2025 IEEE 16th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON) | 979-8-3315-6505-3/25/$31.00 ©2025 IEEE | DOI: 10.1109/IEMCON67450.2025.11381261

On the Effectiveness of End-Users' Data Backup Practices Against Data Corruption

Lalchandra Rampersaud
Florida International University
lramp004@fiu.edu

Behzad Ousat
Florida International University
bousa001@fiu.edu

Caitlin Brown
Florida International University
cbrow202@fiu.edu

Selcuk Uluagac
Florida International University
suluagac@fiu.edu

Amin Kharraz
Florida International University
mkharraz@fiu.edu

***Abstract*—With the rising threat of data loss, data backup has become an important part of defense mechanisms. Data corruption attacks such as ransomware have become more consequential when victims cannot retrieve their data and are more likely to pay the ransom fee. In this study, we evaluate end-users' readiness in protecting their data and potential issues with their data protection plans. We investigate how users protect their data and what their data recovery plans look like. We identified some contradictory and surprising findings. For instance, we observed that many users were confused about who was responsible for data protection. Only 57 (11.6%) of the participants reported that data protection is a user's responsibility. We asked participants to give a dollar value on how much their data might be worth. While 4 (0.8%) reported values of more than one million and 114 (23.1%) reported less than $100, 233 (47.3%) participants believed that their data was worth at least $1,000. However, 103 (20.9%) of those participants did not use any viable backup solution to protect their data.**

***Index Terms*—Data Backup Practices, Data Protection Strategies, Ransomware Awareness**

I. Introduction

Data corruption has been one of the primary objectives of modern attacks (e.g., ransomware). This strategy has been an important part of the attack chain because it allows malware campaigns to force victims, who do not have proper backup plans, to pay the ransom fee. In response to the increasing number of ransomware incidents, end-users are often advised to create backups of their critical data [13], [32], [40], [43]. However, several reports [19], [31], [14] reveal that many users do not usually follow recommendations and often do not take the necessary steps to safeguard their critical data.

So far, little attention has been given to users' decision-making process regarding their data protection plans. It is not quite clear why users are not prepared yet for these attacks after so many public announcements and recommendations on data backups. What are the main factors in the users' decision-making process that prevent them from being prepared for such attacks? How prepared are they for data loss? How might they react to a ransomware attack? In this paper, we shed light on this problem space and formalize the underlying issues in users' decision-making about their data protection plans.

Our study is guided by a set of primary research questions. First, how well were normal users prepared to protect their data against such attacks? Second, what are possible issues with their decision process that might make them vulnerable to such attacks? We recruited a total of 493 participants through Prolific [5], an online survey platform widely used in academic research for recruiting diverse and reliable study samples. We recruited 493 participants from 56 different professions or areas of study and various degrees of interest in computer science topics. In the following, we highlight important takeaways from our study.

First, insecure data protection practices are prevalent. We observed that 273 (55.4%) of the participants did not have any data protection plan or use the free tier of synchronization services such as Dropbox [1] or iCloud [3] that offer only 2-5 GBs of storage. Although 453 (91.9%) participants realized the importance of data backup in defending against attacks such as ransomware, only 220 (44.6%) participants had a sufficient security solution in place for protecting against such incidents. We consider such a plan to be one that offers more than just the basic free storage option. Typically, these solutions provide ample storage space for backing up critical files, have automatic version control, deploy encryption, and provide the ability to restore files in the event of a failure or malicious attack [4], [2].

Second, it is not always clear to users who is responsible for data backup. Participants, in general, had a vague understanding of which entity is responsible for their data safety plan. We note that 230 (46.6%) reported that the OS should be responsible for data protection, as it has always been considered a classic OS service. 57 (11.6%) reported that third-party software services are responsible for data protection. The remaining 206 (41.8%) base their choice on software performance. For many users, differentiating the environment in which the data was produced and consumed was not trivial.

Third, paying the ransom fee was considered a quick solution for a large group of participants. As a part of the experiment, we were interested to know how participants would react to a ransomware attack. We asked them if they

would be willing to pay the ransom fee to gain access to their data. While 353 (71.6%) of the participants mentioned they would not pay the ransomware fee and would not contribute to such businesses, 140 (28.4%) of the participants found that it is a quick way to gain access to the data and were willing to pay the ransom fee if the requested amount was less than $1,000 or if they could pay the ransom fee in multiple installments.

Finally, the problem of enabling users with robust backup solutions is not as trivial as it seems. In fact, reluctance to pay a subscription fee and have a reliable backup is just one issue. We noticed that many users often have trouble differentiating what data requires backup. Many of the users are left with an illusion that their operating system or the software tools they use on a daily basis are responsible for all their data protection. Another issue is that a large number of users underestimate the true cost of their data loss. This has a significant impact on their decision-making process with regard to their data protection. That is, users become significantly less loss-averse and fail to properly evaluate decisions about their data protection plans. We acknowledge that more scientific work is required to understand cognitive biases on users' decision-making in this domain. We hope this work serves to raise awareness about the importance of using foundational work in other human-centric areas [30] in assisting users to make better security decisions. We also hope our approach will prove useful to the security community and open the door for future user-oriented solutions, education, and awareness trainings.

II. Background and Motivation

In this section, we briefly explain some of the recent attacks and discuss contemporary defense mechanisms available to users and businesses.

