To Cloud or not to Cloud: A Qualitative Study on Self-Hosters' Motivation, Operation, and Security Mindset

Lea Gröber, *CISPA Helmholtz Center for Information Security and Saarland University;* Rafael Mrowczynski, *CISPA Helmholtz Center for Information Security;* Nimisha Vijay and Daphne A. Muller, *Nextcloud;* Adrian Dabrowski and Katharina Krombholz, *CISPA Helmholtz Center for Information Security*

https://www.usenix.org/conference/usenixsecurity23/presentation/grober

This paper is included in the Proceedings of the 32nd USENIX Security Symposium.

August 9–11, 2023 • Anaheim, CA, USA

978-1-939133-37-3

To Cloud or not to Cloud: A Qualitative Study on Self-Hosters' Motivation, Operation, and Security Mindset

Lea Gröber[1,2] Rafael Mrowczynski[1] Nimisha Vijay[3] Daphne A. Muller[3]
Adrian Dabrowski[1] Katharina Krombholz[1]

[1]CISPA Helmholtz Center for Information Security
[2]Saarland University
[3]Nextcloud

Abstract

Despite readily available cloud services, some people decide to self-host internal or external services for themselves or their organization. In doing so, a broad spectrum of commercial, institutional, and private self-hosters take responsibility for their data, security, and reliability of their operations.

Currently, little is known about what motivates these self-hosters, how they operate and secure their services, and which challenges they face. To improve the understanding of self-hosters' security mindsets and practices, we conducted a large-scale survey (N_S=994) with users of a popular self-hosting suite and in-depth follow-up interviews with selected commercial, non-profit, and private users (N_I=41).

We found exemplary behavior in all user groups; however, we also found a significant part of self-hosters who approach security in an unstructured way, regardless of social or organizational embeddedness. Vague catch-all concepts such as *firewalls* and *backups* dominate the landscape, without proper reflection on the threats they help mitigate. At times, self-hosters engage in creative tactics to compensate for a potential lack of expertise or experience.

1 Introduction

The year is 2023 A.D. The Internet is entirely occupied by commercial cloud services. Well, not entirely... One small minority of indomitable self-hosters still holds out against the invaders.[1] Cloud computing has been on the rise for the past decade, and is popular with both individuals and organizations for its scalability, affordability, and accessibility [29]. On the flip side, commercial clouds are criticized for posing privacy risks to consumers [31, 59, 60]. The associated concentration of user data also carries security risks, such as increased attractiveness for attackers due to the proximity of the data [45]. Tim Berners-Lee criticizes the current centralization of the Internet and its services by a few companies as the creation of *data silos* where users' data is locked away. Not only has the user considerably less control, but they also need to trust the service- and the data center operator [39]. Self-hosting is sometimes promoted as an opposition to this development [32], promising to protect and secure one's own data by regaining autonomy. The term *self-hosting* describes maintaining the hard- and/or software for internal and external services on your own as opposed to buying access to these services from a third party [45]. A wide range of self-hostable software covers file synchronization, streaming, calendars, password managers, messaging, and many more [4]. Additionally, self-hosting allows for a diverse set of deployment and configuration options, and, with respect to different threat models, security strategies.

Anecdotal evidence suggests that self-hosters commonly find themselves thrown in at the deep end of suddenly being responsible for an Internet-facing service [63]. In this context, self-hosters represent a special population, as they become administrators without necessarily having the relevant technical expertise nor experience. They are an intermediary group between end-users and professional administrators.

To shed light on security challenges within the complex self-hosting ecosystem, we investigate the security mindset and practices of people with varying levels of technical expertise who host in personal, organizational, or non-profit contexts. To do so, we combine a large-scale survey (N_S=994) with semi-structured interviews involving selected survey participants (N_I=41). All participants are users of Nextcloud, a well-known self-hostable cloud office suite that covers a wide range of functionality with a variety of apps. The Nextcloud community is a suitable test bed to study the self-hosting phenomenon, as it has a large and active community covering a broad variety of use cases. Hence, with a combination of qualitative and quantitative methods, we answer the following research questions:

RQ1: *What motivates people to self-host?* Uncovering reasons to self-host helps to understand self-hosters' goals and might explain why they make certain (security-relevant) decisions.

RQ2: *How do self-hosters operate?* Understanding admin constellations and social embeddedness will help to

[1] If this paper were a French comic about Romans [25].

uncover unique roadblocks that self-hosters face, and the resources they rely on to overcome problems. Understanding the context of operations is necessary to improve adoption, support, and administrative tools.

RQ3: *What are perceived threats and how do self-hosters manage them?* Analyzing security practices, including attacker modeling, risk perception, and selection of defensive mechanisms is a crucial step to uncover structural gaps in self-hosters' security mindset.

RQ4: *How do self-hosters maintain their operations?* Understanding maintenance, as a facet of security practices, helps to explore self-hosters' security mindset.

RQ5: *In how far does the multidimensional space of self-hosting create tension?* Understanding what problems certain (combinations of) individual characteristics, such as knowledge or motivation, cause and how they affect security outcomes can help people make better decisions.

We found that a lack of it-expertise does not prevent people from self-hosting, especially if they are driven by normative values. To overcome their inexperience, they may enter special operational constellations, such as knowledge barter arrangements, or embed themselves in online communities. Certain motivational factors can impact how participants approach security. The results are meant to guide the development and deployment of helpful advice, information sources, and tools for the self-hosting community.

Replication Package We provide a full replication package and artifact repositories to support open science, reproducibility, and follow-up studies.[2]

2 Self-Hosting

Self-hosting refers to running and maintaining services or software under one's own control for personal or organizational use, rather than relying on shared services from third parties. Most of the time, this means the services run on-premise, i.e., on the service owner's own property, but can also mean putting their own servers in a third-party data center (co-location) or renting servers there. Renting servers falls on a spectrum of various levels of control over hardware and software (e.g. dedicated servers enabling hard- and software configurations; virtual private servers enabling software choices within the virtual machine). It can also include cases where customers have dedicated installations on rented servers with limited access (shared hosting). These Software-as-a-Service (SaaS) instances are included as an edge case, but a typical cloud service without dedicated installations for a single user is not. The three cornerstones for self-hosting are: (i) user control over hardware, (ii) control over software including the operating system and configuration, (iii) and a dedicated installation for the user or organization. Self-hosting is not limited to open-source software and can include closed-source products like game servers.

Nextcloud For this study, we research the self-hosting ecosystem on the example of Nextcloud [49]. Nextcloud developed from a mere file-syncing tool similar to Dropbox to a content collaboration platform with support for office documents, calendars, contacts, forms, and workflow management. They are installed on around 400,000 servers [48] and entail a large online community to tap into. This allows us to study the self-hosting population from a holistic point of view, as Nextcloud is adopted by private, commercial, non-profit, and governmental organizations. All of which we captured in the survey and most of them in the qualitative interviews.

3 Related Work

We study security practices of self-hosters. While there is no directly related work on self-hosting security practices, as a counter horizon we review security, privacy, and human factors research on cloud computing.

Security and Privacy in the Cloud Monlar et al. [45] discuss the security implications for organizations that move their infrastructure from self-hosting to third-party clouds. Our participants share the concerns that cloud service providers make attractive targets for attackers. Similarly, there is multiple research on the security of commercial cloud computing discussing potential privacy violations [31,59,60]. We find that privacy and autonomy are central motivating factors that drive people to self-hosting. While there is research on how privacy in clouds can be achieved by client-side encryption [1,18], Van et al. [67] argue that cryptography alone cannot solve privacy issues in cloud computing.

End User Perceptions on the Cloud Users that rely on a third-party cloud storage are a counter horizon to self-hosting. There is multiple research on cloud adoption and its influencing factors [2,23,24], user perceptions of cloud-services [5,17,61,68,69], and technology to assist users with data management in third-party clouds [11,33,34]. A common theme is that people lack awareness of which data is stored in clouds and that they have a need to take control. Tabassum et al. researched users' understanding of smart home devices and concerns regarding privacy-risking data practices [61]. They found that knowledge of smart homes did not impact their threat models and protection behavior. This is an interesting finding in the context of what motivates people to self-host. We found that IT knowledge alone does not predict if people will become self-hosters.

End Users and Administrator Security Practices People without technical experience may self-train to become self-hosters without having the knowledge or experience of professional admins to operate and secure their instances. Therefore, self-hosters constitute a special population as they are an intermediary group between end users and professional administrators. In the context of end users, we review studies

[2] https://github.com/usrgroup/USENIX23-selfhosting

on people who use their home network for more than just internet access. There is work on the complexity, and challenges of administrating home networks [9,13,20,28,62], and design guidelines for better home network management tools [52]. Bly et al. [9] investigated the overhead, or *"problem-time"* people have to invest when dealing with network devices. Similarly, our participants identified a lack of expertise as a major roadblock to self-hosting. There are several studies that compare end users' and experts' mental models of threats and defensive mechanisms [3,7,19,36]. Especially relevant in the context of hosting is the deployment of HTTPS. Krombholz et al. [36] investigated end users' and administrators' comprehension of the mechanism and found differences in the level of abstraction and perceived security benefits. Similarly, we find that, especially non-experts under- or overestimate the level of protection of security mechanisms. In a study about experts' and non-experts' mental models on VPN [7], Binkhorst et al. found that even experts have misconceptions about the security aspects of VPNs. Although we did not examine specific security measures, our research showed that possessing security knowledge does not necessarily equate to implementing a systematic security approach, such as threat modeling.

4 Methodology

Since little is known about the phenomenon of self-hosting, we carefully combined different qualitative and quantitative methods to explore the topic broadly. Our study consists of two parts: **(1)** a Nextcloud community **online survey** (N_S=994) covering demographic information about the instance, as well as motivations, use cases, and a coarse security assessment. **(2) semi-structured interviews** (N_I=41) with selected participants of the Nextcloud community survey, focusing on self-hosting as a socially-embedded activity, operator constellations, maintenance practices, threat-modeling, and defensive measures. Figure 1 provides an overview of the methodology. We compensated interviewees with 30€[3]; participants of the community survey were not compensated, as the survey was a joint effort with the community.

4.1 Study Population and Recruitment

Self-hosting is a broad concept with a broad variety of use cases, motivations, and approaches. Nextcloud is a suitable test bed to study the phenomenon, as it covers a variety of use cases, is open-source, and exhibits a large and active community. Although our findings are in detail Nextcloud specific, we expect generalizability for overarching concepts such as motivation to self-host, structural issues stemming from operator embeddedness, and security assessments of self-hostable solutions. We supposed in advance that there might be differences between personal and institutional usage of self-hosting. To study the phenomenon holistically and to identify areas of tension, we examine the following user groups: personal, commercial, non-profit, and government (the latter is only for the survey and not the interview).

