their underlying infrastructure. These can be exploited by attackers to compromise the confidentiality, integrity, or availability (CIA Triad) of data and resources. For example, one widely studied vulnerability on the client side is Cross-side Scripting (XSS) [16, 51, 64, 66]. This vulnerability allows attackers to inject their code into a website, performing tasks in the context of a victim who clicked a malicious link. However, such an XSS payload can also be stored on the server (stored XSS), potentially impacting every user of that particular website. Other common examples of server-side vulnerabilities include SQL injections (SQLi) [53], Insecure Direct Object References (IDOR) [54], or path traversal [52].

To understand the prevalence, workings, and evolution of vulnerabilities, researchers routinely conduct measurement studies by requesting data from thousands of websites and analyzing it on the client side. This research on the client side can be conducted in a safe manner running on the researchers-controlled user agent. For the server side, we define a *server-side scan* (3S) as an automated series of requests to systematically check a large number of web servers / websites for specific server-side behavior, such as potential vulnerabilities like SQL injections.

As mentioned, these 3S projects pose ethical questions. Predicting the potential for harm is challenging due to the unpredictable behavior of servers. Potential consequences are manifold and may include server crashes or information leaks, all potentially causing additional work and anxiety for the operators of websites and servers.

Additionally, legal uncertainties are a significant problem. Previous work indicates that security researchers often avoid certain types of research, e. g., 3S, due to the fear of legal repercussions [28]. In fact, legal action has been taken against researchers in the past [6, 34, 47].

2.2. Legal Risks

The legal permissibility of 3S research is highly complex, as such scans are typically performed at large scale over the Internet, potentially affecting many jurisdictions. Since we interviewed legal experts from Germany, we outline selected provisions of German law as an example of the legal problems researchers may face when conducting 3S. Germany has a civil law system rooted in Roman tradition, thus relying on legal codes as its main source of law.

Criminal Offenses. German criminal law is mainly laid down in the German Criminal Code (Strafgesetzbuch, StGB [30]); Bohlander [12] provides an introduction. Offenses potentially committed via 3S activity include data espionage (Sec. 202a StGB), which is the unauthorized circumvention of access protection to data not intended for the perpetrator and *"specially protected against unauthorized access."* Sec. 202b criminalizes "phishing," defined as *"[the unauthorized interception of data] not intended for [the perpetrator] by technical means from non-public data transmission"* [...]. Sec. 202c penalizes preparatory actions to data espionage and phishing. Due to its wide applicability that also covers the creation of software that could be used for hacking with malicious intent, it has been dubbed the "hacker provision" and widely criticized [19, 32]. The potential for harm caused by 3S research opens the door to offenses related to property damage. Sec. 303a (data manipulation) punishes *"[w]hoever unlawfully deletes, suppresses, renders unusable or alters data [Sec. 202a (2)]"* and 303b (computer sabotage) interference *"with data processing operations [...] of substantial importance"* conducted by means including acts under Sec. 303a (1) or 202a (2) or *"destroying, damaging, rendering unusable, removing or altering a data processing system or a data carrier"*.

Damages. Further, researchers could be held liable under civil law for harm caused through 3S activity. Damages can be issued based on German tort law, as codified in Sec. 823 of the German Civil Code (Bürgerliches Gesetzbuch, BGB [29]). This requires an intentional or negligent unlawful violation of another person's right, including property, or alternatively, a breach of a statute intended to protect another person, such as criminal law provisions.

Other. Aside from these fundamental legal risks, server-side research could be at odds with other more specialized areas of law, including data protection law. In Germany, this comprises European laws including the General Data Protection Regulation (GDPR) [26], but also data protection laws at the federal and state level, which, in turn, may constitute additional offenses and legal bases for liability.

2.3. Ethics

Ethical considerations have always been an integral yet challenging aspect of conducting research, particularly when human subjects are involved. To support researchers with these challenges and to discuss morally acceptable boundaries, various guidelines exist, such as the Belmont Report [69] or the computer security-focused Menlo Report [7]. Despite these frameworks, the nuances of ethical considerations are complex and invite diverse views.

In an attempt to support researchers struggling with these complexities, many universities have established Institutional Review Boards (IRB). IRBs serve as independent committees that evaluate research methods involving human subjects. They review research proposals, provide constructive feedback, and suggest improvements in the method to ensure ethical standards. The significance of IRBs is growing, as many academic conferences now ask for IRB approval for research involving human subjects.

Still, security research often encounters unique ethical dilemmas that extend beyond direct human interactions [45, 55]. Due to the nature of the digital world, these studies may have broader, indirect implications for individuals. One example is the Encore paper [14], which was IRB approved, yet raised ethical concerns in the review process, as the research method potentially led to browsers of individuals living in suppressive regimes sending requests to censored websites [17, 42]. Similarly, the Hypocrite Commits paper [72], in which researchers introduced vulnerabilities into the Linux kernel, triggered extensive ethical debates and was finally withdrawn from the conference's proceedings [37].

In response to such incidents, leading computer security conferences like the USENIX Security Symposium (USENIX Security) [15] and the IEEE Symposium on Security and Privacy (IEEE S&P) [38] have implemented Research Ethics Committees (REC). Equipped with a more profound understanding of the ethical nuances in security research than IRBs, these committees discuss moral aspects of submitted papers and help clarify ethical questions before making a decision on a paper's acceptance. In contrast to IRBs, RECs only evaluate submitted research and thus do not prevent potentially harmful research in the first place.

Still, this step towards more ethical security and privacy research is essential and has prompted the research community to put greater focus on ethical considerations [45, 55]. Yet, to the best of our knowledge, discussions largely remain abstract and often do not consider operators' perspectives.

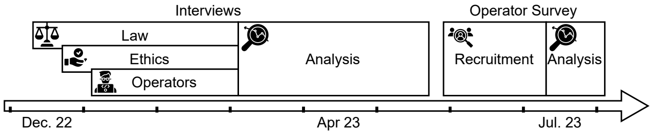

Figure 1. Study overview and timeline.

3. Methods

Our study aims to provide guidance for server-side scanning (3S), focusing on use cases in web security. We investigate the boundaries of 3S from three perspectives, law (RQ1), ethics, represented by REC members (RQ2), and operators (RQ3). Our insights are grounded in interviews and an online survey.

To gain an in-depth understanding of these three perspectives, we first conducted semi-structured problem-centered interviews [71] with people from all three groups (see Figure 1). Our interviews with legal professionals and REC members had the character of expert interviews [11] aiming at the reconstruction of their specialist knowledge related to our RQs. Since legal regulations differ from one country to another, we narrowed the law-related scope of this study by focusing on German law. For the operators, semi-structured interviews provided insights into their mindsets and opinions on server-side research. Additionally, with far more operators existing than ethics and legal experts, we complement our findings with a quantitative survey study with operators to broaden the scope of opinions. As a key element to all interviews and the survey questionnaire, we use example 3S scenarios, presented in form of vignettes [59].

3.1. Vignettes of 3S Scenarios

Vignettes are "short stories about hypothetical characters in specified circumstances, to whose situation the interviewee is invited to respond" [27]. This instrument allowed us to elicit interviewees' considerations and judgments on specific 3S scenarios that are relevant to answer our RQs [3, 5]. For the survey, they let us approximate how well operators' opinions can be generalized on a larger scale.

We designed the vignettes to cover a range of typical 3S scenarios in web security research, which we identified by reviewing widely exploited real-world vulnerabilities known as Common Weakness Enumeration (CWE). Furthermore, the vignettes were meant to challenge participants' assumed mental models and therefore consider edge cases pointed out in ethics [7] and German legal [8] literature, e. g., the limits of access control, the importance of researchers' intent, and privacy law concerns. We also made sure to cover all dimensions of the CIA Triad. Each of the scenarios is identified by the name of its fictional protagonist (see Table 1). The full texts for all scenarios are provided in Appendix B. For the survey, we slightly modified the wordings to increase clarity; these versions are included in the full questionnaire in the supplementary material [36].

TABLE 1. SCENARIOS USED IN THE INTERVIEWS AND SURVEY.

Scenario	Vulnerability	CWE Rank		CIA
Alice	SQL injection	3	CWE-89	CA
Bob	Invalid HTTP header	6	CWE-78	A
Charlie	Insecure Direct Object Reference	22	CWE-269	CI
Daisy	Stored XSS	2	CWE-79	CI
Eve	Path traversal	8	CWE-22	C

3.2. Pre-Registration Proposal

Another integral part of both the problem-centered interviews and our survey questionnaire was our proposal for how to improve the status quo for 3S research. We propose a pre-registration process through a trusted third party (TTP) to provide a concrete framework within which 3S research can be carried out in a reliable and transparent way. The idea initially came up with insights from the first interviews and matured in internal discussion in our project team. In the later interviews and the survey, we briefly outlined the idea and presented our full proposal in Section 6.2.

3.3. Problem-Centered Interviews

Our problem-centered interviews called for different recruitment strategies for each group of interviewees.

For legal experts, we compiled a list of professional roles we wanted to interview, e. g., legal scholars as well as legal practitioners like lawyers and prosecutors. We included the perspective of academics, as (1) they are the most frequent contributors to *legal commentaries*, an important source of knowledge for all German legal professionals; (2) their *scholarly opinions* are regularly quoted in judicial decisions [44]; and (3) they also initiate debates that inspire legislative policies. Once the list was completed, we emailed legal experts who we identified via an online search and also followed further recommendations by our interviewees.

For operators, we followed a similar approach, creating a list of roles along with desired attributes (e. g., company in the security sector, small company). We found suitable operators online and emailed them, recruited via social media (Twitter and LinkedIn), and also used personal contacts.

Lastly, for the ethics perspective, we invited members from the RECs of USENIX Security and IEEE S&P, as they are experts who assess the ethical validity of security research. To avoid conflicting too many potential REC members for our own submission, we invited one expert after another instead of sending out multiple invites at once.

For each of the three groups, we continued recruiting until we had filled all desired roles and attributes. If two people in one category had very diverse opinions, we recruited more people of this category until opinions saturated.

The guides we crafted for our interviews begin with general questions about web security, followed by more detailed questions tailored for each specific group. All five 3S scenarios and our improvement proposal were embedded in the guides; they are available as supplementary material [36].

