Strategies and Perceived Risks of Sending Sensitive Documents

Noel Warford, *University of Maryland;* Collins W. Munyendo, *The George Washington University;* Ashna Mediratta, *University of Maryland;* Adam J. Aviv, *The George Washington University;* Michelle L. Mazurek, *University of Maryland*

https://www.usenix.org/conference/usenixsecurity21/presentation/warford

This paper is included in the Proceedings of the 30th USENIX Security Symposium.

August 11–13, 2021

978-1-939133-24-3

Open access to the Proceedings of the 30th USENIX Security Symposium is sponsored by USENIX.

Strategies and Perceived Risks of Sending Sensitive Documents

Noel Warford[‡], Collins W. Munyendo[§], Ashna Mediratta[‡], Adam J. Aviv[§], and Michelle L. Mazurek[‡]
[‡] *University of Maryland,* [§]*The George Washington University*

Abstract

People are frequently required to send documents, forms, or other materials containing sensitive data (e.g., personal information, medical records, financial data) to remote parties, sometimes without a formal procedure to do so securely. The specific transmission mechanisms end up relying on the knowledge and preferences of the parties involved. Through two online surveys ($n = 60$ and $n = 250$), we explore the various methods used to transmit sensitive documents, as well as the perceived risk and satisfaction with those methods. We find that users are more likely to recognize risk to data-at-rest after receipt (but not at the sender, namely, themselves). When not using an online portal provided by the recipient, participants primarily envision transmitting sensitive documents in person or via email, and have little experience using secure, privacy-preserving alternatives. Despite recognizing general risks, participants express high privacy satisfaction and convenience with actually experienced situations. These results suggest opportunities to design new solutions to promote securely sending sensitive materials, perhaps as new utilities within standard email workflows.

1 Introduction

Users are often required to send sensitive information — such as personally identifiable information (PII), medical information, or financial information — to remote parties. The approaches people use to send this information can vary based on personal skill level, available tools, the situational context in which this information is required, and, importantly, the perceived sensitivity of the data involved and the trust in the remote party receiving the data [36, 44, 47].

Significant prior work has focused on why users do (or do not) adopt specific private communications channels, such as end-to-end encrypted messaging, as well as how to make these channels more usable and transparent [2, 5, 11, 34, 35]. However, users who are required to send specific sensitive information to possibly unfamiliar recipients, perhaps in a new context, may not have the same tools at their disposal, or are unaware of their availability or applicability. Little is known about how or why people choose *specific channels* for secure or private transmission of sensitive data.

In this paper, we explore how users cope when required to send sensitive information in the digital age. In particular, we sought to answer three key research questions:

RQ1: What methods do people choose when sending sensitive information, and why?
RQ2: Are participants satisfied with their current approaches, particularly in terms of whether they offer sufficient privacy? Why or why not?
RQ3: What risks are people most concerned about when sending sensitive information?

To address these questions, we conducted two online surveys. In the first survey (Survey 1, $n = 60$), we asked participants to provide primarily open-ended responses about the communication methods they used, or would expect to use, to send sensitive documents. We asked for responses to nine different scenarios, such as applying for a mortgage or an apartment, or opening a bank account. Participants reported on their satisfaction with the transmission methods, from both privacy and convenience perspectives, as well as their perception of potential risks and ways to mitigate these risks. Twenty participants responded to each scenario, and participants described 11 different methods, including delivering documents in person, physical mail, email, fax, and direct messaging.

We designed a second survey (Survey 2, $n = 250$), containing predominantly closed-item questions with answer choices derived from Survey 1 responses. While Survey 1 was scenario-driven, Survey 2 was method-driven. Participants identified at least one of eight most frequently cited methods from Survey 1 that they had used successfully to transmit sensitive information. They were then asked to describe a specific situation where one method was used successfully, followed by multiple-choice and Likert-type responses about the people and data involved and their satisfaction with the method. We also asked about privacy and risk, such as the

comparative risk at the end-points or in transit.

We find that many participants typically deliver sensitive information using "offline" means — most frequently in person, but also via physical mail and phone calls. Unsurprisingly, the most common digital approach is to use online forms or portals provided by the recipient; many participants also use standard (unencrypted) email. For a few Survey 1 tasks, such as sharing a password, there was a higher preference for direct messaging or phone calls, but little appetite for using secure technologies. Survey 2 indicates that while participants have largely heard of secure technologies, relatively few have used them. These results suggest that if a predetermined online form is not available or not appropriate, email is the only other widely used digital option. Nonetheless, in both surveys the vast majority of participants expressed satisfaction with both the convenience and privacy of their method.

Both surveys revealed interesting patterns in participants' perceptions of risks. Survey 1 participants, answering open-ended questions, did not prioritize the risk to sensitive information *in transit*, but rather *what happens after it arrives*, either due to malicious action by the recipient or simply because the recipient did not take appropriate care with the data. When prompted with multiple choice questions, Survey 2 participants weight in-transit risks and risks at the recipient similarly, but discount risk to the data at the sender (namely, themselves, e.g., whether their own email storage is at risk).

Our findings illuminate opportunities to both improve end-user education and design new, transparent solutions for securely sending sensitive information. These tools could include building connections to secure-document-sending into existing communications modes like email, and improving retrospective privacy tools to help people delete sensitive content persisting at-rest once they are no longer needed.

2 Related Work

This paper builds on extensive research on secure communication. In 1999, Whitten and Tygar's classic paper, *Why Johnny Can't Encrypt: A Usability Evaluation of PGP 5.0* [45], described numerous user-facing issues that make encrypted email impractical for many users, and similar problems persist in PGP 9.0 [38]. Follow-up research suggests that usability challenges in encrypted email continue [5, 15, 18, 34, 42], despite many attempts to automate the process [35].

More recently, secure, end-to-end messaging applications (e.g., Signal, WhatsApp, Telegram) have proliferated as a straightforward and transparent way for users to communicate privately. Secure messaging adoption is driven largely by peer influence, rather than its security properties [2, 11], and users may have misconceptions about the security properties, sometimes believing outside parties can read these encrypted communications, or that methods like SMS are *more* secure [1, 2]. Inaccurate mental models of security may contribute to these misconceptions [48], and we also find that our participants do not strongly grasp secure communication.

A particular challenge in secure communication is how to indicate when transmissions are (not) secure. Numerous researchers have investigated the effectiveness of different indicators, including website authentication indicators [7, 30, 37, 41] and phishing warnings [4, 27]. Making these indicators intelligible and noticeable, without impeding workflows or engendering habituation, remains an open challenge.

Even after transmission, data may continue to reside on servers, at the sender and/or receiver. Cloud storage poses a particular problem [39] as many users have incorrect models of the longevity and location of cloud-based information. Clark et al. [8] and Khan et al. [17] found that when shown data stored in the cloud, most users find at least one item they wish to delete. Users also lack urgency to delete cloud-stored information [31] and express interest in tools designed to do this [24]. This contradicts a common user perception that they have nothing to hide [40]. This problem is only amplified over time, with an increasing number of messaging platforms and other services relying on cloud storage and computing.

Difficulty protecting information may arise in part because users (and even experts) often have difficulty defining what information is sensitive in the first place [40]. Different users may also have different standards for what does or does not fall into the category of sensitive information, as this is highly reliant on context and personal preference [36, 44, 47].

Researchers have also investigated how people learn about digital security and privacy, and how they develop associated behaviors. People's mental models for security often focus on direct and visible threats [14, 44], and people tend to adopt behaviors based on where the behavior was learned, rather than its content [32]. As might be expected, convenience-security tradeoffs also play an important role in adoption of security behaviors [13]. Other work suggests that social factors, such as observing others performing a particular security behavior, can motivate users to take more security or privacy precautions [9, 10]. We observe that many of these factors — convenience, social expectations, and directness of expected threat — also play a role in our participants' choices when sending sensitive information.

3 Methods

We designed and conducted two online surveys exploring participants' current practices and perceptions related to sharing sensitive documents. In Survey 1, $n = 60$ participants commented on three (randomly selected from nine) scenarios where they would need to communicate sensitive information or documents to another party. Questions in this survey were primarily open-ended, in order to obtain a wide range of responses about participants' experiences and perceptions.

Survey 2 builds on the results of Survey 1 with a larger sample, $n = 250$. Participants were randomly assigned to answer questions about their experiences with and perceptions of one transmission method they had successfully used. Survey 2 used primarily closed-item questions with answer choices drawn from the qualitative analysis of Survey 1. Both surveys were approved by the University of Maryland IRB, and participants were recruited using Prolific.

3.1 Survey 1

Survey 1 consisted of four sections, described below. [1]

1. *Instructions*: Participants were briefed about the purpose of the survey and provided consent.
2. *Scenarios (x3)*: Each participant was surveyed about three different, randomly chosen scenarios in which someone might send sensitive documents. For each scenario, participants were asked whether they had experienced the scenario before, and if so, how they provided the required information. Alternatively, participants were asked to *imagine* how they would transmit the information in such a scenario. We also asked, on a five-point Likert scale for each scenario, about their satisfaction with the communication method overall and in terms of privacy/security and convenience. Each closed-item question had an open-ended followup question.
3. *Risks and Mitigation*: Participants were asked to identify and describe two risks (or concerns) with providing sensitive documents and two precautions they would take to reduce those risks, as well as if they have ever taken these precautions. All these questions were open-ended.
4. *Demographics*: Finally, we asked about demographics, including IT/CS background, income level, and experience working with a security clearance or in a sector with data privacy regulations (e.g., health care, law). Other demographic information was obtained directly from Prolific rather than via survey questions.