We worked with the Nextcloud community to create a voluntary community survey; see Section 4.2 for details. Participants were also asked to indicate whether they would like to be contacted for possible further questions. Based on the survey's records, we shortlisted potential interview partners. We manually selected interesting participants covering a broad set of traits. That is, we accounted for (self-declared) security expertise, team size, reasons to use Nextcloud, and security concerns regarding their instance. We reached out to shortlisted candidates via email and invited them to online video interviews. During the course of the study, we updated the shortlist with complementing candidates according to our recruitment success, until we achieved coverage for the traits. The positive response rate per shortlist was 50% (personal), 26% (commercial), and 29% (non-profit/gov), and no one from governmental users.

4.2 Community Survey (N_S=994)

The survey was created in collaboration with the Nextcloud community. Community members started a discussion on the Nextcloud forum about how details of the community are unknown, and the idea arose to gather questions in a shared document [41]. This document was public, so everyone was able to collaborate. The community then reached out to Nextcloud employees who took over the operational aspects of the survey construction and distribution. In addition to the questions from the community, we added complementing questions about motivation, security perceptions, and operator constellation. Finally, Nextcloud's marketing department distributed the survey invitation via their newsletter, and a community member shared it in the forum [42]. After 1000 entries, we closed the survey, and a few open sessions increased the total returns to 1015, resulting N_S=994 after data cleanup. We collected data in September and October 2021. The survey enables us to get a bird-eye perspective of a self-hosting population. It served as a starting point for analysis and provided the basis to select a broad range of participants for interviews (see Section 4.3). The survey contained 21 questions in free-text and multiple-choice format focusing on Nextcloud instance-specific information (see replication package). Individual characteristics were subject to the interviews, not the survey. Questions group into three categories: (1) **technical details** such as server type (SaaS, Home Server, Dedicated Server, Virtual Private Server (VPS), Colocation), CPU architecture, Nextcloud version, security concerns (free text), (2) **operator constellation** including team size, and number of people with security background, (3) **details on use case** such as number of users , population (drop-down: personal, commercial, saas, non-profit, governmental), apps installed (free text), reasons to use Nextcloud, additional self-hosted services. The survey also

[3] Some waived compensation, as the study serves the open-source cause.

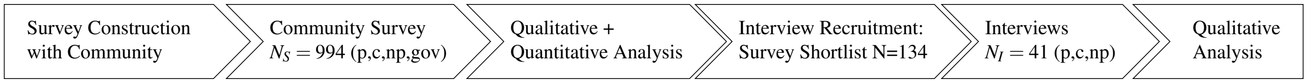
Survey Construction with Community
Community Survey $N_S = 994$ (p,c,np,gov)
Qualitative + Quantitative Analysis
Interview Recruitment: Survey Shortlist N=134
Interviews $N_I = 41$ (p,c,np)
Qualitative Analysis

Figure 1: Overview of the study process, and user groups involved in each step (**p**ersonal, **c**ommercial, **n**on-**p**rofit, **gov**ernment).

included some Nextcloud-specific questions about app usage and development requests that are out of scope for this paper.

4.3 Interviews (N_I=41)

We conducted 45 in-depth semi-structured interviews (total: 50 hours and 49 minutes; average: 67 minutes) to complement the survey's findings with rich qualitative data but removed four as they did not meet our definition of self-hosting (Section 2), i.e., they were predominantly hosting for third parties, not for themselves. For an overview of the selection process see Figure 7 in the Appendix. Talking to selected survey participants enabled us to tie the in-depth insights from the interviews to the large-scale but coarse picture that the community survey yields and vice versa. Thus, allowing us to explore interesting concepts that surfaced in the survey and fill in the gaps. Those 17 personal, 11 commercial, and 13 non-profit users give insight into the reasons that led them to self-host their service, the social ties in which they operate, and how they maintain and secure their instances. The analysis of the survey informed the development of the interview guide, enabling us to complement the instance-specific data of the survey with concrete technical challenges in a socially embedded context that surfaced during interviews. The resulting interview guide consisted of four parts containing questions that were tailored to the different user groups (personal, commercial, non-profit). See the replication package for the full interview guide.

In the first block, we talked about **reasons for adoption, and areas of application**. To open the conversation, we invited participants to tell us about their professional and educational backgrounds and history. Then we asked about their story of how they became Nextcloud users. Both questions gave us context and doubled as ice-breaker questions. We continued with the participants' privacy notions, how they use Nextcloud, and their technical setup. For organizational users (commercial, non-profit), we additionally asked how the self-hosted service is socially and technically embedded in daily operations, e.g. when working with clients. In the second block, we talked about **maintenance practices**. Participants reported their approach to maintenance, regarding different components of the software and hardware stack. Third, we inquired about **threat models and defensive mechanisms**. Participants told us about any past incidents, their approach to securing their instance, which defensive mechanisms they deployed, who they try to protect from, and where they think their system could be vulnerable. Additionally, we asked organizational users if they have any security policies or guidelines for their infrastructure. Lastly, we complemented missing **demographical information** if not mentioned in the first block. We recorded the age (bracket), gender, country, occupation, and technical & security background. For organizational users (commercial, non-profit) we also asked about the size of the entire organization, sector of operation, and size of the operational & security team.

4.3.1 Interview Pre-tests

We conducted two pre-tests to ensure the questions were suitable for IT-savvy and non-savvy participants and understood correctly.The first pre-tester hosts Nextcloud on a home server without a technical background. The second pre-tester studied computer science and hosts Nextcloud instances for commercial and personal use on virtual private servers and home servers. The pre-tests led to minor rephrasing and changes to the order of questions to improve the flow of the interview.

4.4 Data Analysis

The qualitative analysis was a multi-step process, involving a total of four coders with different backgrounds (two computer scientists, one designer, and one sociologist). We followed an iterative procedure combining the "top-down" approach of qualitative content analysis [37, 43, 56, 57, 66] with the "bottom-up" strategy inspired by "open coding" in Grounded Theory [15, 40, 58]. First, two coders constructed codebooks for the survey. Based on these, we conducted thematic analysis [12, 66], grouping codes into core themes and concepts. Second, all coders worked together and iteratively analyzed the interview data. For each research question, we tied the coarse findings of the survey to the detailed insights of the qualitative interview analysis. The rich interview data allowed us to explore, confirm, and extend the themes and categorizations of the survey. While analyzing, we always re-read the corresponding transcript segments to make sure the analysis is grounded in data. The following sections provide a detailed description for the survey and interviews.

4.4.1 Community Survey

For the analysis of the two qualitative questions about reasons to self-host, and security concerns, we only considered entries that contained an answer to at least one of the two questions resulting in 912 records (see Table 2).Two researchers with different backgrounds constructed a separate codebook for each question following the *open coding* approach. The lead author is a computer scientist with a focus on security and privacy who constructed the initial codebooks based on 10% of the dataset. The second author has a background in design and used the initial codebooks to independently code the same percentage of the dataset. The coders discussed their coding and adjusted the codebooks accordingly. Subsequently, they proceeded to iteratively code and discuss portions of the dataset until they reached saturation [8, 64]. Saturation was reached after three iterations, taking into account 27% of the dataset. We conducted two additional rounds of coding where no new high-level concepts emerged. The inter-coder reliability Krippendorff's alpha [35] was between 0.69 and 0.869 for each codebook version. The remaining dataset was split in half among the coders to be analyzed with the final codebooks. Afterward, the coders discussed if new concepts emerged or if any changes were needed. They agreed that the codebooks were stable and needed no further alterations. The final codebooks contain 35 codes and are provided in the supplementary material referenced in Section 1.

4.4.2 Interviews

We analyzed the interviews starting with some initial thematic codes derived from the survey findings and the interview guide, but enriched and specified them by open coding. For the initial codebook construction, we picked five interviews constituting the presumably most contrastive cases in our dataset. Two coders (computer scientist, sociologist) coded the interviews independently. They discussed their coding and merged the codebooks into one. We started a *documentary* analysis [10, 51] at this point, where we derived themes, concepts, and how the different self-hosters react to similar problems, based on case comparisons. We then proceeded to iteratively code selected interviews to test and contrast patterns in our analysis. That way we reached a stable codebook after coding 25% of the dataset. We proceeded that way, involving two additional coders (computer scientist, designer) until having coded 50% of the dataset and we agreed that we reached saturation. We jointly discussed the codings and identified six axial categories (knowledge, motivation, social embeddedness, it-operations, security mindset, use cases) with respective sub-themes. All coders worked together to write concise summaries of all interviews, containing quotes and references to the raw data, for the sub-themes of the axial categories. This type of analysis does not need an inter-coder agreement calculation, as all codings were jointly discussed and resolved resulting in a hypothetical agreement of 100%. The final codebook contains 585 codes and is provided in the supplementary material referenced in Section 1.

4.5 Ethical Considerations

This study got approval from the Universität des Saarlandes ethical review board. Self-hosting is a sensitive topic since it revolves around personal data. In particular, the interviews expose personal security and privacy choices and provisions. We made sure that participants were informed about data collection practices prior to taking part in the study. Additionally, before each interview, we thoroughly explained the process in order to obtain informed consent.

4.6 Limitations

Generalizability We recruited participants through the Nextcloud newsletter. We selected Nextcloud as a case study because of its widespread use and engaged community. While we have no indication that self-hosters of other projects feel and behave differently, we also cannot deny the possibility. Nonetheless, many of our participants also self-host other projects (see Section 5), indicating a large overlap and a similar mindset. Questions about product specifics (e.g., the type of update mechanism) are obviously not generalizable to other projects.

Selection Bias Community-based recruitment limits our view to successful installations. Furthermore, users would need to be invested enough in the topic that they subscribe to the newsletter and volunteer to the interviews. While we asked about installation problems, we were missing out on potentially fatal roadblocks to self-hosting.

Recall Bias With an interview and questionnaire methodology, self-report experiences might be several years old. Future work could use more controlled lab or diary studies for details on current, e.g., installation problems, but would miss out on the mindset of the experienced user base.

Social Desirability Bias We mitigated social desirability bias by stressing that the study's goals are to improve Nextcloud and identify roadblocks to self-hosting. In recruitment, participants were not primed for security. During interviews most people did not hesitate to discuss security choices, and were seeking guidance or feedback (which we offered after the interview to avoid bias). Against this background, we assume that our interviewees were relatively open to talk about weaknesses and vulnerabilities of their instances.

5 Results

Direct survey$^{S\text{-}x,g}$ and interview$^{I\text{-}x,g}$ quotes are translated verbatim into English where necessary. x denotes the participant id within the dataset (survey or interview), and g adds the group (personal, commercial, non-profit, government).

