“Just a tool, until you stab someone with it”: Exploring Reddit Users’ Questions and Advice on the Legality of Port Scans

TEMIMA HRLE, MARY MILAD, JINGJIE LI, and DANIEL W. WOODS, University of Edinburgh, United Kingdom

Users, particularly amateurs, face uncertainties about technology law related to both interpretation and enforcement. This uncertainty can have a chilling effect on how users experiment with technology. However, little is known about the precise uncertainties that users face and what kind of advice is available. Our paper focuses on user questions and advice surrounding the legality of port scanning, a dual-purpose technique used in both defensive and offensive security. We identified and analyzed 414 pieces of advice, in response to questions about the legality of port scanning from 36 Reddit threads. We find that users ask two types of questions: (1) reactive questions in which they have scanned and are concerned by the consequences; and (2) proactive questions in which they ask about legality and seek ways to comply with the law. We found no consensus in the advice about legality or the likelihood of prosecution. In justifying advice, users deployed a range of anecdotes, analogies, and URLs. Subtle variations on the analogy between port scanning and physical building security are used to explain why it is both legal and illegal. Users also reason from individual cases, such as arguing prosecution is unlikely because the user had not personally been prosecuted or arguing prosecution is likely because Aaron Swartz was prosecuted. Finally, the most influential URL was a “Legal Issues” page maintained as part of an open-source project. We reflect on how these results can inform forum moderation and public-policy dissemination.

Additional Key Words and Phrases: cybersecurity law, offensive security, security and privacy discourse, content analysis

ACM Reference Format:
Temima Hrle, Mary Milad, Jingjie Li, and Daniel W. Woods. 2024. “Just a tool, until you stab someone with it”: Exploring Reddit Users’ Questions and Advice on the Legality of Port Scans. 1, 1 (August 2024), 22 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Security failures are often caused by systems not aligning with users’ everyday practices and understanding [1–4], in part due to uncertainty about how technology works [5–8]. However, uncertainty about security technology can also result from the law. Prior work has focused on how technology law impacts professional security practitioners [9–11] and the impact of legal risks on vulnerability researchers [12, 13]. The passage of the GDPR “made business more privacy aware” [9], and confidentiality rules impact how digital forensics practitioners communicate and document findings [10]. However, less is known about how the law impacts amateur users — such as individuals who engage with technology and security practices without formal training or professional experience. This matters because legal scholars have raised the possibility that broad technology laws risk criminalizing routine computer interactions [14].

We focus on one such routine activity, namely port scanning, which is a dual-use technique used to identify exposed network assets. The legality of port scanning has been a gray area since it emerged as a common practice in the

Authors’ address: Temima Hrle, temima.hrle@ed.ac.uk; Mary Milad, mary.milad@ed.ac.uk; Jingjie Li, jingjie.li@ed.ac.uk; Daniel W. Woods, daniel.woods@ed.ac.uk, University of Edinburgh, United Kingdom.

late 1990s [15]. Broad phrasing and technical jargon in cybersecurity legislation make it difficult for individuals to understand the legal boundaries of their actions [14]. While there is no law that explicitly addresses port scanning, several laws stipulate the illegality of accessing unauthorized material on a computer, including references to damage in the context of impairment or compromise of data and program integrity [16]. In the United States, port scanning could fall under the *Computer Fraud and Abuse Act* (CFAA), which criminalizes accessing a computer without authorization, or exceeding authorized access [16]. Similarly, Section 3(1) of the UK's *Computer Misuse Act* (CMA) considers any unauthorized act in relation to a computer an offense, especially if one causes a computer to perform a function with the intent of accessing programs or data held on it [17, 18].

To understand legal uncertainty about port scanning from a user perspective, we conducted a content analysis of online forum posts. Prior usable security research has studied online forums to identify user uncertainties about security and privacy technologies [19–21]. Online forums allow us to study a rare community who are sufficiently interested in cybersecurity to read and discuss the legality issues of security online. Further, our observational field study allows us to study user responses prompted by real-world uncertainties, rather than an artificial prompt created by researchers. We collected and analyzed 467 comments found across 36 threads related to the legality of port scanning from Reddit, the most popular English-speaking online forum. Our content analysis analyzed the following research questions:

- **RQ1** What uncertainties do users have about the legality of port scanning?
- **RQ2** What advice do users provide about the legality of port scanning?
- **RQ3** How do users justify advice?

Findings. Most threads on the legality of port scanning were posted after mid 2020. Our research indicates that users provide contradictory advice on both the legality of port scanning and the likelihood of prosecution upon scanning. Most users view legality as contingent on some other factor, most commonly permission. Users predominantly cite technical websites, as opposed to legislation, in communicating the law to others. Users frequently used analogies to explain why they believe scanning to be legal or illegal, frequently likening it to physical security.

Concerns about potential legal repercussions are prevalent, with many users expressing fear of prosecution or legal trouble due to unclear legal guidelines and inconsistent enforcement across different jurisdictions. This ambiguity leads to reliance on informal community-driven anecdotes rather than formal legal advice. This underscores the need for accessible legal education and advice tailored to amateurs, particularly those who are starting to experiment with real-world systems.

2 BACKGROUND

Port scanning is a network analysis technique widely used by both defenders and attackers. Simply stated, port scans attempt to make connections with different ports on a target server to identify which services it is running. Servers either respond with a packet indicating that the target server is listening or closed, or the server sends no response [22]. An open port can be used to communicate with the server. This makes port scanning dual-use in that the results of a port scan can be used by defenders to monitor a network and identify exposed assets, but the same information can be used by attackers as reconnaissance before an attack [23].

This potential for malicious activity necessitated novel legal frameworks to address the emerging challenges of computer crimes, which were initially unregulated. The United States responded with the Computer Fraud and Abuse Act (CFAA) in 1986, an amendment to the Comprehensive Crime Control Act of 1984 [24]. In parallel, the United Kingdom established the Computer Misuse Act (CMA) in 1990, aiming to protect online privacy by criminalizing unauthorized computer access [25]. Prior to the CFAA, the legal landscape for addressing unauthorized computer access was vague, primarily relying on laws designed for physical trespass, fraud, or theft [26]. The original 1984 bill that led to the CFAA targeted only government and financial institution computers, but was deemed too narrow. Subsequent amendments expanded its scope to include any "protected computer," now defined broadly enough to cover any computer involved in interstate or foreign commerce [27].

Despite these expansions, the term "without authorization" remains ambiguous. This leads to varied interpretations and applications in court, particularly in cases involving gray hat hackers, whose intentions may not be clearly malicious [28]. The broad language of the CFAA has been criticized for potentially criminalizing benign computer interactions [14]. This is illustrated by the controversial United States v. Swartz case, where Aaron Swartz faced severe penalties under a broad interpretation of the law [29]. Many years later when the Supreme Court finally ruled on a CFAA case, the court adopted a narrow view of what constitutes "unauthorized access" [30]. This ruling suggests that actions like port scanning, which do not involve accessing or altering data, might not meet the threshold of unauthorized access under the CFAA, provided there is no intent to further access or manipulate protected systems [31].

3 RELATED WORK

In exploring how the legality of portscanning is perceived by users, our article contributes to research into the legality of offensive security research (Section 3.1) and studies on online security and privacy discourse (Section 3.2).

3.1 Legality of Offensive Security

The legality of offensive cyber operations has been studied from various perspectives. International law researchers have considered questions like what constitutes cyber war and how the laws of armed conflict apply to cyber operations [32]. However, this body of law applies to states rather than end-users. For users, criminal and contract law is more relevant. At one end of the spectrum, criminologists study illegal activities. Research goals include characterizing criminals [33, 34], understanding crimes [35–38], and studying market places [39–41]. Such crimes lead to widespread financial losses [42, 43]. This motivates enforcement interventions that aim to deter and disrupt cyber crime [44, 45].

At the other end of the spectrum of legitimacy, efforts to disrupt and prosecute cyber crime can impact defenders. Hantke et al. [46] explored the ethics and legality of client-side scanning research. Even though a majority of operators (57%) had a positive view of such research, a review of German law revealed ambiguity due to the lack of judicial decisions [47]. This becomes even more uncertain when considering multiple jurisdictions, which motivates legislative action to clarify the law [46]. Given that legal experts were uncertain about the legality of non-malicious scanning [46], end-users with no legal expertise should at best be uncertain and at worst mistaken. This motivates our research questions into users' legal questions (**RQ1**), advice (**RQ2**) and evidence (**RQ3**). To find answers, we adopt methodologies from usable security [48, 49], which we survey in the next section.

3.2 Online Security and Privacy Discourse

Online forums are a common data source to understand how users talk about security and privacy [50]. Forums provide a space for users focusing on the same technology to share ideas and resolve problems, by discussing with others in their domain. The benefit of this is transparent advice by those who face the same issues and ask the same questions. Researchers have studied forums like Reddit [19], Quora [51], and Stack Overflow [20]. To date, most of the research has focused on understanding developer practices [19, 52], user privacy perspectives [21], and security and privacy issues [53, 54] from a technical perspective. Such work has found that developers primarily discuss privacy concerns in response to external events [19], that developers encounter challenges in crafting privacy policies [20], and that users' security and privacy considerations evolve based on adoption phases and product factors [21].

These studies uncover themes like the trade-offs users are willing to make between convenience and privacy, the gap between users' perceived and actual security risks, and privacy awareness of users. The previous research on how mental models relate to port scanning is technical by nature [55], and show that there is a research gap in understanding the legal perception of offensive security by users. This paper bridges this gap by providing insight on both perceived legality and technical advice provided by users in circumventing perceived illegality. This ties in with prior work, in providing a comprehensive outline of the sources users rely on in providing advice, and their justification.