Transmission scenarios We developed nine scenarios for sending sensitive documents based on vignettes used in prior work [2, 47] and based on the authors' anecdotal experiences. Scenarios included: applying for a mortgage, sharing a password, participating in a background check (e.g., for volunteering with children), applying for an apartment, creating a checking account, sharing a password-protected document, enrolling a child in a new school, seeing a new doctor, and sending financial documents to a tax accountant. Each participant viewed three scenarios, randomly selected with counterbalancing, resulting in 20 participant responses per scenario. Full text descriptions of each are provided in Appendix A.

[1] The full questionnaire is given in the extended paper (see Appendix C).

Updating for current events Survey 1 was administered in two rounds, before COVID-19 and after. In round two, as part of the scenario section, we asked two additional questions about whether the participant had experienced the scenario before or after social distancing and whether social distancing had changed their (real or imagined) approach.

3.2 Survey 2

We designed a second survey (Survey 2) to explore some of the results of Survey 1 in more detail. In contrast to Survey 1, which was structured around scenarios, participants in Survey 2 were randomly assigned to answer questions about particular transmission methods, later describing a *real* scenario in which they had used that method. [2]

We considered eight methods that participants in Survey 1 commonly reported: email, direct messaging, in-person, online form or portal, document sharing service (e.g., Dropbox), phone call, fax, and physical mail. A participant would first identify which of these methods they had successfully used to transmit sensitive information in the past. We also asked participants to specify any other unlisted methods they had used in a follow-up free-response question.

We then assigned the participant one of their "successful" methods, with the rest of the survey relating to that method. Only methods that were included in our initial list were used for further questions, in order to ensure standardized and consistent questions across conditions. We continually weighted the random assignment of successful methods toward less popular methods based on the results of Survey 1 and the current Survey 2 recruitment in order to keep distribution among methods relatively even.

We asked participants to recall a specific scenario where they successfully used the assigned transmission method and describe the type of information sent, the recipient, and why this method was selected (e.g., did they or the recipient choose it?). These questions were closed-item, with answer choices based on common answers in Survey 1 and an option to write in an "other" response. As in Survey 1, we also asked about privacy and convenience satisfaction using a Likert scale.

We then asked other questions about privacy and risk, with answer choices also drawn from themes we identified in Survey 1. These included potential risks such as a recipient revealing data by accident or on purpose, as well as interception in transit. We also asked about whether the participant believed the recipient could keep their data safe, whether the participant could themselves take action to keep their data safe, and whether the information would be received by the intended recipient. We asked about the likelihood of specific risks, including reputational damage, physical harm, and identity theft. Finally, we asked about the level of risk to the data at the sender, at the recipient, and in transit using Likert-type scales. Lastly, we collected the same demographics.

[2] The full questionnaire is given in the extended paper (see Appendix C).

Table 1: Reliability statistics for qualitative coding, including number of rounds required to reach agreement.

Question	Rounds	Alpha
Survey 1		
Methods used to send	2	0.94
Satisfied with method	2	1.00
Satisfied with privacy	1	0.93
Satisfied with convenience	1	0.86
Potential risks	3	1.00
Survey 2		
What is being sent - Other	1	0.82
Methods used to send - Other	1	0.92

3.3 Recruitment

We recruited via Prolific, and participants were required to reside within the U.S., have a 95% approval on Prolific, be at least 18 years old, and self-report fluency in English. We used free-response questions to validate participants' answers were on-topic and responsive, discarding only one potential participant in Survey 1 and six in Survey 2.

We recruited $n = 60$ participants for Survey 1 and $n = 250$ for Survey 2. Participants were compensated $4.00; Survey 1 took on average 17.4 minutes, while Survey 2 averaged 11.7 minutes. Survey 1 data collection took place in early February and then May 2020, Survey 2 data collection took place in November and December 2020.

3.4 Data Analysis

For most open-ended answers (primarily but not exclusively Survey 1), we used an open-coding content analysis approach [20]. Two researchers worked together to develop an initial codebook for each question, using 10% of the provided answers. They then independently applied the codebook to an additional 10% of the data per round, iteratively updating the codebook between rounds until strong reliability (Krippendorff's Alpha ≥ 0.8) was obtained [19, 22]. At that point, all data was recoded by a single coder using the final codebook. Reliability values are given in Table 1. For open-ended questions with 20 or fewer responses, this approach was impractical; instead, two researchers coded each answer collaboratively. [3]

We pre-planned our quantitative analysis for Survey 2 around an ordinal logistic regression designed to identify factors associated with privacy satisfaction (on a five-point Likert scale) [23]. We tested a range of potential covariates, selecting a final model based on minimum Akaike Information Criterion (AIC) [3]. Complete details are given in Section 4.2.

For other comparisons of Likert-type variables, we use Kruskal-Wallis H-tests to identify differences among three or more items, followed by post-hoc Mann-Whitney U (MWU) pairwise-tests with the Holm-Šidák correction.

[3] See extended paper (Appendix C) for complete codebooks.

3.5 Limitations

Our study has a number of limitations typical of exploratory survey research. First, data in Survey 1 was collected without the opportunity to ask follow-up questions (as would be the case in semi-structured interviews). As a result, some coded responses may not fully portray the nuances of participants' methods and perceptions. To compensate, we designed Survey 2 to validate those results with a larger population.

Free-response questions may suffer from satisficing, in which participants mention the first item that comes to mind rather than answering comprehensively [21]; participants who fail to mention something may simply not have included it, rather than explicitly disagreeing. As such, counts of participants should be considered a lower bound reflecting top-of-mind concerns, rather than absolute prevalence. Survey 1 also asked participants to imagine actions if a scenario was unfamiliar, which could also lead to satisficing, as well as other biases related to self-reporting and imagining hypotheticals. Our design for Survey 2 sought to address this by only asking in depth about successfully used transmission methods, so that participants could report on real experiences instead of imagined ones.

Data collection in Survey 1 was bifurcated due to COVID-19, potentially biasing participant responses. We added questions in the second round of Survey 1 addressing COVID-19 and found few differences, and so we did not focus on COVID-19 effects in Survey 2.

There are inherent limitations in using crowdsourcing platforms like Prolific. Prior work has shown that these platforms provide reasonable samples for security- and privacy-relevant questions [33], and Prolific has been shown to provide high-quality crowdsourced data [25].

We focused only on U.S. participants, as we are most familiar with common data transmission scenarios in the U.S. Our participant recruitment, as is generally the case for crowdsourcing platforms, tended to be more male and younger than the U.S. population as a whole. We neither expect nor claim the data to be fully representative; however, we believe we obtained a reasonably broad view of transmission approaches and associated perceptions in the U.S. Future work could examine similar norms in other countries and cultures.

4 Results

We first report on our participants, and then the results of both surveys. The results are organized by research question, as defined in Section 1.

Participants Demographics for both surveys are provided in Table 2 and are based on both self-reported data provided

Table 2: Demographics of participants in both surveys. Excludes "no answer" and "prefer not to say" options. "Sensitive Information" indicates whether a participant had encountered the listed types of information in a professional context.

		S1#	S2#	S2%
Gender	Female	21	111	44.4
	Male	39	132	52.8
	Non-binary	0	7	2.8
Age	18–30	38	84	33.6
	31–40	13	68	27.2
	41–50	6	35	14.0
	51–60	1	17	6.8
	61+	1	10	4.0
Income	< $50K	23	121	48.4
	$50K-$100K	24	86	34.4
	> $100K	11	35	14.0
Education	No high school	N/A	4	1.6
	HS or equiv.	N/A	72	28.8
	Bachelor or associate	N/A	121	48.4
	Advanced degree	N/A	52	20.8
CS Experience	No	45	190	76.0
	Yes	12	54	21.6
Security Clearance	No	55	214	85.6
	Yes	2	19	7.6
Sensitive Information	Credit card	18	90	36.0
	HIPAA	18	62	24.8
	Social Security number	17	85	34.0
	FERPA	6	20	8.0

Table 3: Distribution of participants across methods (Survey 2). Participants were randomly assigned one method among those they reported having used successfully.

Method	#Part.
In person	37
Online form/portal	35
Email	31
Physical mail	32
Phone call	34
Document sharing service	27
Fax	30
Direct messaging	24

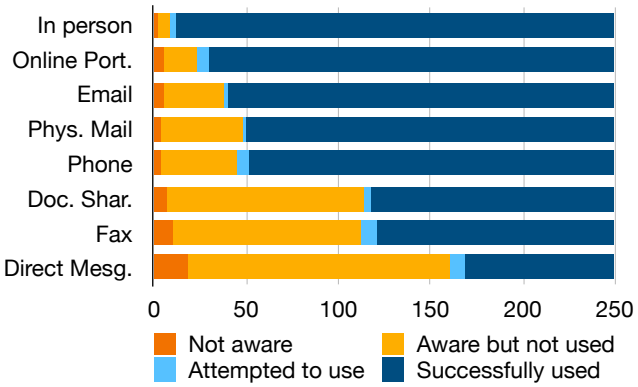

Figure 1: Methods previously used by participants to send sensitive information (Survey 2)

by Prolific and on direct questions from our survey. Most participants do not have a CS background (75% Survey 1, 76% Survey 2) and few previously (or currently) have a security clearance (two and one participant in Survey 1 and Survey 2, respectively). Many describe having worked in roles where they may have handled sensitive information, e.g., Social Security numbers or health information. Participants skew younger and more male, as noted in Section 3.5.