We conducted one pilot interview per group to test the interview process and to fine-tune our guides. The interviewees were experts from our institutions. As this pre-study went smoothly, we kept this interviewing strategy.

Interviews took between 37–116 minutes and were conducted either in-person or via video conference between January and May 2023 (see Figure 1). With each new interview, minor adjustments were made to the guides, e. g., we incorporated new insights from earlier interviews. Most of the interviews were joined by two members of the research team. One person led the interview, while the other asked additional questions and handled the audio recording. We leveraged the Amberscript service for transcription. One interviewee did not consent to audio recording, so the (only) interviewer took notes during the interview and expanded them based on their memories afterwards.

3.4. Operator Survey

We conducted an online survey to validate our findings with a larger sample of operators. Like the interviews, the questionnaire was built around the five 3S scenarios and our pre-registration proposal for the future of 3S studies. To obtain more detailed insights into the boundaries of what types of 3S research operators are comfortable with, we leveraged ideas from the interviews to create two variants for each scenario: a less invasive one, which was shown to participants with a neutral or negative response to the base scenario, and a more invasive variant for neutral to positive respondents. The full questionnaire and all scenario texts are available in the supplementary material [36].

The questionnaire first asked about participants' background: if (Q1) and how many (Q2) servers they operated, their main role with regard to the operation of websites or servers (Q3), and the size (Q4) and location (Q5) of the largest organization for which they were currently operating them. We then defined 3S and asked for participants' general comfort level with researchers performing such scans on participants' servers, using five-point Likert scales (Q6). Next, we let participants assess the five 3S scenarios in the same way, presenting them in random order and asking for an assessment of the base scenarios (A1–E1); if applicable, one or both variants (A2–E2 / A3–E3); and an explanation of these assessments (A4–E4). The survey proceeded to present our proposal of pre-registration through a trusted third party. Participants were asked to rate the competence of different potential TTPs to assess 3S studies (Q7) and again to indicate their comfort level with such pre-approved scans (Q8). The section concluded with questions about factors that would be important to operators in such a pre-registration procedure (Q9) and if (Q10) and in what way (Q11a–b) it would change their opinion about scenario assessments. The survey ended with questions about previous security training (Q12–13), age (Q14), gender (Q15), and general feedback about 3S and our study (Q16). Participants could voluntarily provide an email address to receive the study results.

We pre-tested the survey in think-aloud sessions and test runs with operators recruited from the authors' social circles until no more comprehension issues arose. We distributed the survey to operators at scale by emailing the generic email address `webmaster@DOMAIN` for websites on the Tranco top 1 million list [46] from June 21, 2023 (ID: X5XVN). We divided this list into 10 popularity bins of 100,000 domains and, once to twice a day, emailed batches of 5,000 domains, with 500 randomly drawn from each bin. This resulted in a total of 200,000 emails sent between June 23 and July 19, 2023, 159,009 bounced emails, and 291 survey accesses.

3.5. Data Analysis

Interviews. We implemented the procedure of qualitative content analysis [48, 60, 61], but in a relatively flexible manner. We derived our initial set of codes from our interview guides. These codes represented topics and sub-topics addressed by our questions.

After the codebook became saturated, a key concept of qualitative research [33, 67], three team members coded all interview transcripts in three consecutive stages. They had varied expertise with two being computer scientists, one also holding a German law degree, and the third being a social scientist specialized in sociology of law. Coding tasks were assigned such that every interview transcript was independently coded by at least two researchers. After each stage, we compared and discussed our codings until we reached an inter-subjective agreement. The final version of the codebook is included in supplementary material [36].

After finishing the coding process, we identified the key categories, i. e., the most important codes and used them to systematically look across different transcripts. In this way, we found commonalities as well as differences in what our interviewees had told us about 3S issues and our scenarios. The process of comparative data interpretation was facilitated by writing memos [33] on the main topics represented by our key categories. These memos served as the foundation for writing up our findings.

Survey. Overall, we registered 291 survey accesses. We removed 86 responses without consent, 85 partial responses that had not been submitted, and one response that had not answered a single question. This left us with a total of 119 final responses, which we analyzed using descriptive statistics. We did not conduct a formal qualitative analysis of the open-ended responses, as an inspection did not yield any new concepts beyond the interviews.

3.6. Research Ethics

Prior to recruiting participants for our interviews and the survey, we took extra care to minimize any potential harm or privacy concerns. We obtained prior approval from both our institution's Ethical Review Board and data protection officer. Both interviewees and survey participants were briefed about the goal of the study before their participation and asked for their consent. They also had the option to withdraw from the study at any time. One interviewee was not comfortable with being recorded, so we respected that wish and resorted to taking notes. The operator survey was distributed to popular websites by contacting only the generic email address `webmaster@DOMAIN`. While this could be considered as unsolicited bulk mail, our use of email generics meant for general public inquiries was found by a data protection authority to be less invasive than more targeted approaches [70]. Each address was only emailed once, without any reminders or confirmations. Email addresses voluntarily provided by survey participants to learn about the study results will only be contacted once to send out a preprint of this paper and deleted afterwards.

3.7. Limitations

The qualitative part of our study has general limitations that characterize this type of research: We can identify different arguments with interviews, but we cannot make generalizing inferences about their distribution within the entire population of legal experts or REC members. For example, we were able to identify arguments rooted in different understandings of research ethics, but we cannot indicate which direction dominates among REC members.

Our inquiry into legal assessments of 3S research is also limited in terms of geographic scope. We exclusively interviewed experts on German law, due to practical reasons. Despite all transnationalization trends of recent decades [10], law remains a fairly national phenomenon and most legal professionals operate within their respective national legal systems [1, 2]. A comprehensive assessment of relevant laws would require a vast multi-national research team and massive resources, which we do not have. Thus, we decided to focus on Germany as an example. We hope that this paper inspires researchers globally to conduct similar studies in their countries, painting a broader picture piece-by-piece.

As for server operators, we combined semi-structured interviews with a survey. This mixed-methods approach aimed to overcome the limitation in the qualitative part. However, our survey also has its own methodological limitations. Despite the random selection of potential respondents, self-selection bias necessarily occurs. In particular, we suppose that security-aware operators were more likely to respond.

4. Results

The insights obtained in 23 semi-structured interviews and 119 survey responses paint a consistent picture of the legal and ethical challenges of server-side scans and the potential problems for operators. In this section, we summarize the participants' views, emphasizing that these should not be misconstrued as legal counsel. First, we describe our participant samples and the three interviewed groups' general stance towards 3S. Afterwards, we go into more detail, focusing on the five 3S scenarios selected to represent common server-side issues in web security. We conclude with participants' suggestions for improvement and opinions on pre-registration.

TABLE 2. INTERVIEWEE DEMOGRAPHICS AND BACKGROUND.

	ID	Role	Gender	Country
Law	1-L	Law professor (criminal law)	M	DE
	2-L	Law professor (privacy law)	M	DE
	3-L	Law professor (criminal law)	M	DE
	4-L	Law professor (privacy law)	M	DE
	5-La	Law professor (criminal law)	M	DE
	5-Lb	Legal research assistant	F	DE
	6-L	Legal practitioner	M	DE
	7-L	Legal practitioner	M	DE
	8-L	Legal practitioner	M	DE
Operators	9-O	CISO of a large company	M	DE
	10-O	Hobbyist and self-hoster	M	DE
	11-O	Operator for web agency	M	DE
	12-O	Pentester and web operator	M	DE
	13-O	Owner of multiple web shops	M	DE
	14-O	CISO in the public sector	M	DE
	15-O	CTO of a small startup	M	DE
	16-O	CEO of a web agency	M	DE
	17-O	CTO of an international company	M	UK
	18-O	Pentester and self-hoster	M	DE
Ethics	20-E	Ethics Committee Member	M	-
	21-E	Ethics Committee Member	M	-
	22-E	Ethics Committee Member	M	-
	23-E	Ethics Committee Member	M	-
	24-E	Ethics Committee Member	M	-

4.1. Participant Samples

Interviews. As shown in Table 2, we interviewed a total of 9 legal experts, 10 operators of various websites, and 5 members of conference ethics committees. Considering the small number of the REC members in general, we opted to omit the country's name to prevent the de-anonymization of any participants. We also spoke to other people who did not fit our three designated groups, e. g., politicians. While not included in our analysis, they still contributed to our overall understanding. Interview 5 was conducted with two participants simultaneously, resulting in 23 interviews but a total of 24 participants.

Despite our efforts to achieve a diverse participant pool by sending out emails to a broad spectrum of people and also advertising on Twitter (>40K impressions) and LinkedIn (∼2K), all but one interviewee were male – a reflection of the unfortunate gender disparity in these fields. Interviewees' ages ranged from mid-twenties to retirement, and they were based in Germany, the UK, the US, and Switzerland.

Participants were diverse in terms of role, professional background, and seniority. Among the legal experts, 6 hailed from academia, ranging from research assistants to senior professors. The practitioners worked as lawyers and prosecutors. The REC members also ranged from junior to senior professors. The interviewed operators held a variety of positions and had a diverse understanding of security. The group included self-hosters, CTOs, CISOs, penetration testers, web developers, and one owner of multiple online shops. Overall, our groups of interviewees covered all the roles we had planned to interview.

Survey. After data cleaning as described in Section 3.5 we were left with 119 valid survey responses. Participants here also predominantly (84.9 %) identified as male and had a mean age of 43.0 years (std 9.8, min 20, med 43, max 72). All but one participant reported to operate a web server or site; within the last three years, they had operated a mean of 343.0 servers (std 1809.5, min 1, med 15, max 16,000), with most being responsible for 2–50 servers. Most (92.4 %) reported to have received prior security training, most frequently via self-teaching (89.1 %) or "learning by doing" (77.3 %). Formal education was less common, led by courses at university or school (48.7 %). Appendix C contains detailed statistics on survey participants' demographics and background in server operation.

4.2. General Assessment of Server-Side Scans

In the first part of each interview, we introduced the interviewees to the topic of 3S and asked what they thought in general about researchers conducting such scans. Depending on their field, this already provided evidence of diverse perspectives and the breadth of affected subject areas. The following sections describe interviewees' general assessments of 3S, while Section 4.3 provides more detailed insights discussing specific 3S scenarios.