4 METHODOLOGY

We focused on users' discussion on Reddit, a major online platform actively used by researchers to study different topics in security and privacy. Our exploratory searches confirmed that most legal questions were posted on Reddit, whereas Quora and Stack Overflow were used for technical questions.

We conducted five rounds of searches, using a range of search terms (as shown in Appendix A, Data Collection) and both Reddit and Google. We terminated each search when additional results no longer yielded relevant threads (see Figure 7 in the Appendix). We iteratively analysed the collected data, and developed a code book, and we verified its reliability by calculating Cohen's Kappa ($0.722, p < 0.05$) using a second coder. Themes were developed inductively, and focused on various stances users had in relation to legality, prosecution and technical advice. The rest of this section explains our methodology in more detail.

4.1 Data Collection

To mitigate biases in our search, we used various search terms and search platforms (see Figure 7) stopping when additional search returned no relevant results. We carefully designed our search terms and Boolean logic through our exploratory searches of discussion threads (we define a *thread* to be the collection of the original post and all subsequent discussion comments). We observed that shorter searches were more efficient, whereas longer searches were more likely to yield additional relevant searches further down the list. Upon finalizing the search terms, we used the same Boolean searches on both Google and Reddit to gather Reddit threads in the order they appeared.

To determine relevance, we applied the following *funnel*:

(1) Initial Screening: We read the original post, and excluded those that did not mention network or port scans.
(2) Technical Exclusion: We excluded threads that were primarily about setting up or troubleshooting scans.
(3) Legal Relevance: We read the entirety of the remaining threads, excluding any that did not discuss legal aspects.

Threads that merely touched on port scanning tangentially or discussed it within the context of unrelated technology topics were excluded as irrelevant to our research focus. All threads written in languages other than English were also excluded to maintain consistency in coding.

Reddit's internal search, which yielded 26 unique threads, was generally more efficient and comprehensive than the Google search, which yielded 22 relevant threads. To sample the whole population one must use both, given the majority (67%) of threads were only found by searching one of those platforms. After removing any duplicates, we identified a total of 36 unique threads and analyzed all 467 comments found within these threads.

Finally, one must be careful with how the platforms deploy Boolean logic. We compared the results from ("law" OR "legality") AND ("port scan" OR "network scan") with the results from four smaller search phrases: ("law" AND "port scan"); ("law" AND "network scan"); ("legality" AND "port scan") and ("legality" AND "network scan"). One of the shorter Boolean searches on Reddit identified a relevant thread that the long search missed, which was added to the dataset. This suggests long Boolean search terms can miss relevant posts.

4.2 Data Analysis

We adopted a qualitative approach by open coding. We first read the entire post to understand the entirety of the users' discussion. We then iteratively built the codebook around three broad themes themes related to our research questions—type of questions, type of advice, and background information. The research team discussed and refined the code book as more refined subthemes and subsubthemes were added, revisiting data and classifications each time.

The primary researcher broke the text down into units of analysis, which captured a distinct meaning with relation to our themes. Technical terms were often small units consisting of a phrase, whereas the units corresponding to an analogy might comprise multiple sentences. When writing the article, we extracted quotes from the raw data, which sometimes combined multiple units of analysis.

To probe reliability of the code book, two coders independently coded 161 units from multiple threads, including both the OP and comments by users. Both researchers independently coded the units. We calculated agreement via Cohen's Kappa, finding a score of 0.722, which suggests the agreement between the two raters is consistently reliable.

After this test, the primary researcher coded the rest of the threads using the code book. All 36 Reddit threads were coded using the code book and were listed in arbitrary order. Most units were coded under only one code, with no overlap with other codes. An exception to this was coding whether a jurisdiction was specified in the provision of legal advice, under the legal advice subcategory, which was coded in addition to another code in this category — what users thought about the legal status of port scanning.

4.3 Limitations

Our study had various limitations. First, the number of questions we analyzed in our study is limited, as topics about legality of port scanning were less common compared to other security issues that general users or developers encounter daily. This is possibly because people are less willing to openly discuss practices in a legal gray area. Additionally, we only included threads in English to avoid potential inaccuracies and misinterpretations that could arise from translation. This decision was made to ensure the reliability and consistency of our analysis, and prevent nuances from being lost in translation.

Nevertheless, the sample size is comparable to prior research using online forums, such as privacy-related advice [52, 56]. In addition, we identified active advice proposed by users on Reddit and observed data saturation in our coding. Secondly, we analyzed anonymous discussions on Reddit, which does not allow us to know users' demographics and verify if their self-disclosure is faithful – we refrain from doing so given the sensitivity of this topic. As such, our analysis emphasizes qualitative results instead of quantitative findings. Nevertheless, we report prevalence of codes that characterize our dataset.

4.4 Ethics

This research project received ethical approval from our institution. It relies on data shared publicly by users on Reddit, as done with prior work [57–59]. We recognize the importance of protecting the privacy and anonymity of users. To this end, we have taken steps to ensure that no sensitive information is disclosed that could put users at risk.

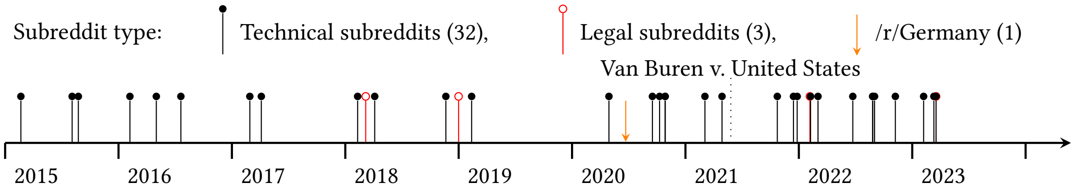

Fig. 1. Timepoints when the questions were posted and in which type of subreddit. Van Buren v. United States was a Supreme Court case dealing with the Computer Fraud and Abuse Act (CFAA).

Some articles alter direct quotes in order to minimise the likelihood of de-anonymising users [60]. Before including a quote in the paper, we evaluated the risk for the user, rewording quotes when we believed there was a high risk. For example, we altered quotes where a user had conducted a scan and doubted if it was legal. However, we did not alter quotes where a user asked if a hypothetical scan would be illegal or when comments provided legal advice without personal anecdotes. This provides a balance between user privacy and conveying the exact sentiment sought to be expressed by users.

5 RESULTS

Section 5.1 identifies user questions about the legality of port scanning — specifically whether questions are proactive or reactive (**RQ1**). Section 5.2 describes the 414 pieces of advice that were provided in response (**RQ2**), which includes advice on legality, technical steps, and the likelihood of prosecution. Section 5.3 answers how the users justified their advice (**RQ3**), which explores the use of URLs, analogies, and anecdotes to support recommendations.

5.1 Questions about Port Scanning Legality (RQ1)

We identified 36 threads, including 467 comments, where people asked questions about legality of port scanning. Figure 1 shows these threads were made between early 2015 and when we stopped collecting data in October 2023. The frequency of threads actually increased over time, with the majority coming after 2020. This is likely due to a combination of a growing Reddit user base and also increasing interest in cybersecurity as a topic. This could also be as a result of the Van Buren Supreme Court case and renewed interest in the application of the CFAA in offensive security as a result of government bodies, such as the Department of Justice, stating that good faith hackers will no longer be prosecuted as of May 2022 [61]. The vast majority of threads were found in technical subreddits, such as *r/hacking* in which 22% of threads were found. Other threads were found in offensive security subreddits with a focus on learning skills (4 threads in *r/HowtoHack* and 2 in *r/HowtoHack*), ethics (2 threads in *r/ethicalhacking*) and specific tools (3 threads in *r/nmap*). A small number of threads were found in defensive security subreddits, 3 in each of *r/cybersecurity* and *r/AskNetsec*. A handful were found specifically in legal subreddits such as *r/AskLawyers*, *r/legaladvice* and *r/LegalAdviceEurope* . The threads in our dataset can be classified into two broad categories according to whether the questions are proactive or reactive about port scanning activities.

(Calm) proactive questions about future activities. The first category is proactive threads about the legality of future activities. Proactive questions comprised 64% of our dataset, and included questions like:

Q1: *"May I perform a network port scan on someone's gateway (WAN side) with their permission?"*

Q2: *"How illegal is doing OSINT? Or passive reconnaissance? And where is active reconnaissance on this*

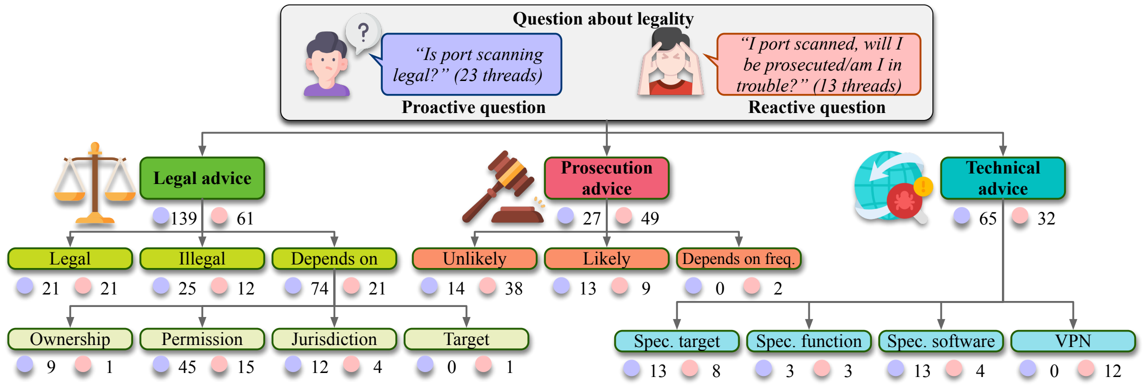

Fig. 2. The types of advice provided in response to the two types of legal questions – proactive and reactive. The numbers adjacent to purple and red dots represent the occurrence of advice for proactive questions and reactive questions respectively.