Participants were evenly and randomly distributed among information-transmission scenarios in Survey 1; 20 participants per scenario. We used frequency weighting to partially balance assignment to transmission methods in Survey 2. The distribution of participants across methods is given in Table 3.

4.1 RQ1: What methods are used and why?

Survey 1 In Survey 1, participants provided open-ended answers about how they sent required information in different scenarios. If they had not experienced the scenario, we asked them to imagine how they would send information in the scenario. We refer to these as *real* and *imagined* responses, respectively. Table 4 describes each identified mode of transmission, as defined in our codebook, with frequency of occurrence across all scenarios. (Table **??** in Appendix B provides the most common transmission methods per scenario for both real and imagined instances.)

By far, the most commonly reported transmission methods were taking the documents in person and sending the documents via email, especially for imagined scenarios. Online forms, direct messages, and phone calls were also common methods, and some responses indicated non-digital transmission methods. For example, P25 made an unprompted reference to not trusting digital methods: "I would fax the documents to them simply because I do not trust sending that information via the internet."

From Survey 1 responses, we wondered whether these methods, which were clearly top of mind, were also the methods participants had the most experience with. We also wondered whether participants knew about certain modes but had chosen not to use them, or were unfamiliar with them at all. Responses also suggested that participants frequently used methods chosen (or required) by the recipient, rather than choosing the method independently. We addressed these questions as part of Survey 2.

Table 4: Transmission methods reported by participants in Survey 1, across scenarios. Counts are provided for all scenario instances, and broken down by real and imagined instances. Participants sometimes indicated more than one method per instance.

	Count				
Code	**Total**	**Real**	**Imag.**	**Description**	**Quote**
In person	85	54	31	Delivering the information by hand to the recipient, whether written down or simply told to them	"I provided the information on an application in the office. It was on paper and when I was done I handed it to them"
Email	52	24	28	Sending the information via email, regardless of email platform or encryption	"I sent the password to the persons private email that I knew for a fact was only accessible by only them."
Online form or portal	32	30	2	Using an institution's site, app, or portal to upload the information	"I applied for a savings account recently, but I did it through their mobile app. Really they had all of my information, but they did ask to confirm questions like social security number and contact information."
Direct messaging	22	14	8	SMS, secure and insecure messaging services, and other similar modes of communication	"I would text them the password and tell them to delete it from their phone after they are done."
Phone call	19	9	10	A direct telephone call	"I provided it to him over a phone call. I do not trust electronic devices with password sending."
Fax	8	3	5	Faxing documents to the recipient	"I would fax the documents to them simply because I do not trust sending that information via the internet."
Sending online (unspec.)	8	6	2	Sending the information online without providing a specific method beyond that.	"I would send all documents online."
Physical mail	8	7	1	US Postal Service, UPS, Fedex, and other services	"I would probably print everything out and snail mail it all to the doctor."
Secure sending online	6	5	1	As "Sending online" above, but with an indication of security while simultaneously remaining nonspecific	"Sending it securely online is a more convenient way to do it for everyone involved."
Document sharing service	3	2	1	Services like Google Drive, Box, or Dropbox where a document is uploaded to a shared location	"Maybe through an app like dropbox with both password and PDF in a shareable link."
Video call	1	1	0	Facetime, Google Meet and other similar platforms	"I think the best way would be through what i already described being email or webcam call or text or a secure form to submit to them."

Survey 2 In Survey 2, participants were asked whether they had used or heard of the most commonly described eight transmission methods from Survey 1. The results, shown in Figure 1, indicate that most participants were aware of most methods, clarifying an uncertainty in Survey 1. Participants were most successful using in person, email, and online forms, aligning with the findings of Survey 1, and also mirroring Survey 1, document-sharing services (e.g., Dropbox or Google Drive) and faxing were relatively uncommon.

There are differences between Survey 1 and Survey 2. Physical mail was rarely mentioned in Survey 1, but participants had high levels of experience with it in Survey 2; on the other hand, while direct messaging was relatively popular in Survey 1, it was the least frequently used method in Survey 2. (Participants also provided other methods used; the resulting qualitative codes appear in Table 9, in the Appendix B.)

Survey 2 participants were randomly assigned a successfully used transmission method, with counter-weighting for balancing (see Table 3). Participants were asked to recall an instance of using their assigned method to send sensitive information and reported sending many types of information (Figure 2), with financial information and Social Security numbers (SSNs) most common top-of-mind instances. "Other" responses, aggregated via open coding, are also weighted heavily toward identifying information. (These are summarized in Table 10 in Appendix B.) There is little variation in what was being sent for each transmission method (Figure 2).

We also asked about the recipient of the information, in categories including an organization (e.g., a bank), a particular professional (e.g., an accountant), a friend or family member, and others. Results are shown in Figure 3. In keeping with the trend toward financial and identity information, the most popular responses were an organization, a government agency or institution, and individual professionals. Governments received mail most often, and direct messages often went to friends and family. Governments and organizations, unsurprisingly, were also most likely to use an online form.

We also asked participants if they or the recipient chose the method (Figure 4, left). Overwhelmingly, participants indicated that the recipient had either suggested or required

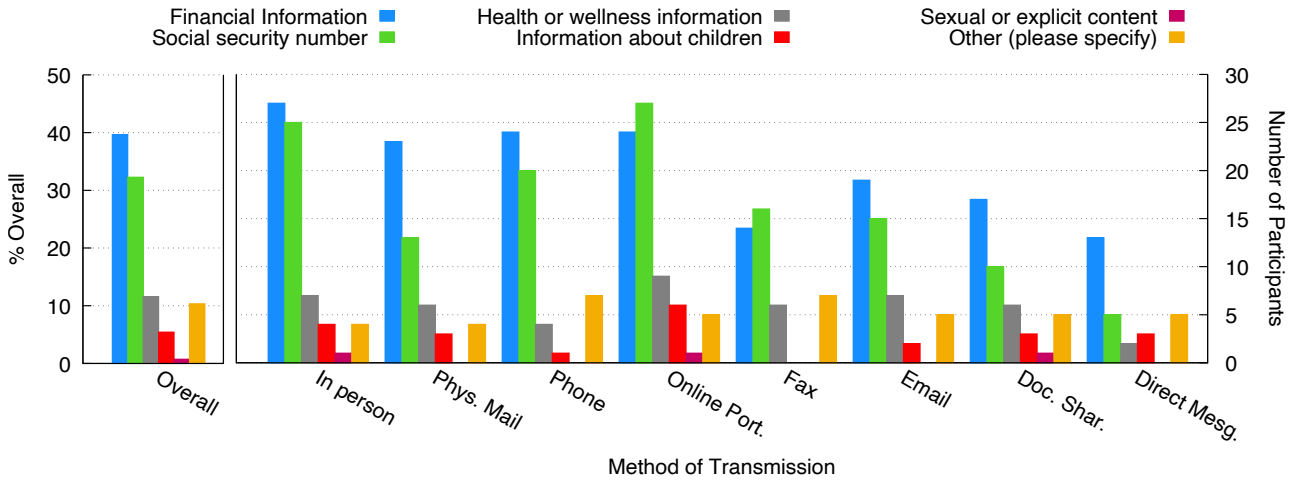

Figure 2: Type of information sent by participants across different methods (Survey 2)

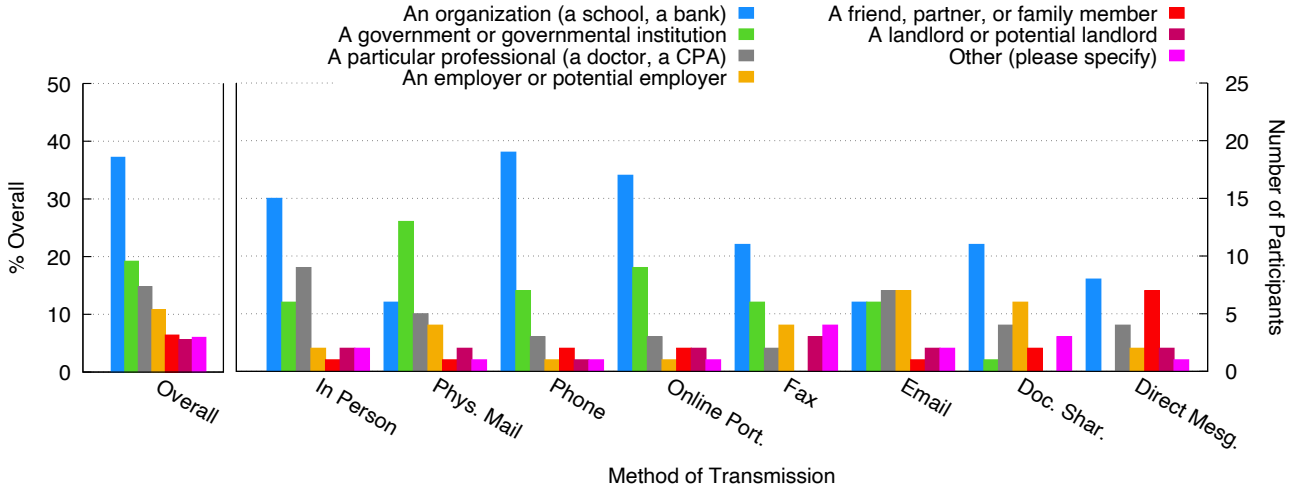

Figure 3: Recipients of sensitive information across different methods (Survey 2)

the method. Recipients required or suggested faxes, online forms, physical mail, and document sharing services, while participants more frequently suggested taking the documents in person or via phone call. When discussing jointly, participants and recipients often landed on email (Figure 4, right).