We interviewed 40 men and one woman. Table 1 provides an overview, more details are presented in Appendix B.1. Participants come from 16 countries across Europe, North America, and Oceania. The participants' professional background is broad and ranges from non-technical occupations (e.g., teachers, journalists, lawyers), over generally IT-related (e.g.,

Table 1: Overview of interview participant demographics.

		p	com	np
Number of participants		17	11	13
Average age	[years]	40.5	47.9	30.8
IT background	[#edu/self/no]	12/4/1	5/3/3	9/4/0
Security bg.	[#edu/self/no]	6/2/9	1/3/7	1/4/8
IT-related occupation	[#yes]	14	5	9
Hosting-rel. occupation	[#yes]	6	4	5

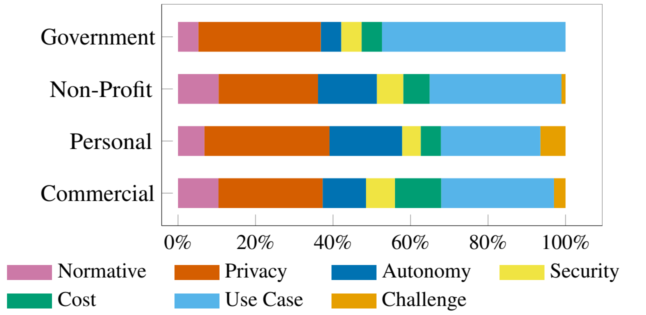

Figure 2: *Survey data:* Relative frequencies of reported motivational factors across user groups.

developers, data specialists), to hosting-related occupations (e.g., system administrators, system engineers, IT-support). For organizational use of self-hosting, we cover a broad spectrum of different industries (commercial: travel agency, law firm, journalists, etc.; non-profit: research institute, university, schools, political party, etc.). Participants relayed on a variety of server types to provider their operations. Figure 3 provides an overview over the interview population in relating participants' it-knowledge to their set-up choices.

The survey data provides insights on admin constellation and other self-hosted services (compare Table 2). Refer to Appendix B.2 for an overview of server types. The majority of participants self-hosts at least one service in addition to Nextcloud. Use cases are broad, spanning from other file storage, synchronization, and file transfer solutions, websites, home automation, communication and messaging tools, password managers, over mail servers, DNS servers, and software development version control, to game servers.

5.1 Motivation

Across all groups, **normative** driven self-hosters practice self-hosting because they see it as the right thing to do. Based on the interview data, we identify two (not necessarily mutually exclusive) sub-types of normative motivation for self-hosting: (1) by general socio-political values (e.g., for the society); (2) by professional-ethical values (e.g., for themselves and their occupation). A common theme for (1) is the appreciation of *“privacy [...] as a fundamental right. I believe that privacy is what protects us from totalitarian states. That means that by exercising or using my right to privacy, I am in a way strengthening democracy.”*$^{\text{I-8p}}$ We find (2) for professionals who strongly rely on trustful relations between individual practitioners and their clients (attorney, journalist): *“I deem it utterly unacceptable from the legal point of view when attorneys use official [third party] cloud systems like iCloud or OneDrive and store client data there because they cannot control at all to what extent confidentiality can be ensured there.”*$^{\text{I-14c}}$ Prominent motivational factors are **privacy** and **autonomy**, which often co-occur: *“I like the idea of having my data being completely private, physically on my own storage devices that I own and I can manage so that I don't have to use other means of encryption or something.”*$^{\text{I-6p}}$ Privacy is about establishing ownership over, and protection of sensitive data. Autonomy refers to the need for independence from third parties (usually commercial vendors), and a need to exert control over technical set-up and configuration. Tightly connected to this, some participants turn to self-hosting to improve **security** in contrast to relying on commercial cloud solutions. Those who mentioned security as a motivation reported to have security expertise either through education or self-taught and express a need for transparency or a lacking trust in third-party vendors. Across all groups, participants turned to self-hosting to save **costs**, usually to avoid the explicit cost of buying in the market, but it can also reduce administrative costs within organizations. Saving costs, however, can also interfere with privacy considerations, e.g. when non-profit organizations have to opt for cheaper server types. Unlike people who are strongly normative driven, for others, the decision to self-host is a pragmatic one. These people are mostly driven to meet a specific **use case**. Last but not least, participants self-host for the fun of it. They enjoy the **personal challenge** or want to learn something about hosting. Figure 2 provides an overview of motivational factors.

Key Takeaways: Based on the survey and interviews, we categorized seven motivational factors that led people to self-hosting. People can exhibit multiple motivational factors simultaneously. Thereby, tensions can arise where one factor can outweigh another.

5.2 Operator Constellations

In this section we describe self-hosting operations as a socially embedded activity. In the interviews, we learned that self-hosting is practiced in different constellations of social actors. The two major dimensions are *digitally mediated (social) interactions* and *IT operations* (Appendix A explains all concepts in detail). We use the social embeddedness of technical operations as the primary structuring category to identify several types of self-hosters and tie them to the survey results on operator constellations (compare Table 2).

Individual operators with family and friends 87.1% of personal self-hosters run their Nextcloud instances on their own without any significant assistance from other individuals. They use them for private purposes and often also host data of their family members, friends, and acquaintances, but they are the only person responsible for the entire operation of the self-hosting infrastructure. Hence, they are usually socially embedded in their digitally mediated interactions (e.g., sharing family photographs, coordinating activities via a self-hosted calendar app, etc.), but they act on their own in the domain of IT operations. The only rudimentary form of social embeddedness in the latter domain is the participation in online forums from where *individual operators* extract needed

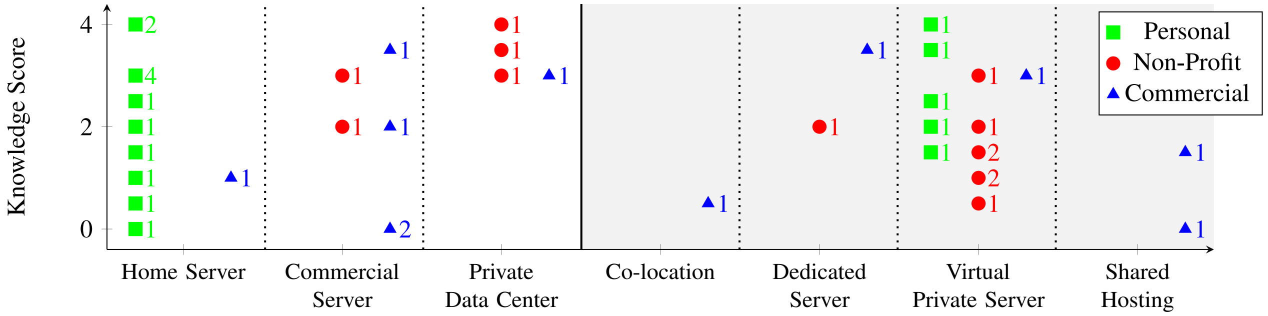

Figure 3: Visualization of the interview population by mapping IT expertise against server types. Knowledge score, as a points sum of self-reported *IT background, security background, IT-related occupation*, and *hosting-related occupation* data; where yes=1, self-taught=0.5, no=0 (compare Table 1). Server types: on-premise facilities in white, third-party hosting in gray.

pieces of IT expertise.

Most individual operators have profound IT knowledge and practical skills providing them with the self-confidence for self-hosting, though only 32.8% of individual operators report having a background in security (through education or self-taught). However, a lack of expertise can lead to a permanent struggle with technical problems. For example, interviewee 5 is strongly driven towards self-hosting by a normative (specifically political) motivation: as a person with a communitarian-socialist mindset, he does not want to use the cloud services of big capitalist companies. But as a humanities teacher, he commands very little IT knowledge. As a result, he is not able to overcome more complex technical problems: *"Basically, networking is what defeats me, I just don't get it. I set up my own server, which is here, that's it."*[I-5p]

Organizationally embedded sole operators They are partly similar to the individual operators since they are the only individuals running the respective self-hosting instance. However, as members of organizations, they take certain (formal) responsibility for the functionality of the IT infrastructure including its self-hosting components. Sole and part-time administrators are typically found in small organizations, with other duties on top of it. According to the survey, 60.4% of commercial, 55.8% of non-profit, and 42.3% of governmental Nextcloud instances are administrated by sole operators. Such constellations are particularly common in civic organizations like clubs or associations, where this task is allocated to an (assumed) domain-competent member. This subjective assumption is not necessarily accurate, and competence is relative to other members.

Table 2: Survey demographics on admin constellation, other self-hosted services in percentages per user group.

		p	com	np	gov
Group size	[#participants]	656	95	131	13
Single admin	[%]	87.1	60.4	55.8	42.3
Single admin w/ security bg.		32.8	42.2	31.2	16.7
Admin teams security bg.	[%]	49.6	38.3	50.1	80.1
Host additional services	[%]	65.9	67.0	67.4	71.4

The major similarity between *individual operators* and *sole operators in organizations*: they make their choices without any substantial interference by others; they determine all key aspects of the self-hosting infrastructure on their own; they also outline all further maintenance strategies including security measures. And if they err, there is virtually none to stop or even to warn them. However, online forums were frequently mentioned as an important source of critical information. Thus, we propose that there is something like virtual social embeddedness besides interpersonal social embeddedness.

Key Takeaways: *Sole operators*, both individual and those in organizations, enjoy the highest degree of leeway in determining their entire IT infrastructure, but they may face serious challenges when they hit the limits of their technical expertise.

Team members within organizations Bigger organizations usually deploy entire teams for IT operations (up to 39.6% of commercial, 44.2% of non-profit, and 57.3% of governmental Nextcloud instances). Hence, self-hosters acting in such organizational contexts are often embedded in a group of people with a certain division of tasks, responsibilities, and expertise. This adds another layer to their social embeddedness besides their general membership in the organization. Members of IT teams usually command extensive expertise which is often the main reason why they became an IT team member in the first place.

We identified two sub-types of specialized teams in which self-hosters can be embedded:

1. Teams with no internal specialization where each individual member is potentially responsible for all IT-related tasks within the organization. This redundancy can provide for continuity in situations when individual team members become temporarily incapacitated. But it may also lead to a confusion of responsibilities, inter alia, in terms of security practices.
2. Teams with internal specialization where each individual member focuses on a limited sub-set of tasks in accordance with their expertise. Individuals maintaining the

organizational self-hosting infrastructure in such constellations may partly resemble sole operators with regard to their independence and autonomy. However, differences arise from their embeddedness in a more complex organization, as detailed below.

Organizational embeddedness has both limiting and enabling implications. A bigger and more complex organization with a specialized IT team can restrict individual choices since other organization members may have a stake in (fundamental) decisions whether to cloud or not to cloud as well as more detailed technological choices. In other words, organizationally embedded self-hosters have to account for the diverse needs, preferences, and interests of different stakeholders.