4.2.1. Legal Experts.

Our interviews with legal experts indicated that topics like 3S-based studies, searching for server vulnerabilities, and benevolent hacking activities constituted a *"niche topic"* in German juridical debates (3-L). Currently, this kind of research can only be done in a *"legal gray zone"* (7-L) – a fact that implies serious legal risks for those who conduct 3S-based studies (7-L; 8-L).

On the one hand, a systematic search for vulnerabilities on remote computer systems, especially if performed at scale, can constitute an act of crime under present-day legislation (5-La). The only legally sound way to perform 3S-based security research would require explicit consent by every single operator whose server is subjected to a scan (7-L). However, 7-L underlines this is infeasible for 3S studies.

On the other hand, as of August 2023, there are no public court rulings on 3S-related cases in Germany (5-La). As a prosecutor interviewee emphasized, no *"white-hat hacker"* has ever been sentenced by a German criminal court, as all such rare cases had been already closed before trial, either because of a lack of "public interest" in prosecution or because of "minor guilt" – both are discretionary prerogatives often available to prosecutors (6-L). This fact, however, has ambiguous implications for 3S researchers: A conviction appears unlikely but there is no guarantee (5-La), and a *"mean-spirited"* prosecutor could still open an investigation, which by itself is already stressful for suspects (e. g., confiscation of their hardware), even if the case is later closed before trial (2-L).

The biggest legal risks and uncertainties for 3S researchers lurk, however, in the civil law domain. If a scan causes harm on remote systems, affected operators can sue for compensation (6-L). Harm can include different things: a need to pay for IT specialists to fix a server, a direct loss of revenue due to the unavailability of an operator's Internet presence, as well as a loss of customers due to reputation damage (6-L). The distinction between intent and negligence, which is vital in criminal cases, or notions of public interest or benevolent motivations are barely relevant in civil law cases (6-L).

4.2.2. REC Members. Besides the 3S topic, we asked the interviewed REC members about the process of ethical review. They are selected to serve on the REC based on their research experience. Although they usually do not have systematic academic training in practical philosophy, some of them do explicitly refer to main ethical traditions, as indicated by one of our interviewees and in literature [45]. Other REC members we interviewed perform their tasks based on implicit understandings of ethical norms without explicitly linking them to broader philosophical debates.

According to our data, there are two main ethical traditions influencing the REC members. The first is *utilitarianism* [9, 50] or *consequentialism* [45]. It focuses on relations between benefits and harms as consequences of human actions, e. g., a 3S-based study. The second philosophical tradition is *deontology* that emphasizes the *categorical* moral obligation to respect the fundamental personal rights [43], e. g., privacy. "[D]ifferent ethical frameworks can lead to different conclusions about right and wrong" when discussing ethical challenges posed by specific research projects [45].

Other ethical frameworks were also mentioned: *"discourse ethics"* and *"ethical colonialism"* (22-E). Proponents of the discourse ethics [4, 35] argue that an adequate ethical assessment can only emerge as a result of a free and open-ended discussion between all stakeholders affected by that action. The term *ethical colonialism* is understood by one of our interviewees as a critique of a situation in which *"people who are in a position of power [are] imposing their worldview on other people"* (22-E). This notion seems to be closely related to what is also denoted as *ethical imperialism* [62] or *moral imperialism* [41].

Our data shows that utilitarian and deontological arguments are frequently made by all interviewed REC members, even if they do not explicitly reference any philosophical sources. Usually, reasoning about ethical problems posed by our scenarios is rooted in both major ethical traditions.

Discussing the 3S research also highlighted the ethical challenges for researchers. Uncertainty about a possible impact of research-related 3S activities on remote systems was mentioned as the key problem (20-E, 22-E). This results in an ethical dilemma: *"Is the morally right decision to not scan because not scanning would then mean that the researchers are not involved in the actual crashes of any machines, or is the morally right decision to scan, thereby identifying vulnerable machines and encouraging them to fix and thereby helping other users?"* (22-E).

There are also other general ethical problems identified by one interviewee: (1) absorbing server resources for processing of numerous requests; (2) binding human resources by generating many log events which can alarm security personnel; (3) researchers lacking resources for a large number of vulnerability disclosures (24-E). Another interviewee (23-E) named three types of 3S activity that constituted absolute red lines from an ethical point of view: (1) *"fuzzing random stuff"*; (2) intentionally damaging someone's business operations; (3) extracting any kind of personal information.

Still, our interviews revealed that *"the red lines are not clear"* (24-E) and *discourse ethics* played an implicit role in all interviews with REC members. Researchers performing 3S should report in their papers any incident caused by their research activities (e. g., server crashes or privacy breaches), because only under this precondition can the REC become aware of harmful side effects of a given study and adequately assess its ethical implications. Researchers should also justify their selected research design by demonstrating that it is best suited to keep up high ethical standards (24-E).

4.2.3. Operators. Overall, the interviewed operators mostly exhibited a positive attitude towards 3S conducted by academic researchers. Six interviewees explicitly mentioned that they saw it as their own responsibility to secure their services once they go public on the Web. They acknowledge that such scans are part of Internet reality and operators should be well prepared for that, because *"[a]t the end of the day, the bad guys do it."* (18-O).

Most operators understood the purpose and necessity of 3S activity, as it contributes to securing the Web in the future. Still, operators voiced concerns about potential harm: *"I think [3S] is absolutely fine. Our system must also be able to withstand this to a certain extent. If you basically say, '[3S] is legal,' I would have a problem with that."* (16-O).

Issues related to infrastructure were noted in all interviews: *"[Y]ou never know what kind of business workflow you're going to trigger behind the scene."* (17-O). This lack of knowledge could lead to unintended consequences such as server crashes. Scans should make as few requests as possible to avoid accidental server overloads or denial of service. A red line would be crossed if researchers risked the system's availability on purpose. *"Anything that goes in the direction of denial of service, performance degradation and the like is problematic"* (11-O).

Another core aspect is the effects related to the end-user: *"[S]ecurity research should not interfere with websites' users"* (13-O). One frequently mentioned red line is the leakage or manipulation of private user information. Researchers should not actively search for personal data. If such data is discovered, they should handle it confidentially.

Operators' third main concern is harm to the organization. This could be reputation damage, when vulnerabilities are leaked to the public (9-O; 13-O), thus, researchers should handle vulnerability information confidentially (9-O). 3S could also cause financial losses for the organization. For example, as cloud services charge by traffic, large scans can cost organizations money (15-O). Many organizations also hire commercial pay-as-you-go CERTs to monitor their infrastructure, which means alarms appearing from 3S could lead to unnecessary costs (15-O).

Seeing the potential damage, operators also mentioned how they would react to 3S activity. Some highlighted their right to block IP addresses; others pointed out they had already implemented mechanisms to automatically block aggressive scans. Survey respondent 1-OS noted, *"I would actively try to stop it from my side. I believe you have the right to try and I have the right to try to not allow it."*

Besides technical reactions to scans, operators mentioned that they would seek communication with the scanning party if necessary, e. g., when noticing aggressive scans or harm done (12-O; 13-O). Also the filing of criminal complaints was mentioned, it appears to be less common and is usually reserved for events of significant harm or threats to reputation. If operators see that researchers are behind a scan, e. g., indicated by a header, direct communication is the preferable first response, with criminal complaints seen as the last resort. Nevertheless, two operators mentioned that in case of significant harm, they would be obligated to file criminal complaints regardless of the scanner's intention as their company's legal team would ask for it (14-O; 15-O).

Although the interviewed operators see the negative consequences, we already pointed out that the majority of operators in our study are positive about such scans, if conducted by academic researchers. They would rather accept the potential harm from researcher-conducted scans, knowing that it would be responsibly disclosed, than risk damage caused by criminal actors. Not only do the interviews indicate these opinions, but also the survey data in Q6. As shown in Figure 2 and Appendix A, more than half of the participants (36.1 % comfortable, 21.8 % somewhat comfortable) were positive about researchers conducting server-side scans, while 9.2 % were neutral and about a third expressed a negative stance (13.4 % somewhat uncomfortable, 19.3 % uncomfortable).

Key Takeaways: Legal assessments of 3S remain problematic due to a lack of criminal precedents. Even more critical is the risk of civil lawsuits. REC members stress the discussion of researchers' ethical decisions and transparency in publications, wishing for a well-argued balance between potential harm and benefits. The operators group is the most positive group in our study when it comes to 3S, as they also benefit from it in the form of vulnerability disclosure. They view legal actions as a last resort in cases of significant harm or if required by the company's legal team.

4.3. 3S Scenarios

At opportune moments during the interviews, we asked the interviewees about their opinions on five typical 3S scenarios. This allowed us to follow their trains of thought in assessing the concrete risks and benefits of 3S. Before reading the assessments, we would like to encourage readers to first think about each scenario and form their own opinion.

4.3.1. Alice – SQL Injection. In the first case study, we asked participants about the use of the SQL `sleep` function to test for the presence of SQL injection vulnerabilities.

The majority of legal experts in our study agreed that these types of scans had a low criticality from a legal standpoint. This is because Alice did not read any sensitive information from the server, which neither constituted the offenses of data espionage or phishing nor violated privacy legislation. Moreover, the potential for harm to the infrastructure is low, reducing the risk of a civil lawsuit. Still, several interviewees noted that a strict interpretation of the criminal code could lead to Alice's action being considered an unauthorized manipulation of data, as one could *"deliberately delay the response now and activate some particular mode in the database [...]"* (3-L).

While the legal experts attested the SQLi approach a low risk level, most concerns raised by the REC members in our study referred to legal issues: *"Alice should also make sure that she obeys all the laws in the country"* (20-E). They were also concerned that Alice had no control over server reactions. While most servers would likely just execute the sleep command, *"how can she make sure that the server does not crash or maybe misbehave"* (20-E). Consequently, they concluded that the REC would urge authors to consider other methods to address the research question and ask: *"[D]o you really need to do this to get your answer or are there safer things to do?"* (21-E). Despite the raised concerns, two of the REC members considered the sleep function *"a good way of minimizing the risks"* (24-E). They would accept such a paper and rather advise authors to disclose any incidents.