> *spectrum? Even identifying targets and vulnerabilities without acting on them?”*
>
> Q3: *“...[Via tutorials, I found] warnings that some things on nmap are illegal ... So I want greater clarity on what should be avoided...and what punishments can be expected for doing the wrong thing.”*

Proactive threads tended to be calmer in tone. The majority of threads simply asked whether they could perform scans, without specifying whether the target was their own or another person's network. Some users wanted to distance themselves from malicious actors by clarifying why they wanted to scan, which included: trying to find vulnerabilities in the user's own router; checking whether a third person is connecting to the user's network; and participating in a bug bounty program. A key theme surrounding such questions is where the line in terms of legality is drawn.

Fear in the reactive questions. The second category is reactive threads after the OP had conducted some form of scanning. These threads represent 36% of our dataset, with specific questions like:

> Q4: *“i used an nmap script ... What should I expect? How illegal is this?”*
>
> Q5: *“I stupidly put a public ip in an NMap command ... Am I in trouble?”*
>
> Q6: *"I tried to scan the website but accidentally scanned the provider's subnet ... can anyone provide advice”*

Reactive threads typically involved fear and regret, as users worried about potential legal consequences, such as OP in Q6 who explains *“now I'm freaking out”*. Two users (threads 4 and 27) asked whether a person would get prosecuted for port scanning, another user (thread 14) asked about the prevalence of prosecution, and one user asked what steps could be taken to mitigate issues with their ISP who had detected their scanning (thread 35).

Law and jurisdiction not specified in questions. The majority of legal questions (78%) did not specify which law or even jurisdiction they sought advice on. A further 15% specifically mentioned legality in the US. The remaining questions included a question about Germany, one about the Netherlands, and one question titled *“are port scans illegal? (Other than USA)”* (thread 7). For unspecified questions, the default assumption is likely US law given most of the subreddits were not country specific, apart from one thread in *r/Germany*, and most of Reddit's user base is located in the US [62].

5.2 Advice on Port Scanning Legality (RQ2)

The three main forms of advice—legal, prosecution and technical—can be seen in Figure 2. Legal advice focused on compliance with the law, whereas prosecution advice focused on the likelihood and risk of being pursued by law enforcement. Technical advice typically concerned how and what to scan, including technical means on how to evade legal issues. The rest of this subsection unpacks the specific advice.

5.2.1 Legal Advice. We identified 200 pieces of legal advice grouped into four sub-codes. The most simple legal advice declared port scanning is either legal or illegal without caveats. Highlighting the ambiguity in the law, we found exactly the same amount of advice (21% of legal advice) opining in either direction. However, users responding to reactive threads were more likely to say it was legal, possibly to calm the OP's worries.

Legality Depending on Circumstances. Advice that legality depends on the circumstances surrounding the scan was the most common form of legal advice. Many users advised that port scanning is either illegal without express permission (63% of dependent advice), or depends on the jurisdiction from which one is scanning (17% of dependent advice). Thus, obtaining authorization was the most common form of actionable legal advice from a user's perspective (30% of all legal advice). Other factors users brought up that impact the legality of scanning included: what is being scanned; the intention; the frequency of the scan; and the impact of the scan. All users agreed that port scanning falls within legal confines when conditions like obtaining permission are satisfied. The details of how permission is to be sought (e.g., whether verbally or written) was not discussed. There was no mention of contracts or other formal agreements for offensive/defensive security purposes.

A minority of users stated that the legality of a scan hinges on its impact. This implies that users engaged in port scanning are aware of the potential outcomes of their scans, perhaps insinuating that the results of such scans are predictable in some cases. However, some users suggested it is more complex:

> *"Non-customary traffic MAY be illegal if it causes adverse effects on the target system and that was the intent, which is more slippery than duck shit. A simple port scan might cripple a Raspberry Pi or equivalent if there are other factors involved. With caveats: no, it's not illegal."*

This indicates that users believe the legality of scanning can be determined retrospectively, based on the actual outcome and impact of the scan. Other users linked legality to the target and jurisdiction in which the scan is being performed including the jurisdiction of the scanee. Examples of this are:

> *"First off, laws differentiate wildly across different jurisdictions. What kind of trouble you may face depends on where you live and where your target is. It's both explicitly legal and not legal in different places."* and *"Depends on the law in your country. I don't know how it works if you scan a server that's in a country where port scanning is illegal"*

Uncertainties about the impact of scanee's jurisdiction. Most users were vague about whether the scanner or scanee's jurisdiction is more important in determining legality, or how one impacts the other. This is most likely due to user's uncertainty in providing cross-jurisdictional advice, especially when in most cases users do not provide which jurisdiction they themselves are located in.

Qualification of Providing Legal Advice. Most advice was provided without a qualifier or disclaimer. However, a minority of users who provided advice made it clear they were not qualified, such as the evocative example: *"So in my completely not a lawyer opinion, and you're basically getting a magic eight ball's reliability here: [advice]."* Other users implied that technical expertise could be a substitute for legal training: *"Not a lawyer (nor Dutch) but I do have extensive experience working in Network Security."* Notably, all qualifiers are related to legality or prosecution, and not to technical advice. This suggests there is little confusion on the technical aspects of port scanning, such as how one could circumvent legal issues through technical means.

Moderators in predominantly technologically centered forums may not be aware of the risks surrounding poor legal advice. For example, *r/AskLawyers* includes the blanket statement *"...we cannot offer legal advice here for a number of reasons"*, followed by *"Also, it's not a good idea to solicit legal advice from random strangers online, despite what you may find elsewhere on Reddit. We do not know all of the facts of your case, and are likely not licensed in the jurisdiction that you're in. A real attorney worth their salt will not comment on your specific legal predicament on an anonymous forum."* This complements a common user sentiment in our sample of advice to seek legal advice: *"the correct answer almost always includes consulting with an attorney"*.

5.2.2 Prosecution. Instead of advising on whether scanning was legal, some users advised on the risk of legal consequences. Across the 76 pieces of prosecution advice, users focused on the likelihood of facing prosecution rather than the severity of the punishment itself. The few pieces of advice about punishment noted how severe it could be. For example, one user noted that the CFAA: *"can carry up to 10 year prison sentence and a hefty fine (the fine is technically unlimited)"*.

Split Opinions on Likelihood of Prosecution. When advice addressed the likelihood of prosecution, the majority suggested it was unlikely (68%). Prosecution advice was almost twice as common in reactive threads where a user had already conducted a port scan, as opposed to proactive threads inquiring about the general legality of port scanning. In reactive threads, only 18% of the advice said that prosecution is likely, as opposed to 48% in proactive threads. It could be that users suggest there is a higher likelihood of prosecution in proactive threads in order to deter users who are uncertain about the legality of their conduct seeking to perform scans.

A minority (30%) of prosecution advice stated that prosecution was a likely outcome. Such advice predominantly included specific cases over other forms of evidence, such as legislation or anecdotes to justify their view. Examples of excerpts relying on precedent are as follows:

> *"Andrew 'weev' Auernheimer. "...he found a major bug that revealed the identies and personal information of AT&T iPad subscribers, gave that list to Gawker, was convicted under the Computer Fraud and Abuse Act (CFAA), and spent 15 months in prison before having his case overturned."* and *"In June 2003, an Israeli, Avi Mizrahi, was accused by the Israeli authorities of the offense of attempting the unauthorized access of computer material."*

Prosecution advice, whether in advising its legality or illegality, was infrequently coupled with supporting evidence that will be explained in **RQ3**. This is potentially because proving a negative is difficult. One user advised that:

> *"I don't think anyone has ever been prosecuted just for running an nmap scan on a remote host. I also suspect any such scan would be lost in the sea of noise which is the Internet, provided it's not to aggressive. But, you could always end up hitting that one host which is watching like a hawk and knows someone who decides that your ass looks ripe for the taking."*

Threads of this nature often attempted to reassure users that their actions would likely go unpunished, but that it depended on the actions of the entity who was scanned. Broadly speaking, the community relies on anecdotal evidence and shared beliefs when advising on the likelihood of prosecution.

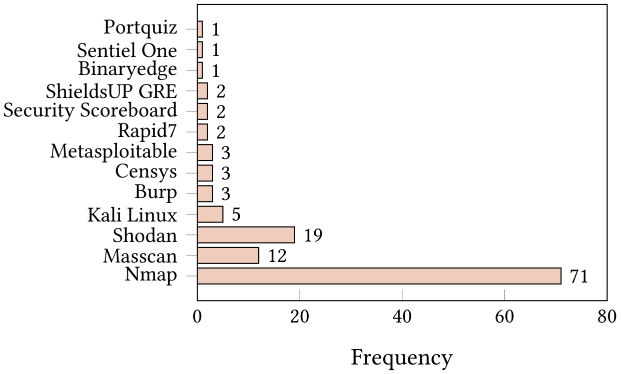

Fig. 3. Frequency of port scanners mentioned in Reddit threads

5.2.3 Technical. Despite the original questions typically focusing on the legality of port scanning, we identified 97 pieces of technical advice across 25 threads. This means technical advice is more common (26% of all advice) than prosecution advice (20% of all advice), possibly because the threads were mainly in technical subreddits (see Figure 1). The technical advice ranged from directing the OP towards targets that are legal to scan through to advanced configuration details to evade detection. Technical advice is more diverse, which can be seen in the share (51%) of technical advice that we classified as miscellaneous (see Figure 2). This is likely because it is tailored to a specific user's situation.