Key findings for RQ1 Email, online forms and taking documents in person are the most common transmission methods. Fax, document sharing services like Google Drive or Dropbox, and direct messages are least common. Participants have heard of these less common methods but not used them as frequently. Recipients are more likely than senders to choose the method of transmission.

4.2 RQ2: Are people satisfied? Why?

Survey 1 Most participants in Survey 1 were generally satisfied with the privacy and convenience of the transmission methods, and this satisfaction was consistent across scenarios (see Figure 5). We asked participants to describe why they were satisfied (or unsatisfied) with the method's privacy (see Table 5), and many (62 instances) described satisfaction due the security of the method. Another common reason for satisfaction (34 instances) laid at the communication endpoint. Participants believe the receiver will maintain security and privacy, and thus they are satisfied with the privacy of the transmission method.

Participants who reported being "unsatisfied" or "very unsatisfied" with the privacy of their transmission method expressed concern that the method being used is insecure (12 occurrences), or mentioned dissatisfaction in general without specifying further (6 occurrences). Some participants clearly described their distrust in a method but not why or how a threat might arise. For example, one participant mentioned the threat of "access by others," but not how or why this would happen. To explore this topic further, the codes from these free responses were used to develop Likert-scale questions

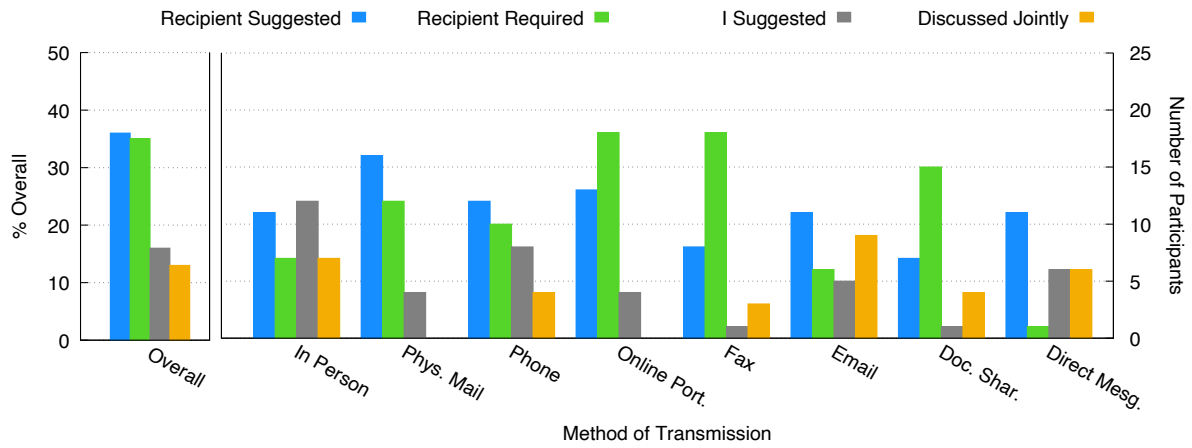

Figure 4: Determinants of the methods used to send sensitive information (Survey 2)

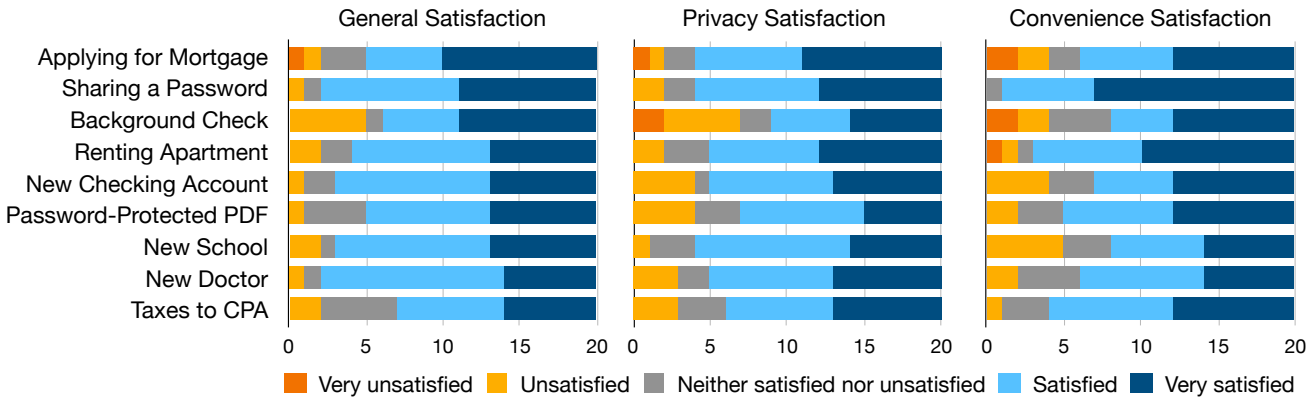

Figure 5: Satisfaction levels across different scenarios (Survey 1)

Table 5: Participants' reasons for being "satisfied" or "very satisfied" with privacy of their transmission modes in Survey 1. Participant answers may have had more than one code.

Satisfied Privacy Response Code	Frequency
My method of sending is secure	62
The recipient will keep my information safe	34
Information received by the intended recipient	21
I am satisfied (no specification)	12
I am unsure about the security of my method	7
The method of sending is insecure	6
I can keep my information safe	5
I am unsatisfied	2
The recipient may unintentionally disclose	2

about satisfaction. [4]

[4] See Survey 2 questions 14–15 in the extended paper (see Appendix C).

Survey 2 In Survey 2, we again observed that a large majority of participants were satisfied with the privacy of their methods (Figure 6, left). Only online forms, taking the documents in person, fax, and email registered any (and very few) "Very dissatisfied" responses.

Regression on privacy satisfaction We ran an ordinal logistic regression (our main planned analysis) to see what factors most correlated with privacy satisfaction when sending sensitive information. We report the results in Table 6. Privacy satisfaction was our outcome variable. Potential covariates included the following:

- Method used
- Type of data
- Identity of the recipient
- Level of trust in the recipient
- Who chose the transmission method
- The reported tech-savviness of both the participant and the recipient
- Likert-type responses for a variety of items, generated based on Survey 1 free responses. Responses were

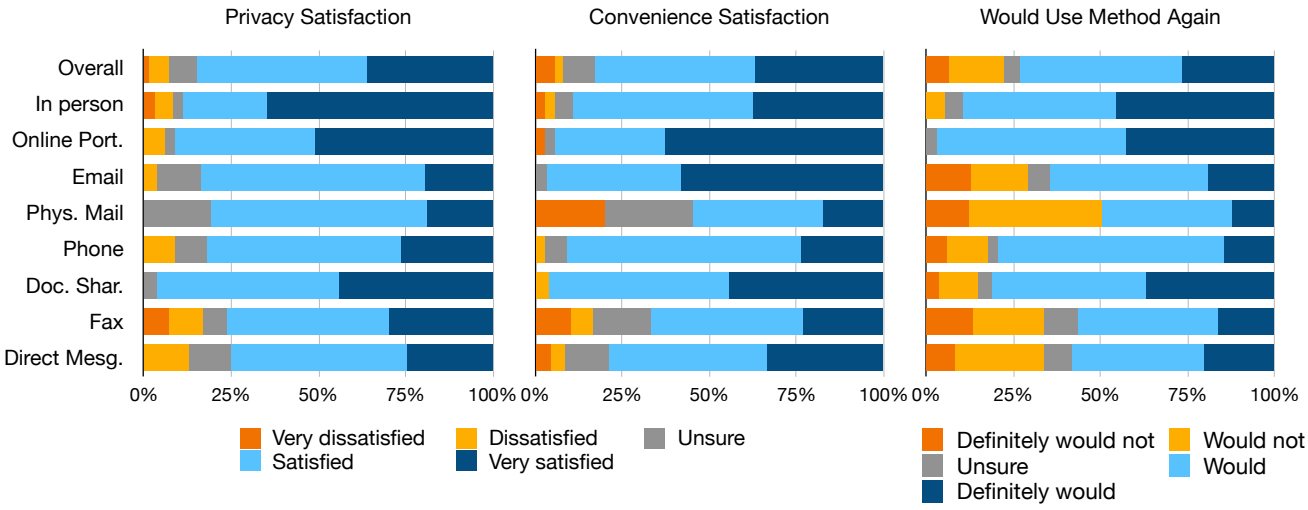

Figure 6: Satisfaction levels and willingness to use different methods again (Survey 2)

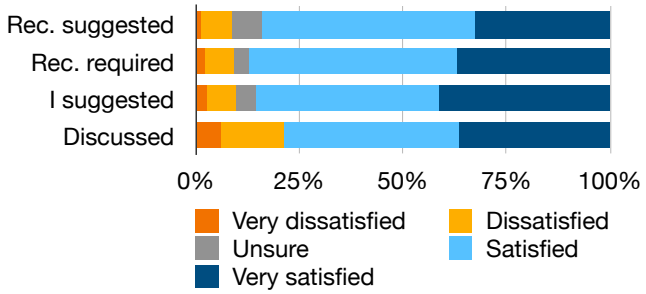

Figure 7: Privacy satisfaction based on the determinant of the method used (Survey 2)

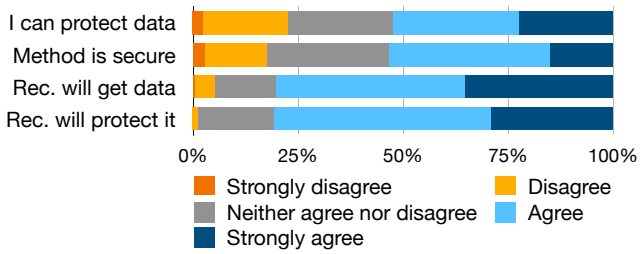

Figure 8: Agreement with reasons for privacy satisfaction across all methods (Survey 2)

binned into binary variables for analysis.[5] :

- The recipient will unintentionally reveal my data.
- The recipient will intentionally reveal my data.
- My data will be intercepted in transit.
- The recipient can keep my data safe.
- I can do something to keep my data safe.
- This method is inherently secure.
- The information would be received as intended.
- The data is at risk on my end.
- The data is at risk in transit.
- The data is at risk at its destination.