Attacks on Users' Data. Numerous successful ransomware attacks have been reported on users, institutions, and critical businesses over the last few years. In fact, the scale and frequency of these incidents have become so large that they do not surprise us anymore. These attacks have traces in massive social-engineering attacks on end-users, institutions, hospitals, and governments over the last few years [24], [9], [23], [28], [29], [15], [22]

Current Defenses. Robust backup solutions have been the most reliable layer of defense against ransomware. There have been several attempts to make a robust defense solution against ransomware, but the offered techniques often do not receive wide acceptance or are too expensive to be used on a large-scale by normal users. For instance, Malwarebytes ransomware protection [20] is now only accessible as part of the Malwarebytes premium service. Acronis Ransomware Protection [7], which used to be a free standalone tool, is currently offered as a premium service. Heiling defense [18] which was offering RansomOff has taken down the service. CyberSight RansomStopper is no longer available [34], and Cybereason RansomFree has been discontinued [17]. This trend has left users with one viable approach, i.e., robust backup services, that had received wide adoption and showed to be effective against modern ransomware attacks. Ransomware is a clear and serious example of data corruption, but our study looks at backup methods as a way to protect against both malicious threats and more typical risks like hardware failures, synchronization issues, or inadvertent deletion.

Focus of This Study. In this study, we first empirically test how well end-users are prepared to protect their data. We believe this is a valid and important question. Data loss due to ransomware attacks has been among the top security threats for several years and plays a critical role in monetizing adversarial businesses. Research on how to minimize victims' contribution to these activities, not paying ransom fees, is critical. Second, we assess end-users' data recovery plans and possible issues in their data protection. How much do victims value their data and what are possible ways to help them make a better decision about their data protection plan? Empirical studies, similar to this paper, can potentially assist to learn more about how normal users view the problem and how to minimize the associated risks of data loss at the time of an attack.

III. Methodology

In this section, we introduce the two main research questions we seek to answer. We then elaborate on how we conducted our research to collect data to answer these questions. In the data collection phase, we conducted the survey using Prolific [5], an online survey platform, to recruit a diverse group of participants and capture their insights about their data protection practices.

Survey Questionnaire. In addition to Appendix A that contains only the question segments, the details of the questions and possible responses are provided at https://anonymous.4open.science/r/dataprotection-F522.

A. Research Questions

Our research tries to answer the following research questions:

RQ1: What is users' level of preparedness to protect their data from modern data loss and corruption threats? There is no lack of evidence that users have been impacted seriously by ransomware attacks, hardware failures, or accidental deletion. A question that arises is how users manage their data protection. What do they think about their roles in protecting their data? We study their preparedness by also considering their technical security background.

RQ2: What are the possible issues in their data protection plans that can make them vulnerable to data loss events whether caused by malicious attacks such as ransomware or by non-malicious failures such as device crashes? While we still do not clearly know how users protect their data, we do not know the effectiveness of their data protection plans either. What would go wrong if a real data loss event occurs on their machines? We study the security posture of data protection plans and evaluate their effectiveness.

B. Ethical Considerations

IRB Waiver. For our data collection and user study, we have obtained exception approval from the university office of research IRB. The experiments to collect data in this study were conducted anonymously. There are few to no ethical concerns with regard to the data collected from respondents. Individual responses did not contain any information which could allow a trace-back to respondents. We did not collect any sensitive data or Personally Identifiable Information (PII) about participants such as names or email addresses. The survey is done online voluntarily with anonymized IP addresses. Data protection procedures have been implemented at all the layers. The responses were aggregated and encrypted on an internal database for querying and charting. These databases were only accessible to members of the team who had passed special IRB training courses.

C. Recruitment

Initial Study and Participant Recruitment. Before running a large scale experiment, we conducted a smaller-scale experiment with 208 participants from the hosting institution using a custom survey website we developed. The goal of the experiment was to update the question catalog and revise the questions for clarifications. After a multi-round experiment, we finalized the questions to make sure they were clear and understandable by a large number of participants with different language skills and technical backgrounds.

Participant recruitment. We conducted the larger-scale experiment through Prolific [5], an online research recruitment platform, on January 22, 2025. In order to avoid the effects of priming, participants were not informed about the key goals of the experiment before taking the survey. The recruitment requirement was that the participants are familiar with web browsers so that they could perform the given tasks correctly and answer the surveys in one sitting. All the experiments were routed through the survey website we designed for this experiment.

A total of 493 participants took part in the survey. Each person received $1.20 USD as their compensation if they answered all the questions and abided by the rules (e.g., avoiding random responses and irrelevant response formats). The survey[1] was meant to take about 10 minutes to fill out, but people actually took anything from 4 to 33 minutes to finish it. We set the sampling parameters so that the sample included people of different sex, age, and ethnicity as shown in table I. There were 254 (51.5%) women and 239 (48.5%) men in the final sample. There were 57 (11.6%) people between the ages of 18 and 24, 87 (17.7%) people between the ages of 25 and 34, 89 (18.1%) people between the ages of 35 and 44, 77 (15.6%) people between the ages of 45 and 54, and 183 (37.0%) people between the ages of 55 and 100.

The majority of the participants identified as White (309, 62.7%), followed by Black (57, 11.6%), Mixed (54, 11.0%), Asian (31, 6.3%), and Other (35, 7.1%), with 7 participants (1.4%) choosing not to disclose their ethnicity. This recruitment technique yielded a representative dataset as per the 2020 US census [11], hence enhancing the reliability and generalizability of the findings. Backgrounds of the participants include computer science, electrical engineering, Music and Art, and Health. Figure 1 shows a summary of the participants' field who contributed to the study.