Choosing an Appropriate Target. Most technical advice focused on choosing an appropriate target to scan, such as scanning dedicated websites or setting up a home lab. 24% of users advised to find scanning targets via HackTheBox (39%), TryHackMe (22%), HackerOne (13%) and Metasploitable (8.7%). Other users (14%) advised setting up dedicated infrastructure: *"Might be time to build yourself a small little lab. Don't need much, a managed switch would be ideal so you could make some different "vlans"."* Target-based advice was justified by notions of avoiding harm, such as *"this way you can find exploits while not harming "anyone business"."*

Recommending Scanning Software. Some users (16%) also advised on which scanning software to use, with some users emphasizing the educational side with imperatives like *"never stop learning"*. Figure 3 shows that Nmap was mentioned over three times as frequently as the next most popular solution, Shodan, and followed by Masscan. This advice implies that using known tools provides safeguards, which mirrors how some users point to the popularity of Nmap and Shodan when providing legal advice. Some users provide advice on how to configure the tools: *"As an exercise, try playing with nmap's -T parameter to avoid being blocked. You'll find different sites have different sensitivities. With most the block will eventually time out."* This kind of advice was most commonly found in subreddits focused on offensive security like *r/hacking* (43%) and *r/howtohack* (29%). Only 2% of users recommended using a VPN as part of the technical advice provided.

5.3 Supporting Evidence and Reasoning in Advice (RQ3)

This subsection explores how participants use supporting evidence and reasoning like analogies, anecdotes, and URLs when providing advice. The amount and balance of each type of supporting evidence depends heavily on the type of advice provided, which can be seen in Figure 4. Technical advice is exclusively supported by URLs, typically linking to

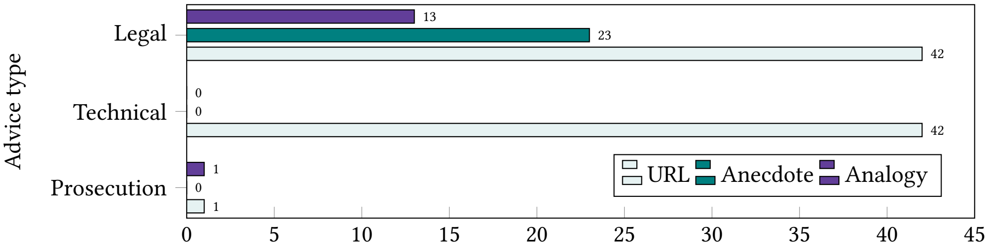

Fig. 4. Distribution of evidence types provided by users across advice categories.

a tool or website. Legal advice includes a mixture of supporting evidence, most commonly relying on URLs. Meanwhile, prosecution advice rarely includes any supporting evidence.

5.3.1 Analogies. Analogies were mostly used to probe intuitions about the legality of port scanning. The most common analogies compared port scanning to different ways of probing the security of buildings. Only one analogy addressed the likelihood of prosecution. No analogies were used in providing technical advice or details.

Comparisons to Non-Invasive and Passive Activities in the Physical Space. Users advising that port scanning was legal used analogies to imply that port scanning is a non-invasive activity like *"walking down a street to see which windows have curtains"* or *"checking to see if something is on or off"*. Others falsely implied that port scanning was a passive activity *"akin to overhearing a conversation in a public place"*. Another analogy supporting legality was *"checking whether a business was open"*, which emphasizes the commercial aspect of the Internet.

Describing Malicious Intent in Analogies. In contrast, 41% of analogies were used to justify why port scanning was illegal. These typically focused on malicious intent. For example, one analogy to home security compared port scanning to *"jiggling a doorknob to see whether the door is open"*. Another analogy emphasized how port scanning could be done at scale by *"going to a business's building and jiggling every door handle to see which ones are locked and which ones aren't"*. Another user used the analogy of speeding to say that although port scanning was illegal, prosecution was unlikely:

> *"it seems to me that it's like speeding on the highway. Thousands of people do it daily for a various reasons - passing a semi, wanted to get home faster, trying out your car, etc. Some of them are being caught by a radar. Some of them get ticket/go to jail (depends if you're speeding too much/drunk, etc.). Chances of having trouble, if you going over the limit only once for very short amount of time, are indeed very small."*

Adding Nuance. Others used analogies to bring nuance to the discussion. One user invoked the idea of dual-use technology by saying port scanning *"is just a tool, until you stab someone with it"*. This shifts the question of legality away from port scanning, and towards the intent and consequences of the scanner.

It is perhaps most notable that users who deploy the same analogy to private homes can arrive at different conclusions on legality. This varies based on whether port scanning is like *"ringing a doorbell"* or *"jiggling a doorknob"*, arguably both are true depending on the protocol. This suggests analogies could cause problems if relied upon to determine legality. Indeed, one user questioned the value of analogies given the immature state of technology law:

> *"none of these analogies are very appropriate, because the laws about what we can do with houses are a lot more developed than the laws about what we can do with computers."*

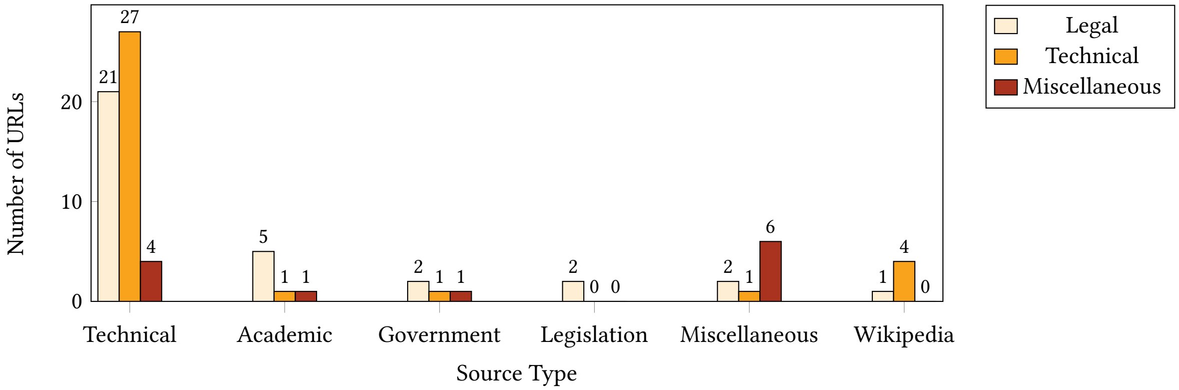

Fig. 5. Types of sources and the type of advice users relied on them for.

5.3.2 Anecdotes. Users include anecdotes to support advice, typically by indicating what is common practice in the security community. Anecdotes can be divided into 68 personal stories about the user (I statements), and 38 community anecdotes about third parties like colleagues and companies (they statements). Sometimes anecdotes were provided without any advice at all.

Personal Experiences of Port Scanning. Many users included personal anecdotes that indicated they regularly port scanned without legal issues. For example, the personal anecdotes *"I port scan machines all the time"* and the more colorful *"I've been down to the docks plenty of times"* were respectively followed by *"and i didn't get into trouble"* and *"and haven't been arrested yet"*. Such anecdotes cast port scanning as a common activity, typically to calm users who are concerned about prosecution. A different approach is to include an extreme personal story, which makes the OP's activity appear benign by comparison. For example, one user said *"I did a port scan of whitehouse.gov once and didn't get into trouble"*. Another described a story from *"back in '97"* in which the user *"successfully took out every printer in the company and the ERP system hosted on an AS400, just from doing a basic port scan"*.

Companies' Regular Scanning Activities. Community anecdotes often pointed out that some companies exist to run scans and sell the results. Users sometimes differentiate between Nmap, a tool which scans specific networks users have selected, and companies like Shodan, a search engine that continuously scans the Internet and collects data about devices, which are searchable in a database. One user made this distinction, saying *"It appears that part of Shodan's strategy is to provide carefully limited service - e.g., the results of portscans."* followed by prosecution advice ‘*"Shodan has existed for long enough in the US (and its operators seem to have enough credibility with white-hat/institutions) that it's unlikely to face prosecution now"*.

5.3.3 Information Sources. Users included links to 78 URLs, the most common form of supporting evidence. However, there was no single source of authority given there were 53 unique URLs and 48 unique domains across the 78 links. The most commonly shared links were Nmap's legality page (shared 17 times), Cornell University's page that replicates the CFAA statute (shared 3 times), and the Stanford Internet Research Data Repository (shared twice). All of the Wikipedia pages provided information about technical topics, such as Wikipedia's "Port scanner", "Shodan", and "List of assigned /8 IPv4 address blocks" pages, with the exception of the "Computer Fraud and Abuse Act" page. Legislation most

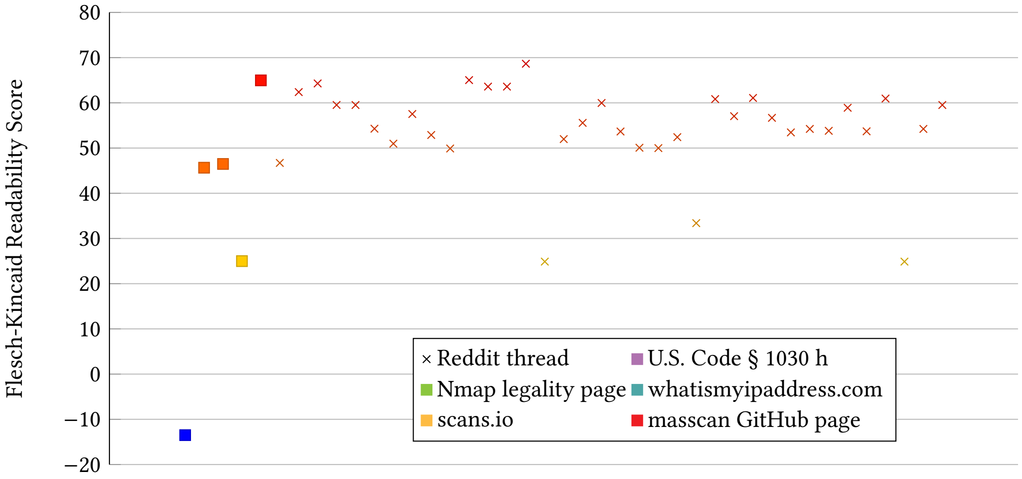

Fig. 6. Readability scores of Reddit threads and most commonly referenced webpages on port scanning legality. Higher scores indicate better readability. The Reddit threads have been ordered from earliest to latest according to the date of the original post.

commonly referenced in user provided URLs are the CFAA and to a lesser extent the UK's CMA. However, these threads did not discuss legislation in detail.