The enabling aspects of organizational embeddedness are exemplified by Interviewee 20, who operates several Nextcloud instances for a night-school program at a German university. Here, he is a member of a small team that can be extended, if needed, by a few student assistants. In addition to that, he can also team up with the employees of the university's central IT services: *"We also have very frequent interactions with the computing center of the university. They also have a Nextcloud instance for the entire university [...]. Whenever they discover a problem, sure, we work then together on a solution. And it is the same the other way around when we discover something."*[I-20n] The third layer of his organizational embeddedness results from institutionalized cooperation between his university's IT specialists and their counterparts at another higher-education institution with an explicitly technical focus. Hence, he is able to easily mobilize additional expertise and workforce available within other segments of a complex organizational structure. This constellation enables him to implement self-hosting solutions that are beyond the reach of individual self-hosters.

Key Takeaways: Organizational team members are more constrained in their choices than self-hosters, but they can more easily receive support from their colleagues whose IT expertise complements their own knowledge and skills.

Collaborative networks refer to the cooperation of multiple individuals based on their personal traits rather than organizational structures. These networks emerge bottom-up as each member decides whether to initiate, continue, or renew cooperation with specific partners. Some participants may take on a pivotal role and lead the network, while others follow their lead.

This social mode of IT operation is vividly exemplified by Interviewee 37: he belongs to a cyber-activist community leaning towards leftist anarchism. He perceives free and open source software and self-hosting as key technologies that make independence from big capitalist companies possible (normative/socio-political motivation). Together with several other tech enthusiasts gravitating around a non-profit radio station, he operates different online services, including a Nextcloud instance for a community of like-minded users. The services are operated by a team that does not constitute any kind of formal organization: *"People freelance [...] and currently, I'd say that there are almost ten people that come and go, but in equivalent full-time, I would say like five people."*[I-17n] Hence, they form a social network connected by their common socio-political cause. The interviewee describes knowledgeable members of the team training novice admins as part of the community activities aiming at the dissemination of IT knowledge.

As an operator of self-hosting infrastructure, being part of a collaborative network has both benefits and drawbacks similar to those of organizational embeddedness. On one hand, the network can limit choices, but on the other hand, it can provide access to additional skills and workforce. However, the constraints of being in a network are less rigid due to its voluntary nature. Individuals who feel too confined can easily leave since there are no formal exit barriers typically found in organizations.

Key Takeaways: Collaborative networks demonstrate that pooling of self-hosting expertise can also occur without a formalized organizational framework. An important prerequisite for this form of collective operation appears to be a shared motivational idea that coalesces people commanding the required skills.

Knowledge barter This term denotes long-term, non-monetary exchanges of knowledge-based services as regular reciprocal favors. It is a form of social embeddedness that involves individuals with a relatively low level of own IT skills, but a strong motivation to self-host on their own.

This relationship is best illustrated by the case of Interviewee 14, a German defense attorney with a very strong professional-ethical (normative) motivation to self-host. Since he lacks sufficient IT expertise for operating a self-hosted instance at his law firm's premises entirely on his own, he relies on crucial technical support from an *"IT nerd"*[I-14c] (as the interviewee repeatedly calls this person) who happens to also be a client of him in need of criminal defense for alleged digital crimes. The fact that law-enforcement agencies were unable to penetrate "IT nerd's" own systems makes him a credible IT expertise provider in the eyes of the attorney. Such knowledge-barter relationships require a high level of general interpersonal trust because the IT-savvy partner acquires access to the most sensitive parts of the counterpart's computer infrastructure: *"This alternative support is the difficult part for me. It's trust-based because I have to let him in very deep into my system."*[I-14c] Hence, Interviewee 14 is aware of the general issue, but specific cyber-security risks of such a relationship were not discussed in detail by him.

Key Takeaways: *Knowledge bartering* can be a way to overcome one's own lack of expertise, but at the same time, it poses far-reaching security risks.

5.3 Maintenance Practices

We asked interviewees about their approach to maintaining their service. Participants broadly regarded updates as a crucial step in providing their operations reliably and securely*[I-20n]*. While some participants have a structured approach to maintenance*[I-40n]*, others update sporadically*[I-30n]*, or admit neglecting it*[I-5p]*.

Structured We identified different building blocks that suggest a structured approach to maintenance. Participants in this category reported on at least one of these, often in response to negative experiences with updating (e.g., data loss, downtime, functionality loss). **(1)** Participants **defined update cycles** (e.g., weekly to bi-annually) that may vary between different software components. Additionally, these participants frequently reported waiting for a stable version of Nextcloud before updating. A common approach to *"save some work"[I-21n]* is to slack *"a couple of minor releases"[I-21n]* behind the current release, especially when new major releases come out*[I-1p]*. Some participants reported treating critical security updates differently, immediately updating once they receive notice*[I-13n]*. Some participants stick to **(2) defined update procedures**, e.g., in the form of self-made technical checklists*[I-2c]*. Update procedures include making snapshots to recover from failed updates*[I-35c]* and prior checking forums for reported issues*[I-21n]*. If the setup allows, some participants carry out the updates step by step, starting with the least critical system (e.g., in case of multiple Nextcloud instances*[I-20n]*). However, not everyone with this option utilizes it: *"I upgrade all of them at once, but this might not be a good strategy. Recalling what has recently happened. Maybe, I'll just upgrade one at a time and see how it works"[I-19p]*. **(3) Testing** Organizational users reported on testing strategies. Sometimes they have dedicated test instances for development and updates. Additionally, they might define a set of use cases and manually test functionality after updates*[I-20n]*.

Best effort Across user groups, people reported on maintenance behavior that we describe as *best effort*. They do updates sporadically when they have time or get a notification. *"So [updates are] a bit of high life as it comes."[I-14c]* For people who work in distributed admin teams or rely on third parties for help, this may lead to a diffusion of responsibilities: *"Otherwise, our approach is: You check from time to time whether there are security patches, or you check: When is the next big update? Or if it's super important, then someone will get in touch. So if it's really, really important, we also have other admins who let us know."[I-13n]*

No Strategy Some personal and organizational participants report not regularly updating, e.g., not having updated the OS since initial set-up*[I-36n]*. The choice of infrastructure can block participants' update abilities. One participant opted for a shared hosting set-up and is now stuck with an outdated database which prevents him from updating Nextcloud *[I-34c]*. Some participants report missing updates regardless of their technical proficiency. Here, one participant identifies his knowledge gaps as a major roadblock: *"No, my maintenance is very poor. I think, I probably reached the limit of my comfortable knowledge with setting it up. [...] and so I'm slightly on a wing and a prayer. I do all the updates, the stable updates, and I keep Debian updated when I remember and just hope for the best, which isn't good, not good at all."[I-5p]*

Key Takeaways: Maintenance practices are inconsistent. For both organizational and private self-hosters, there are examples of structured and unstructured approaches, with elaborate strategies predominantly found in the organizational context.

5.4 Security Mindset

In the survey and in the interviews, different perspectives on cybersecurity emerged. Most participants expressed either a *fatalistic* or *pragmatic* security mindset, which are two sides of the same coin. There is a broad understanding that *"security is [a] prerequisite for everything else"[S-957g]*, and that without good security, self-hosting is a lost cause*[S-879p]*. Similarly, they usually share the view that no software*[I-17n]* or system can ever be 100% secure*[I-20n]*. People with a *fatalistic* mindset conclude that therefore a skilled attacker can break into any system, so they *"wouldn't even try [defending]"[I-44c]*. In contrast to that, people with a *pragmatic* mindset acknowledge threats, but conclude that security is achievable*[I-2c]* when following state-of-the-art security recommendations. These mindsets are relative to attacker models, e.g., people can be *pragmatic* when defending against untargeted external attacks, and *fatalistic* with respect to state actors. We were not able to identify a candidate factor that correlates with the security mindset, e.g., security knowledge does not seem to influence if people are *fatalistic* or *pragmatic*. There is also a third group of people who did not comment on security because they lack the confidence due to a lack of expertise, they completely outsourced security to a third party, or they neglected the topic based on their self-perceived unworthiness as a hacker target*[I-15c]*.

5.4.1 Attacker Models

In the following, we describe concrete and unspecific attacker models that surfaced during the interviews.

Targeted State Actor A lawyer*[I-14c]* and an investigative journalist*[I-15c]* in our dataset explicitly framed state actors as their most important attacker model. Both have concrete institutions and their capabilities in mind against which they want to protect. Their threat model is based on public knowledge of how these intuitions are legally allowed to operate, and of their own and their colleagues' experiences in dealing with and defending against them. They especially define the threat of them gaining physical access via search warrants. Neither of the two is tech-savvy, so they rely on a *knowledge-barter* (see Section 5.2) constellation to secure their operations. However, both have a *pragmatic* mindset with regard to defending against the state actors they defined. This is because they believe self-hosting is ultimately the only way to protect their data, and they have trust in the capabilities of their security operators: *"[the operator] is a former client of mine. And no law enforcement agency in the world had managed to penetrate [their] systems"*[I-14c]. Across all user groups, people are aware of state actors (technical or legal, like above). While most don't view themselves as targets, others explicitly state that even if they were, they would not stand a chance: *"it would be game over against a national security service. I don't think someone at my level can defend against that, so I wouldn't even try"*[I-44c]. Interestingly, these participants don't refer to concrete capabilities or attack vectors. They seem to view state actors as omnipotent, omniscient adversaries.

Targeted External Attacker Only organizations identified targeted attacks from non-state actors, such as business competitors, opponents to their cause, or personal enemies. Attacker's goals varied: While rivaling artists use hacking as a form of dialogue[I-21n], globally operating energy corporations seek to spy on and sabotage climate activists[I-17n]. Similarly to *state actors*, participants predominately have a *pessimistic* mindset about successfully defending instances: *"Any kind of attacker that can spend on one person that is skilled/motivated for some months would be able to access data. So this is my rough estimation, which is based on nothing"*[I-17n].

Untargeted External Attacker The most prominent attacker model across user groups was untargeted external attacks. Participants frequently referred to automated bots and *"Script Kiddies"*[S-447p], who *"poke around the Internet for the fun of it"*[I-5p]. However, they can also work as a first-stage reconnaissance to select easy targets for ransom or extortion. Most participants rank this to be the top threat they need to address. Although people usually were *pragmatic* in defending against these automated untargeted attackers, especially people with low technical expertise struggled in identifying adequate means of protection: *"My security is probably woeful"*[I-5p]. Additional mismatch emerges when the security mindset is borrowed from the end-user domain, e.g., a personal self-hoster who thinks his Ubuntu server is safe because *"ransomware usually targets Microsoft, not Linux"*[I-5p].

Internal Attacker Few participants also mention a need to protect from internal attackers, such as malicious admins[S-822n]. In the case of off-premise instances, participants often identify the hosting provider as a potential attacker, accessing their data[I-32n]. One participant describes users as a potential threat with regard to data theft[I-40n]. There is a broad understanding that users are not trustworthy[I-4p], but they think it is their incompetence that makes them a risk, not malicious intent: *"[I know my users], so it's unlikely that there would be malicious intent"*[I-44c]. Personal self-hosters do not report users as potential attackers, possibly because their user base mostly consists of friends and family.