[1]https://anonymous.4open.science/r/dataprotection-F522

TABLE I: **Demographics of Survey Participants.**

Demographic	Cohort	Count (%)
Ethnicity	Asian	31 (6.3%)
	Black	57 (11.6%)
	Mixed	54 (11.0%)
	Other	35 (7.1%)
	White	309 (62.7%)
	N/A	7 (1.4%)
Age (Years)	18–24	57 (11.6%)
	25–34	87 (17.7%)
	35–44	89 (18.1%)
	45–54	77 (15.6%)
	>54	183 (37.0%)
Sex	Male	239 (48.5%)
	Female	254 (51.5%)
Degree Level	Associate or High School	202 (41.0%)
	Graduate	103 (20.9%)
	Undergraduate	183 (37.0%)
	None or N/A	5 (1.1%)
Profession	Business & Management	79 (16.0%)
	Science	64 (13.0%)
	Health	47 (9.5%)
	Other	303 (61.5%)
Courses	Computer Science	60 (12.2%)
	Cybersecurity	15 (3.0%)
	IT	62 (12.6%)
	Combination of Above	106 (21.5%)
	No technical field	250 (50.7%)
Technical Interest	Yes	336 (68.2%)
	No	157 (31.8%)

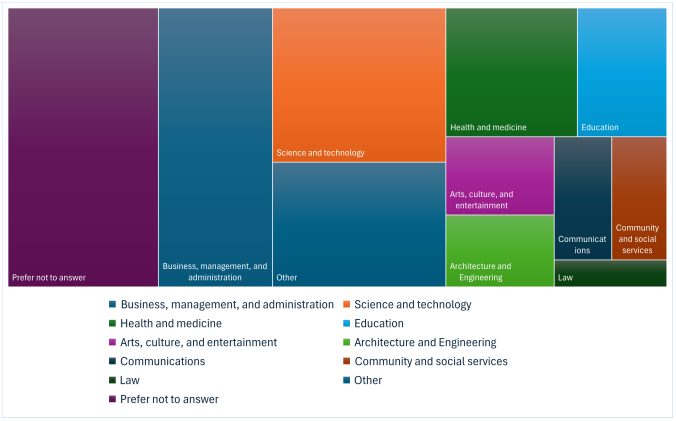

Fig. 1: Distribution of Participant's Fields

IV. Studying Users' Decisions

Armed with survey data from 493 participants, we answer the research questions discussed in Section III-A. Each of

the research questions was broken down into multiple sub-questions to gain sufficient insights into the users' decision process. We first answer how well users are prepared for potential data losses. What are their current practices for this particular problem? We then discuss some of the more common issues in their practices and what makes them vulnerable to attacks. Lastly, we evaluate a few ways in which their decision-making process could be improved without imposing too much overhead in the training process or enrolling in a viable solution.

A. RQ1: How well are users prepared for potential data losses?

Protecting users from ransomware attacks can be done in various ways. One of the most effective practices has been the integration of reliable backup solutions, where the service offers sufficient backup storage with quick and reliable rollback options. These two factors have a significant impact on the cost and effectiveness of the service. The first question we asked the participants was about their active data protection plans: whether they have any active backup plan and what that plan entails. We observed that 273 (55.4%) participants did not use any backup solution or use the free version of cloud-synchronized services such as Dropbox, OneDrive, and iCloud. These synchronized services are offered in multiple forms and packages. The free version of these services only provides between 2 GB to 5 GB of data backup which is not likely sufficient to fully protect these users in case of an attack. Our analysis showed that only 48 (9.7%) participants had purchased a premium plan with one TB data protection plan.

Who is responsible for data backup? 206 (41.8%) of the participants did not have any response to this question. In particular, they mentioned that they care more about the service and its usability rather than who provides the service. We observed that many participants were confused about who was responsible for backup. More than 230 (46.6%) of the participants believed that the first-party services such as OS vendors and software providers who offer the service should also offer data protection for their services. Only 57 (11.6%) of the participants reported that data protection is the responsibility of the user of computing systems. Perhaps the takeaway of this experiment was that users, in today's modern computing landscape, are the consumers of various data-intensive software systems that offer background data backup capabilities. This has defined an implicit assumption that they are the last entity that should care about backup.

What is the role of technical knowledge? The fact that only 220 (44.6%) participants had a reliable backup plan is concerning. We asked five security-related technical questions (see the first five questions in table V) to test how well people knew the basics of computer science and cybersecurity. Results from 493 participants show that over half, 262 (53.1%) could not answer any questions correctly, and just 7 (1.4%) received a perfect score. Approximately 22% of the participants scored between 1–2 correct answers, suggesting limited or partial understanding of the topics. This distribution shows that there is a skew towards low technical proficiency, even within a group of people from different backgrounds. To get insights into the relationship between technical skill and cybersecurity practices, participant performance on a five-question technical assessment was cross-referenced with their reported utilization of backup services across different tiers as shown in table II. The findings show a trend where individuals with no correct answers had a higher incidence of not using any backup solution. On the other hand, people who paid for backup services, especially those who paid for higher storage spaces, were more technically competent (at least 3/5). For instance, only one of the participants who received a perfect score (5/5) said they didn't use backup services at all. The rest were found in paid tiers that ranged from 50 GB to over 1 TB. Participants who scored perfectly almost always used some form of backup, especially at higher capacities which may indicate awareness of digital risk and mitigation strategies. Users of paid services across all levels were also more likely to have moderate to high proficiency (3–5 accurate responses), which may support the premise that being technically literate may lead to proactive security practices.