Reliance on Technical Sources. Figure 5 shows that links to technical websites were most common across all source types, except miscellaneous information from miscellaneous source types. This holds true even when the URL supports legal advice. The reliance on technical information sources could result from most threads being posted in technical subreddits, where users are familiar with technical websites as opposed to laws. The academic URLs were a mix of papers that clarify the law as opposed to specifically outlining the legality of scanning. For example, users cite a USENIX Security Symposium paper [63] that focuses on the technical aspects of Internet-wide scanning, analyzing data to identify patterns, tools used for scanning, and the types of protocols targeted. The website frequently referenced from Cornell University provides legal context by detailing 18 U.S. Code § 1030 from the CFAA.

Given how often it was referenced, we applied our entire code book to Nmap's legality page and found that the majority of the webpage falls under the miscellaneous legal category [64]. It evades any blanket statements that state port scanning is either entirely legal or illegal. Instead, Nmap addresses the scenario of an ISP getting complaints based on users' actions and the possibility of getting banned and having to switch ISPs. The page emphasizes that port scanning cases are rare, but that laws differ across jurisdictions, and provides case law from various countries where civil cases were dismissed, which are coded by the authors under the legal advice category. Towards the end, the webpage advises users to use specific functions to ensure scanning is light so as to not crash a device.

Readability Challenges in Understanding Legal and Technical Information. We compared readability scores of the Nmap legality page and 18 U.S. Code § 1030 of the CFAA, in addition to URLs users shared across Reddit threads two or more times. We used the Flesch Reading Ease Score readability test. This test measures the readability of a text based on the average length of sentences and the average number of syllables per word. Scores of around 70 are easily understood by 13 year olds, meanwhile scores below 50 are considered college level [65]. Negative scores are possible for dense technical content.

Figure 6 illustrates the readability of each individual Reddit thread, as well as common information sources cited by users. Nmap's legality page has a readability score of 45.65, indicating the text is moderately difficult to read. This is typical for technical topics such as legal discussions on software tools, where precise language is necessary to accurately convey information. Most Reddit threads fell within a similar, slightly higher readability range, with the exception of one outlier with lower readability score, due to poor wording and presentation in the thread. In contrast, 18 U.S. Code § 1030 has a score of -13.50, indicating the text is very difficult to understand for the average reader.

The contrast between 18 U.S. Code § 1030 and other sources, including the entirety of individual Reddit threads illustrates the contrast in readability and inaccessibility of legislation to users. The reason for the low readability score of 18 U.S. Code § 1030 highlights an important issue: legal documents, particularly those dealing with topics like tech law, are often significantly more challenging for the average reader to understand compared to more informal types of communication, such as Reddit threads. This can be attributed to inaccessible and complex language used in legislation, that users in online forums make up for with discussions on the law.

6 DISCUSSION

Section 6.1 first unpacks the implications for users, technology vendors, forum moderators, and policy-makers. Section 6.2 makes recommendations. Section 6.3 outlines next steps for future work.

6.1 Implications

The Legality of Port Scanning. A naive conclusion from our study is that online forums are an unreliable source of advice on technology law. This could be justified by pointing out that roughly as much advice saying port scanning is legal as there is advice saying it is illegal (see Figure 2). This interpretation would mirror prior work that found online forums contain insecure code snippets that are often copied into code bases [66–68].

However, the story is more complicated. Whereas a code snippet can be classified into a binary of secure vs insecure, it is far more difficult to classify legal advice into accurate vs inaccurate, particularly when it concerns broad statues like the CFAA. Ambiguity results from incomplete descriptions of the facts—most court cases involve hundreds of pages of documents, whereas a forum thread contains a few sentences—and the lack of precedent with regards to the CFAA. For these reasons, advice that avoids the legal vs illegal binary is arguably more justified. For example, the majority of legal advice stated that legality "depends on" factors like jurisdiction and permission, which are both accurate statements [30]. Most of the technical advice tried to direct the OP towards "safe spaces" like dedicated machines and home networks, and rarely recommended methods to evade detection like using a VPN. This suggests most Reddit advice guides users towards lawful activity.

Another function of Reddit advice is calming users who are worried about the consequences of scanning. Advice in response to such questions was more likely to say port scanning was legal and/or unlikely to be prosecuted (see Figure 2). The prosecution angle is perhaps a natural response given the CFAA is so broad that it is hard to say a given activity is unambiguously legal [14], instead various users imply that enforcement is rare, often pointing to examples of tools and companies that exist to scan.

How Users Reason about Technology Law. Given that port scanning is just one security technique among many, our findings related to how users came to their conclusions have more general implications. For example, policy-makers may want to raise awareness of the law and new enforcement guidelines. It is therefore relevant that users relied on technical sources when providing evidence about the law (see Figure 4). In particular, one legal page maintained as part of the open-source Nmap project was cited five times as often as the CFAA statute. To take advantage of this, one potential innovation would be for policy-makers to work with vendors and project maintainers to embed legal guidelines in the documentation of projects.

Turning to another type of supporting evidence, the reliance on analogies when explaining legal advice has implications for communication of the law. Physical security is the most common form of analogy, which is interesting given the CFAA was introduced because laws designed for physical trespass could not be applied to computer security [26]. The core problem is that subtle shifts in the analogy to physical security, such as whether port scanning counts as ringing the doorbell or jiggling a door knob, can lead to different conclusions in legality. Lawyers could potentially exploit these differences in trying to influence juries, who are unlikely to understand the technical details. Given the jury's judgment could lead to criminal charges, this setting raises the stakes of mental models of technology, a common topic for HCI research [69].

6.2 Recommendations

Users. Given the uncertainty in the legal advice, it is tempting to prevent users seeking advice on online forums, such as by recommending moderation policies that remove threads about legal issues. However, such a policy would eliminate a popular and readable information source. It is far from certain that users would find an alternative, let alone a more accurate one. Instead we recommend a strategy that improves both the moderation of online forums *and* the accuracy and usability of the information sources that underpin advice.

Forums. Forum moderators should seek to maintain the role of online forums in triage and sign-posting. In particular, forums are valuable in addressing easy questions where there is little uncertainty. One such example is threads in which the user wants to gain experience in port scanning. It would be absurd for users to pay for a lawyer to provide this advice, given that users already provide appropriate suggestions like using a dedicated sites like HackTheBox. However, moderation is needed for threads where scanning has already taken place. Arguably, these threads should be in dedicated legal forums, given those communities have already debated the right balance between initial advice on forums and recommendations to seek professional advice. Technical subreddits should either create a similar policy or direct users to subreddits where a policy has been created.

Software and Tool Providers. Providers of port scanning tools can offer support and guidance, as exemplified by the Nmap Legal Issues page [70]. Ideally, this would involve offering clear, user-friendly information on how to use their tools legally and securely. This can be achieved through improved user interfaces, comprehensive help/support documentation, and user education initiatives [71]. Information could include links and summaries of prior cases, best practices for obtaining permission before scanning, and explicit warnings about the potential legal consequences of unauthorized scanning. In doing so, providers must avoid providing legal advice as this is a regulated activity, and should be done by a professional. Legal ethics stipulate that lawyers conduct themselves with integrity and honesty in the provision of legal services, as it is the core tenet of the profession [72]. Providers should instead focus on directing users to appropriate legal resources and professionals.

Policy-makers. Uncertain advice on Reddit is just a symptom of an uncertain legal landscape. The core problem is that the CFAA is a broad law [14], and the issue of broad legislation is reflected across other jurisdictions [46]. While this helped to future proof the law against technological advances, evidence it was perhaps too broad can be found in the Supreme Court adopting a narrow interpretation of where authorization is necessary [30]. This is compounded by high-profile prosecutions that create a chilling effect, for example both the Aaron Swartz [73] and Weev [74] cases were cited as evidence that prosecution was likely. While this may have a positive effect in deterring cyber crime, it also deters users experimenting with security tools.

Policy-makers should think about how to provide guidance on what is permissible under the CFAA, such as the 2022 clarification that the Department of Justice will not pursue "good-faith security research". While the message is useful, we recommend that policy-makers try to send it in novel channels. One promising example is partnering with tool providers and embedding guidance in the tool's documentation, which is motivated by the widespread citations of Nmap's legal issues page [64]. For the guidance to be credible, law enforcement must adhere to stated policy, which the Electronic Frontier Foundation has called into question [75].