Unspecific Attacker Model Across all user groups, participants elicited vague ideas of who could be an adversary to their system. They made unspecific claims that *"everything is a threat"*[I-40n], or that they are not protecting from anyone specifically[I-22c]. Both people with and without IT background or security expertise lack explicit attacker models. Likewise, this attitude is found across all use cases, and particularly concerning in contexts where one would expect elevated threat models, such as schools[I-32n].

Key Takeaways: We identified four attacker models in the interview data. However, only a few participants explicitly analyzed threats and threat actors prior to deciding on defensive mechanisms. The majority had unclear perceptions of attackers' capabilities.

5.4.2 Self-Hosters Security Perceptions

78% of survey participants explicitly stated that the security of their Nextcloud instance is a concern to them. They are concerned for a variety of reasons. First, because they think they are an attractive target to attackers based on who they are (e.g., government institution, independent media organization, lawyers), or because of the kind of data they possess (e.g., sensitive private or business data, client data). In particular, personal self-hosters are worried about their *"digital identity"*[I-14c]. As a consequence of a breach, organizations anticipate reputation damage and losing customers' trust[I-13n]. Personal users additionally worry about letting down their family and friends. Second, security in the sense of reliability concerns them as service is a critical infrastructure within their organization. Any downtime or loss of access would negatively impact the organization's day-to-day operations. If the use of the instance is perceived as non-critical *"it's a hobby project"*[S-893p], this can have the opposite effect. Third, taking adequate security measures might be a legal obligation for certain organizations, e.g., if they process personal data that is under GDPR protection. Participants said, them failing to secure their instance makes them liable to prosecution[I-14c].

Perceived Risks Both in the survey and interviews participants expressed their concerns about a variety of risks that they associate with self-hosting.

1. **Hosting on-premise.** Participants referred to the possibility of physical theft or confiscation of data, e.g., in the case of dealing with state actors. Family and friends who also had physical access were not a concern, because of a trustful relationship. Also, participants identified a need to maintain hardware components, e.g., to avoid data loss due to aging hard drives.
2. **Hosting on the public Internet.** One of the most pressing security concerns are Internet-facing instances which

participants perceive as the primary entryway for attacks: *"it needs to be accessible easily which is (but does not have to) sort of contradictory to being secure"*[S-810c]. Similarly, the secure configuration of software components such as web servers, databases, and all attached services is understood as the first line of defense. Simultaneously, the potentially complex interactions between software components leave ample room for mistakes: *"I learn what I can [...], but server security feels like a bottomless pit"*[S-118p]. Participants acknowledge the importance of maintaining the set-up, posing a security risk if updates are not rolled out regularly throughout the software stack. This gets complicated if services demand different versions of dependencies, or apps within a service block the update process because they are not compatible with the service's latest version, as can be the case with Nextcloud. Also, participants are concerned with the rapid update cycle of Nextcloud, feeling overwhelmed when trying to keep up[S-655p].

3. **Software.** Participants are aware of risks, that are generally associated with any software product, such as vulnerabilities in the code, and corresponding 0-days: *"I am afraid of 0-days at all levels of my Nextcloud/Linux system as state-sponsored attackers have access and vulnerabilities for all types of infrastructures and software"*[S-149p]. Moreover, they worry about supply chain attacks, especially when it comes to using pre-configured yet unsigned docker containers[I-19p]. In the case of Nextcloud, some worry about the underlying substrate as *"PHP has a reputation for security problems"*[S-235p]. Many participants view third-party apps for Nextcloud as one possible entryway into their system[S-628p].
4. **Admin capabilities.** With self-hosting, as opposed to relying on hosting providers, participants mainly identified two risks: First, knowledge gaps with respect to general server setup, the configuration in general, and security expertise in specific: *"[I] only have pro-amateur know-how"*[S-555p], *"I am no expert, so it could leak any moment"*[S-123p]. Second, they acknowledge a lack of resources, e.g., time and team size to properly secure the instance: *"As much as I don't trust Silicon Valley with my data. I always have to think they have more people working on security than I could have."*[S-15p].
5. **Users.** Users are broadly viewed as a risk to the system. Participants usually see them in a passive role where they fall victim to malware[S-586p], viruses[S-656c], and ransomware[S-710c]. *"I am more afraid of the users being stupid than the box being hacked"*[S-1002p].

Trust Anchors In both the survey and interviews, participants named several factors that alleviate their security concerns. We distinguish these *trust anchors* from actively deployed defensive mechanisms like 2FA, HTTPS, or security training for end users, as they are things out of participants' control or tools they use to assess security. Open-source software is a central trust anchor, because of its transparency, especially if a large community is involved. Here, participants also rely on social proofs to manage their security needs: *"I rely on the community average needs [for security]"*[S-2c]. In this context extensive documentation, including guidelines also builds confidence: *"[I] rely on well-documented software that I can trust"*[S-29p]. Participants are aware about yet undetected software vulnerabilities, but Nextcloud's bug bounty program helps to establish trust. When using third-party hosting providers, users are less worried depending on the server's applicable jurisdiction, e.g. EU[I-32n]. Participants often rely on security audits as feedback channels, e.g., automated scanners[I-5p], or more seldomly on penetration tests[I-31c]. We found that in particular non-professionals widely acknowledge audits as useful: *"That would be what I'd pay for, is a security audit."*[I-5p]. If these certify a good score, it relieves participants' security concerns. However, this can also lead to a false sense of security, e.g., if people rely on outdated or incomplete scanners[I-25c]. Having security knowledge was a trust factor for some[I-28p]. Others shifted admin responsibilities, e.g., by relying on external maintenance, such as NextcloudPi[S-370p], however, often security remains a concern.

Key Takeaways: The security of their operations is a concern for the majority of participants. Regardless of technical expertise, they are creative with identifying potential risks spanning across hardware, software, network, and human factors. Measures that create transparency, and social proofs are important trust anchors.

5.4.3 Perceptions of Defensive Mechanisms

During the interviews, we discussed concrete mechanisms participants use to secure their operations. Across all user groups, participants report following security advice[I-5p], best practices[I-2c], and documentation[I-21n]. One participant wrote their own security mechanisms to protect against and monitor suspicious activity and explained why: *"I want to know what's going on in the software. I only trust software as far as I can see what's happening."*[I-44c] While an overview of all defensive mechanisms is presented in Appendix Figure fig:securitymechanism, we describe selected, controversial ones below. **Firewalls** are very popular with our participants. They use them to separate subnets[I-40n] from each other, and to restrict access from outside to selected ports[I-2c], giving them a secure feeling: *"I don't have to pay attention to what services are running and what ports they have open"*[I-35c]. One person combined them with self-written intrusion detection[I-44c]. While firewalls are *"the most important thing"*[I-1p] for some, others leave them out completely: *"I believe that the firewall issue is simply misunderstood in the vast majority of cases. A firewall [...] only does port filtering in 99% of the installed instances. [...] But that doesn't make any difference if you simply check what else is running on the instance and sim-*

*ply shut down these services"*I-33p. This participant prefers a simple set-up because he had a negative experience with a firewall appliance that broke down and shut him out of his instance when he was abroad. **End-to-end encryption (E2EE)** is requested by many participants, but Nextcloud's implementation is not feature complete [47]. Self-hosters without their own hardware see this as a way to protect their data from unauthorized access by hosting providers*I-32n*. As a consequence of E2EE, participants have concerns about complexity of key management, increased computing load on the server, reduced recovery options in the case of data loss, and users not understanding the mechanism and its implications*I-40n*. Interestingly, participants often have misconceptions about the security benefits E2EE provides over other measures (e.g., over HTTPS*I-5p*, server-side, and hard-drive encryption*I-15c*). **Backups** are considered to be very important by most participants. They often make participants feel safe, even if they are aware of their poor security strategy. Interestingly, this is also the case if participants do not perform regular backups*I-5p*. **Two-factor authentication** is an example of how participants deploy different levels of security on user or instance level*I-13n*, e.g., enforcing 2FA for administrators or users with access to sensitive data*I-33p*. In this context, participants report on challenges explaining the concept of 2FA to non-tech-savvy users*I-1p*. In Nextcloud 2FA is implemented in apps (add-ons). Participants reported being locked out of their instance when the 2FA app did not work after an update*I-2c*: *"If I have to turn off two-factor authentication, I don't think that it means tomorrow somebody's going to get hacked but it means that people have trouble logging in, because all of a sudden their method for logging in changed and that is when people take shortcuts that leads to security risks."*I-2c Because of these issues, some participants reported being reluctant to adopt 2FA.

Key Takeaways: Participants find it difficult to select suitable security mechanisms. There is a tendency to pick supposed "catch-all" defenses, while the actual effectiveness and security guarantees are often unclear.

6 Discussion

We discuss security mindset and practices in relation to administrator constellations and identify high-level gaps in participants' reasoning. Moreover, we relate individual characteristics to participants' server-type choices. Last, we discuss areas of tension and outline recommendations.

6.1 Gaps in Security Mindsets (RQ 3,4)

Participants have contradicting perceptions about the security of self-hosting. Some turn to self-hosting because they believe staying off big commercial clouds is inherently more secure, e.g., because data proximity makes big vendors an attractive target for attackers. Researchers likewise identified this threat in the context of cloud computing [45]. Other participants argue that they cannot compete with the knowledge and resources of big tech companies, thus making their instances inherently less secure than commercial alternatives. While it is difficult to compare security across instances and organizations, research on end user and expert perceptions of threats reveals different levels of abstraction and comprehension [7, 36]. We found gaps and inconsistencies in the security mindsets of both personal and organizational self-hosters. Thereby, neither the technical expertise nor administrator constellations imply a structured approach to security (a.k.a. threat modeling). This suggests, that even experts who are socially embedded into organizations struggle with a systematic approach to security. Understanding gaps in security mindsets will allow academics and practitioners to develop tooling and targeted information sources to help users in securing their instances. Based on our findings, we discuss four major gaps in participants' security mindsets:

(1) Attacker models are often unclear or non-existent. Most participants do not actively model attackers, even if they have the technical expertise. This is true for both single operators and people working in teams in an organizational context. When asked, participants were not confident in identifying potential attackers and which capabilities they would have. Unawareness of possible attackers or their capabilities is detrimental, as it is an essential step in modeling threats and implementing effective security mechanisms.

(2) The data suggests that most participants find it difficult to **prioritize risks**. While participants are in general comfortable identifying potential risks, they tend to lose track in the face of the multitude of potential vulnerabilities. For some, this gives the impression that all efforts are wasted and security cannot be achieved. This is especially an issue for self-hosters who cannot draw on additional resources or who have a limited time budget to secure their operations.