Definitions and levels for the above factors can be found in Table 11 in Appendix B.

We used a Variance Inflation Test (VIF) to check multicollinearity in the initial model with all of the above factors. All variables were well below the threshold value of 5 except for the "Other (please specify)" option for the type of data being sent. Since the types of data being sent were each independent binary factors, rather than a single categorical choice, we excluded this factor from our model selection process.

We then compared a set of potential models, keeping method used and type of data (except for the factor we removed) in every model but testing all possible combinations of the other covariates, not including interaction factors. We excluded interaction factors because we did not have sufficient power to include all the potential combinations. For parsimony, we selected the model with minimum Akaike Information Criterion (AIC) [3]. The final model, shown in Table 6, exhibits a pseudo-R^2 of 0.55 using the Aldrich-Nelson method, as evaluated by Hagle and Mitchell [16], indicating a fairly strong fit. Odds ratios above 1 indicate an increase in dissatisfaction relative to the baseline, as dissatisfaction was much less common than satisfaction. The model identifies several covariates that significantly correlate with privacy dissatisfaction, as follows.

Transmission methods Relative to the baseline of in-person transmission — selected because it is the only method

[5] These statements are slightly abbreviated; full text can be found in questions 14, 15, and 21 in the extended paper (see Appendix C).

Table 6: Final selected ordinal logistic regression model for participants' privacy dissatisfaction. Odds ratios above 1 indicate more dissatisfaction, relative to the baseline. The baseline for method is "taking the documents in person"; other baselines are false, disagree, and unlikely. Pseudo-R^2: 0.55

Variable	Value	Odds Ratio	Conf. Int.	p-value
Method	*In person*	—	—	—
	Online form	1.2	[0.4, 3.4]	0.761
	Email	3.7	[1.3, 11.3]	0.017*
	Mail	3.8	[1.4, 10.9]	0.012*
	Phone	3.0	[1.1, 8.5]	0.034*
	Doc sharing	1.1	[0.4, 3.4]	0.876
	Fax	3.4	[1.2, 9.9]	0.026*
	DM	1.9	[0.6, 6.3]	0.274
Financial	True	1.9	[1.1, 3.6]	0.032*
SSN	True	0.5	[0.3, 0.8]	0.007*
Health	True	1.1	[0.5, 2.2]	0.891
Children	True	1.0	[0.4, 2.6]	0.978
Explicit	True	4.4	[0.4, 39.0]	0.178
Risk at dest.	Agree	1.9	[1.1, 3.3]	0.032*
Recip't keep safe	Agree	0.4	[0.2, 0.7]	0.006*
Method secure	Agree	0.2	[0.1, 0.3]	<0.001*
Recip't share on purpose	Likely	2.8	[1.0, 8.4]	0.056

that does not require communications infrastructure — physical mail is associated with a 3.8× higher likelihood of more privacy dissatisfaction.[6] Email, phone calls, and faxes similarly exhibited odds ratios greater than or equal to 3. No other method was significantly different from in-person.

Type of data Several types of data being transmitted were also significantly correlated with privacy dissatisfaction. Because participants were allowed to select multiple potential options for data type, data types are modeled in the regression as independent boolean factors (baseline is false). Participants reported significantly more privacy dissatisfaction (odds ratio: 1.9) when financial information was included in the transmission. Surprisingly, they reported less dissatisfaction (odds ratio: 0.5) when transmitting Social Security numbers. This effect appears to be driven by an unusually large number of participants reporting "very satisfied" for transactions involving Social Security numbers.

Likert factors Figure 8 illustrates responses to some of the Likert-type questions relating to reasons for privacy satisfaction (questions based on Survey 1 responses). On the whole, participants were confident recipients would receive and protect data but less confident that transmission methods were secure or that they themselves could protect data.

Four Likert-type statements appear in the final regression model for privacy satisfaction: agreeing/disagreeing that the data is at risk at the destination, that the recipient can keep data safe, and that the method is inherently secure; as well as likelihood that the recipient will intentionally reveal data.

Participants were 1.9× as likely to report more privacy dissatisfaction when they agreed that data was at risk at the destination. In contrast, participants reported lower dissatisfaction when they agreed the recipient could keep their data safe (odds ratio: 0.4) or agreed the method was inherently secure (odds ratio: 0.2). All of these results are intuitive and match participants' comments from Survey 1.

Other factors None of the other factors we tested appeared in the final model, indicating that they are not meaningfully correlated with privacy satisfaction. Somewhat to our surprise, these non-factors included whether the recipient or the participant chose the method; this result is illustrated in Figure 7.

Convenience and Reuse We also asked participants whether they were satisfied overall with the convenience of their method, and whether they would use the method again.

Much like Survey 1, large majorities of participants were satisfied with convenience (see Figure 6, center). Post-hoc, pairwise MWU comparisons (see Section 3.2) indicate participants found physical mail significantly less convenient than in-person, online portal, email, and document sharing, and found faxing significantly less convenient than online portals or email. (Full details are given in Table 12 in Appendix B.)

Despite the overall satisfaction with privacy and convenience, we saw somewhat more variance when the participants were asked if they would use the method again (Figure 6, right). Post-hoc, pairwise MWU comparisons find that participants were most likely to want to use an online portal again (significantly more than email, physical mail, phone, fax, or direct messages). In-person was also significantly more popular for reuse than physical mail or fax. (Full details are given in Table 13 in Appendix B.)

Key findings for RQ2 Participants are overwhelmingly satisfied with the privacy of their methods, even when they did not choose the transmission method. Reasons for this largely depend on the *recipient* keeping data safe as well as confidence in the inherent security of their method. Both taking the documents in person and using an online portal — despite seemingly being quite different from each other — are perceived as providing a good overall tradeoff among privacy and convenience.

4.3 RQ3: What risks are people most concerned about?

Survey 1 We asked participants to describe potential risks associated with transmitting sensitive data generally, not in the context of a specific scenario. Participants overwhelmingly referred to risks to the data at rest, after transmission, rather than risks in transit, as shown in Table 7.

[6] We note that this sample was collected in the U.S. shortly after the 2020 presidential election, during which the reliability and security of the postal service received significant negative attention.

Table 7: Perceived risks of sending sensitive documents in Survey 1. Participant answers may have contained more than one code.

Risk	Frequency
The data at rest is at risk	50
Unspecified "malicious intent"	25
Identity theft	15
The data in transit is at risk	12
The data will be lost or misplaced	9
COVID-related concern	3
Monetary damage	2
Sending to the wrong person	2

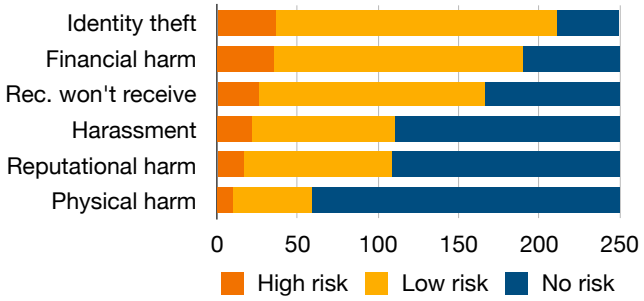

Figure 9: Reported risk levels of various types of harm across all methods (Survey 2)

Some examples include P5, who says, "A facility or institution misplacing, losing, or selling my information to a 3rd party can be worrisome." P30 worries about "Not knowing if the information will be kept safe." and P9 notes that "the place I give these documents stores or disposes of them", presumably indicating that if this storage and disposal is done improperly, their data will be at risk. It is notable that participants almost always identified risks at the recipient, rather than risks involving themselves.

In general, participants did not provide many specific details when asked to identify risks. We used the broad categories that they identified as well as concerns they raised about the data at rest to inform our design for Survey 2. This allowed us to collect more details on the perceived risks of sending sensitive documents.

Survey 2 We asked participants Likert-type questions based on the risks reported in Survey 1, as well as additional risks that we considered interesting or important. First, we asked whether — for the specific incident we had asked them to recall — they believed there was high risk, low risk, or no risk for a set of consequences, such as financial harm, reputational harm, or harassment. Participants overwhelmingly reported no or low risk (Figure 9). Slightly more risk was reported for identity theft and financial harm than for other concerns, which aligns with the prevalence of sending financial information and SSNs. Post-hoc, pairwise MWU comparisons (see Section 3.2) indicate that participants found identity theft and financial harm to be significantly more likely to be harmful. (Full details are given in Table 14 in Appendix B.)