6.3 Future Work

We identify the following opportunities for future research. Our results only speak to user perceptions about the legality of port scanning, which we chose because it was on the benign end of the offensive security spectrum. It would be interesting to compare the results to user perceptions about vulnerability scanning and collections of tools like Kali Linux and pwntools for Python. One would expect advice about illegality and prosecution to be more frequent. It would be interesting to see if technical mitigations are mentioned more often.

Given that analogies were frequently used in the provision of legal advice, a larger study on the use of analogies and possibly metaphors utilized by users in discussing offensive security could provide a more comprehensive view of user sentiment towards similar conduct. This research could explore how these rhetorical devices shape perceptions and influence user behavior regarding legality and ethics in cybersecurity practices.

Future work could additionally investigate the role of professional and technical guidance in shaping user behavior and perceptions of legality. This could involve studying how cybersecurity professionals interpret and implement legal advice compared to novices. It would be valuable to study how effective technical solutions and professional guidance are in helping users comply with legal requirements. This means looking at how well different tools and advice from experts help users follow the law while performing port scans.

Another direction for research is to explore advice among a different population to Reddit users, who are likely to skew towards novices. Exploring perceptions among practitioners, particularly those who scan for a living, could be interesting given they are likely to have access to professional legal advice. Another direction would be to explore perceptions among criminal groups, who may be more likely to discuss evasion techniques.

7 CONCLUSION

Broad technology laws risk criminalizing legitimate computer activities [14], which may prevent amateur users from experimenting and learning about cybersecurity. To fill in the gap of prior research that focuses on people's perceptions about the technical aspects of security and privacy, we focused on perceptions of the legality of port scanning, a technique used by both attackers and defenders to examine exposed network assets.

To explore the topic, we conducted a qualitative content analysis of users' discourse about port scanning legality from Reddit, a popular online platform that has been actively studied by researchers on topics related to security and privacy. Our results show there are questions about legality both before scanning has taken place, and emotionally charged reactive questions about legal consequences after scanning. Most reactive questions suggest an innocent mistake was made, and most proactive threads explain non-malicious intents.

In response, users provided a mixture of legal, technical and prosecution advice. We discovered contradictory advice regarding whether port scanning is legal, although most advice appears to guide users towards lawful activity. In supporting their advice, users leveraged various arguments and evidence, predominantly technical URLs, analogies, and anecdotes. We caution against reasoning with analogies given the same comparison to physical security can lead to differing conclusions on legality.

Our case-study highlights considerable uncertainty surrounding technology law. This is especially concerning given port scanning is a relatively benign technique. Technology policy makers must be conscious of the potential for legislation to deter legitimate computer activity, and actively pursue communication strategies that address uncertainty. This could help reduce the reliance on potentially outdated or over-simplified interpretations of the law found on technical sites.

REFERENCES

[1] Martina Angela Sasse, Sacha Brostoff, and Dirk Weirich. Transforming the 'weakest link'—a human/computer interaction approach to usable and effective security. *BT Technology Journal*, 19(3):122–131, 2001.

[2] Lorrie Faith Cranor. A framework for reasoning about the human in the loop. In *Proceedings of the 1st Conference on Usability, Psychology, and Security*, UPSEC'08, USA, 2008. USENIX Association.

[3] Pedro Leon, Blase Ur, Richard Shay, Yang Wang, Rebecca Balebako, and Lorrie Cranor. Why Johnny can't opt out: A usability evaluation of tools to limit online behavioral advertising. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 589–598, 2012.

[4] Alexander Vetterl and Richard Clayton. Bitter harvest: Systematically fingerprinting low-and medium-interaction honeypots at internet scale. In *12th USENIX Workshop on Offensive Technologies (WOOT 18)*, 2018.

[5] Ruogu Kang, Laura Dabbish, Nathaniel Fruchter, and Sara Kiesler. "My data just goes Everywhere" User mental models of the internet and implications for privacy and security. In *Eleventh Symposium on Usable Privacy and Security (SOUPS 2015)*, pages 39–52, 2015.

[6] Justin Wu and Daniel Zappala. When is a Tree Really a Truck? Exploring Mental Models of Encryption. In *Fourteenth Symposium on Usable Privacy and Security (SOUPS 2018)*, pages 395–409, 2018.

[7] Sergej Dechand, Alena Naiakshina, Anastasia Danilova, and Matthew Smith. In encryption we don't trust: The effect of end-to-end encryption to the masses on user perception. In *2019 IEEE European Symposium on Security and Privacy (EuroS&P)*, pages 401–415. IEEE, 2019.

[8] Veroniek Binkhorst, Tobias Fiebig, Katharina Krombholz, Wolter Pieters, and Katsiaryna Labunets. Security at the end of the tunnel: The anatomy of VPN mental models among experts and non-experts in a corporate context. In *31st USENIX Security Symposium (USENIX Security 22)*, pages 3433–3450, 2022.

[9] "It may be a pain in the backside but..." insights into the resilience of business after GDPR, author=Buckley, Gerard and Caulfield, Tristan and Becker, Ingolf, booktitle=Proceedings of the 2022 New Security Paradigms Workshop, pages=21–34, year=2022.

[10] Daniel W Woods, Rainer Böhme, Josephine Wolff, and Daniel Schwarcz. Lessons lost: Incident response in the age of cyber insurance and breach attorneys. In *Proceedings of the 32nd USENIX Security Symposium, Anaheim, California*, 2023.

[11] Alexander Vetterl, Richard Clayton, and Ian Walden. Counting outdated honeypots: Legal and useful. In *2019 IEEE Security and Privacy Workshops (SPW)*, pages 224–229, 2019.

[12] Alexander Gamero-Garrido, Stefan Savage, Kirill Levchenko, and Alex C Snoeren. Quantifying the pressure of legal risks on third-party vulnerability research. In *Proceedings of the 2017 acm sigsac conference on computer and communications security*, pages 1501–1513, 2017.

[13] Zakir Durumeric, Eric Wustrow, and J Alex Halderman. ZMap: Fast internet-wide scanning and its security applications. In *22nd USENIX Security Symposium (USENIX Security 13)*, pages 605–620, 2013.

[14] Orin S Kerr. Vagueness Challenges to the Computer Fraud and Abuse Act. *Minn. L. Rev.*, 94:1561, 2009.

[15] Mark Allman, Vern Paxson, and Jeff Terrell. A brief history of scanning. In *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, IMC '07, page 77–82, New York, NY, USA, 2007. Association for Computing Machinery.

[16] Computer Fraud and Abuse Act. 18 U.S.C. § 1030, 1986. Amended 2008.

[17] Computer Misuse Act 1990 - section 3(1).

[18] Computer Misuse Act 1990 - sections 1(1)(a) and (b).

[19] Tianshi Li, Elizabeth Louie, Laura Dabbish, and Jason I Hong. How developers talk about personal data and what it means for user privacy: A case study of a developer forum on Reddit. *Proceedings of the ACM on Human-Computer Interaction*, 4(3):1–28, 2021.

[20] Mohammad Tahaei, Kami Vaniea, and Naomi Saphra. Understanding privacy-related questions on stack overflow. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pages 1–14, 2020.

[21] Jingjie Li, Kaiwen Sun, Brittany Skye Huff, Anna Marie Bierley, Younghyun Kim, Florian Schaub, and Kassem Fawaz. “it’s up to the consumer to be smart”: Understanding the security and privacy attitudes of smart home users on Reddit. In *IEEE Symposium on Security and Privacy (SP)(SP)*, pages 380–396. IEEE Computer Society Los Alamitos, CA, 2023.

[22] Marco de Vivo, Eddy Carrasco, Germinal Isern, and Gabriela O. de Vivo. A review of port scanning techniques. *SIGCOMM Comput. Commun. Rev.*, 29(2):41–48, apr 1999.

[23] Elena Silenok Cynthia Bailey Lee, Chris Roedel. Detection and characterization of port scan attacks. *Department of Computer Science & Engineering University of California, San Diego*, 2003.

[24] Jessica R Herrera-Flanigan and Sumit Ghosh. Criminal regulations. In *Cybercrimes: A Multidisciplinary Analysis*, pages 265–308. Springer, 2010.

[25] Andrew Cormack. Can CSIRTs Lawfully Scan for Vulnerabilities. *SCRIPTed*, 11:308, 2014.

[26] Laura Bernescu. When is a Hack not a Hack: Addressing the CFAA’s Applicability to the Internet Service Context. *U. Chi. Legal F.*, page 633, 2013.

[27] Mary M Calkins. They Shoot Trojan Horses, Don’t They-An Economic Analysis of Anti-Hacking Regulatory Models. *Geo. LJ*, 89:171, 2000.

[28] Sangkyo Oh, Kyungho Lee, et al. The need for specific penalties for hacking in criminal law. *The Scientific World Journal*, 2014, 2014.

[29] Sarah A Constant. The Computer Fraud and Abuse Act: A Prosecutor’s Dream and a Hacker’s Worst Nightmare—The Case Against Aaron Swartz and the Need To Reform the CFAA . *Tul. J. Tech. & Intell. Prop.*, 16:231, 2013.

[30] Orin S Kerr. Focusing the CFAA in Van Buren. *The Supreme Court Review*, 2021(1):155–184, 2022.

[31] Samantha Hourican. CFAA and Van Buren: A Half-Measure for a Whole-ly Ineffective Statute. *Seton Hall Legis. J.*, 47:30, 2023.

[32] Daniel W Woods and Jessica Weinkle. Insurance definitions of cyber war. *The Geneva Papers on Risk and Insurance-Issues and Practice*, 45:639–656, 2020.

[33] David Décary-Hétu and Benoit Dupont. Reputation in a dark network of online criminals. *Global Crime*, 14(2-3):175–196, 2013.