Participants were asked which cyber attacks (given the following list: Man-in-the-middle (MitM), Phishing, SQL injection, Cross-site-scripting (XSS), Ransomware, Denial-of-service (DoS)/Distributed Denial-of-Service (DDoS), and CSRF) they were familiar with and were presented with seven common attacks. Only 56 (11.4%) did not know of any, 362 (73.4%) knew of at least 1 to 3 attacks, and 75 (15.2%) knew of at least four. Among participants who reported no knowledge of any cyber attack type, the majority either used no backup service (3.4%) or only used free-tier services (4.7%). In contrast, individuals who were more familiar with cyber attacks, especially those who reported 1–3 types, were more evenly spread out over all tiers, with a clear trend toward paid backup options. In this category, 28.6% of participants used free services, while 33.1% used paid backup solutions with different storage capacity. Among participants who were aware of four or more types of cyber attacks, the utilization of strong backup solutions became increasingly evident, with 5.0% employing free services and 11.3% selecting premium services, particularly higher-capacity tiers such as "over 1 TB" (3.0%) and "50–100 GB" (1.0%). These results indicate a trend where enhanced awareness of various cyber attack types is associated with improved personal data protection practices, suggesting that such awareness may influence security-focused decision-making. We conducted statistical tests to examine the relationship between technical knowledge and backup use. A chi-square test showed a significant association between technical knowledge and backup category ($\chi^2(4) = 22.0$, $p < 0.001$), indicating that users with higher technical knowledge were more likely to adopt paid backup services. Logistic regression further confirmed this: each additional correct answer in our technical knowledge assessment increased the odds of using a reliable (paid) backup plan by 40% (OR = 1.40, 95%

TABLE II: Participants were asked 5 technical questions related to computer science or cybersecurity. This table shows the distribution of correct answers with respect to the tier of backup service used. Participants with greater technical knowledge were more inclined to adopt paid backup services.

Tier	# of Correct Technical Answers	Count
No backup	0	61 (12.4%)
	1	9 (1.8%)
	2	8 (1.6%)
	3	3 (0.6%)
	4	2 (0.4%)
	5	1 (0.2%)
Free service	0	110 (22.3%)
	1	35 (7.1%)
	2	19 (3.9%)
	3	12 (2.4%)
	4	13 (2.6%)
	5	0 (0.0%)
<50 GB	0	17 (3.4%)
	1	8 (1.6%)
	2	10 (2.0%)
	3	5 (1.0%)
	4	9 (1.8%)
	5	1 (0.2%)
50–100 GB	0	23 (4.7%)
	1	11 (2.2%)
	2	9 (1.8%)
	3	6 (1.2%)
	4	2 (0.4%)
	5	2 (0.4%)
100–250 GB	0	18 (3.7%)
	1	4 (0.8%)
	2	6 (1.2%)
	3	1 (0.2%)
	4	5 (1.0%)
	5	0 (0.0%)
250–500 GB	0	5 (1.0%)
	1	5 (1.0%)
	2	5 (1.0%)
	3	4 (0.8%)
	4	1 (0.2%)
	5	0 (0.0%)
500–1000 GB	0	6 (1.2%)
	1	2 (0.4%)
	2	2 (0.4%)
	3	2 (0.4%)
	4	1 (0.2%)
	5	2 (0.4%)
>1 TB	0	22 (4.5%)
	1	11 (2.2%)
	2	5 (1.0%)
	3	7 (1.4%)
	4	2 (0.4%)
	5	1 (0.2%)

CI [1.22, 1.60], $p < 0.001$).

Security knowledge about social engineering was elicited by asking participants to define the term. There were 213 (43.2%) correct answers and 280 (56.8%) did not know. When combined with the use of backup services, the majority of respondents (125, 25.3%) reported utilizing free backup services without any prior knowledge of social engineering. This was followed by free-tier users who did report knowledge of social engineering (64, 13.0%). Only 26 (5.3%) of respondents who know about social engineering reported having no backup solution, while 58 (11.8%) of respondents without social engineering knowledge indicated that they did not use any backup service. This suggests that at least some form of backup use is correlated with knowledge of social engineering, as the proportion of knowledgeable participants who did not have backup was proportionally reduced. These results suggest that awareness of social engineering is associated with more consistent adoption of backup services, including paid solutions at both lower and higher storage tiers. This finding reinforces the idea that security awareness training not only increases recognition of threats but may also promote proactive investment in data protection measures.

Do security awareness programs have an impact? Another question we were interested to answer was the effect of prior security awareness training on participants' decisions. As one of the last questions of the survey, we asked participants if they have attended any security training in the past. We intentionally asked this question in the last section of the questionnaire when almost all the critical questions had been answered in order to reduce the risk of priming – folks might have responded differently if they had been asked whether they had attended a security training program. Our analysis showed that only 74 (15.0%) participants had taken part in a security training program. Those who attended the training sessions reported that the training was provided as part of their current job or a workshop they had attended. We observed that 48 (9.7%) of these participants had a reliable data protection plan when participating in our study. Due to the small size of the participants with prior security training, we cannot elaborate on the significance of this observation.

One immediate question that arose was the role of participants' knowledge of ransomware. We observed that 293 (59.4%) respondents knew ransomware and eloquently described the associated risks and consequences. Out of this group, only 143 (29.0%) had a reliable backup plan while 111 (22.5%) used the free backup option. Furthermore, 117 (23.7%) of the participants, reported that it is *somewhat likely* that they can be infected by ransomware attacks in their normal web activities. However, only 60 (12.2%) reported using a reliable backup plan, 44 (8.9%) used the free version and 13 (2.6%) did not use any backup services. Participants were asked to rate the importance of backup security, 409 (82.9%) noted that it was at least *somewhat important*. However 199 (40.4%) of these use a reliable backup plan, 156 (31.6%) uses the free version and 54 (10.9%) does not use any backup

services. Table III summarizes the awareness of users and their approach to data protection.

Other awareness variables (ransomware, social engineering, prior training) did not reach statistical significance when using the chi-squared test. This suggests that concrete technical proficiency plays a stronger role in shaping backup adoption than general awareness or exposure to training.