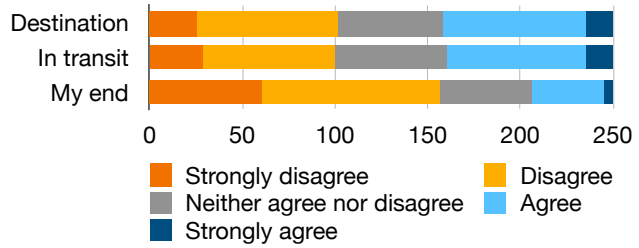

Figure 10: Where the risk is when sending sensitive information across all methods (Survey 2)

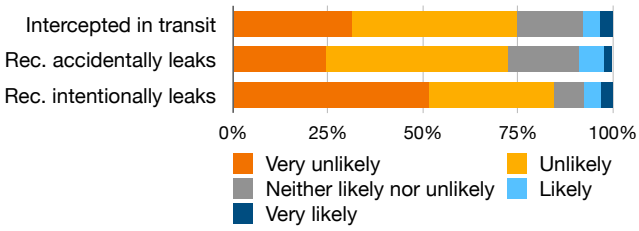

Figure 11: Likelihood of data to be leaked in various ways across all methods (Survey 2)

Figure 11 reports participant perception of how likely data might be to be leaked when transmitting sensitive documents based on statements derived from free responses in Survey 1. "Intercepted in transit" refers to data being intercepted between the source and destination, while the other two statements refer to the recipient revealing sensitive information to a third party, either deliberately or by accident. Most of these scenarios are viewed as "very unlikely" or "unlikely". Post-hoc, pairwise MWU comparisons found that participants considered the recipient intentionally leaking their information to be significantly less likely than the data being intercepted or leaked by accident. (Full details are given in Table 15 in Appendix B.)

We also investigated where in the transmission process participants view risk. Figure 10 demonstrates that in general, participants identify more risk at the destination and in transit than in their own stewardship. Similar post-hoc pairwise tests (Table 8) confirm that risk is perceived to be significantly greater at the destination and in transit than at the participant.

Key findings for RQ3 Participants are primarily concerned about financial harm and identity theft rather than risks of harassment or reputational damage. When unprompted, participants are most concerned with what happens to the data at its destination; after prompting, they express concern about risk in transit but do not identify risk at the sender (themselves).

Table 8: Post-hoc comparisons of risks using pairwise Mann-Whitney U-test with Holm-Šidák correction. (Omnibus Kruskal-Wallace test significant, $H = 44.23$, $p < 0.001$)

Comparison	p
My end vs. in transit	< 0.001*
My end vs. the destination	< 0.001*
In transit vs. the destination	0.922

5 Discussion

In two surveys, we explored how users transmit sensitive information when required to do so, their privacy satisfaction with their transmission methods, and the risks associated with these interactions. In Survey 1, we presented participants with three scenarios and asked them to qualitatively describe a transmission method they used or imagined they would use. Building on those responses, in Survey 2, participants were randomly assigned to a transmission method — among eight methods identified in Survey 1 — that they had previously used successfully. We then asked them to recall a specific instance of using that method to send sensitive information and answer closed-item questions about their privacy satisfaction, convenience, and risk factors. These questions were also derived from our qualitative coding of the results of Survey 1. In both surveys, participants generally described high satisfaction with both the convenience and privacy of their transmission methods and primarily described low risks. In most cases, the majority of participants indicated they would use the same transmission method again.

In this section, we explore larger themes and implications of the results, particularly around how participants see risks in transmitting sensitive information and choose a transmission method, as well as design implications and recommendations.

Familiarity and use are different In Survey 1, email, online forms, and taking the documents in person were the most common methods participants suggested without prompting. This raised an important question: do participants deliberately choose these methods over others, or have they simply not heard of alternatives?

The results from Survey 2 answer this question: Large majorities of participants had heard of all of the transmission methods. Further, participants tended to be satisfied with their transmission methods, regardless of whether they were prescribed by the recipient. This suggests that targeting recipients of sensitive data — like tax professionals and school personnel — for education and advocacy could have a positive impact on the security and privacy of these transmissions.

Additionally, there may be significant benefits to actively encouraging the use of tools that *already* exist to perform this task, such as document-sharing services. Our results suggest that participants know these options exist but simply do not choose to use them often. However, participants who discussed these methods did generally believe they are secure and were usually satisfied with them. Making these services more salient — perhaps by evangelizing them to common document recipients — could provide useful benefits. We also scoped this study to sending documents as a discrete transaction rather than continuous collaboration, which is an interesting but separate use case for which document sharing services might be more commonly used. Continuous collaboration on sensitive documents is a promising avenue for potential future work.

Only some information is considered sensitive When prompted in Survey 2 to recall a situation where they sent sensitive information, participants overwhelmingly selected financial information and Social Security numbers. Very few participants' exemplar scenarios included other identifying or secret information (contact details, passwords, etc.), suggesting that this information is considered less sensitive, or is at least less likely to be top-of-mind when imagining sensitive data. This aligns with prior work that finds people have different standards for what information is considered sensitive [36, 44, 47]. It also illuminates a potential gap, in which people may be transmitting sensitive information without realizing the need to take precautions. Future work could more directly examine what triggers people to recognize "sensitive" situations and consider communications privacy.

Risks at the endpoints In Survey 1, participants primarily focused on data leaks at the recipient, rather than in transit or at the source (e.g., from the user's email account). In Survey 2 — when prompted with specific choices — participants identified risks in transit with similar frequency to risks at the destination, but risks at the sender remained unrecognized. This raises two key points.

First, the usable privacy community has primarily focused on risks to data in transit. This includes studies of secure email and messaging adoption [1, 2, 11, 35], as well as challenges in conveying proper and secure transmission, particularly with respect to certificate warnings and phishing [4, 7, 27, 30, 37, 41]. This aligns with our finding that people were not entirely confident their transmission method was secure. While risks in transit are clearly important, our results suggest more attention should also be paid to risks at the endpoints, including, e.g., how to convey meaningful assurance that data is being handled properly at the destination.

Second, this finding accords with prior work showing that retrospective risks related to sensitive data left in one's own possession (often after sending it to someone) are opaque to end users [8, 17, 39]. Further work is needed to develop tools for both senders and recipients to clean up no-longer-needed data, and educational interventions that teach about secure communications should make sure to point out potential risks at the source as well as in transit and at the destination.

Design implications and recommendations Our findings suggest opportunities to improve the design of current transmission methods for sensitive content. In particular, methods should take into account both endpoints, not just security in transit, and the security mechanisms should be as transparent as possible to the user to reduce overhead of using the method. Below, we outline where these results can be applied to certain application spaces.

Document sharing services As mentioned above, document sharing services like Google Drive or Dropbox may provide a convenient and secure method to send sensitive documents. Further research is needed into why these services are used less frequently and what can be done to increase their use. Our participants who had tried them tend to believe they are secure and convenient, but many have not tried.

Confidential mode One attempt to improve transmission of sensitive data is Gmail's existing *confidential mode*.[7] This service encodes an email as an image so the content cannot be printed and will be automatically deleted at a later time, after which the recipient will not be able to view the content. While we know of no direct research on the efficacy of this method, the approach of interceding during the email process has promise, as both surveys and prior work [6, 24] suggest that email is a common approach for sending sensitive content, particularly when an alternative is unknown to either party.

Researchers should examine how to best intercede with the user workflow when opting for email based transmission of sensitive data. The user could then be prompted to apply a better mechanism first. However, the design and frequency of these interventions need to be carefully considered as prior work [12, 28] suggests that very frequent security warnings are likely to be ignored by users. Such interventions need to map to peoples' risk models to be most effective. If a user does not see (or understand) a risk, they are unlikely to make the right choice [46].

Secure message deletion Our study and other recent work [24] show that users are concerned about their data even after it arrives at the destination. One suggestion applicable to email is for senders to use short-lived encryption keys per message that can expire or be revoked [24], similar to popular chat applications such as WhatsApp and Signal [26]. While promising, this idea inherits significant key management challenges [5, 15, 18, 34, 42] related to the decentralized nature of email, and it remains unknown if users will simply copy or screenshot the emails outside the secure email system to retain access. This is a promising area of future work.

No-effort privacy While interventional approaches, such as prompting a user to use confidential mode, are important, an even better approach would be to offer users a transparent way to send sensitive information. This is analogous to incorporating end-to-end encryption into already popular messaging tools. For example, when including a potentially sensitive attachment, the document could automatically be conveyed via a secure document-sharing service, then automatically retrieved at the destination. This would allow the workflow at both endpoints to continue unchanged. This could parallel existing processes in email services that partially or entirely automatically send large attachments via cloud storage links rather than directly via email. As part of providing this service, additional work might be needed to convey the additional privacy benefits; demonstrating when communication is private has proven challenging in domains from web browsing to secure messaging [29, 37, 43].

Retrospective privacy Our results also confirm the previously identified need [17] for retrospective privacy. Email, cloud storage, and document sharing service providers could offer automated suggestions for deleting older content — both sent and received — and automatic message expiration options [24]. Providers could offer an option to mark sensitive content when it is created, to allow for review and potential deletion in the future. This could be modeled on approaches that allow users to "snooze" an email for future action or nudge users to revisit content that has not been accessed in a while. Elements like these could help users protect content at rest, even if it was not protected at transmission time.