The interviewed operators mentioned arguments similar to those of the REC members. Three of them pointed out that researchers could neither predict how the infrastructure would respond to a request nor the implications on log files or caching. For instance, 9-O warned that if the payload was more than a sleep call, *"internal attackers [...] are suddenly able to get stuff [from the log]."* Operator 11-O expressed concerns about an even less invasive variant using `SELECT 1+1` instead of `sleep`. It potentially disrupted caches and, consequently, *'some subpage is no longer the title of the post [...] but 2"* (11-O), impacting end users. Regarding the initial sleep payload, another concern was that Alice did not know how time-critical a service was (17-O). In some real-time services like stock trading, a one-second delay could already cause significant financial damage. However, 16-O challenged this argument, stating that the natural latency on the Internet could often exceed one second. Despite these concerns, the common consensus was that a short sleep was acceptable, as it did not significantly impact end users. The `1+1` variant was even better received by the majority of operators, yet a red line would be crossed as soon as personal information was leaked, with some already considering knowledge of the database structure as critical.

These views are further supported by our survey data, as shown in Figure 2 and Table 3 in Appendix A: 69.7 % of the surveyed operators reported being (somewhat) comfortable with the initial sleep scenario (question A1), while 22.7 % were (somewhat) uncomfortable with it. The less invasive `1+1` case (A3) yielded a (somewhat) comfortable assessment by an additional 4.2 %, with 18.5 % still being (somewhat) uncomfortable. The more invasive variant in which Alice reads the database structure (A2) still led to some degree of comfort with 37.8 % of respondents and 29.6 % being (somewhat) uncomfortable with it.

4.3.2. Bob – Invalid HTTP Request. In this scenario, Bob sends invalid HTTP requests, accidentally crashing a server.

The interviewees from law all agreed that it came down to Bob's intent and potential negligence. For the judge in a criminal case, it was important to understand whether Bob knew and expected what would happen or not: *"[I]t depends very much on [...] the probability [of a crash]"* (3-L). If Bob had known that the server might crash and still sent an invalid request, or if this even was his intention, it could be argued that he committed computer sabotage. Regardless of the legal status of the act of sending such a request, civil law enables operators to demand compensation for caused harm or file an injunction to stop any 3S activity against their servers if Bob acted with intent or negligence.

The interviewed REC members expressed similar views and highlighted the importance of due diligence to demonstrate good intent. *"[M]any reviewers would first ask the question [...] what diligence did Bob do to mitigate harms beforehand?"* (22-E). Interestingly, they *all* asked this question, concluding that *"[i]f [Bob]'s not aware of this problem at the beginning, then it would be ethical"* (24-E), as his intent was not to crash a server. However, as soon as Bob's monitoring indicated a crash, he should stop scanning, disclose the issue to the operators, and mention the incident in the paper such that others would not make such mistakes.

In the same vein, the interviewed operators agreed that sending invalid HTTP requests would be fine. In the real world, something similar could also happen due to bugs in end-user software like browsers (9-O). Thus, operators should be prepared for such behavior and researchers should be allowed to test for it. *"I understand that's how research works, you can't make everything 100 % safe"* (15-O). The operators in our study further mentioned that they could also learn from such crashes and prevent them in the future. For example, Survey participant 97-OS wrote, *"If it is a crash that is reliably replicated [...], I care more about knowing about it than the service being temporarily down."* However, most operators would only be fine with a crash if it occurred just once. Thereafter, they would expect researchers to stop.

The survey responses paint a similar picture, as 54.6 % of participants would be (somewhat) comfortable with Bob's scan in the base scenario (see Figure 2 and Table 3). However, 37.8 % expressed (some) discomfort, probably because this scenario involved a crash. The less invasive variant (B3) only convinced another 10.1 % of participants and still was viewed negatively by 25.2 % of participants. On the other hand, the more invasive scenario (B2) still elicited a (somewhat) comfortable response in 21.8 % of surveyed operators, with 31.9 % being (somewhat) uncomfortable.

4.3.3. Charlie – Insecure Direct Object Reference. In this scenario, Charlie first changes a GET request that contains his user ID to instead use the ID of another user. This checks if it is possible to receive information from other users. In the more invasive variant, Charlie changes the ID in a POST request to check if he can modify other users' data.

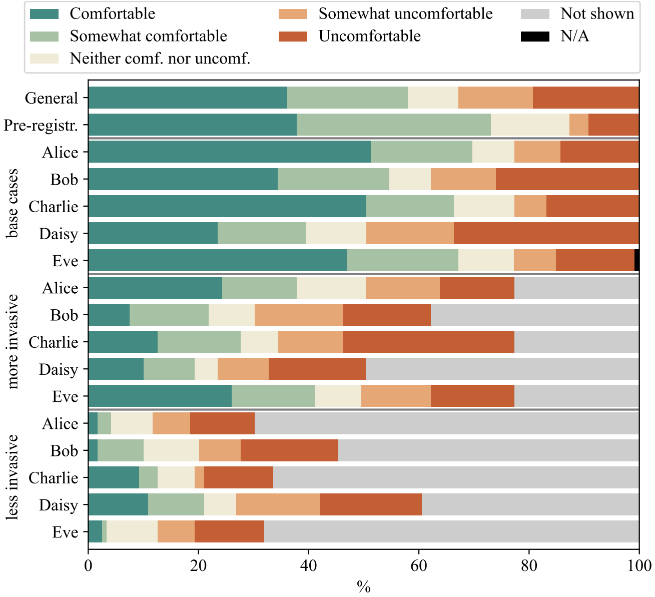

Figure 2. Surveyed operators' level of comfort with 3S in general (Q6, Q8) and 3S scenarios (A1–E3). The less and more invasive scenarios were only shown to participants who already were uncomfortable / comfortable with the base scenario, respectively (see Section 3.4).

For the GET request, all interviewed legal experts agreed that from a criminal law perspective, the possibility of charges for data espionage depended on whether an ID could be considered access protection. *"The question is, is it already enough as access security that a standard Internet user [...] cannot do that?"* (L-3). Six of them concluded that a predictable ID by itself could not be considered as sufficient access protection but pointed out that other experts might disagree. Thus, Charlie's first scan would *"probably [go] unpunished, because you cannot circumvent any access protection"* (L-6).

However, the POST request that modified data would unquestionably be illegal, even without bypassing any access protection. As 1-L exemplified: *"If I were to leave my laptop here unencrypted and you [...] made an A into a B, that would already be a [illegal] data manipulation, regardless of whether there is password protection or not."*

Beyond criminal law, for both scans, the interviewees referred to privacy laws and their complexity. Some regulations, such as Article 89 GDPR [26], hold exceptions for researchers, meaning that with sufficient justification the processing of users' personal data for research is legal.

The REC members also considered the GET request to be less harmful. Two interviewees would accept such a method in a paper without discussion, *"[h]e just needs to make sure that the handling of [...] the sensitive information [...] is also okay and then also describe it in the paper."* (E-20). Yet, this would likely lead to discussion in the committee. For the modifying request, all members agreed that it would cross a line. However, they all would consider alternative research designs, asking: *"Is there a reason Charlie didn't create two accounts that he then tries to change between the two?"* (22-E). All interviewed REC members stated they would be fine with such research if Charlie instead used a dummy account to evaluate his experiments against. This would not harm any individual users and, thus, would ethically be acceptable. Nevertheless, participant 22-E brought up the question of potential terms of service violations. Asked about this, legal expert 5-La expressed a similar view, acknowledging that while the use of a dummy account might not constitute a criminal offense, it could potentially violate the service's terms and conditions.

Asked about Charlie's experiments, no operator mentioned terms of service. Three of them were fine even with the data modification, and five with the GET request, as every operator should check for such flaws. Three operators mentioned privacy concerns that would speak against the scans, because *"as soon as I notice that users' data is being leaked, I have to run to the data protection authority [...]. I don't feel like doing that."* (10-O). Operators also supported the dummy account solution, especially if they saw a connection between two accounts, e. g., similar names.

Survey data (Figure 2) confirms the interview sentiments, with 66.4 % of participants being (somewhat) comfortable with Charlie's GET request (C1) and 22.7 % expressing (some) level of discomfort. As expected, comfort levels were lower for the modifying POST request (C2): 42.9 % of operators expressed (some) discomfort with this, but still 27.7 % felt (somewhat) comfortable. By contrast, Charlie only accessing his own dummy profiles (C3) yielded an additional positive response of 12.6 % but still caused some level of discomfort with 14.3 % of surveyed operators.

4.3.4. Daisy – Stored XSS. Daisy enters an XSS payload into an input field on a website, subsequently storing it on the server. This payload potentially reaches all site users and, upon execution, sends a ping back to Daisy's server.

The legal experts in our study viewed this case as problematic. All but two of them pointed out that Daisy would be infringing on privacy laws by collecting users' IP addresses. *"The IP address, whether static or dynamic, is personal data"* (4-L). While the Article 89 GDPR does make exceptions for researchers, any advancement of scientific research must be weighed against the individual's right to privacy. The experts were not convinced of the benefits of Daisy's research. From a criminal law perspective, all but one expert considered uploading data explicitly deemed to be an attack payload an illegal manipulation of the server-side data: *"[T]his code is stored somehow [...], individual bits and bytes are actually changed without the user's consent"* (1-L). As Daisy's actions did not only target a server but potentially all of its users, she may be subject to prosecution for multiple offenses, against the server and each individual client.

The interviewed REC members assessed the scenario similarly, stating that researchers should not store potentially harmful code on a web server. They expressed concern that once the code was stored, it left a lasting sign of an attack on the server. Like the jurists, they found it very concerning to execute code on multiple clients. Only one interviewee (24-E) said they would be fine with Daisy's experiment if the method was the only viable means to answer the research question and potential benefits outweighed the potential harm. This participant would require Daisy to provide more information explaining that her research was ethical and her method represented the least risky approach, as *"ultimately, it's the responsibility of the authors to demonstrate that their research is ethical"* (24-E). If Daisy could revise her design to ensure she does not inadvertently attack unknowing users, most REC members we interviewed would not reject the paper on ethical grounds. One option could be a check added to the payload that ensured only browser(s) controlled by Daisy send a ping back. Other browsers would still execute the confirmation code, but the harm would be acceptable for most of the interviewed REC members.