(3) Participants struggled with **identifying defensive mechanism** that are suitable for their use case and set-up. The mapping of perceived risks to defensive mechanisms is especially hard for them, all the more if an understanding of potential attacker models is missing. Most participants were somewhat confident in naming risks they thought could apply to their operations. However, when it came to defensive mechanisms, some take the view of not having enough expertise to judge those. A few (even commercial users) turn to knowledge barter relationships to manage the situation. Others try to find help in online communities. In general, most participants have a *more is better* mindset regarding the deployment of security mechanisms (e.g., wanting E2EE, but not knowing how it would protect them). Only two experts chose an approach to keep their set-up and defensive mechanisms simple (e.g., not deploying a firewall, but making sure ports are closed). Moreover, misconceptions of security benefits can lead to adopting inadequate security practices. For example, it was a frequent notion that data is safe because there are backups, which was occasionally also used to justify a lack of security.

(4) Maintenance, most notably regular updates, are not performed by all participants. Both personal and organizational self-hosters lack adequate maintenance practices, with some not having done updates since installation. This might indicate that some people see security more as a one-time action item, while others view it as a continuous effort.

6.2 Impact of Individual Characteristics and Social-Embeddedness (RQ 1,2)

We found that expertise alone is not enough to predict the server type that people opt for. One might expect that people who have less IT knowledge would prefer managed servers that give them less control, but have dedicated people working on security, thus balancing privacy needs and work effort. However, this is not the case. People who are strongly normative or challenge-driven might go for the on-premise setup even if it potentially causes great struggles for them due to lacking expertise. Similarly, cost constraints may overrule privacy needs and steer people toward hosting providers. Our data suggests, that the server type choices can negatively impact security outcomes, e.g. when cheap server types block software updates. This highlights how *motivation*, *operation*, and *security practices* are connected and we conclude, that we need to take people's motivation and use cases into account when making server-type suggestions.

A major roadblock for people is a lack of expertise and resources, e.g., time, across all user groups. This is in line with research on the complexity and challenges of administrating home networks [9, 13, 20, 28, 62], especially when having to configure network devices [9]. Depending on their social embeddedness, self-hosters choose different ways to overcome their inexperience. *Individual operators* might tap into online communities, while *organizationally embedded operators* sometimes enter *knowledge barter* relationships. No participant voiced any concerns or reported negative experiences regarding these two forms of support. Especially the Nextcloud community was universally described as friendly and helpful, unlike other online communities known for toxic interactions between users [6, 16, 21]. Our data hints at a gender skew towards men in the self-hosting population, although future work needs to validate this.

Another way to overcome lacking expertise and resources, is to go for a server type that requires less maintenance (e.g., with a full-managed hosting provider), or rely on ready-made software solutions to ease the burden of maintenance (e.g., preconfigured or managed docker containers). However, balancing knowledge requirements and automation requires tradeoffs: (1) The goal that people have in mind might not be compliant with outsourcing hardware/software maintenance, (2) Relying on solutions that make adoption easier, might later complicate maintenance. Future work can explore how concrete set-up choices influence roadblocks people have to cope with when adopting self-hosted services, and possible impacts on security practices, such as update experiences.

6.3 Areas of Tension (RQ 5)

Self-hosting can involve a wide range of motivations, use cases, set-ups, admin capabilities, and social embedding. Therefore, no one-size-fits-all solution exists and personalized information sources are necessary. Participants especially need help with securing their operations. Here, even people who have concrete attacker models, talk about problems identifying attackers' capabilities and realistic threat models. Additionally, people often struggle to identify adequate defensive mechanisms. Information sources tailored to their specific use-case and set-up could assist self-hosters in identifying potential attackers, corresponding risks, and defenses. While security scanning tools are valued, participants may require assistance in selecting a reliable one. Automation is often seen as a promising solution to improve security, but it cannot be the sole solution, since self-hosters are responsible for both set-up and maintenance. Yet, tools like set-up wizards might mitigate the major roadblock that is admin capabilities. However, people actively avoid solutions like ready-made docker containers to reduce complexity, both for ease of maintenance and to lower security risks. Attempting to keep things simple and transparent can also pose risks, as exemplified by the participant who writes all security tools themselves.

7 Conclusion

This study explores and connects three dimensions of self-hosting: *motivation*, *operation* in the form of self-hosters' social embeddedness, and *security* mindset. A need for privacy, autonomy, and security together with the belief that self-hosting is the right thing to do are prominent motivational factors. Yet, the decision to self-host is frequently a pragmatic one influenced by cost considerations and the availability of high-quality self-hostable solutions. Motivational factors don't exist in a vacuum but are enabled or constrained by the resources participants can rely on. For instance, participants report varying levels of technical proficiency and work in different *operator constellations*, some of which are deliberately entered to cope with lacking expertise, especially when self-hosting is a hard requirement for them due to their professional-ethical values. Strongly normative-driven self-hosters might opt for server types that allow for a maximum level of hardware and software control, although their lacking expertise turns hosting into a cumbersome task. Others find themselves in the conflict between their need for privacy and cost constraints, causing them to rent third-party servers from large tech companies. Security is often approached in an unstructured fashion. Only a few — even commercial and organizational users with a dedicated admin team — invest in a threat and attacker analysis. Without such an analysis, security features are chosen more spontaneously than reflected.

References

[1] Lilian Adkinson-Orellana, Daniel A Rodríguez-Silva, Felipe Gil-Castiñeira, and Juan C Burguillo-Rial. Privacy for google docs: Implementing a transparent encryption layer. In *International Conference on Cloud Computing*, 2010.

[2] Noa Aharony. An exploratory study on factors affecting the adoption of cloud computing by information professionals. *The Electronic Library*, 2015.

[3] Simon Anell, Lea Gröber, and Katharina Krombholz. End user and expert perceptions of threats and potential countermeasures. In *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, 2020.

[4] awesome-selfhosted (pseudonym). "awesome-selfhosted". Github, https://github.com/awesome-selfhosted/awesome-selfhosted, accessed 2023-02-07.

[5] Benett Axtell and Cosmin Munteanu. Back to real pictures: A cross-generational understanding of users' mental models of photo cloud storage. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2019.

[6] Nicole A Beres, Julian Frommel, Elizabeth Reid, Regan L Mandryk, and Madison Klarkowski. Don't you know that you're toxic: Normalization of toxicity in online gaming. In *Proceedings of the 2021 CHI conference on human factors in computing systems*, 2021.

[7] 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 *USENIX Security Symposium*, 2022.

[8] Melanie Birks and Jane Mills. *Grounded theory: A practical guide*. Sage, 2015. ISBN 9781446295786.

[9] Sara Bly, Bill Schilit, David W McDonald, Barbara Rosario, and Ylian Saint-Hilaire. Broken expectations in the digital home. In *CHI'06 extended abstracts on Human factors in computing systems*, 2006.

[10] Ralf Bohnsack, Nicolle Pfaff, and Wivian Weller, editors. *Qualitative analysis and documentary method in international educational research*. 2010.

[11] Will Brackenbury, Andrew McNutt, Kyle Chard, Aaron Elmore, and Blase Ur. Kondocloud: Improving information management in cloud storage via recommendations based on file similarity. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, 2021.

[12] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006.

[13] AJ Brush. It@ home: Often best left to professionals. In *Position paper for the CHI 2006 Workshop on IT@ Home*, 2006.

[14] Manuel Castells. Informationalism, networks, and the network society: a theoretical blueprint. In Manuel Castells, editor, *The Network Society*. 2004.

[15] Kathy C. Charmaz. *Constructing Grounded Theory*. Sage, 2014. ISBN 9780857029140.

[16] Justin Cheng, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. Antisocial behavior in online discussion communities. In *International AAAI conference on web and social media*, 2015.

[17] Jason W Clark, Peter Snyder, Damon McCoy, and Chris Kanich. " i saw images i didn't even know i had" understanding user perceptions of cloud storage privacy. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 2015.

[18] Gabriele d'Angelo, Fabio Vitali, and Stefano Zacchiroli. Content cloaking: preserving privacy with google docs and other web applications. In *Proceedings of the 2010 ACM symposium on applied computing*, 2010.

[19] Alexander De Luca, Sauvik Das, Martin Ortlieb, Iulia Ion, and Ben Laurie. Expert and non-expert attitudes towards (secure) instant messaging. 2016.

[20] W Keith Edwards and Rebecca E Grinter. At home with ubiquitous computing: Seven challenges. In *Ubicomp: Ubiquitous Computing: International Conference*, 2001.

[21] Mai ElSherief, Shirin Nilizadeh, Dana Nguyen, Giovanni Vigna, and Elizabeth Belding. Peer to peer hate: Hate speech instigators and their targets. In *International AAAI Conference on Web and Social Media*, 2018.

[22] Raspberry Pi Foundation. Raspberry pi web site. https://www.raspberrypi.org/, accessed 2023-07-27.

[23] Gary Garrison, Carl M Rebman Jr, and Sang Hyun Kim. An identification of factors motivating individuals' use of cloud-based services. *Journal of Computer Information Systems*, 58(1), 2018.

[24] Sigi Goode. Keeping the user in the cloud: a cognitive social capital antecedent to use continuance and trust-commitment in personal cloud storage services. *Behaviour & Information Technology*, 38(7), 2019.

[25] René Goscinny and Albert Uderzo. Asterix comics. https://asterix.com/, accessed 2023-02-07.

[26] Mark S. Granovetter. The strength of weak ties. *The American Journal of Sociology*, 78(6):1360–1380, 1973.

[27] Mark S. Granovetter. Economic action and social structure: The problem of embeddedness. *American Journal of Sociology*, 91(3):481–510, 1985.

[28] Rebecca E Grinter, W Keith Edwards, Mark W Newman, and Nicolas Ducheneaut. The work to make a home network work. In *ECSCW: European Conference on Computer-Supported Cooperative Work*, 2005.

[29] Ibrahim Abaker Targio Hashem, Ibrar Yaqoob, Nor Badrul Anuar, Salimah Mokhtar, Abdullah Gani,

and Samee Ullah Khan. The rise of “big data” on cloud computing: Review and open research issues. *Information systems*, 47:98–115, 2015.

[30] Robert W. Helsley and William C. Strange. Knowledge barter in cities. *Journal of Urban Economics*, 2004.

[31] Wenjin Hu, Tao Yang, and Jeanna N Matthews. The good, the bad and the ugly of consumer cloud storage. *ACM SIGOPS Operating Systems Review*, 44(3), 2010.

[32] John Kehayias. Taking back the internet one server at a time, 09 2021. Vice, https://www.vice.com/en/article/pkb4ng/meet-the-self-hosters, accessed 2023-02-07.

[33] Mohammad Taha Khan, Maria Hyun, Chris Kanich, and Blase Ur. Forgotten but not gone: Identifying the need for longitudinal data management in cloud storage. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018.

[34] Mohammad Taha Khan, Christopher Tran, Shubham Singh, Dimitri Vasilkov, Chris Kanich, Blase Ur, and Elena Zheleva. Helping users automatically find and manage sensitive, expendable files in cloud storage. In *USENIX Security Symposium*, 2021.