B. RQ2: What are the common issues in the users' decision-making process?

Our analysis of the previous part shows that many users are not very well-prepared in case of a data loss. While we discussed some of the common characteristics of their approach to data and data protection, it was not very clear to us why we had those observations. The fact that only 220 of the participants out of 493 had a reliable data protection plan shows some potential flaws in users' decision process about data protection. As we mentioned, we observed that not all the participants may know who is responsible for data backup, especially in today's mixed computing environment, where some applications have specific in-cloud backup support (e.g., Zoom [6]). Another possibility is that participants might underestimate the value of their data – the monetary cost of having a backup solution is more than the value of their data. In the following, we asked them about the value of their data. In particular, we asked the participants to estimate the value of their data on their machines.

How much is your data worth? We were interested to know the participants' opinions about the value of their data. One hypothesis was that a significant number of participants do not have any active protection plan probably because they think their data is not worth enrolling in a premium protection service. We asked participants to give a dollar value on how much their data might be worth. We observed that the reported responses were very diverse. For instance, 4 (0.8%) participants mentioned that their data would be worth more than $1 million. On the other hand, 114 (23.1%) of the participants reported that their data is worth less than $100. Data valued in the range of $100 to less than $1000 was reported by 133 (27.0%) participants. We observed that 229 (46.4%) of the participants believed that their data was worth between $1,000 and $1 million inclusive.

We did not specifically ask the participants how they valued their data, but in 13 (2.6%) cases, we noticed that the participants did not completely understand the question. For instance, in three instances, we observed that the reported value was unavailable or given extreme values such as "priceless". There were ten cases where the value was undisclosed by the participants because the participants highlighted that they either did not know how to value their data or refused to provide a value. However, 47.3% of the participants mentioned that their data would be worth at least $1,000. Figure 2 shows the CDF of the amounts reported by participants about their digital assets.

To gain a better understanding of how much users value their data and what it would take for them to have a reliable plan, the following question was asked: *A reputable data protection company offers a guaranteed full data protection for $200 per year. You don't need to pay the subscription now. You pay the subscription fee after a year (12 month) or after the first time you use the service to recover your data within the 12 month plan, whichever comes first. Would you enroll to such subscription plan for $200 per year?* The majority (67.3%) of the participants indicated that they would not choose this option, giving various reasons. Many respondents (20.5%) indicated the plan was too expensive or "not worth the cost," especially given that payment was due only upon recovery or after a year. This reflects a high sensitivity to pricing in data protection decisions. Other responses vary from no interest or perceived need, already have backup solutions, or ambiguous answers (72.8%) such as "nah" and "maybe later". It's possible that the ambiguous answers were due to either low engagement, survey fatigue [21] since this is one of the final questions or the need for a more structured prompt.

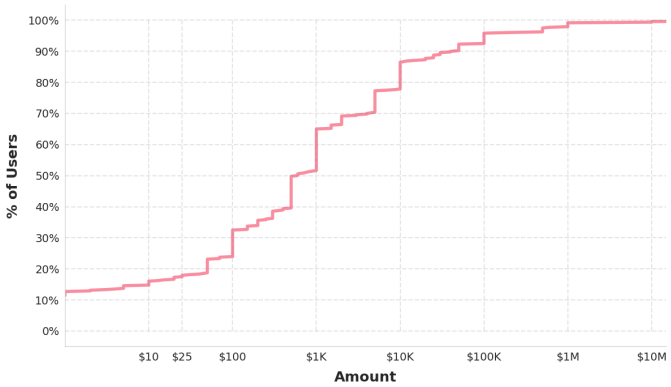

Fig. 2: **CDF of Estimated Data Worth.** 47.3% of participants valued their digital assets at $1,000 or more.

How much are you willing to pay as a ransom fee? As a follow-up question, we asked the participants how much they are willing to pay as a ransom fee in case their systems get infected and their data became encrypted. We observed that 353 (71.6%) participants reported they would not pay the ransom fee under any circumstances. The remaining 140 (28.4%) participants provided different responses, which were surprising to us. For instance, 13.8% of the participants were willing to pay the ransom fee if the amount was less than $500. The rest of those participants answered they would pay the ransom fee if the amount is less than $1,000 or they could pay in multiple installments. We observed that 62 (12.6%) of those participants who were willing to pay the ransomware fee did not have any backup solution or were using free subscription plans for cloud synchronization services (only 2-5 GB).

V. Discussion

While our analysis shows some fundamental flaws in users' decision-making on how to protect their data, it is not immediately clear how the security community should use these findings to protect users from ransomware and other emerging forms of attacks. In this section, we discuss the implications

TABLE III: **Backup tiers and participants' knowledge on data corruption attacks.** While 59.4% of respondents knew about ransomware, 30.4% were using limited or no backup services.

Backup Tier	Knows About Ransomware	At Least Somewhat Likely to be Infected	Rate Backup as At Least Somewhat Important to Backup Data
No backup services	39 (7.9%)	13 (2.6%)	54 (10.9%)
Free backup service	111 (22.5%)	44 (8.9%)	156 (31.6%)
Paid backup service	143 (29.0%)	60 (12.2%)	199 (40.4%)
Total	**293 (59.4%)**	**117 (23.7%)**	**409 (82.9%)**

of our investigation and make recommendations for potential routes forward.

Moving towards democratized protection has been slow. One direction to protect user data, in addition to data backup, has been to augment the operating system with data protection service [25], [27], [16], [35], [33]. The core insight in these research efforts is that data protection has cross-application appeal and must therefore be centrally positioned and supported by the operating system. That is, the defense mechanism should be well-integrated into the OS and deployed across target machines using a library, accessible to any authorized OS user. This approach not only simplifies wide adoption and promotes democratized protection against such attacks but also removes the need to introduce disjoint programs across target platforms. We understand that defining a solution that is compatible with every use case in today's heterogeneous computing environment is non-trivial. However, new design principles help to build pre-trained and generic defense models that can be optimized locally based on the workload and specification of target machines. Similar approaches have been used in other domains and have been successful. For instance, the latest state-of-the-art in face detection such as Face ID authentication [12] has taken an enormous leap forward by incorporating pre-trained models on mobile phones.