Operators in our study were also mostly comfortable with the proposed alternative research design. *"It is okay. The way you are choosing the payload shows your intention, I think"* (17-O). In the base scenario, operators expressed discomfort with a malicious payload being stored on their servers, similar to the REC members; yet, they were more troubled by the privacy implications pointed out by the legal experts. This sentiment is echoed in the survey data (Figure 2 and Table 3), with 49.6 % of participants in the base case (D1) expressing some level of discomfort with such a scan and only 39.5 % being (somewhat) comfortable with it. As expected, the more invasive variant (D2) was only viewed favorably by 19.3 % of respondents, while 26.9 % had a negative stance. If Daisy only sent pings from her own accounts (D3), these rates were 21.0 % and 33.6 %, respectively.

4.3.5. Eve – Path Traversal. Eve measures the prevalence of supposedly inaccessible files inadvertently exposed to the public, e. g., `.git`. This was the least divisive scenario among interviewees.

For criminal law, the interviewed legal experts focused on data espionage, as researchers were not manipulating any data in this scenario. The legality of scanning for confidential files depended on whether the researchers bypassed any access protection. Legal expert 3-L explained that Eve's action did not constitute the offense of data espionage, as *"the mere intention that something is secret is not enough to secure access; I need some objective barrier to access"*. Six other legal experts leaned in a similar direction, explaining that if the target file was hidden behind a random file name, like a UUID, this could be argued as an implementation of access control. However, if researchers were only looking for well-known files like *.git*, most agreed that this could not be considered an access control bypass. Regarding privacy laws, the experts repeated that researchers must have a valid reason for processing any user data: *"If it is personal data, then of course you have a processing of personal data. Consequently, it is somehow a data protection right and needs* *the usual justification"* (2-L). If no sensitive information is expected, there should be no privacy concerns.

The REC members in our study also saw the privacy aspect as the most critical one. However, they all agreed that such a scan would be acceptable if researchers did their utmost to minimize data processing. As participant 22-E put it in our interview, *"[W]hat I would try to do is try to develop a mechanism that minimizes the need for humans to look at data [...] that is sensitive."*

All but two operators we interviewed shared a similar perspective. Participant 13-O stated that the critical aspect was *"not knowing what happens to the information"*. On the other hand, 15-O said about the data: *"It is public! Whether I wanted it or not, it is just public, and now even the bad guys see that!"* All but two interviewed operators agreed that they would accept such scans.

In the survey, all Eve variants received similarly high approval (see Figure 2 and Table 3): 67.2 % exhibited (some) degree of comfort with the scan in the base case (E1), while 21.8 % were (somewhat) uncomfortable with it. 41.2 % still answered positively in the more invasive scenario (E2), while 27.7 % a provided an assessment that tended towards negative. The less invasive case (E3) did not lead to notably higher acceptance with initially skeptical operators; here 3.4 % where (somewhat) comfortable and 19.3 % still (somewhat) uncomfortable with Eve's scan..

Key Takeaways: Throughout all scenarios, legal experts often avoided making absolute statements, only leaning into one possible direction, underlining the complexity of the legal landscape. REC members occasionally referred to the legal side, yet focused more on users' privacy and considering alternative research designs, e. g., using dummy accounts. Operators mentioned a few red lines during the interviews, e. g., data leakage and modification, yet many were open towards 3S research, provided it is accompanied by a proper risk assessment and a responsible disclosure process.

4.4. Participants' Suggestions for Improvement

Throughout our interviews, legal experts frequently highlighted that 3S-based research currently occupied a gray zone, with benevolent researchers carrying the risks (7-L). Thus, most interviewees from law suggested legislative changes to add exceptions for security researchers (3-L; 7-L), akin to exceptions in drug research. 1-L also suggested a system for approving pre-registered projects. A counterargument raised against both ideas is that the state cannot simply allow actions that could cause direct harm (6-L). In response to this, 5-La noted that in fact many potentially harmful activities, such as the operation of nuclear power plants, are already legally allowed. The more critical issue lies in defining who qualifies as a *researcher* to prevent the law from being abused by actors with malicious intent. However, legal experts also recognized that such legislative measures at the nation-state level would not extend to other jurisdictions, underscoring the need for a global solution.

The interviewed REC members proposed that security research institutions should establish their own ethical review boards or enhance IRBs with IT knowledge to raise concerns prior to conducting studies. Nonetheless, they advocated for ethical guidelines and procedures like *"ethics modeling"*, mentioned by 22-E with reference to literature [45]. These guidelines should not be strict, but rather outline a process that considers various ethical perspectives to address each stakeholder's interests (20-E, 21-E). In their publications, researchers should then discuss these considerations and justify how the chosen research designs minimized the potential for harm while still being able to answer the research questions. Lastly, interviewees admitted that there may be a lack of communication between RECs and the community, raising transparency concerns. They suggested that more information about the REC process and educational material such as practical ethics scenarios, like our vignettes, could address these concerns (20-E; 23-E).

Transparency is important not only for ethics boards but also for the operators in our study who desire clarity regarding scans performed on their systems. Measures such as identifying fixed IP areas from which scans originate (16-O; 20-O), including headers that indicate the scanning actor's identity (16-O), or fixed scanning time slots were suggested as a means to differentiate between legitimate security scans and malicious activities (13-O; 16-O). Nevertheless, due to the challenges of justifying such measures to operators, almost all operators said they would prefer to be asked for consent or informed in advance about a scan. Then they would know who is behind a scan and could prepare their website, e. g., via *"declar[ing] clearly on the homepage that in the next two hours, in the next 24 hours, there may be errors"* (13-O). In general, the operators in our study wished to be actively involved in the process, and they consistently highlighted the importance of disclosing identified vulnerabilities responsibly.

4.5. Opinions on Pre-Registration

We also asked interviewees for their opinion on a pre-registration procedure overseen by a trusted third party (TTP), proposed in more detail in Section 6.2. Some had already suggested such a process themselves before we brought it up.

All academic legal experts we interviewed were positive about the idea, yet L-4 noted that pre-registration would not prevent civil lawsuits, e. g., in case of harm. Practitioners were less optimistic, pointing out that significant legislative change would be required. Nevertheless, the general feedback was optimistic, with the German Federal Office for IT Security (BSI) viewed as a potential TTP. The REC members that we interviewed also welcomed the proposal but raised concerns about a national institution serving as the TTP due to the global nature of the scans. They would prefer a globally coordinated and community-driven initiative. The interviewed operators also disliked the idea of a governmental agency being in charge of a TTP, as they assumed bureaucracy would unnecessarily delay research. Instead,

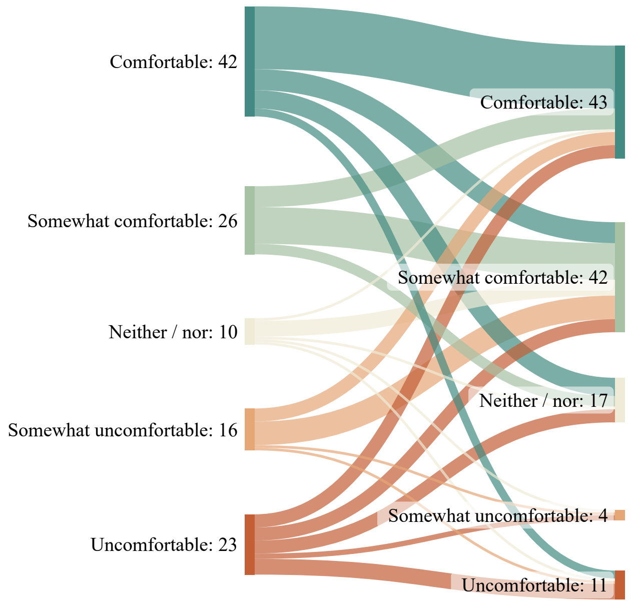

Figure 3. Changes in surveyed operators' stance towards 3S between Q6 (first general assessment; left) and Q8 (with pre-registration; right)

they favored an entity from the IT community (9-O; 18-O) or industry (9-O; 16-O). Still, the majority appreciated the idea of a TTP that would bring transparency to 3S research.

In the survey, we also asked participants about their general assessment of 3S; first before the scenarios and the proposal (Q6) and later in the context of our proposed pre-registration with a TTP (Q8). Figure 3 compares the respective answers, detailed numbers are listed in Appendix A. Between the two questions, we observe a shift in opinion towards a more comfortable position. Those whose opinion changed from "comfortable" to a less positive position often explained this with unnecessarily hindering research (*"it doesn't make sense to restrict the security researchers"* [105-OS]), as malicious actors would perform scans anyway.

Asked to rate the competence of different entities to serve as a TTP in such a pre-registration process (Q7) on a Likert scale from competent (score: 5) to incompetent (1), we find that the surveyed operators assigned higher competence to NGOs / white-hat hacker organizations (mean 4.15, std ±1.1, med 4.0) and academic institutions (mean 3.79, std ±1.13, med 4.0), while industry organizations (mean 2.67, std ±1.36, med 3.0), government agencies (mean 2.78, std ±1.38, med 3.0), and international organizations (mean 2.67, std ±1.36, med 3.0) were considered less competent.

5. Discussion

Our findings reveal numerous challenges when planning and conducting 3S studies. We now discuss the main insights from our interviewees and propose how to move forward.

5.1. Legal Roadblocks

Our project aimed to determine the legal boundaries of 3S research (RQ1). While we analyzed these boundaries through the lens of German law, conversations with international experts indicated some similarities in other countries. Among numerous aspects to consider from our interviews, we found three to be the most frequently raised.

No judicial decisions. Currently, there are hardly any judicial decisions on 3S or similar hacking-related research topics. Not only does this leave researchers with uncertainty about the legal situation, but it also makes it difficult for legal experts to assess how the courts would interpret the law and decide specific cases. On the contrary, our expert interviews left us with the impression that German legal practitioners deliberately try not to pursue such cases or to settle them out of court, apparently to consider the importance of white-hats and security research but ultimately stalling jurisprudential progress.