6 Conclusion

This paper reports on two surveys of users' experiences sending sensitive information: Survey 1 using common scenarios drawn from prior work as prompts ($n = 60$, 180 total scenario instances) and Survey 2 ($n = 250$) asking more detailed questions based on results from Survey 1. We found that users most frequently expect to deliver documents in person or to use email; in reality, they typically use these methods as well as online portals or forms provided by institutional recipients. We also found that participants report high levels of satisfaction with the privacy and convenience of their existing methods, while recognizing that there are possible risks associated with transmitting this information, particularly risks of data leaking after being received at the destination. These results suggest new opportunities for tools and user interventions designed to make secure transmission of documents simpler and more transparent, and supporting retrospective privacy by nudging users to delete no-longer-needed content.

Acknowledgements

We gratefully acknowledge support from a UMIACS contract under the partnership between the University of Maryland and DoD. The views expressed are our own.

We'd also like to thank the reviewers for their insightful comments and feedback, as well as Kelsey Fulton, Omer Akgul, Nathan Reitinger, and the other members of the SP2 and GWUSEC labs for their help and support.

[7] https://support.google.com/a/answer/7684332 (viewed Feb 3, 2020)

References

[1] Ruba Abu-Salma, Elissa M. Redmiles, Blase Ur, and Miranda Wei. Exploring user mental models of end-to-end encrypted communication tools. In *FOCI 2018: Workshop on Free and Open Communications on the Internet*, 2018.

[2] Ruba Abu-Salma, M. Angela Sasse, Joseph Bonneau, Anastasia Danilova, Alena Naiakshina, and Matthew Smith. Obstacles to the adoption of secure communication tools. In *S&P 2017: Symposium on Security and Privacy*, 2017.

[3] Hirotogu Akaike. Information theory and an extension of the maximum likelihood principle. In *Selected papers of Hirotugu Akaike*, pages 199–213. Springer, 1998.

[4] Devdatta Akhawe and Adrienne Porter Felt. Alice in warningland: A large-scale field study of browser security warning effectiveness. In *Security 2013: USENIX Security Symposium*, 2013.

[5] Wei Bai, Doowon Kim, Moses Namara, Yichen Qian, Patrick Gage Kelley, and Michelle L. Mazurek. An inconvenient trust: User attitudes toward security and usability tradeoffs for key-directory encryption systems. In *SOUPS 2016: Symposium on Usable Privacy and Security*, 2016.

[6] Olha Bondarenko and Ruud Janssen. Documents at hand: Learning from paper to improve digital technologies. In *CHI 2005: ACM Conference on Human Factors in Computing Systems*, 2005.

[7] Cristian Bravo-Lillo, Saranga Komanduri, Lorrie Faith Cranor, Robert W. Reeder, Manya Sleeper, Julie Downs, and Stuart Schechter. Your attention please: Designing security-decision UIs to make genuine risks harder to ignore. In *SOUPS 2013: Symposium on Usable Privacy and Security*, 2013.

[8] 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 *CHI 2015: ACM Conference on Human Factors in Computing Systems*, 2015.

[9] Sauvik Das, Laura A. Dabbish, and Jason I. Hong. A typology of perceived triggers for end-user security and privacy behaviors. In *SOUPS 2019: Symposium on Usable Privacy and Security*, 2019.

[10] Sauvik Das, Adam D.I. Kramer, Laura A. Dabbish, and Jason I. Hong. Increasing security sensitivity with social proof: A large-scale experimental confirmation. In *CCS 2014: ACM Conference on Computer and Communications Security*, 2014.

[11] Alexander De Luca, Sauvik Das, Martin Ortlieb, Iulia Ion, and Ben Laurie. Expert and non-expert attitudes towards (secure) instant messaging. In *SOUPS 2016: Symposium on Usable Privacy and Security*, 2016.

[12] Serge Egelman, Lorrie Faith Cranor, and Jason Hong. You've been warned: An empirical study of the effectiveness of web browser phishing warnings. In *CHI 2008: ACM Conference on Human Factors in Computing Systems*, 2008.

[13] Michael Fagan and Mohammad Maifi Hasan Khan. Why do they do what they do?: A study of what motivates users to (not) follow computer security advice. In *SOUPS 2016: Symposium on Usable Privacy and Security*, 2016.

[14] Kelsey R. Fulton, Rebecca Gelles, Alexandra McKay, Yasmin Abdi, Richard Roberts, and Michelle L. Mazurek. The effect of entertainment media on mental models of computer security. In *SOUPS 2019: Symposium on Usable Privacy and Security*, 2019.

[15] Simson L. Garfinkel and Robert C. Miller. Johnny 2: A user test of key continuity management with S/MIME and Outlook Express. In *SOUPS 2005: Symposium on Usable Privacy and Security*, 2005.

[16] Timothy M. Hagle and Glenn E. Mitchell. Goodness-of-fit measures for probit and logit. *American Journal of Political Science*, pages 762–784, 1992.

[17] 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 *CHI 2018: ACM Conference on Human Factors in Computing Systems*, 2018.

[18] John S. Koh, Steven M. Bellovin, and Jason Nieh. Easy email encryption with easy key management: Why Joanie can encrypt. In *EuroSys 2019: EuroSys Conference*, 2019.

[19] Klaus Krippendorff. Estimating the reliability, systematic error and random error of interval data. *Educational and Psychological Measurement*, 30(1):61–70, 1970.

[20] Klaus Krippendorff. *Content analysis: An introduction to its methodology*. Sage Publications, 1989.

[21] Jon A. Krosnick. Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5(3):213–236, 1991.

[22] J. Richard Landis and Gary G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, pages 159–174, 1977.

[23] Peter McCullagh. Regression models for ordinal data. *Journal of the Royal Statistical Society: Series*

B (Methodological), 42(2):109–127, 1980.

[24] Tyler Monson, Scott Ruoti, Joshua Reynolds, Daniel Zappala, Trevor Smith, and Kent Seamons. A usability study of secure email deletion. In *EuroUSEC 2018: European Workshop on Usable Security*, 2018.

[25] Eyal Peer, Laura Brandimarte, Sonam Samat, and Alessandro Acquisti. Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, 70:153–163, 2017.

[26] Trevor Perrin and Moxie Marlinspike. The double ratchet algorithm. https://signal.org/docs/specifications/doubleratchet/, 2016.

[27] Justin Petelka, Yixin Zou, and Florian Schaub. Put your warning where your link is: Improving and evaluating email phishing warnings. In *CHI 2019: ACM Conference on Human Factors in Computing Systems*, 2019.

[28] Adrienne Porter Felt, Elizabeth Ha, Serge Egelman, Ariel Haney, Erika Chin, and David Wagner. Android permissions: User attention, comprehension, and behavior. In *SOUPS 2012: Symposium on Usable Privacy and Security*, 2012.

[29] Adrienne Porter Felt, Robert W. Reeder, Alex Ainslie, Helen Harris, Max Walker, Christopher Thompson, Mustafa Embre Acer, Elisabeth Morant, and Sunny Consolvo. Rethinking connection security indicators. In *SOUPS 2016: Symposium on Usable Privacy and Security*, 2016.

[30] Adrienne Porter Felt, Robert W. Reeder, Hazim Almuhimedi, and Sunny Consolvo. Experimenting at scale with Google Chrome's SSL warning. In *CHI 2014: ACM Conference on Human Factors in Computing Systems*, 2014.

[31] Kopo Marvin Ramokapane, Awais Rashid, and Jose Miguel Such. "I feel stupid I can't delete...": A study of users' cloud deletion practices and coping strategies. In *SOUPS 2017: Symposium on Usable Privacy and Security*, 2017.

[32] Elissa M. Redmiles, Sean Kross, and Michelle L. Mazurek. How I learned to be secure: a census-representative survey of security advice sources and behavior. In *CCS 2016: ACM Conference on Computer and Communications Security*, 2016.

[33] Elissa M. Redmiles, Sean Kross, and Michelle L. Mazurek. How well do my results generalize? Comparing security and privacy survey results from MTurk, web, and telephone samples. In *S&P 2019: Symposium on Security and Privacy*, 2019.

[34] Scott Ruoti, Jeff Andersen, Scott Heidbrink, Mark O'Neill, Elham Vaziripour, Justin Wu, Daniel Zappala, and Kent Seamons. "We're on the same page": A usability study of secure email using pairs of novice users. In *CHI 2016: Conference on Human Factors in Computing Systems*, 2016.

[35] Scott Ruoti, Nathan Kim, Ben Burgon, Timothy Van Der Horst, and Kent Seamons. Confused Johnny: When automatic encryption leads to confusion and mistakes. In *SOUPS 2013: Symposium on Usable Privacy and Security*, 2013.

[36] Scott Ruoti, Tyler Monson, Justin Wu, Daniel Zappala, and Kent Seamons. Weighing context and trade-offs: How suburban adults selected their online security posture. In *SOUPS 2017: Symposium on Usable Privacy and Security*, 2017.

[37] Stuart E. Schechter, Rachna Dhamija, Andy Ozment, and Ian Fischer. The emperor's new security indicators. In *S&P 2007: IEEE Symposium on Security and Privacy*, 2007.