[35] Klaus Krippendorff. *Content analysis: An introduction to its methodology*, chapter Reliability, page chap. 11. Sage, 2004. ISBN 9780761915454.

[36] Katharina Krombholz, Karoline Busse, Katharina Pfeffer, Matthew Smith, and Emanuel Von Zezschwitz. " if https were secure, i wouldn’t need 2fa"-end user and administrator mental models of https. In *2019 IEEE Symposium on Security and Privacy (SP)*, 2019.

[37] Udo Kuckartz. Qualitative content analysis: From kracauer’s beginnings to today’s challenges. *Forum: Qualitative Sozialforschung / Forum: Qualitative Social Research*, 20(3):Art. 12, 2019.

[38] Nan Lin. Social networks and status attainment. *Annual Review of Sociology*, 25:467–487, 1999.

[39] Steve Lohr. He created the web. now he’s out to remake the digital world, 01 2010. New York Times, https://www.nytimes.com/2021/01/10/technology/tim-berners-lee-privacy-internet.html, accessed 2023-02-08.

[40] Corbin Juliet M. and Anselm L. Strauss. Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative sociology*, 19(6):418–427, 1990.

[41] marcelklehr (pseudonym). Re: Creating a community survey. Nextcloud Forum, https://help.nextcloud.com/t/123092/1, accessed: 2023-01-28.

[42] marcelklehr (pseudonym). Who are we? - take part in the nextcloud community survey. Nextcloud Forum, https://help.nextcloud.com/t/124056/18, accessed: 2023-01-28.

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

[44] Jeremy Mitchell. Hierarchies. introduction. In Grahame Thompson, Jennifer Frances, Rosalind Levačić, and Jeremy Mitchell, editors, *Markets, Hierarchies and Networks*. Sage Publications, 1998.

[45] David Molnar and Stuart E Schechter. Self hosting vs. cloud hosting: Accounting for the security impact of hosting in the cloud. In *WEIS*, pages 1–18, 2010.

[46] Victor Nee. Norms and networks in economic and organizational performance. *American Economic Review*, 88(2):85–89, 1998.

[47] Nextcloud GmbH . Threat model & accepted risks. https://nextcloud.com/security/threat-model/, accessed: 2023-01-28.

[48] Nextcloud GmbH. Nextcloud about page. https://nextcloud.com/about/, accessed: 2023-01-28.

[49] Nextcloud GmbH. Nextcloud online collaboration plattform. https://nextcloud.com/, accessed 2023-01-30.

[50] Douglass C. North. *Institutions, Institutional Change and Economic Performance*. 1990. ISBN 9780511808678.

[51] Axel Philipps and Rafael Mrowczynski. Getting more out of interviews. understanding interviewees’ accounts in relation to their frames of orientation. *Qualitative Research*, 21(1), 2021.

[52] Erika Shehan Poole, Marshini Chetty, Rebecca E Grinter, and W Keith Edwards. More than meets the eye: transforming the user experience of home network management. In *ACM conference on Designing interactive systems*, 2008.

[53] Alejandro Portes and Julia Sensenbrenner. Embeddedness and immigration: Notes on the social determinants of economic action. *American Journal of Sociology*, 98(6):1320–1350, 1993.

[54] Walter W. Powell. Neither market nor hierarchy: Network forms of organization. *Research in Organizational Behavior*, 12:295–336, 1990.

[55] Cenice Prendergast and Lars Stole. Barter relationships. In Paul Seabright, editor, *The Vanishing Rouble*. Cambridge University Press, 2000.

[56] Margaret R. Roller. A quality approach to qualitative content analysis: Similarities and differences compared to other qualitative methods. *Forum: Qualitative Sozialforschung / Forum: Qualitative Social Research*, 20(3):Art. 31, 2019.

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

[58] Anselm L. Strauss and Juliet M. Corbin. *Grounded theory in practice*. Sage, 1997.

[59] Dan Svantesson and Roger Clarke. Privacy and consumer risks in cloud computing. *Computer law & security review*, 26(4):391–397, 2010.

[60] Nestori Syynimaa and Tessa Viitanen. Is my office 365 GDPR compliant?: Security issues in authentication and administration. In *International Conference on Enterprise Information Systems*, 2018.

[61] Madiha Tabassum, Tomasz Kosinski, and Heather Richter Lipford. I don't own the data: End user perceptions of smart home device data practices and risks. In *SOUPS, Symposium on Usable Privacy and Security*, 2019.

[62] Peter Tolmie, Andy Crabtree, Tom Rodden, Chris Greenhalgh, and Steve Benford. Making the home network at home: Digital housekeeping. In *European Conference on Computer-Supported Cooperative Work*, 2007.

[63] u/muchTasty (pseudonym). r/selfhosted: Beginner guide: How to secure your self-hosted services. Reddit, https://www.reddit.com/r/selfhosted/comments/pufhs0/beginner_guide_how_to_secure_your_selfhosted/, accessed 2023-02-07.

[64] C. Urquhart. *Grounded theory for qualitative research: A practical guide*. Sage, 2012.

[65] Brian Uzzi and Ryon Lancaster. Embeddedness and price formation in the corporate law market. *American Sociological Review*, 69(3):319–344, 2004.

[66] Mojtaba Vaismoradi and Sherrill Snelgrove. Theme in qualitative content analysis and thematic analysis. *Forum: Qualitative Sozialforschung / Forum: Qualitative Social Research*, 20(3):Art. 23, 2019.

[67] Marten Van Dijk and Ari Juels. On the impossibility of cryptography alone for privacy-preserving cloud computing. *HotSec*, 10(1):8, 2010.

[68] Lucian L Visinescu, Olajumoke Azogu, Sherry D Ryan, Yu "Andy" Wu, and Dan J Kim. Better safe than sorry: A study of investigating individuals' protection of privacy in the use of storage as a cloud computing service. *International Journal of Human–Computer Interaction*, 32(11), 2016.

[69] Dominik Wermke, Christian Stransky, Nicolas Huaman, Niklas Busch, Yasemin Acar, and Sascha Fahl. Cloudy with a chance of misconceptions: exploring users' perceptions and expectations of security and privacy in cloud office suites. In *SOUPS, Symposium on Usable Privacy and Security*, 2020.

Appendix

A Background: Social Science and Sociology

Our analysis draws on social-scientific concepts and ideas to understand the broader social contexts of self-hosting. In addition to an IT dimension, self-hosting also entails various forms of interactions between humans taking place under specific social-structural conditions.

Social embeddedness encapsulates the idea that all individual human actors (also self-hosters) are involved in various social relations [27, 53, 65]. We distinguish between two main forms of social embeddedness found in self-hosting: (1) *Digitally mediated interaction* are all sorts of social interactions mediated by IT. Examples are sharing photographs with family members or collaborating on a paper draft. Here, social embeddedness means the social constellation at large in which given digitally mediated interactions take place. (2) *IT operation*, focuses on interpersonal and broader social constellations in which activities, specifically aimed at the operation of the IT infrastructure, take place. These two dimensions can overlap since IT operators often use digital means of communication to coordinate their activities.

Organizational embeddedness is a focal point in self-hosting. Herein, we understand organizations as clearly defined and coherently acting groups of humans that are meant to exist over longer periods of time, pursuing a specified goal. They also have sets of explicit behavioral rules for their members, e.g., duties, explicit commands, membership fees, or general loyalty expectations. Very often, organizational rules include hierarchical relations [44, 50]. However, even most formalized organizations are full of informal relations. Some of those may improve professional interactions, while others can circumvent or even undermine official goals [46].

Collaborative networks is a concept that builds on the broader notion of "social networks" in sociology [14, 26, 38, 54]. It denotes frequent interactions between social actors (individuals or groups) based on relevant characteristics of these actors who cooperate without forming an official organization. For example, a particular game programmer and a particular graphic designer frequently cooperate on different projects because of mutual trust in their abilities "to do a good job", but without establishing a formal organization.

Knowledge Barter describes an exchange relationship between at least two social actors (usually individuals) who directly trade knowledge-based services (assistance) without using money as a transactional medium [30]. It is a derivative of the broader economic-sociological term *barter* used to denote moneyless exchanges of goods [55]. Knowledge-barter exchanges often employ a *delayed reciprocity*: actor A does not immediately reciprocate a helping act by actor B, but rather offers their assistance to B when the latter is really in need of it and vice versa – often described as "helping each other out" or "exchanging favors." This concept is a different interaction than "community participation" (e.g., in an open-source community), since it refers to relationships between specific individuals who perceive their mutual obligations in a personalized manner, i.e. as obligations vis-à-vis specific humans and not vis-à-vis an entire community as such.

