"I was told to buy a software or lose my computer. I ignored it": A study of ransomware

Camelia Simoiu
Stanford University

Christopher Gates
Symantec

Joseph Bonneau
New York University

Sharad Goel
Stanford University

Abstract

Ransomware has received considerable news coverage in recent years, in part due to of several cases against high-profile corporate targets. Little is known, however, about the prevalence and characteristics of these attacks on the general population. Using a detailed survey of a representative sample of 1,180 American adults, we estimate that 2%–3% of respondents were affected over a 1-year period between 2016 and 2017. The average payment amount demanded was $530 and only a small fraction of affected users (about 4% of those affected) reported paying. Perhaps surprisingly, cryptocurrencies were typically only one of several payment options, suggesting that they are not a primary driver of ransomware attacks. Nevertheless, given the high payment amounts, our results suggest that American users may be paying on the order of $100 million per year to attackers. We conclude our analysis by developing two risk-assessment models, one based on self-reported security habits and a second based on detailed, individual-level web browsing patterns.

1 Introduction

Ransomware is a particularly pernicious form of malware that restricts an individual's access to their computer (e.g., by encrypting their data) and demands payment to restore functionality. While the first documented ransomware attack dates to 1989, ransomware remained relatively uncommon until the mid 2000s [19]. Since then, the attack has been automated and professionalized. It is believed to be highly lucrative, with previous damages estimated at hundreds of millions of dollars per year. For example, the damages caused by a single ransomware variant, CryptoWall3, were estimated to be over $320 million in 2015 alone [1].

Consumers are the most common victims of ransomware, accounting for 57% of infections from January 2015 to April 2016 [3]. While most attacks are thought to be untargeted, consumers are often less likely to have robust security in place, increasing the likelihood of falling victim to ransomware [3].

Despite the harm ransomware can inflict, relatively little is known about the prevalence and characteristics of such attacks in the general population. Various government and industry organizations have attempted to document the phenomenon, but results are often inconsistent, in part due to the non-representative data they are based on. Industry reports are typically published by security firms based on users of their software products, while government agencies typically report rates based on voluntary victim reports that likely underestimate the true rate [27]. For example, one industry report estimates that two million U.S. households were victims of ransomware from January 2015 to April 2016 [3], while the FBI reported only 2,673 ransomware attacks in 2016 based on victim complaints filed with the Bureau's Internet Crime Complaint Center [2]. Reliable estimates of the prevalence of ransomware infection are necessary both for understanding the nature of today's threat landscape, as well as for longer-term comparison and analysis.

Apart from the difficulty in characterizing the extent of the problem, little is known about the factors and behavioral patterns that place individuals at risk of such attacks. Devising accurate risk assessment methods to identify the vulnerable population is particularly relevant for ransomware attacks, as there is often little recourse for victims who need to recover their data other than to pay the ransom. Once identified, information about the vulnerable population can be used to establish proactive strategies to mitigate the effects of ransomware attacks. For example, security vendors can fine-tune and re-prioritize defense mechanisms to offer additional protection layers or differentiated services to at-risk users. In the same vein, the vulnerable population can be influenced through several means, including personalized educational resources and training, or discounted offers for services to mitigate the effects of infection (e.g., cloud-based data backup services).

We make two key contributions. First, we report the results of a representative online survey of 1,180 U.S. adults that queried respondents' security experiences with particu-

lar focus on ransomware attacks. Second, we propose two user-centric risk assessment models for ransomware victimization, the first based on self-reported security habits, and a second based on individual-level web browsing behavior collected by a large antivirus provider. Using time-stamped web browsing history as well as telemetry data on blocked ransomware attacks, we construct a machine learning model to predict future infection status for an anonymized sample of over 4 million users.

2 Related work

The classical paradigm to defend against malware attacks has traditionally been victim-agnostic and reactive, with defenses focusing on identifying the attacks or attackers (e.g., phishing emails, malicious websites, and files) [15]. For example, several studies propose technical, automated solutions to prevent ransomware attacks [19, 28, 18, 32, 6].