Need for legislative action. Given this lack of court decisions, even experts from law struggled to provide clear answers to our legal questions. To provide researchers with a reliable regulatory environment, the interviewed experts suggested to introduce exemptions for researchers into German law to avoid prosecution and limit liability[1]. However, this often raised the question of who such an exemption would apply to – only academic researchers, white hat hackers, and who else? Exemptions for security researchers are also of political importance, as underlined by this topic being included in the current German coalition agreement [39]. For example, the Netherlands [18] have already established similar exceptions.

International dimension. Although some countries already implemented legal exceptions for researchers, an international agreement is necessary, as the global nature of the Web and its vulnerabilities confronts 3S researchers with a multitude of jurisdictions. This further complicates the task for legal experts to provide reliable guidance regarding such scans. With modern web applications heavily relying on cloud hosting and content delivery networks, a server's location might be hard to determine and/or subject to change, making it hard to predict which jurisdiction applies.

5.2. Ethical Challenges

We sought to understand how RECs form their decisions and how 3S research can align with ethical standards (RQ2).

Ethical discourse. Our interviews revealed that REC decisions emerge from in-depth discussions with authors. Instead of having a preconceived opinion, they try to understand the authors' intentions and figure out the benefits of the research. They weigh them against potential harm, which can be challenging to predict given the Web's complexity. The interviewees often stated that their aim was to help the

[1] After finalizing this paper, the German federal ministry of justice issued a statement on modernizing the criminal law [31]. While they address §202 StGB to enable security research, they omit §303 in their statement, which our interviewees also highlighted as equally critical.

authors rather than hinder them. For example, the REC at S&P 2022 gave a “Reject” recommendation only for two out of 67 papers flagged for ethics (from 1,006 in total) [13].

Missing guidelines. Our interviewees highlighted the absence of practical ethical guidelines. While the Menlo report [7] is a well-known standard, it is largely theoretical and lacks practical examples. Moreover, ethical understanding varies globally, suggesting that a single standard may be insufficient. Instead, we should learn from previous REC decisions, e. g., if published anonymized. REC members also repeatedly said that they disliked fixed rules for ethics due to edge cases and preferred to maintain their discourse approach to consider each case individually. They welcomed our 3s scenarios as a possible step towards creating ethical teaching material that fosters a fair discourse.

Post-mortem. REC decisions are made during peer review of a paper after the research has been conducted. This means that if an experiment went wrong, the harm had already been done, leaving the REC with no option to prevent it. Some interviewees emphasized they would prefer a mechanism to review research ex-ante. Others disagreed, arguing that this would entail too much work and that authors should be trusted. Examples such as the Tor Research Safety Board [68], medical research, or IRBs show that such a pre-approval of research projects is feasible if well designed [20]. We discuss this in more detail in Section 6.2.

5.3. Operator Perspectives

For our third question (RQ3), we wanted to get server operators’ views on 3S, as they are the ones directly affected.

Legal action. Our interviews with operators revealed that most of them would not consider legal action against researchers for their scans. They see this only as the last option if direct communication fails to result in a solution or in an event of significant financial or reputational damage. Still, some operators might be obligated to file criminal complaints at the request of their company’s legal team, regardless of the intention behind the scan.

Consent. Most interviewed operators were willing to support researchers, but to prepare for potential harm and know who to contact, many preferred to be asked for consent before the scan. However, 3S-based research typically needs to be conducted at scale for meaningful results. As shown by prior work [65] and the bounce rate of our survey invitations, reaching the operators of a large number of websites is a persisting challenge, let alone asking for their permission. Approval requests could also introduce observer effects and selection bias towards security-aware operators. As much as this approach would be favored, it is simply not practical.

Harm acceptance. 3S research has the potential to cause harm. Still, many interviewees stated that their log files already indicated such scans regularly occur in the wild. They do not expect additional harm caused by researchers. If researchers run non-invasive and well-pre-tested scans likely to get lost in the Internet’s noise, the operators in our study might be willing to accept them. They further would find value in researchers responsibly disclosing vulnerabilities.

6. Recommendations

Our work confirms that server-side scanning is challenging and comes with the potential for harm. However, the majority of our interviewees agreed on the need to enable 3S-based research, as malicious actors also conduct such scans without prior notice but with ill intent. Based on our findings, we now introduce best practices for 3S research and discuss our proposal for a pre-registration board.

6.1. Best Practices for Server-Side Scans

Our interviews aimed to find out how 3S research can be conducted safely and without violations of legal or ethical norms. One core aspect was to first evaluate alternative research designs that also answer the research question but minimize the potential for harm, e. g., dummy accounts to test against instead of real users (see Section 4.3.3). If only a 3S-based design is viable, researchers should try to follow a few rules that partially extend prior recommendations [24] to ensure the benefits exceed the potential harm:

Laboratory pre-study. Before running the scans on the Web, there should be a laboratory pre-study. This way, researchers can catch potential issues early in the process and minimize the chance of accidentally crashing any servers.

Data minimization. The experiment should collect and store as little data as possible. Additionally, the process of validating data should be automated so that researchers minimize the risk of viewing personal information.

Data manipulation. Even though any request by definition triggers the server-side manipulation of some bits, researchers should limit data manipulation to the bare minimum and refrain from altering user-related data.

Resource minimization. The experiment should be designed to require as few resources on the server side as possible to not overload a server. If many requests are needed, they should be scheduled over a longer time period.

Monitoring. Researchers should always monitor the status and results of their scans. This means that not only should the scanner be monitored, but one process should also check that the scanned websites remain online after a scanning action. If a server crash is detected, the scan should be paused for closer investigation. For transparency, any incidents should be mentioned in the resulting publication.

Transparency. For the acceptance of 3S it is important to allow operators to learn about the purpose and scope of the scans occurring on their servers and the study goals. For this, we recommend creating a study website and linking to it in a custom header in all requests to the server. A reverse lookup of the scanner’s IP address should lead to the same website. Some operators also would welcome a signature-based method to verify that scans really originated from the purported institution. We leave this to future work.

Fixed IP address. Not only should the scanner’s IP address point to the study website, but it is equally important that the IP address(es) used for scanning remain the same throughout the entire study. This allows operators to add them to their allow list or opt out by blocking them.

Allow explicit opt-out. Operators should always have the option to explicitly opt out of the study. Thus, the study website should provide clear instructions on how to opt out and, ideally, a form to make this process as easy as possible.

6.2. Pre-registration Board

Based on our findings, we propose to establish a *pre-registration* process overseen by a *trusted third party* (TTP) for future 3S-based research. Although this approach cannot provide absolute legal certainty due to the global scope of 3S, it can offer proof of researchers' intentions, removing some subjective elements in laws such as Germany's Sec. 303b StGB on computer sabotage.

The proposed TTP would review submitted research proposals that should outline the research question, expected outcomes, and the chosen study design, as well as a discussion on why alternative methods are not feasible. This will allow the TTP's ethics reviewers to weigh the potential for harm against the benefits, engaging in dialogue with the researchers. Ideally, the TTP should be supported by legal advisors, given that ethics reviewers cannot be expected to know the applicable laws of all jurisdictions.

To avoid excessive simultaneous testing of systems, the approval of research proposals should include the assignment of multiple time slots, during which the scans can be performed. The TTP should also maintain a publicly available list of all ongoing scans, the associated IP addresses, and points of contact, enabling operators to get information about current approved scanning activities.

As our survey suggests, the TTP should be established as an NGO, possibly with support from academics from all over the globe to represent diverse ethical views. While we are aware that establishing such an entity is an enormous task, we believe it would greatly benefit future research.

7. Conclusion and Call to Action

We explored the normative boundaries of security and privacy research based on server-side scanning (3S) through interviews and a survey. While ethical and legal risks for researchers have led to reluctance in conducting 3S research, our findings reveal that the slight majority of operators in our study might be open to 3S, yet there is a need for global legislation and ethical guidance early in the research process to provide researchers with a clear and reliable framework. Our hope is that the best practices and pre-registration approach proposed as the outcome of our study will provide guidance toward this goal and stimulate productive discourse within the security and privacy research community.

As this topic is also a political one, we also hope to raise awareness beyond academia and to provide more normative clarity to non-academic researchers and ethical hackers around the globe. We encourage other researchers to build upon our work and conduct similar studies in their own countries. This would allow our community to gain a more comprehensive understanding of the normative boundaries of 3S and security research on an increasingly larger scale. The envisioned result could be a global set of research best practices and, ultimately, more research-friendly adjustments of policy decisions worldwide.

Acknowledgments

We would like to thank all anonymous reviewers and our shepherd for their constructive feedback in the process of improving this paper. In addition, we would like to express our gratitude to all interview and survey participants for their valuable insights and thoughts. This work was conducted in the scope of a dissertation at the Saarbrücken Graduate School of Computer Science.

References

[1] Richard L. Abel, Ole Hammerslev, Hilary Sommerlad, and Ulrike Schultz, editors. *Lawyers in 21st-Century Societies: Volume 1: National Reports*. Hart, 2020.

[2] Richard L. Abel, Ole Hammerslev, Hilary Sommerlad, and Ulrike Schultz, editors. *Lawyers in 21st-Century Societies: Volume 2: Comparisons and Theories*. Hart, 2022.

[3] Cheryl S. Alexander and Henry Jay Becker. The use of vignettes in survey research. *Public Opinion Quarterly*, 42, 1978.

[4] Karl-Otto Apel. *Diskurs und Verantwortung: Das Problem des Übergangs zur postkonventionellen Moral*. Suhrkamp, 1990.

[5] Christiane Atzmüller and Peter M. Steiner. Experimental Vignette Studies in Survey Research. *Methodology*, 6(3), 2010.

[6] Eric Auchard. Researcher Found Security Holes at Smartphone-Only Bank N26. https://web.archive.org/web/20220929184415/https://www.reuters.com/article/us-cyber-fintech-n26-idUSKBN14H1EM, September 2022.

[7] Michael Bailey, David Dittrich, Erin Kenneally, and Doug Maughan. The Menlo Report. *IEEE Security & Privacy*, 10(2), 2012.

[8] Silvia et al. Balaban. Rechtslage der IT-Sicherheitsforschung (in German; English: "Legal Status of IT Security Research"). https://sec4research.de/assets/Whitepaper.pdf, 2022.