Users are left with vague ideas about data protection. Our analysis suggests that the participants had vague ideas about who should be in charge of providing data protection. Various data protection policies in third-party software systems have made the situation even more ambiguous. Users are left with the illusion that the provider of the service is the entity that is responsible for their data security. While we believe services will be promoted because of their robustness in safeguarding user data, it can also have some important consequences that have not been very evident, especially for less sophisticated users who are more likely to underestimate their role in maintaining their security posture. We acknowledge that more scientific work needs to be done in this domain to understand cognitive biases and systematic errors in users' perceptions of the modern cybersecurity landscape.

Improving users' decision-making through a behavioral science lens. The experiments suggest that users with insufficient data backup are not only vulnerable to ransomware but are likely to contribute to malware campaigns' businesses by paying the ransom fee. This work also shows that there is a lack of rigorous scientific methods in the cybersecurity space to understand and analyze users' judgment and their decision-making process on security and privacy issues. It is important to develop low-cost solutions that assist users to understand what their common mistakes are when it comes to decision-making and what needs to be done to influence users' behavior toward better decision-making. Proven techniques in other domains such as behavioral economics and psychology could be helpful in this domain as well. For instance, the concept of nudging (e.g., reminders, risk-framed messages, progress dashboards) [39] and other behavioral science concepts that helped people become better investors and avoid common mistakes about their money can be used in this context to influence users' perceptions for better decision-making.

VI. Limitations

In this section, we briefly explain the limitations of our study and their potential impacts on the overall experiments.

One limitation of this study was the sample size. While the participants are from different fields of study and technical skill sets, age and sex, a larger sample size would likely strengthen confidence in the observed patterns. A larger sample size would give us a more accurate view of how participants view the value of their data and how the value changes if the sample size grows dramatically. Furthermore, estimating the effectiveness of online awareness training would be less challenging with larger sample sizes. In addition, while Prolific offered a varied sample, participants might possess greater technological proficiency [41] than the general population, hence introducing some selection bias [8]. To reduce these biases in future work, we suggest using offline approaches (such as surveys done in person) together with Prolific recruitment.

We cannot extrapolate how real users might react to a real-world ransomware attack since this would require performing a behavioral analysis of the users when they are exposed to real attacks. Our study was mainly estimating if a participant is vulnerable to a ransomware attack and if common ransomware attack strategies, forcing the victim to pay, would be successful. We answered this question by examining whether there was a viable backup solution and if the participant was willing to pay. We did not investigate nor have the data to analyze the responses in a real-world setting.

VII. Related Work

User data protection has always been a critical topic in the computer science domain. In the following, we explain the trends that had formed over the last few years in the industry and academia, as well as important related work on the proposed solutions to safeguard users. The research is segmented into four categories: Ransomware, Phishing, Malware, and Scams in which surveys or observations were conducted on the human aspect of each event.

Users' Reactions to Ransomware. Camelia Simoiu, et al [38] conducted a study on 1,180 American adults and they estimated 2%-3% of respondents were affected and responded to attacks. Moreover, authors reported ransom fee on average was $530, and 4% of respondents paid this amount. Other work by Choi, et al [37] uses publicly available data to evaluate the impact of ransomware. They present a case study of such attacks against police departments. A Cyber-Routine Theoretical method was employed in this work to explain why attacks using ransomware have grown so widespread. The study suggested that online habits are a key contributor to ransomware victimization,

Users' Reactions to Phishing and Scams Websites. Users' ability to identify attacks has also been the topic of research in many prior works. For instance, Schechter, et al [36] evaluated users' ability to identify possible phishing websites. The study showed that users who disregarded HTTPS indicators and site-authentication graphics were rated ineffective. The study consisted of 67 participants. Alsharnouby, et al [10] took the approach of simply asking 21 users to identify phishing websites. The study discovered that even when primed to detect phishing websites, users only spotted 53% of them and they spent very little time looking at security indicators compared to website content when making assessments. A similar study was conducted on scam pages. Kirlappos et al [26], studied how much protection trust seals like VeriSign and TRUSTe provided to internet customers. They conducted a survey in which 60 experienced online buyers scored six websites depending on how trustworthy they considered them to be, with and without trust seals. They concluded that trust seals do not offer effective protection against scam websites. In another study [42], participants were exposed to a fake Google login page. 51 Participants contributed to this experiment. The study shows that only 10% of the participants recognized the fraudulent page.

VIII. Conclusions

In this paper, we conducted a study on users' preparedness for data corruption attacks. The experiment involved 493 participants from different fields of study and technical backgrounds. The results indicated that a significant number of participants were not prepared to protect their data in case of data loss. Most of the participants gave different responses about which entity should be in charge of their data. Participants who had undergone some form of awareness training appeared somewhat more inclined toward resilient data protection, though the effect was modest and not statistically significant given the small sample. We also observed that a significant number of participants could be potentially paying victims of ransomware attacks. That is, these participants did not have any viable backup solution and were willing to pay a ransom of less than $1,000. Finally, we provide perspectives on how we might proceed more effectively to respond to such threats in the future.

Acknowledgment

We would like to thank the anonymous reviewers for their thoughtful feedback. This project was supported by Microsoft AI Security, US National Security Agency (NSA) under Grant No. H98230-21-1-0324, US National Science Foundation (NSF) under Grant No. 2219920, and Intergovernmental Personnel Act Independent Research & Development Program. Opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding agencies.