More relevant to our work are studies that adopt victim-centric approaches to identify the vulnerable population and behaviors that predispose users to malware infections. Ngo et al. [25] apply the general theory of crime and routine activities [10] to assess the effects of individual and situational factors on seven types of cybercrime victimization—among them, a computer virus. They administer an online survey of self-reported cybercrime victimization to 295 students in the U.S., and find that the odds of obtaining a computer virus increased for users who were younger, non-white. Perhaps unintuitively, they also find that individuals who frequently opened any unfamiliar attachments or clicked on web-links in the emails that they received, opened any file or attachment on their instant messengers, and frequently clicked on a pop-up message that interested them, had lesser odds (by about 35%) of obtaining a computer virus. Bossler et al. [5] conducted a survey of 788 college students to study the risk factors of data loss caused by malware infection. The factors studied include "deviant" behavior (e.g., pirated media downloads, visiting adult websites), routine behaviors (e.g., social media use, programming, shopping), guardianship measures (e.g., having AV software, sharing passwords), and computer skills. Drawing on results from a logistic regression model, the authors find that being employed and being female increased the odds of malware victimization. Engaging in deviant behavior was generally not a strong predictor of malware infection—the only form of personal deviance that increased the risk of malware infection was pirating media. Personal and physical guardianship played small roles in explaining infections; and strong computer skills and careful password management did not reduce estimated threat of malware victimization.

More recently, researchers have turned to large-scale, data-driven approaches to predict the risk of various cyber threats. Drawing on a dataset of ADSL customers, Carlinet et al. [8] profiled customers who were more likely to generate malicious traffic based on their network usage with respect to different types of applications. They found that the risk of virus infections increased with usage of web and streaming, while no such link could be established for peer-to-peer and chat applications.

Maier et al. [23] examine whether the risk of generating malicious traffic is correlated with security hygiene using DSL data logs of anonymized network traces. They find that having a good security hygiene (e.g., applying operating system software updates) has little correlation with being at risk, while accessing blacklisted URLs more than doubles risk. Sharif et al. [29] propose a system that enables proactive defenses at the level of a single browsing session. By observing three months of HTTP traffic generated by 20,645 users of a large cellular provider, the system can predict whether users will be exposed to malicious content on the web seconds before the moment of exposure, thus opening a window of opportunity for proactive defenses.

Levesque et al. [21, 22] observe malware exposure of 50 subjects over a 4-month period using instrumented computers from the clinical trial of an antivirus product. They find that malware victimization is correlated with a high self-reported level of computer expertise, increased file downloads and application installations, and high browsing volume. The authors find mixed results with respect to the age of the user and the content categories of websites.

Using Symantec telemetry for a subset of 1.6 million users over an 8-month period, Ovelgonne et al. [26] study the relationship between the number of attempted malware attacks detected and user profiles. The authors classify users into 4 categories (gamers, professionals, software developers, others), and find that software developers are more at risk of engaging in risky cyber-behaviors and that there is a subpopulation of gamers with especially risky behavioral patterns.

Using a similar telemetry dataset from Symantec, Canali et al. [7] use data on URLs visited by more than 100,000 users during a period of three months to predict their risk of visiting a malicious Web site. A logistic regression model based on 74 features related to web browser usage confirm that users with abundant and varied browsing behavior experience higher risks. With the exception of adult content, the category was generally not found to be significant.

Yen et al. [31] and Bilge et al. [4] study individual user-level malware encounters in an enterprise setting. Yen et al. draw on web proxy logs, user demographics, and VPN logs from a large, multi-national enterprise. The authors investigate features related to categories of web sites visited, aggregate volumes of web traffic, and connections to blocked or low-reputation sites. They use a logistic regression model for inferring the risk of hosts encountering malware and find that among the three feature categories, user demographics is the most powerful at indicating risk, followed by VPN behavior. Counterintuitively, web activity contributed marginally

to the overall model and they reasoned that this is due to the fact that only 3% of the hosts encountered malware from the web. Bilge et al. [4] develop a predictive model of malware infections that uses per-machine file appearance logs collected from 600,000 machines belonging to 18 enterprises. Although focused on enterprises, they use a similar data mining approach to our work, constructing a prediction model based on machine profiles. Profiles consist of 89 features that are based on the volume of events, their temporal patterns, application categories, rarity of files, patching behavior, and past threat history. They find that temporal file download and creation behavior, as well as the volume and diversity of file creation activities have the highest impact on identifying risky machines.