[34] Sergio Pastrana, Alice Hutchings, Andrew Caines, and Paula Buttery. Characterizing eve: Analysing cybercrime actors in a large underground forum. In *Research in Attacks, Intrusions, and Defenses: 21st International Symposium, RAID 2018, Heraklion, Crete, Greece, September 10-12, 2018, Proceedings 21*, pages 207–227. Springer, 2018.

[35] Chris Grier, Lucas Ballard, Juan Caballero, Neha Chachra, Christian J Dietrich, Kirill Levchenko, Panayiotis Mavrommatis, Damon McCoy, Antonio Nappa, Andreas Pitsillidis, et al. Manufacturing compromise: the emergence of exploit-as-a-service. In *Proceedings of the 2012 ACM conference on Computer and communications security*, pages 821–832, 2012.

[36] Kirill Levchenko, Andreas Pitsillidis, Neha Chachra, Brandon Enright, Márk Félegyházi, Chris Grier, Tristan Halvorson, Chris Kanich, Christian Kreibich, He Liu, et al. Click trajectories: End-to-end analysis of the spam value chain. In *2011 ieee symposium on security and privacy*, pages 431–446. IEEE, 2011.

[37] Danny Yuxing Huang, Maxwell Matthaios Aliapoulios, Vector Guo Li, Luca Invernizzi, Elie Bursztein, Kylie McRoberts, Jonathan Levin, Kirill Levchenko, Alex C Snoeren, and Damon McCoy. Tracking ransomware end-to-end. In *2018 IEEE Symposium on Security and Privacy (SP)*, pages 618–631. IEEE, 2018.

[38] Ben Collier, Richard Clayton, Alice Hutchings, and Daniel Thomas. Cybercrime is (often) boring: Infrastructure and alienation in a deviant subculture. *The British Journal of Criminology*, 61(5):1407–1423, 2021.

[39] Kurt Thomas, Damon McCoy, Chris Grier, Alek Kolcz, and Vern Paxson. Trafficking Fraudulent Accounts: The Role of the Underground Market in Twitter Spam and Abuse. In *22nd USENIX Security Symposium (USENIX Security 13)*, pages 195–210, 2013.

[40] Rolf Van Wegberg, Samaneh Tajalizadehkhoob, Kyle Soska, Ugur Akyazi, Carlos Hernandez Ganan, Bram Klievink, Nicolas Christin, and Michel Van Eeten. Plug and prey? measuring the commoditization of cybercrime via online anonymous markets. In *27th USENIX security symposium (USENIX security 18)*, pages 1009–1026, 2018.

[41] Michele Campobasso and Luca Allodi. Impersonation-as-a-service: Characterizing the emerging criminal infrastructure for user impersonation at scale. In *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*, pages 1665–1680, 2020.

[42] Ross Anderson, Chris Barton, Rainer Böhme, Richard Clayton, Michel JG Van Eeten, Michael Levi, Tyler Moore, and Stefan Savage. Measuring the cost of cybercrime. *The Economics of Information Security and Privacy*, pages 265–300, 2013.

[43] Casey Breen, Cormac Herley, and Elissa M Redmiles. A large-scale measurement of cybercrime against individuals. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–41, 2022.

[44] Ben Collier, Daniel R Thomas, Richard Clayton, and Alice Hutchings. Booting the booters: Evaluating the effects of police interventions in the market for denial-of-service attacks. In *Proceedings of the Internet Measurement Conference*, pages 50–64, 2019.

[45] Arman Noroozian, Elsa Turcios Rodriguez, Elmer Lastdrager, Takahiro Kasama, Michel Van Eeten, and Carlos H Gañán. Can ISPs Help Mitigate IoT Malware? A Longitudinal Study of Broadband ISP Security Efforts. In *2021 IEEE European Symposium on Security and Privacy (EuroS&P)*, pages 337–352. IEEE, 2021.

[46] 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. In *2024 IEEE Symposium on Security and Privacy (SP)*, pages 104–104. IEEE Computer Society, 2024.

[47] Ferdinand Grieger. Port scans in the light of the law: a criminal and civil law analysis. *International Cybersecurity Law Review*, 2:297–316, 2021.

[48] Mary Ellen Zurko and Richard T Simon. User-centered security. In *Proceedings of the 1996 Workshop on New Security Paradigms*, pages 27–33, 1996.

[49] Simson Garfinkel and Heather Richter Lipford. *Usable security: History, themes, and challenges.* Morgan & Claypool Publishers, 2014.

[50] John Karat, Clare-Marie Karat, and Carolyn Brodie. Human-computer interaction viewed from the intersection of privacy, security, and trust. In *The Human-Computer Interaction Handbook*, pages 665–684. CRC Press, 2007.

[51] Ayako A Hasegawa, Mitsuaki Akiyama, Naomi Yamashita, Daisuke Inoue, and Tatsuya Mori. Analysis of non-experts' security-and privacy-related questions on a q&a site. *IEICE TRANSACTIONS on Information and Systems*, 106(9):1380–1396, 2023.

[52] Mohammad Tahaei, Tianshi Li, and Kami Vaniea. Understanding privacy-related advice on stack overflow. *Proceedings on Privacy Enhancing Technologies*, 2022.

[53] Ze Shi Li, Manish Sihag, Nowshin Nawar Arony, Joao Bezerra Junior, Thanh Phan, Neil Ernst, and Daniela Damian. Narratives: the unforeseen influencer of privacy concerns. *arXiv e-prints*, pages arXiv–2206, 2022.

[54] Swaathi Vetrivel, Veerle Van Harten, Carlos H Gañán, Michel Van Eeten, and Simon Parkin. Examining consumer reviews to understand security and privacy issues in the market of smart home devices. In *32nd USENIX Security Symposium (USENIX Security 23)*, pages 1523–1540, 2023.

[55] Moitrayee Chatterjee, Prerit Datta, Faranak Abri, Akbar Siami Namin, and Keith S Jones. Cloud: A platform to launch stealth attacks. In *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*, pages 1558–1563. IEEE, 2020.

[56] Nicholas Proferes, Naiyan Jones, Sarah Gilbert, Casey Fiesler, and Michael Zimmer. Studying Reddit: A systematic overview of disciplines, approaches, methods, and ethics. *Social Media+ Society*, 7(2):20563051211019004, 2021.

[57] Nicholas Botzer, Shawn Gu, and Tim Weninger. Analysis of moral judgment on Reddit. *IEEE Transactions on Computational Social Systems*, 10(3):947–957, 2022.

[58] Lu Xiao and Humphrey Mensah. How does the thread level of a comment affect its perceived persuasiveness? a Reddit study. In *Science and Information Conference*, pages 800–813. Springer, 2022.

[59] J Nathan Matias. Going dark: Social factors in collective action against platform operators in the Reddit blackout. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 1138–1151, 2016.

[60] Miranda Wei, Sunny Consolvo, Patrick Gage Kelley, Tadayoshi Kohno, Tara Matthews, Sarah Meiklejohn, Franziska Roesner, Renee Shelby, Kurt Thomas, and Rebecca Umbach. Understanding help-seeking and help-giving on social media for image-based sexual abuse. *arXiv preprint arXiv:2406.12161*, 2024.

[61] American Banker. The DoJ is no longer prosecuting good faith hackers, 2023. Accessed: 2023-01-10.

[62] Statista. Reddit.com: monthly visits 2022, by country. Statista, 2022. Accessed: 2023-04-01.

[63] Zakir Durumeric, Michael Bailey, and J Alex Halderman. An Internet-Wide view of Internet-Wide scanning. In *23rd USENIX Security Symposium (USENIX Security 14)*, pages 65–78, 2014.

[64] Nmap legal issues. https://nmap.org/book/legal-issues.html#:~:text=While%20Nmap%20is%20open%20source,Nmap%20also%20carries%20no%20warranty. Accessed: 2023-04-6.

[65] Rudolph Flesch. A new readability yardstick. *Journal of Applied Psychology*, 32(3):221, 1948.

[66] Yasemin Acar, Michael Backes, Sascha Fahl, Doowon Kim, Michelle L Mazurek, and Christian Stransky. You get where you're looking for: The impact of information sources on code security. In *2016 IEEE Symposium on Security and Privacy (SP)*, pages 289–305. IEEE, 2016.

[67] Felix Fischer, Konstantin Böttinger, Huang Xiao, Christian Stransky, Yasemin Acar, Michael Backes, and Sascha Fahl. Stack overflow considered harmful? the impact of copy&paste on Android application security. In *2017 IEEE Symposium on Security and Privacy (SP)*, pages 121–136. IEEE, 2017.

[68] Tianyi Zhang, Ganesha Upadhyaya, Anastasia Reinhardt, Hridesh Rajan, and Miryung Kim. Are code examples on an online Q&A forum reliable? A study of API misuse on Stack Overflow. In *Proceedings of the 40th International Conference on Software Engineering*, pages 886–896, 2018.

[69] Xinlan Emily Hu, Mark E Whiting, and Michael S Bernstein. Can online juries make consistent, repeatable decisions? In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–16, 2021.

[70] Fyodor and the Nmap Project. Legal issues. Accessed: 2024-07-19.

[71] Julie Haney, Yasemin Acar, and Susanne Furman. " it's the company, the government, you and i": User perceptions of responsibility for smart home privacy and security. In *30th USENIX Security Symposium (USENIX Security 21)*, pages 411–428, 2021.

[72] Eugene R Gaetke. Expecting too much and too little of lawyers. *U. Pitt. L. Rev.*, 67:693, 2005.