References

[1] Dropbox. https://www.dropbox.com, Accessed: 09-15-2025.
[2] Dropbox plans. https://www.dropbox.com/plans, Accessed: 09-15-2025.
[3] icloud. https://www.icloud.com/, Accessed: 09-15-2025.
[4] Microsoft onedrive compare plans. https://www.microsoft.com/en-us/microsoft-365/onedrive/compare-onedrive-plans, Accessed: 09-15-2025.
[5] Prolific. https://www.prolific.com/, Accessed: 09-15-2025.
[6] Zoom. https://www.zoom.com/, Accessed: 09-15-2025.
[7] Acronis Inc. Ransomware Protection and Removal. https://www.acronis.com/en-us/solutions/ransomware-protection/, 2021.
[8] Billur Aksoy and Saggi Nevo. Unlocking insights into prolific: Research implications, participant behavior and motivations. *Participant Behavior and Motivations (March 21, 2025)*, 2025.
[9] Allison Sylte. Cyberattack involves 300,000 CU records, university won't pay ransom. https://www.9news.com/article/news/local/university-of-colorado-cyberattack/73-61a82ad4-abbb-459a-9b9e-e72cce0e179d#:~:text=BOULDER\%2C\%20Colo.,to\%20some\%20social\%20security\%20numbers., Accessed: 08-03-2021.
[10] Mohamed Alsharnouby, Furkan Alaca, and Sonia Chiasson. Why phishing still works: User strategies for combating phishing attacks. *International Journal of Human-Computer Studies*, 82:69–82, 2015.
[11] ANY.RUN. United states census. https://data.census.gov/profile/United_States?g=010XX00US, Accessed: 09-11-2025.
[12] Apple Inc. Logging a User into Your App with Face ID or Touch ID. https://developer.apple.com/documentation/localauthentication/logging_a_user_into_your_app_with_face_id_or_touch_id, Accessed: 07-18-2021.
[13] Armit Singh. Ransomware: How to Prevent or Recover From an Attack. https://www.backblaze.com/blog/complete-guide-ransomware/, Accessed: 07-28-2021.
[14] Carrie Reber. Why Ransomware Is a Business Continuity Issue? https://www.n-able.com/blog/why-ransomware-business-continuity-issue, Accessed: 08-20-2020.
[15] Catalin Cimpanu. Airplane maker Bombardier data posted on ransomware leak site following FTA hack. https://www.zdnet.com/article/more-than-290-enterprises-hit-by-6-ransomware-groups-in-2021/, Accessed: 08-01-2021.
[16] Andrea Continella, Alessandro Guagnelli, Giovanni Zingaro, Giulio De Pasquale, Alessandro Barenghi, Stefano Zanero, and Federico Maggi. Shieldfs: a self-healing, ransomware-aware filesystem. In *Proceedings of the 32nd Annual Conference on Computer Security Applications*, pages 336–347, 2016.
[17] CyberReason Inc. https://www.cybereason.com/hubfs/ransomfree-EOL-message.pdf, Accessed: 06-17-2021.
[18] Heiling Defense Inc. Heiling Defense: Developing innovative defensive solutions today to defeat tomorrow's advanced threats. https://www.heidef.com/, Accessed: 07-17-2021.

[19] HelpNet Security. While nearly 90% of companies are backing up data, only 41% do it daily. https://www.helpnetsecurity.com/2020/04/03/back-up-data/, Accessed: 11-02-2020.

[20] Malwarebytes Inc. Malwarebytes anti-ransomware technology. https://www.malwarebytes.com/business/solutions/ransomware, Accessed: 07-23-2021.

[21] Dahyeon Jeong, Shilpa Aggarwal, Jonathan Robinson, Naresh Kumar, Alan Spearot, and David Sungho Park. Exhaustive or exhausting? evidence on respondent fatigue in long surveys. *Journal of Development Economics*, 161:102992, 2023.

[22] Jimena Tavel. University of Miami hit with ransomware attack, private info of medical patients posted online. UniversityofMiamihitwithransomwareattack,privateinfoofmedicalpatientspostedonline, Accessed: 08-02-2021.

[23] Jordan Williams. University of california victim of ransomware attack. https://thehill.com/policy/cybersecurity/546335-university-of-california-victim-of-ransomware-attack, Accessed: 08-01-2021.

[24] Kartikay Mehrotra and William Turton . CNA Paid $40 Million in Ransom After March Cyber Attack. https://www.insurancejournal.com/news/national/2021/05/21/615373.htm, Accessed: 07-19-2021.

[25] Amin Kharraz and Engin Kirda. Redemption: Real-time protection against ransomware at end-hosts. In *Proceedings of the 20th International Symposium on Research in Attacks, Intrusions and Defenses (RAID)*, 9 2017.

[26] Iacovos Kirlappos, M. Angela Sasse, and Nigel Harvey. Why trust seals don't work: A study of user perceptions and behavior. In Stefan Katzenbeisser, Edgar Weippl, L. Jean Camp, Melanie Volkamer, Mike Reiter, and Xinwen Zhang, editors, *Trust and Trustworthy Computing*, pages 308–324, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[27] Eugene Kolodenker, William Koch, Gianluca Stringhini, and Manuel Egele. Paybreak: Defense against cryptographic ransomware. In *Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security*, pages 599–611, 2017.

[28] Lawrence Abrams. Flagstar Bank hit by data breach exposing customer, employee data. https://www.bleepingcomputer.com/news/security/flagstar-bank-hit-by-data-breach-exposing-customer-employee-data/, Accessed: 07-19-2021.

[29] Lisa Vaas. DarkSide Hits Toshiba; XSS Forum Bans Ransomware. https://threatpost.com/darkside-toshiba-xss-bans-ransomware/166210/, Accessed: 08-01-2021.