[38] Steve Sheng, Levi Broderick, Colleen Alison Koranda, and Jeremy J. Hyland. Why Johnny still can't encrypt: Evaluating the usability of email encryption software. In *SOUPS 2006: Symposium On Usable Privacy and Security*, 2006.

[39] Peter Snyder and Chris Kanich. Cloudsweeper: Enabling data-centric document management for secure cloud archives. In *CCSW 2013: ACM Cloud Computing Security Workshop*, 2013.

[40] Daniel J. Solove. 'I've got nothing to hide' and other misunderstandings of privacy. *San Diego Law Review*, 44:745, 2007.

[41] Joshua Sunshine, Serge Egelman, Hazim Almuhimedi, Neha Atri, and Lorrie Faith Cranor. Crying wolf: An empirical study of SSL warning effectiveness. In *Security 2009: USENIX Security Symposium*, 2009.

[42] Michael Sweikata, Gary Watson, Charles Frank, Chris Christensen, and Yi Hu. The usability of end user cryptographic products. In *InfoSecCD 2009: Information Security Curriculum Development Conference*, 2009.

[43] Joshua Tan, Lujo Bauer, Joseph Bonneau, Lorrie Faith Cranor, Jeremy Thomas, and Blase Ur. Can unicorns help users compare crypto key fingerprints? In *CHI 2017: ACM Conference on Human Factors in Computing Systems*, 2017.

[44] Rick Wash and Emilee Rader. Too much knowledge? Security beliefs and protective behaviors among united states internet users. In *SOUPS 2015: Symposium on Usable Privacy and Security*, 2015.

[45] Alma Whitten and J. Doug Tygar. Why Johnny can't encrypt: A usability evaluation of PGP 5.0. In *Security 1999: USENIX Security Symposium*, 1999.

[46] Michael S. Wogalter. Purposes and scope of warnings. In Michael S. Wogalter, editor, *Handbook of Warnings*. Lawrence Erlbaum Associates, 2006.

[47] Allison Woodruff, Vasyl Pihur, Sunny Consolvo, Laura Brandimarte, and Alessandro Acquisti. Would a privacy fundamentalist sell their DNA for $1000... if nothing bad happened as a result? The Westin categories, behavioral intentions, and consequences. In *SOUPS 2014: Symposium On Usable Privacy and Security*, 2014.

[48] Justin Wu and Daniel Zappala. When is a tree really a truck? Exploring mental models of encryption. In *SOUPS 2018: Symposium on Usable Privacy and Security*, 2018.

Appendix

A Scenarios

The following are the nine scenarios we presented to the subjects, exactly as they were shown in the questionnaire.

Applying for a Mortgage You are applying for a mortgage so you can purchase a new home. You must send the bank the following information.

- Proof of income - W-2 forms and two most recent payroll stubs or other income information
- 60 days worth of bank statements
- Monthly debt payment information - car payments, student loan payments, credit card debt payments
- Rent payment for the past twelve months
- Divorce decree, if applicable

Sharing a Password A trusted friend needs access to an email account the two of you share. You need to send them the password to this account.

Background Check You are interested in doing some volunteer work, and the group you are working for has asked you to do a background check. You must send the volunteer group the following information.

- Full name
- Social Security Number
- Date of birth
- All addresses where you have lived in the past 5 years
- Names and contact information for two personal references

Applying for an Apartment You are applying to rent an apartment and are preparing your paperwork. You are required to send the landlord all of the following documents or information.

- Basic demographic information - name, email, phone number
- Emergency contacts
- Social Security Number

Opening a Checking Account You are opening a checking account at a new bank. The bank requires you to send them the following information.

- Social Security Number and date of birth of all account holders
- Phone number and email address
- Physical U.S. address (no post office boxes)
- Debit card or account information for funding your new account

Sharing a Password-Protected Document You have a password-protected PDF that is encrypted. You need to share both the PDF and the password to open it with a trusted friend.

New School Your child is starting at a new school, and you must send the school copies of the following documents about your child.

- Birth certificate
- Proof of custody/guardianship
- Proof of residency like one of the following: current property tax bill, current rental lease, current utility
- Immunization record
- Social Security card

Seeing a New Doctor You are going to see a new doctor for the first time. You are asked to send the new doctor's office the following information.

- Current insurance information
- An image of your driver's license or other valid photo ID
- A list of any medication you are currently taking
- Your health history

Sending Tax Documents You are getting ready to prepare your taxes and have hired a Certified Public Accountant (CPA). They ask you to send them the following information.

- A copy of your Social Security card
- All income-related tax documents - W-2, 1099, etc.
- All expense-related tax documents - 1098, rental expenses, etc.

B Additional Tables

Table 9: Free responses to "What other methods have you used?" (Survey 2). Many participants repeated methods that were provided in the closed-answer question.

Transmission Method	#Part.
Online form or portal	20
In person	14
Physical mail	11
Email	9
Direct messaging	8
Online (no further spec.)	5
Courier service	4
Email (mentions encryption)	4
Fax	3
Phone call	3
Direct messaging (mentions encryption)	2
Document sharing service	2
Via flash drive	2
Encryption (no further specification)	1
Live chat support	1
Used a VPN	1
Via encrypted flash drive	1

Table 10: Free responses to "What kind of information were you sending: Other" (Survey 2)

Data Type	#Part.
Home address	12
Identity documents	12
Demographic details	9
General personal information	6
Financial information	5
Contact information	4
Login credentials	3
Work documents	3
Titles and deeds	2
Insurance documents	1

C Extended Appendices

An extended version of the paper including the full text of each survey and the qualitative codebook for Survey 1 can be found at https://arxiv.org/abs/2105.14619.

Table 11: Factor levels for inputs to ordinal logistic regression before model selection. Asterisks (*) indicate baselines.

Factor	Levels
Method Categorical	In person* Physical mail Direct messaging Online form or portal Phone call Fax Email Document sharing service
What was being sent Binary for each option as participants could choose multiple. Baseline for each was false	Financial info Social Sec. number Info about children Info about health Sexual or explicit content Other (please specify)
Recipient Categorical	An organization* A particular professional A friend, partner, or family member An employer or potential employer A landlord or potential landlord A gov't or gov't institution Other (please specify)
Who chose the method Categorical	I suggested* Recipient suggested Recipient required Discussed jointly
Trust in recipient Binned from 5pt Likert	Don't trust* Trust
Agreement with statements Each below binned from 5pt Likert *I am tech-savvy* *The rec. is tech-savvy* *Risk is at my end* *Risk is at dest.* *Risk is at their end*	Disagree* Agree
Likelihood of leaks Each below binned from 5pt Likert *Rec. will accidentally leak* *Rec. will intentionally leak* *Data intercepted in transit*	Unlikely* Likely

Table 12: Post-hoc comparisons of convenience satisfaction (Survey 2, Q13) across transmission methods using pairwise Mann-Whitney U-test with Holm-Šidák correction. (Omnibus Kruskal-Wallace test significant, $H = 42.56$, $p < 0.001$)

	In Person	**Online Port.**	**Email**	**Physical Mail**	**Phone**	**Doc. Share**	**Fax**
In Person	—						
Online Port.	0.461	—					
Email	0.630	0.952	—				
Physical Mail	0.024*	< 0.001*	< 0.001*	—			
Phone	0.909	0.054	0.101	0.089	—		
Doc. Share	0.909	0.884	0.909	0.006*	0.630	—	
Fax	0.461	0.012*	0.020*	0.884	0.785	0.149	—
Direct Mesg.	0.909	0.285	0.384	0.362	0.964	0.832	0.887

Table 13: Post-hoc comparisons of likelihood to reuse a given transmission method (Survey 2, Q16) using pairwise Mann-Whitney U-test with Holm-Šidák correction. (Omnibus Kruskal-Wallace test significant, $H = 38.53$, $p < 0.001$)

	In person	**Online Port.**	**Email**	**Physical Mail**	**Phone**	**Doc. Share**	**Fax**
In Person	—						
Online Port.	0.989	—					
Email	0.067	0.032*	—				
Physical Mail	0.001*	< 0.001*	0.950	—			
Phone	0.105	0.043*	0.977	0.453	—		
Doc. Share	0.965	0.955	0.676	0.078	0.864	—	
Fax	0.018*	0.006*	0.982	0.977	0.925	0.365	—
Direct Mesg.	0.081	0.041*	0.989	0.960	0.971	0.676	0.986

Table 14: Post-hoc comparisons of severity of risks of the participant's transmission method (Survey 2, Q18) using pairwise Mann-Whitney U-test with Holm-Šidák correction. (Omnibus Kruskal-Wallace test significant, $H = 249.69$, $p < 0.001$

	Harassment	**Identity theft**	**Financial**	**Physical**	**Reputational**
Harassment	—				
Identity theft	< 0.001*	—			
Financial	< 0.001*	0.158	—		
Physical	< 0.001*	< 0.001*	< 0.001*	—	
Reputational	0.723	< 0.001*	< 0.001*	< 0.001*	—
Not Received	< 0.001*	< 0.001*	0.045*	< 0.001*	< 0.001*

Table 15: Post-hoc comparisons how likely sensitive data is to be leaked (Survey 2, Q14) using pairwise Mann-Whitney U-test with Holm-Šidák correction. (Omnibus Kruskal-Wallace test was significant, $H = 32.31$, $p < 0.001$)

	Intercepted in transit	**Recipient leaks by accident**
Intercepted in transit	—	
Recipient leaks by accident	0.259	—
Recipient leaks on purpose	< 0.001*	< 0.001*