ID	Use Case	Country	Sector of Operation	Occupation	IT Occupation	Hosting Occupation	IT Background	Sec Background	Servertype	Motivation	Social Emb.
13	Non-Profit	Germany	Research Institute	IT System Administrator	●	●	●	○	Private Data Center	A, U, C	Org:Team
17	Non-Profit	France	Freelancer Association	∅	●	●	●	○	Virtual Private Server	N, C	Coll. Network
20	Non-Profit	Germany	University	Developer/ System Administrator	●	●	●	●	Private Data Center	U, P	Org:Team
21	Non-Profit	Germany	Art Preservation	Art Conservator	○	○	⊙	⊙	Virtual Private Server	U, A, P	Org:Team
23	Non-Profit	Germany	Bicycle Club	IT Consultant	●	●	●	⊙	Private Data Center	C, P	Org:Team
24	Non-Profit	Italy	EU Project	Technical Translator	●	○	⊙	○	Virtual Private Server	N	Org:Team
26	Non-Profit	Germany	Theater Club	Student	○	○	●	○	Virtual Private Server	C	Org:Sole
29	Non-Profit	Germany	Sports Club	CTO in Telecom Company	●	○	●	○	Virtual Private Server	P, A, C	Org:Sole
30	Non-Profit	Germany	School	Media Designer	○	○	⊙	○	Virtual Private Server	C, N, U	Org:Sole
32	Non-Profit	Spain	School	Teacher	○	○	●	⊙	Virtual Private Server	P, N	Individual
36	Non-Profit	Germany	Crisis Line	Professor Computer Science	●	○	●	○	Dedicated Server	U, N, A	Org:Sole
37	Non-Profit	Slovenia	Dataprotection Community	Web Developer	●	○	●	○	Commercial Server	P, C, A	Coll. Network
40	Non-Profit	Switzerland	Political Party	System Engineer	●	●	⊙	⊙	Commercial Server	P, S, A	Individual
2	Commercial	U.S.	Production	CEO	○	○	●	●	Commercial Server	C, A, S, P	Individual
12	Commercial	Germany	IT Consulting	IT Consultant	●	●	●	⊙	Dedicated Server	N, P, A	Org:Sole
14	Commercial	Germany	Law Firm	Lawyer	○	○	○	○	Commercial Server	U, A, P	Knowledge Barter
15	Commercial	Sweden	Journalism	Investigative Journalist	○	○	⊙	○	Co-location	A, P	Knowledge Barter
22	Commercial	France	Consulting	Public Policy Consultant	○	○	○	○	Commercial Server	U, N	Individual
25	Commercial	Canada	Consulting	Consultant	○	○	●	○	Home Server	U	Individual
31	Commercial	Netherlands	Production	System Administrator	●	●	●	⊙	Commercial Server	N	Org:Sole
34	Commercial	France	Travel Agency	Tour Guide	○	○	○	○	Shared Hosting	F, N	Org:Sole
35	Commercial	Germany	Media Design	Freelancer	●	○	⊙	○	Shared Hosting	P, U	Org:Sole
44	Commercial	Netherlands	IT Consulting	IT Support	●	●	⊙	⊙	Virtual Private Server	U, A, P, N	Org:Sole
45	Commercial	Netherlands	Architecture	IT Professional	●	●	●	○	Private Data Center	F, A, U	Org:Sole
1	Personal	New Zealand	∅	IT Project Manager	●	○	●	⊙	Home Server	A, P, F	Individual
3	Personal	U.S.	∅	Software Engineer	●	○	●	○	Virtual Private Server	P, A, N	Individual
4	Personal	U.S.	∅	Networking Systems Engineer	●	●	●	●	Virtual Private Server	C, A, S	Individual
5	Personal	U.K.	∅	Teacher	○	○	○	○	Home Server	F, A, N	Individual
6	Personal	Italy	∅	Student	●	○	●	●	Home Server	F, P, A, S	Individual
7	Personal	Germany	∅	Doctoral Student	○	○	⊙	○	Home Server	U, F	Individual
8	Personal	Germany	∅	Cloud Architect	●	●	●	●	Home Server	N, P, S, U	Individual
9	Personal	Czech Republic	∅	Data Specialist	●	○	●	○	Home Server	U, P, A	Individual
10	Personal	Germany	∅	IT Administrator	●	●	●	○	Home Server	P, A, C, F	∅
11	Personal	Hungary	∅	DevOps Engineer	●	●	●	⊙	Virtual Private Server	P, S	Individual
19	Personal	Finland	∅	Kernel Programmer	●	○	●	●	Home Server	U, A, P	Individual
28	Personal	U.S.	∅	Software Engineer	●	○	●	●	Home Server	U, A	Individual
33	Personal	Germany	∅	IT Consulting	●	●	●	●	Home Server	F, N	Individual
38	Personal	U.S.	∅	Software Engineer	●	○	⊙	○	Virtual Private Server	N, U	Individual
39	Personal	U.S.	∅	System Engineer	●	●	⊙	○	Virtual Private Server	U	Individual
41	Personal	Germany	∅	Journalist	○	○	●	○	Home Server	F, N	Individual
43	Personal	France	∅	Software Engineer	●	○	⊙	○	Home Server	U	Individual

Table 3: Interview demographics (four interviews were excluded as they did not match our criteria). ∅=no answer given; Self-reported IT proficiency: *IT-related Occupation*, *Hosting-related Occupation*, *IT Background*, *Security Background* where ●=yes, ⊙=self-taught, ○=no. *Motivational factors* according to Section 5.1: N = Normative, P = Privacy, A = Autonomy, S = Security, C = Cost, U = Use Case, F = Personal Challenge or Fun. *Social Embeddedness* c.f. Section 5.2: Individual = individual operators with family & friends, Org:Sole = organizationally-embedded sole operators, Org:Team = team members within organizations, Coll. Network = collaborative networks

B Demographics

B.1 Interview

We analyzed 41 interviews, from 40 men & one woman, who self-host primarily in a personal (N_p=17), commercial (N_c=11), or non-profit (N_n=13) context. Participants come from 16 countries across Europe, North America, and Oceania (see Table 3). Ages range from 23 to 62 years, with an average and median of 42 years. The participants' professional background is broad and ranges from non-technical occupations (N=13: teachers, journalists, lawyers, art conservator, travel agency, media designer, consultants), over generally IT-related (N=13: developers, data specialists, it consultants, computer science professor), to hosting-related occupations (N=15: system administrators, system engineers, IT-support). Likewise, the participants' educational background has a similar broad spread and partly reflects different educational opportunities that were available at the time caused by the wide age range. For 10 participants, high school is their highest level of edu-

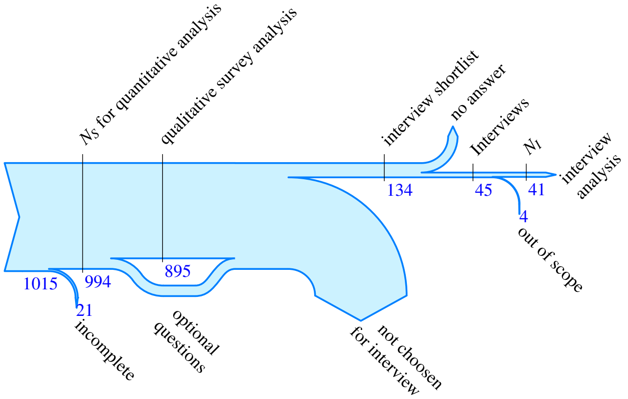

Figure 4: Overview of survey and interview population.

cation, four have completed an occupational apprenticeship, and 27 have a university education (BS, MS, PhD, diploma, state examination). While 26 participants report having an educational IT background, 11 people claim to be self-taught, and four say they have no technical background. The distribution (edu/self/none) across user groups is as follows: personal (12/4/1), commercial (5/3/3), and non-profit (9/4/0). Moreover, we asked participants about their security background. Eight participants reported having a security background either obtained through education or extensive work experience, Nine reported being self-taught, and 24 said they had no background.The distribution across use cases is as follows: personal (6/2/9), commercial (1/3/7), and non-profit (1/4/8).

For organizational use of self-hosting, we cover a broad spectrum of different industries. The commercial users split into an architectural office, consulting agencies, a travel agency, a law firm, journalists, media production, and companies producing physical goods. The non-profit users are a research institute, a university, schools, a political party, an art collective, a freelancer association, different clubs (theater, bicycles, sports), a European-Union-sponsored project, a data protection community, and a crisis helpline.

Participants provide their operations on diverse set-ups (Figure 7). In the interviews, they told us about everything from under-the-table Raspberry Pis [22]$^{I\text{-}5p}$, repurposed or upcycled hardware$^{I\text{-}28p}$, to private data centres$^{I\text{-}20n}$, and hosting on third-party clouds$^{I\text{-}2c}$.

B.2 Survey

The survey indicates that the majority of instances are administrated by single admins (per user group: com 60.37%, p 87.09%, np 55.79%, gov 42.28%). For organizational self-hosting, admin teams between two and three people are common, however, the biggest team with 11 admins was reported by a personal self-hoster. It is not a given that admin teams have members with a security background (on average per user group: com 38.33%, p 49.59%, np 50.07%, gov 80.09%). For single admins, less than half of the people report having a security background (per user group: com 42.18%, p 32.76%, np 31.16%, gov 16.66%). The majority of participants self-hosts at least one service in addition to Nextcloud (per user group: com 66.98%, p 65.89%, np 67.39%, gov 71.42%).

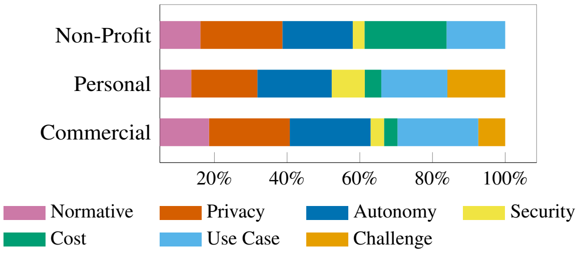

Figure 5: *Interview data:* Relative frequencies of reported motivational factors across user groups.

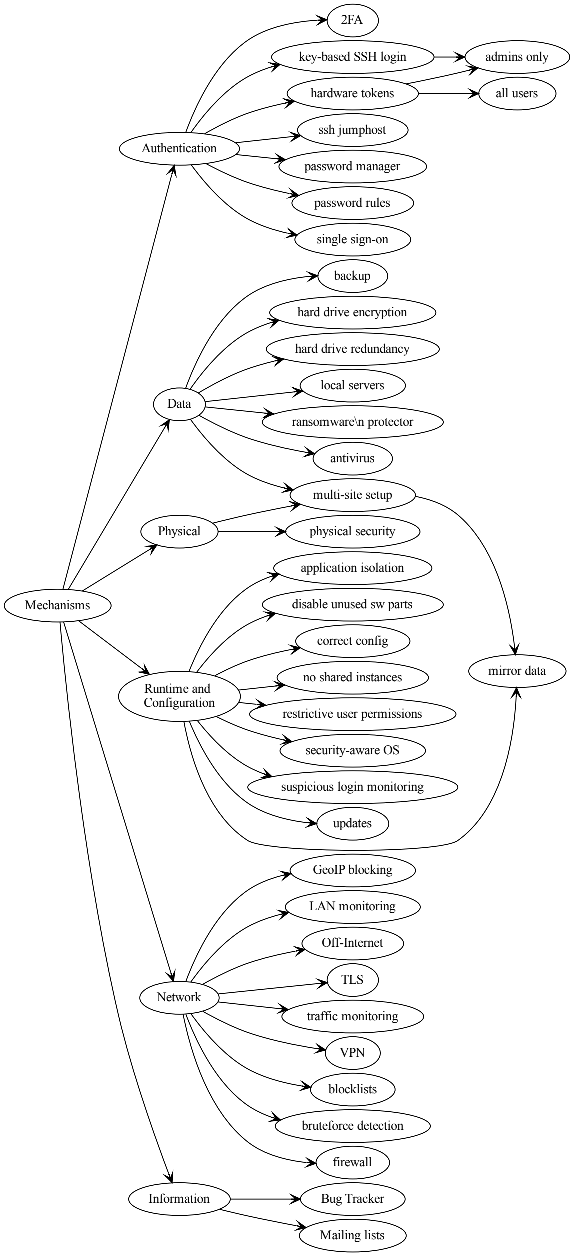

Figure 6: Excerpt of interview code book for security mechanisms deployed by the participants

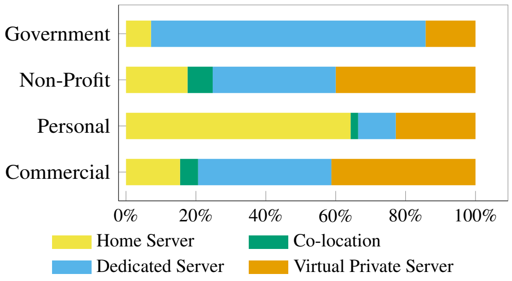

Figure 7: *Survey data:* Relative frequencies of reported server types across user groups.