[30] Vincent Lombardi, Sarah Ortiz, Jen Phifer, Tomas Cerny, and Dongwan Shin. Behavior control-based approach to influencing user's cybersecurity actions using mobile news app. In *Proceedings of the 36th Annual ACM Symposium on Applied Computing*, SAC '21, page 912–915, New York, NY, USA, 2021. Association for Computing Machinery.

[31] Mark Gill. 10 Shocking data loss and disaster recovery statistics. https://www.comparitech.com/data-recovery-software/disaster-recovery-data-loss-statistics/, Accessed: 07-18-2021.

[32] David R Matos, Miguel L Pardal, Georg Carle, and Miguel Correia. Rockfs: Cloud-backed file system resilience to client-side attacks. In *Proceedings of the 19th International Middleware Conference*, pages 107–119, 2018.

[33] Shagufta Mehnaz, Anand Mudgerikar, and Elisa Bertino. Rwguard: A real-time detection system against cryptographic ransomware. In *International Symposium on Research in Attacks, Intrusions, and Defenses*, pages 114–136. Springer, 2018.

[34] Mike Williams . CyberSight RansomStopper review. https://www.techradar.com/reviews/cybersight-ransomstopper, Accessed: 07-24-2021.

[35] Nolen Scaife, Henry Carter, Patrick Traynor, and Kevin RB Butler. Cryptolock (and drop it): stopping ransomware attacks on user data. In *2016 IEEE 36th International Conference on Distributed Computing Systems (ICDCS)*, pages 303–312. IEEE, 2016.

[36] Stuart E. Schechter, Rachna Dhamija, Andy Ozment, and Ian Fischer. The emperor's new security indicators. In *2007 IEEE Symposium on Security and Privacy (SP '07)*, pages 51–65, 2007.

[37] Kyung shick Choi, Theresa M Scott, and Daniel P. LeClair. Ransomware against police: Diagnosis of risk factors via application of cyber-routine activities theory. 2016.

[38] Camelia Simoiu, Joseph Bonneau, Christopher Gates, and Sharad Goel. "i was told to buy a software or lose my computer. i ignored it": A study of ransomware. In *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*, pages 155–174, Santa Clara, CA, August 2019. USENIX Association.

[39] R.H. Thaler and C.R. Sunstein. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. A Caravan book. Yale University Press, 2008.

[40] Tim Rettig. 2020's Worst Ransomware Attacks and Why Backups Should Lead the Protection in 2021. https://www.intrust-it.com/2020s-worst-ransomware-attacks-and-why-backups-should-lead-the-protection-in-2021/, Accessed: 07-27-2021.

[41] Anne M Turner, Thomas Engelsma, Jean O Taylor, Rashmi K Sharma, and George Demiris. Recruiting older adult participants through crowdsourcing platforms: Mechanical turk versus prolific academic. In *AMIA annual symposium proceedings*, volume 2020, page 1230, 2021.

[42] Enis Ulqinaku, Hala Assal, Abdou AbdelRahman, Sonia Chiasson, and Srdjan Capkun. Is real-time phishing eliminated with fido? social engineering downgrade attacks against fido protocols. In *Proceedings of the 30th USENIX Security Symposium (USENIX Security 21)*, pages 3811–3828. USENIX Association, 2021.

[43] Joobeom Yun, Junbeom Hur, Youngjoo Shin, and Dongyoung Koo. Cld-safe: an efficient file backup system in cloud storage against ransomware. *IEICE TRANSACTIONS on Information and Systems*, 100(9):2228–2231, 2017.

APPENDIX

EVALUATION QUESTIONS

TABLE IV: Education & Interests survey questions.

Education & Interests	
1.	Which degree are you currently pursuing?
2.	Have you taken courses in any of the following subjects?
3.	Have you been employed in any of the following areas?
4.	Do you take interest in cybersecurity, computer science, or information technology?
5.	Which of the following media outlets have you visited during the last week?
6.	Which ad-blocker do you use?
7.	Which Password Manager do you use?
8.	How important is backup security to you?
9.	Which operating system is your main operating system?
10.	Give a score to your current cybersecurity skills.

TABLE V: Technical Background questions.

Technical Background	
1.	What programming language is the most susceptible to buffer overflow attacks?
2.	SQL injection can happen because _____ has vulnerabilities.
3.	Cross site scripting is a _____ attack.
4.	ASLR is a defense mechanism against
5.	Consider the following command where the user is connecting to a remote server: ssh user1-remotehost -p 12345. Is something wrong with the syntax used?
6.	When was the last time you used SSH to connect to a remote server?
7.	Which of the following programming languages are you familiar with? (More than a year of experience)
8.	Which of the following cyber-attacks are you familiar with?
9.	Ransomware does which of the following?

TABLE VI: Knowledge, Experience, and Preferences questions.

Knowledge, Experience, and Preferences	
1.	Have you been a victim of a ransomware attack and if so, did you pay the ransom?
2.	If you answered yes to the last question and did not pay the ransom, what was your reasoning for not paying the ransom?
3.	How important do you think it is on a scale of 1-5 to implement ransomware protection on your personal electronic devices (PC, smartphone, tablet)?
4.	How likely do you think it is that you will be targeted by a ransomware attack in the near future?
5.	If your critical data (e.g., code, documents, projects, personal photos and videos) is encrypted by ransomware and you don't have a backup, are you willing to pay the ransom fee?

TABLE VII: Security Practices questions.

Security Practices	
1.	Do you use antivirus security on your PC? If so, how long have you used it?
2.	Do you currently have ransomware protection installed on your personal electronic devices (PC, smartphone, tablet)?
3.	Do you use any sort of backup service on your PC? (e.g. Google Drive, iCloud, Acronis True Image) If so, for how long?
4.	What tier of backup security service do you use?
5.	Which of the following backup services have you used?
6.	What is your opinion of automated backups vs. manual backups?
7.	What is your preferred backup method?