3 Representative survey

We administered a survey on ransomware experiences to a sample of 1,180 U.S. adults. Participants were recruited by YouGov, an online marketing firm, and reimbursed for their participation. In order to derive nationally representative estimates, YouGov provided weights for the full sample of 1,180 respondents. Summary statistics detailing the demographic and socioeconomic characteristics of respondents are provided in Table A1 in the Appendix. Prior to running the study, the survey tool was piloted on Amazon's Mechanical Turk. Five pilot tests with 100 participants each were run. During each of these phases, the tool was updated based on feedback from respondents. All aspects of our study were approved in advance by the university's Institutional Review Board (IRB protocol number 40466).

3.1 Defining a ransomware attack

We define ransomware as the class of malware that attempts to defraud the user by restricting access to the user's computer or data, typically by locking the computer or encrypting data. There are thousands of different ransomware types in existence today, ranging in design and sophistication. Some ransomware strains can be trivially circumvented (e.g., by a normal restart), while others employ a variety of advanced tactics. For example, they may utilize payload persistence, ensuring the ransomware persists after a restart; use strong encryption methods that are nearly impossible to reverse, or disable system restore functionality (e.g., delete Windows shadow copies) in order to prevent encrypted data from being restored to an older, unencrypted version [13].

Yet another class of ransomware, sometimes referred to as "fake ransomware", informs infected users that their data has been encrypted or their computer locked, however does not actually do these things. These types of attacks are less sophisticated from a technical perspective, are usually relatively easy to circumvent, and rely on scare tactics to coerce the user into paying the ransom amount. We ask respondents

Figure 1: Examples of ransomware screenshots shown to respondents: a strain impersonating the FBI and two encryption ransomware variants, with and without a timer.

to report all types of ransomware, and distinguish between different types of ransomware post-response.

3.2 Establishing victimization status

In order to ensure the accuracy of self-reported ransomware attacks, respondents progressed through a series of ten questions and information pages describing typical ransomware attacks and their characteristics. Respondents were initially shown an information screen with the following definition of ransomware: *Ransomware is a type of malware that will either lock your computer screen or encrypt your files. If you've been infected with ransomware, you will see screens like the examples below, informing you that you must pay a ransom to re-gain access to your computer and/or files, providing instructions on how to do so.*

Three screenshots of ransomware variants were shown as examples: a strain of ransomware impersonating the FBI, and two encryption ransomware variants, with and without a timer (Figure 1). In order to distinguish ransomware attacks from malware with similar characteristics, respondents saw a question page explaining that ransomware is different from technical support scams, where a misleading application alerts the user to a security issue or vulnerability on their computer and then prompts them to call a tech support number to download or purchase anti-virus software in order to resolve the issue.

A series of additional questions were then used to confirm respondents' self-reported victimization status. Three questions asked whether they experienced various characteristics commonly found in ransomware attacks, namely: (1) whether they had seen similar images notifying them that their computer was locked or files/data encrypted; (2) whether their files were encrypted and they saw files with names such as "HOW TO DECRYPT FILES.TXT", "DE-

CRYPT INSTRUCTIONS.HTML", or with unusual extensions such as ".locky"; and (3) if they saw a timer counting down and messages indicating that if payment is not completed before the time expires, the ransom amount will increase or the encryption key will be deleted. The ransomware definition above was then repeated, and respondents were asked whether they had experienced a ransomware attack, and could answer, *Yes*, *No*, or *I am not sure*. If respondents indicated that they were not sure or if their initial response was inconsistent with the three questions on ransomware characteristics (e.g., they checked off at least one of the three characteristics, but concluded that they did not have ransomware), they progressed through a series of further clarification questions and information pages, which ultimately culminated with a question asking them to confirm whether or not they had ransomware.