[73] United States District Court for the District of Massachusetts. United States v. Aaron Swartz. Court Documents, 2011. Criminal Case No. 11-CR-10260, District of Massachusetts.

[74] United States District Court for the District of New Jersey. United States v. Andrew Auernheimer. Court Documents, 2012. Criminal Case No. 2:11-cr-00470-SDW, District of New Jersey.

[75] Andrew Crocker. Is the justice department even following its own policy in cybercrime prosecution of a journalist?, 2024. Accessed: 2024-05-13.

A DATA COLLECTION

The exact search terms corresponding to Figure 7 are as follows:

Search term 1: ("law" OR "legality") AND ("port scan" OR "network scan")

Search term 2: ("legal" OR "law" OR "regulate") AND "Nmap"

Search term 3: ("legal" OR "regulation" OR "crime" OR "illegal") AND "Nmap scan"

Search term 4: (law OR legality OR illegal OR crime OR regulate OR regulation OR prosecute) AND (Nmap OR port

scan OR network scan OR shodan)
Search term 5: (“law” OR “legality” OR “illegal” OR “crime” OR “regulate” OR “regulation” OR “prosecute” OR “prosecution”) AND (“Nmap” OR “Angry IP” OR “Dipiscan” OR “Masscan” OR “NetCrunch” OR “Zmap” OR “Fing” OR “Zenmap” OR “Advanced IP Scanner” OR “port scan” OR “network scan” OR “Shodan” OR “Censys” OR “Criminal IP” OR “ZoomEye” OR “IVRE” OR “FOFA”)

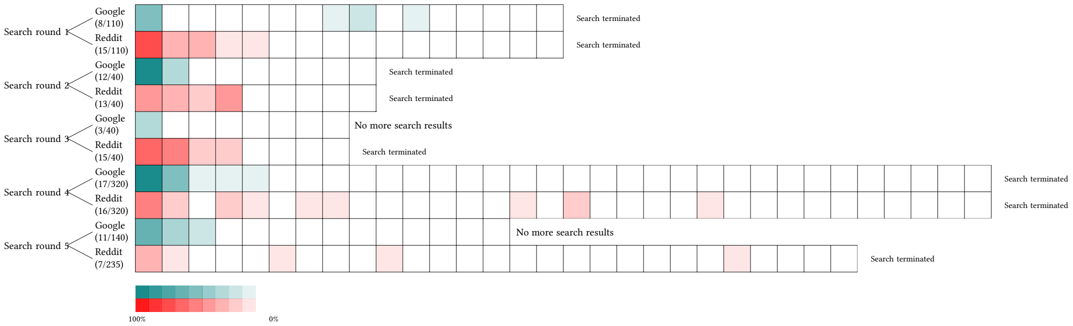

Fig. 7. Figure showing Reddit and Google saturation for five search terms. Darker shading in boxes indicated higher frequency of relevant search terms per 10 search results. White boxes indicate no relevant search results. The numbers below Google and Reddit indicate how many relevant searches were found of total search results for that term. Reddit’s internal search function reached saturation more efficiently than Google Search. However, both search engines identified relevant posts that the other did not.

B CODE BOOK

1 Questions

Any questions being asked within the thread. This includes questions in the body of OP’s question (the post), as well as any questions posed by users – whether in user answers or discussions (see above). Includes any questions not directly related to the discussion of legality. This further includes ‘questions’ without question marks due to careless typing, error, etc. For example “Is port scanning illegal or not please tell me”. Rhetorical questions are included, for consistency. This includes any expression or desire to gain additional knowledge or clarification on a matter, i.e., “I’m hoping to gain some additional information as if I did anything wrong/illegal I certainly want to know.”, “Curious about how this would play out in court.”

- **Legal** – Questions of a legal nature.
- **Technical** – Any questions of technical nature.
- **Prosecution** – Any questions pertaining to prosecution.
- **Rhetorical** – Any questions of rhetorical nature – this includes sarcasm.
- **Miscellaneous** – Any questions thar do not fit under the aforementioned categories, including off-topic questions.

2 Advice

Advice and suggestions are covered under this. This ranges from tentative suggestions to confident and definitive advice. These are usually coded in larger chunks of text (among other overlapping codes) to preserve the context of the advice.

- **Legality** – Any implicit or explicit advice pertaining to the legality of port scanning suggesting/telling OP whether it's legal or not, or making suggestions on how to make port scanning fall within legal confines. A common piece of legal advice concerning whether the scan is authorized. This may be objective or subjective.

  - **General/generic legality** – Any advice about the legality of port scanning not specific to any jurisdiction – just general 'generic' legality.
  - **Jurisdiction specific legality** – Any advice about legality of port scanning relating to a specific jurisdiction, e.g., U.S., Germany.
  - **Legal port scanning sentiment** – Any sentiment expressing or indicating that port scanning is legal.
  - **Illegal port scanning sentiment** – Any sentiment expressing or indicating that port scanning is illegal.
  - **Exceptions to legality** – Any exceptions provided as a part of advice, or indication that legality is contingent on other factors, e.g., "Port Scanning in most countries (USA) is NOT illegal in most cases. The exception is when the scan results in damage (say a denial of service) then there can be problems."
  - **Miscellaneous** – i.e., "Asking for legal advice of the Internet is probably a bad plan. If you want legal advice, ask a lawyer."
- **Technical** – Any advice, suggestions, commands or orders on technical measures to take/not to take. This usually follows 'legal' under 'Advice' (above), whereupon technical advice is afforded on how to circumvent any potential legal issues through change of setup, software, change of technique etc. This covers advice in the form of a command.

  - **Advising to use a VPN**
  - **Advising to check/visit/utilize specific websites** - e.g., "Checkout vulnhub or HTB for practice machines" and "tryhackme is good too" or HackTheBox
  - **Advising to use specific functions** - e.g., "As an exercise, try playing with nmap's -T parameter to avoid being blocked."
  - **Advising to use specific software** - this includes Nmap/Shodan etc.
  - **Miscellaneous**
- **Prosecution** – Any advice on prosecution and whether/how likely/unlikely this is, including implicit reference thereto, e.g., "…and no ones going to do anything about it."
  - **General/generic prosecution** – Any advice about the likelihood of prosecution of port scanning not specific to any jurisdiction.
  - **Jurisdiction specific prosecution** – Any advice about likelihood of prosecution of port scanning relating to a specific jurisdiction, e.g., U.S., Germany.
  - **Prosecution likely sentiment** – Any sentiment expressing or indicating that there is a likelihood of prosecution.
  - **Prosecution unlikely sentiment** – Any sentiment expressing or indicating that prosecution is unlikely.
  - **Exceptions to prosecution** – Any exceptions or indication that prosecution is contingent on other factors.
  - **Miscellaneous**
- **Uncertainty/qualifiers** – Any expression of uncertainty weaved into advice given, or general uncertainty regarding legality. This can also be by 'qualifying' advice, e.g., "I'm not a lawyer but…"

- **Miscellaneous** – Any advice expressed that does not fall under any other advice sub-code, e.g., "If your ISP asks say it was malware and that it's been cleaned up since. End of the day they're not going to be in a rush to terminate their contract with you since they want your money." Or "Just be careful"

3 Supporting Evidence

- **Analogies** – Comparisons and parallels drawn between port scanning and another things, e.g., "port scanning is like trying doors to see whether they're locked".
- **Community anecdotes** – These include norms and social experiences (usually third person, with the exception of first person "we") and references to common practices, i.e., "There is also a lot of academic research using port scans that can be used as a guide of the legal status. Major institutions aren't going to take the risk without a well thought-out or advised position on legality" and "Realistically, I think most reasonable people in this field would say that a single portscan does not create damage or loss."
- **URLs** – Any URLs /links including any hyperlinked text/words fall under this category, even if unrelated to question. This includes legal and non-legal links, e.g., Nmap legality page, and technical software websites to download software. Quotes and excerpts are also coded under this, as it is also a form of evidence.
- **Personal experiences** – Stories personal to users

  - **Legal** – Any personal user experiences, thoughts, opinions or observations specifically pertaining to legality of port scanning that is not covered by legal advice. This is oftentimes used as the basis for advice. "And that brings me to my point about CFAA. I don't know much about UK law, so I can't really speak to the Andrew Cormack thing, but CFAA is really, really bad. It is a very broadly-written, and frankly, very bad law, because it sets forth punishments that include jail time for unauthorized access that causes damage…", "I did a port scan of whitehouse.gov once and didn't get into trouble."
  - **Technical** – Any user experiences of technical nature, i.e., technical measures users have taken to try and protect themselves from potential liability, general user experiences or other technical user circumstances "I think with modern network devices this really doesn't matter anymore. My WAF knows nmap's techniques to be stealthy (or not)." Or "I should note it was a Class B network."
  - **Social details** – Details pertaining to user's circumstances/situation including sub-comments, that aren't of either legal or technical nature but are incidental to the legality of port scanning, e.g., "I moved to Dresden recently, but I have an apartment in Kaiserslautern. I came here two days ago to attend an exam at TU Kaiserslautern.", "I am a foreign student and the last thing I want is to get into trouble, but the issue with my network speed is getting out of hand."
- **Technical 'facts'** – User statements purported as facts, without proof provided; details taken at face value, or 'technical reasoning'.
- **Legal 'facts'** – User statements purported as facts, without proof provided; details taken at face value, or 'legal reasoning'.
- **Miscellaneous** – Under this umbrella code we include anything that is linked to the interaction between forum users and/or isn't of substance or doesn't aid in answering the research question. "Edit : This post is generating more controversy than expected, so I tracked down one the places I've seen this issue discussed." "Very involved question, but let me see".