[9] Jeremy Bentham. *An Introduction to the Principles of Morals and Legislation*. The Athlone Press, 1970.

[10] Paul Schiff Berman. *Global Legal Pluralism: a Jurisprudence of Law beyond Borders*. Cambridge University Press, 2012.

[11] Alexander Bogner, Beate Littig, and Wolfgang Menz, editors. *Interviewing Experts*. Palgrave Macmillan, 2009.

[12] Michael Bohlander. *Principles of German Criminal Law*. Hart, 1st edition, 2008.

[13] Joseph Bonneau. Oakland REC Annual Summary 2022. https://docs.google.com/document/d/15x5Q

d1UTaoMSRZgRRvurPbgRg4SWWLQKhG41ouYV0TY.

[14] Sam Burnett and Nick Feamster. Encore: Lightweight Measurement of Web Censorship with Cross-Origin Requests. In *SIGCOMM 2015*. ACM, 2015.

[15] Kevin Butler and Kurt Thomas. Message from the USENIX Security '22 Program Co-Chairs. https://www.usenix.org/sites/default/files/sec22_message.pdf, 2022.

[16] Ahmet Salih Buyukkayhan, Can Gemicioglu, Tobias Lauinger, Alina Oprea, William Robertson, and Engin Kirda. What's in an Exploit? An Empirical Analysis of Reflected Server XSS Exploitation Techniques. In *23rd International Symposium on Research in Attacks, Intrusions and Defenses*. USENIX Association, 2020.

[17] John W. Byers. Encore: Lightweight Measurement of Web Censorship with Cross-Origin Requests – Public Review. https://conferences.sigcomm.org/sigcomm/2015/pdf/reviews/226pr.pdf, 2015.

[18] Centre for Cyber Security Belgium. Vulnerability Reporting to the CCB. https://ccb.belgium.be/en/vulnerability-reporting-ccb, February 2023.

[19] Chaos Computer Club. Clause 202c of German penal code endangers German IT industry. https://www.ccc.de/en/updates/2008/stellungnahme202c, July 2008.

[20] Andy Cockburn, Carl Gutwin, and Alan Dix. HARK No More: On the Preregistration of CHI Experiments. In *CHI '18*. ACM, 2018.

[21] Jedidiah R. Crandall, Masashi Crete-Nishihata, and Jeffrey Knockel. Forgive Us Our SYNs: Technical and Ethical Considerations for Measuring Internet Filtering. In *SIGCOMM Workshop on Ethics in Networked Systems Research*. ACM, 2015.

[22] Adam Doupé, Ludovico Cavedon, Christopher Kruegel, and Giovanni Vigna. Enemy of the State: A State-Aware Black-Box Web Vulnerability Scanner. In *USENIX Security 2012*. USENIX Association, 2012.

[23] Drupal.org. Drupalgeddon. https://www.drupal.org/project/drupalgeddon.

[24] Zakir Durumeric, Eric Wustrow, and J. Alex Halderman. ZMap: Fast Internet-wide Scanning and Its Security Applications. In *USENIX Security 13*. USENIX Association, 2013.

[25] Benjamin Eriksson, Giancarlo Pellegrino, and Andrei Sabelfeld. Black Widow: Blackbox Data-Driven Web Scanning. In *SP 2021*. IEEE, 2021.

[26] The European Parliament and the Council of the European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Official Journal of the European Union, L 119/1, April 2016.

[27] Janet Finch. The Vignette Technique in Survey Research. *Sociology*, 21(1), 1987.

[28] Alexander Gamero-Garrido, Stefan Savage, Kirill Levchenko, and Alex C. Snoeren. Quantifying the Pressure of Legal Risks on Third-Party Vulnerability Research. In *CCS 2017*. ACM, 2017.

[29] German Federal Ministry of Justice. German Civil Code (Bürgerliches Gesetzbuch – BGB). Non-binding English translation: https://www.gesetze-im-internet.de/englisch_bgb/englisch_bgb.html, 2021.

[30] German Federal Ministry of Justice. German Criminal Code (Strafgesetzbuch – StGB), 2021. Non-binding English translation: https://www.gesetze-im-internet.de/englisch_stgb/index.html.

[31] German Federal Ministry of Justice. Eckpunkte des Bundesministeriums der Justiz zur Modernisierung des Strafgesetzbuchs. German: https://www.bmj.de/SharedDocs/Downloads/DE/Gesetzgebung/Eckpunkte/1123_Eckpunkte_Modernisierung_Strafrecht.pdf3, 2023.

[32] Scott Gilbertson. Germany Outlaws Hacking, Cripples Security Industry. https://www.wired.com/2007/08/germany-outlaws-hacking-cripples-security-industry/, August 2007.

[33] Barney G. Glaser and Anselm L. Strauss. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine Publishing Company, 1967.

[34] Antonia Groß. The CDU's leaky campaign app. https://www.berliner-zeitung.de/en/the-cdus-leaky-campaign-app-li.176310, August 2021.

[35] Jürgen Habermas. *Justification and Application: Remarks on Discourse Ethics*. Polity Press, 1993.

[36] Florian Hantke, Sebastian Roth, Rafael Mrowczynski, Christine Utz, and Ben Stock. Where Are the Red Lines? Towards Ethical Server-Side Scans in Security and Privacy Research – Supplementary Material. https://github.com/cispa/Ethical-Server-Side-Scanning, 2023.

[37] Thorsten Holz and Alina Oprea. IEEE S&P'21 Program Committee Statement Regarding The "Hypocrite Commits" Paper. https://www.ieee-security.org/TC/SP2021/downloads/2021_PC_Statement.pdf, May 2021.

[38] IEEE. IEEE Symposium on Security and Privacy 2022 – Call for Papers. https://www.ieee-security.org/TC/SP2022/cfpapers.html, 2021.

[39] Insight EU Monitoring. Dare More Progress: Unofficial English Translation of German Coalition Agreement Added. https://portal.ieu-monitoring.com/editorial/dare-more-progress-agreement-of-germanys-new-coalition-now-online/363687, January 2022.

[40] Bahruz Jabiyev, Steven Sprecher, Kaan Onarlioglu, and Engin Kirda. T-Reqs: HTTP Request Smuggling with Differential Fuzzing. In *CCS 2021*. ACM, 2021.

[41] Ryan Jenkins. Moral imperialism. In Deen K. Chatterjee, editor, *Encyclopedia of Global Justice*. Springer, 2011.

[42] Ben Jones, Roya Ensafi, Nick Feamster, Vern Paxson, and Nick Weaver. Ethical Concerns for Censorship Measurement. In *SIGCOMM NS Ethics '15*. ACM, 2015.

[43] Immanuel Kant. *Groundwork of the Metaphysics of Morals: A German-English Edition*. Cambridge University Press, 2011.

[44] Matthias Kilian and Ulrike Schultz. Germany: Resistance and Reactions to Demands of Modernisation. In Richard L. Abel, Ole Hammerslev, Hilary Sommerlad, and Ulrike Schultz, editors, *Lawyers in 21st-Century Societies*. Hart, 2020.

[45] Tadayoshi Kohno, Yasemin Acar, and Wulf Loh. Ethical frameworks and computer security trolley problems: Foundations for conversations. 2023.

[46] Victor Le Pochat, Tom Van Goethem, Samaneh Tajalizadehkhoob, Maciej Korczyński, and Wouter Joosen. TRANCO: A Research-Oriented Top Sites Ranking Hardened Against Manipulation. In *NDSS '19*. Internet Society, 2019.

[47] Market Research Telecast. Data leak at Modern Solution: House search instead of bug bounty. https://marketresearchtelecast.com/data-leak-at-modern-solution-house-search-instead-of-bug-bounty/182043/, October 2021.

[48] Philipp Mayring. Qualitative content analysis. *Forum: Qualitative Sozialforschung / Forum: Qualitative Social Research*, 1(2):Art. 20, 2000.

[49] Sean McAllister, Engin Kirda, and Christopher Kruegel. Leveraging User Interactions for In-Depth Testing of Web Applications. In *Recent Advances in Intrusion Detection: 11th International Symposium*. Springer, 2008.

[50] John Stuart Mill. *Utilitarianism, On Liberty, Considerations on Representative Government, Remarks on Bentham's Philosophy*. Everyman, 1993.

[51] OWASP. Cross Site Scripting (XSS). https://owasp.org/www-community/attacks/xss/, 2023.

[52] OWASP. Path Traversal. https://owasp.org/www-community/attacks/Path_Traversal, 2023.

[53] OWASP. SQL Injection. https://owasp.org/www-community/attacks/SQL_Injection, 2023.

[54] OWASP. Testing for Insecure Direct Object References. https://owasp.org/www-project-web-security-testing-guide/latest/4-Web_Application_Security_Testing/05-Authorization_Testing/04-Testing_for_Insecure_Direct_Object_References, 2023.

[55] Eric Pauley and Patrick McDaniel. Understanding the Ethical Frameworks of Internet Measurement Studies. *Workshop on Ethics in Computer Security (EthiCS)*, 2023.

[56] Giancarlo Pellegrino, Martin Johns, Simon Koch, Michael Backes, and Christian Rossow. Deemon: Detecting CSRF with Dynamic Analysis and Property Graphs. In *CCS 2017*. ACM, 2017.

[57] Stijn Pletinckx, Kevin Borgolte, and Tobias Fiebig. Out of Sight, Out of Mind: Detecting Orphaned Web Pages at Internet-Scale. In *CCS 2021*. ACM, 2021.

[58] Ponemon Institute. Costs and Consequences of Gaps in Vulnerability Response. https://media.bitpipe.com/io_15x/io_152272/item_2184126/ponemon-state-of-vulnerability-response-.pdf, 2019.

[59] Peter H. Rossi. Vignette analysis: Uncovering the normative structure of complex judgments. In Robert K. Merton, James S. Coleman, and Peter H. Rossi, editors, *Qualitative and quantitative social research*. Free Press, 1979.

[60] Margrit Schreier. *Qualitative Content Analysis in Practice*. Sage, 2012.

[61] Margrit Schreier. Qualitative Content Analysis. In Uwe Flick, editor, *The SAGE Handbook of Qualitative Data Analysis*. Sage, 2014.

[62] Michael W. Small. Ethical Imperialism. In Robert W. Kolb, editor, *The SAGE Encyclopedia of Business Ethics and Society*. Sage, 2018.

[63] Cristian-Alexandru Staicu and Michael Pradel. Freezing the Web: A Study of ReDoS Vulnerabilities in JavaScript-based Web Servers. In *USENIX Security 2018*. USENIX Association, 2018.

[64] Marius Steffens, Christian Rossow, Martin Johns, and Ben Stock. Don't Trust The Locals: Investigating the Prevalence of Persistent Client-Side Cross-Site Scripting in the Wild. In *NDSS '19*, 2019.

[65] Ben Stock, Giancarlo Pellegrino, Christian Rossow, Martin Johns, and Michael Backes. Hey, You Have a Problem: On the Feasibility of Large-Scale Web Vulnerability Notification. In *USENIX Security 16*. USENIX Association, 2016.

[66] Ben Stock, Stephan Pfistner, Bernd Kaiser, Sebastian Lekies, and Martin Johns. From Facepalm to Brain Bender: Exploring Client-Side Cross-Site Scripting. In *CCS 2015*. ACM, 2015.

[67] Anselm L. Strauss and Juliet M. Corbin. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Sage, 3rd edition, 2008.

[68] The Tor Project. Research Safety Board. https://research.torproject.org/safetyboard/, 2022.

[69] United States National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research*, volume 1. Department of Health, Education, and Welfare, 1978.

[70] Christine Utz, Sabrina Amft, Martin Degeling, Thorsten Holz, Sascha Fahl, and Florian Schaub. Privacy Rarely Considered: Exploring Considerations in the Adoption of Third-Party Services by Websites. *Proceedings on Privacy Enhancing Technologies*, 2023(1), 2023.

[71] Andreas Witzel and Herwig Reiter. *The Problem-centered Interview: Principles and Practice*. Sage, 2012.

[72] Qiushi Wu and Kangjie Lu. On the Feasibility of Stealthily Introducing Vulnerabilities in Open-Source Software via Hypocrite Commits. https://github.com/QiushiWu/QiushiWu.github.io/blob/main/papers/OpenSourceInsecurity.pdf, February 2021.

[73] Yiming Zhang, Mingxuan Liu, Mingming Zhang, Chaoyi Lu, and Haixin Duan. Ethics in Security Research: Visions, Reality, and Paths Forward. In *EuroS&PW '22*, 2022.

Appendix A.
3S Assessments in the Survey

TABLE 3. Surveyed operators' (n = 119) general assessment of 3S, all scenarios and more or less invasive variants, and our pre-registration proposal.

									Likert scores		
Q#	Scenario	Comfortable (5)	Somewhat comfortable (4)	Neither comf. nor uncomf. (3)	Somewhat uncomfortable (2)	Uncomfortable (1)	Not shown	N/A	mean	std	median
		%	%	%	%	%	%	%			n
Q6	3S General	36.1	21.8	9.2	13.4	19.3	–	0	3.42	1.55	4.0
Q8	Pre-registration	37.8	35.3	14.3	3.4	9.2	–	0	3.89	1.22	4.0
A1	Alice (base)	51.3	18.5	7.6	8.4	14.3	–	0	3.84	1.48	5.0
A2	Alice (more inv.)	24.4	13.4	12.6	13.4	13.4	22.7	0	2.54	1.91	3.0
A3	Alice (less inv.)	1.7	2.5	7.6	6.7	11.8	69.7	0	0.66	1.21	0
B1	Bob (base)	34.5	20.2	7.6	11.8	26.1	–	0	3.25	1.64	4.0
B2	Bob (more inv.)	7.6	14.3	8.4	16.0	16.0	37.8	0	1.68	1.71	1.0
B3	Bob (less inv.)	1.7	8.4	10.1	7.6	17.6	54.6	0	1.05	1.44	0
C1	Charlie (base)	50.4	16.0	10.9	5.9	16.8	–	0	3.77	1.53	5.0
C2	Charlie (more inv.)	12.6	15.1	6.7	11.8	31.1	22.7	0	1.98	1.75	1.0
C3	Charlie (less inv.)	9.2	3.4	6.7	1.7	12.6	66.4	0	0.96	1.66	0
D1	Daisy (base)	23.5	16.0	10.9	16.0	33.6	–	0	2.80	1.61	3.0
D2	Daisy (more inv.)	10.1	9.2	4.2	9.2	17.6	49.6	0	1.36	1.76	1.0
D3	Daisy (less inv.)	10.9	10.1	5.9	15.1	18.5	39.5	0	1.61	1.76	1.0
E1	Eve (base)	47.1	20.2	10.1	7.6	14.3	–	0.8	3.76	1.50	4.0
E2	Eve (more inv.)	26.1	15.1	8.4	12.6	15.1	22.7	0	2.56	1.95	2.0
E3	Eve (less inv.)	2.5	0.8	9.2	6.7	12.6	68.1	0	0.70	1.23	0

Appendix B.
3S Scenarios

This appendix contains the scenarios we presented to the interviewees. Each participant had the option to ask follow-up questions clarifying technical misunderstandings. For the survey, we slightly modified the wordings to increase clarity, based on insights from the interviews. These versions are included in the full questionnaire in the supplementary material [36].

Alice: SQL Injection. Alice checks web servers for vulnerable database queries (e. g., via SQL injection). She uses a function to delay the database response (e. g., the MySQL function "SLEEP"). This allows her to verify whether the server is vulnerable or not.

Bob: Invalid HTTP Request. Bob sends a non-standard HTTP request to a web server. This causes the server to crash *unintentionally*. The result is that the server must now be restarted by the website operator's IT department.

Charlie: Insecure Direct Object Reference. Charlie changes his own user ID in a (1) GET or (2) POST request and to (1) receive and (2) change data from another user.

Daisy: Stored XSS. Daisy exploits a stored XSS (cross-site scripting) vulnerability to deliver its crafted code to potentially *all* users of a website. This code is executed on those users' end devices. It sends a confirmation message back to Daisy's server.

Eve: Path Traversal. Eve modifies a link to a web page to read information that is supposed to be confidential but can be publicly viewed due to server-side configuration issues (e. g., a path traversal).

Appendix C.
Survey Participants

The following table provides statistics on the surveyed operators' demographics and background. $^{\text{M}}$indicates multiple-choice questions for which response counts can sum up to more than 100 %. Percentage values are relative to the total number of survey responses (n = 119). For the type of security training received and the number of servers operated (indented lists), percentage values are relative to to the number of participants who indicated to have received security training / to operate servers.

TABLE 4. SURVEY PARTICIPANTS' DEMOGRAPHICS AND BACKGROUND.

	Demographics	n	%
Age	≤ 20	1	0.8
	21–30	11	9.2
	31–40	33	27.7
	41–50	38	31.9
	51–60	21	17.6
	≥ 61	5	4.2
	N/A	10	8.5
Gender$^{\text{M}}$	Woman	3	2.5
	Man	101	84.9
	Non-binary	0	0.0
	Self-described	3	2.5
	Prefer not to disclose	10	8.4
	N/A	5	4.2
Security Training$^{\text{M}}$	Yes	110	92.4
	University / school	58	48.7
	Employer training	42	35.3
	Certifications	35	29.4
	Other courses	35	29.4
	"Learning by doing"	92	77.3
	Professional network	49	41.2
	Personal network	50	42.0
	Self-taught	106	89.1
	Other	9	7.6
	N/A	10	8.4
	No	8	6.7
	Don't know	1	0.8

	Background	n	%
Responsibility for servers	Yes	118	99.2
	1	3	2.5
	2–5	28	23.5
	6–10	25	21.0
	11–50	40	33.6
	51–100	6	5.0
	101–1,000	9	7.6
	≥ 1,001	5	4.2
	N/A	3	2.5
	No	0	0.0
	Don't know	1	0.8
Main role in server operation	Security Engineer	9	7.6
	System Engineer	23	19.3
	Database Engineer	0	0.0
	Frontend Developer	1	0.8
	Backend Developer	6	5.0
	Full-stack Developer	25	21.0
	DevOps	19	16.0
	Management	14	11.8
	IT Consultant	6	5.0
	Penetration Tester	1	0.8
	Content Creator	2	1.7
	Other	13	10.9
Size of largest company	1	10	8.4
	2–5	6	5.0
	6–10	6	5.0
	11-50	22	18.5
	51–100	5	4.2
	101-1,000	24	20.2
	1,000–10,000	14	11.8
	≥ 10,000	11	9.2
	Don't know	21	17.6
Country of largest company	Germany	36	30.3
	USA	21	17.6
	Austria	5	4.2
	United Kingdom	5	4.2
	Belgium	4	3.4
	Canada	4	3.4
	Switzerland	4	3.4
	Other	36	30.3
	N/A	4	3.4

Appendix D.
Meta-Review

The following meta-review was prepared by the program committee for the 2024 IEEE Symposium on Security and Privacy (S&P) as part of the review process as detailed in the call for papers

D.1. Summary

This work takes a principled approach to constructing a normative framework for server-side scanning (3S) in the context of academic research, and legally under German law in particular. They perform a two-phase study: (1) interviews of legal scholars, research ethics committee members, and operators for qualitative understanding of ethical and legal concerns, and (2) a qualitative survey of system operators. The authors suggest that server-side scanning is generally well accepted by relevant parties and ultimately offer ethical guidelines to the community for future research.

D.2. Scientific Contributions

- Independent confirmation of important results with limited prior research
- Addresses a long known issue
- Provides a valuable step forward in an established field

D.3. Reasons for Acceptance

1) This paper independently confirms previously published ethical guidelines through a principled bottom-up approach based on stakeholder interviews/surveys.
2) This paper addresses the long known issue of server-side scanning, which has become increasingly common in the academic research community.
3) This work provides a valuable step forward in the ethics of security research by surfacing actionable results, some of which can generalize to other forms of Internet measurement.

D.4. Noteworthy Concerns

1) Shallow discussion of proposed trusted third-party (TTP) solution. The paper does not discuss the many obstacles to successfully implementing a TTP, thus misrepresenting its feasibility.