Respondents indicating they had experienced a ransomware attack either in the first or second prompt progressed to a series of questions soliciting information about the attack. They were asked to describe the ransomware attack in their own words, with prompts to include the contents of the message or instructions, the appearance of the screen, and if any functionality of their computer was disabled. They were also asked a series of multiple choice questions detailing: the month and year of the attack, the name of the ransomware variant, how much ransom (money) was demanded, the method of payment, whether they paid the ransom and why (or why not), whether access was restored after payment (if applicable), which strategies, if any, they attempted to remove the ransomware and restore access to their computer, whether they sought help in removing the ransomware, whether they were able to remove the ransomware without loosing their data, how the ransomware was eventually removed, and whether they notified the authorities.

These questions served a dual purpose: apart from allowing us to distinguish between strains with differing attributes (e.g. encryption, screen lock, impersonation of law enforcement), they provided an additional means of validating that the reported incident was indeed a ransomware attack. If respondents had experienced more than one ransomware attack, they were instructed to respond to questions based on the last attack.

3.3 Re-classification

As is the case with all surveys, our results are subjects to limitations as they are based self-reported infection rates rather than upon actual detections of malware. Despite our efforts to ensure that respondents understood what a ransomware attack is, we cannot be certain that ransomware attacks or their attributes were correctly identified. For example, if respondents did not remove the ransomware themselves or did not try relevant troubleshooting strategies (e.g., changing the extensions of files back to the original in order to test whether the encryption was real or not), we cannot know with certainty whether strong encryption was used.

The responses of self-reported ransomware victims were independently reviewed by two researchers and ambiguous responses were re-classified when necessary. Each response was classified under two regimes: a *conservative* regime and an *inclusive* regime. The inclusive regime excludes cases where the respondent described a different type of malware attack (e.g., scareware, pop-up announcing that they were the winner of a contest, or tech support scam), or admitted that they did not remember the details of the attack. The conservative regime also excludes the above cases, and in addition, all cases where the description of the attack was ambiguous or no description was provided. Sample responses and their classification under the two regimes are listed in Table 1.

Throughout our analysis, we use the respondent weights provided by YouGov to compute prevalence[1]. In particular, prevalence is estimated via the expression:

$$r = \sum_i \frac{I_i(\text{victim}) w_i}{w_i} \tag{1}$$

where $I_i(\text{victim})$ is an indicator function representing whether the respondent experienced ransomware or not, and w_i is the weight.

4 Survey results

4.1 Rate of ransomware victimization

Originally, 153 respondents (14%) reported that they had experienced a ransomware attack at some time in the past (Table 2). Following re-classification, we estimate that the overall proportion of the U.S. population reporting a ransomware infection at any time in the past ranges between 6% (s.e.=1%) under the conservative regime, and 9% (s.e.=1%) under the inclusive regime. We similarly estimate that between 2% (s.e.=0.4%) and 3% (s.e.=0.5%) of the U.S. population were affected over the one-year period between 2016 to 2017 under the conservative, and inclusive regimes, respectively.

4.2 Ransomware attributes

As noted above, estimated prevalence was similar under both the inclusive and conservative classification schemes. For this reason, in what follows, results are reported for victims classified under the inclusive regime only for ease of exposition.

Ransomware strains that lock the computer appear to be significantly more common than those that employ encryption — 74% of victims reported experiencing screen locks, while only 33% reported experiencing strains that encrypted

[1] Weighted rates were typically quite close to the raw proportions.

Respondent's description	Screen lock	Encryption	Law enforcement	Timer	Inclusive	Conservative
"Illegal files detected. FBI has locked your computer. purchase a prepaid VISA and pay fine online . (included a fake web cam window)	•		•		Ransomware	Ransomware
FBI - YOU HAVE BEEN WATCHING PORN OR GAMBLING OR BOTH, YOU MUST PAY $200 TO MONEYGRAM"	•	•	•		Ransomware	Ransomware
"The screen looked like one you previously displayed. It encrypted just about all my files except *.exe files and a few others. I lost everything on my PC and external hard drive. I ended up reformatting and starting from scratch. I could run programs, but could not access any of my saved working files. A screen would display telling me to call a number, pay the ransom and they would decrypt my files. They wanted $500."		•			Ransomware	Ransomware
"i don't remember what the message said it just prevented me from getting to any of the stuff on my computer and then i started it in safe mode and got rid of it"	•				Ransomware	False Positive
"telling me to go purchase a gift card from walgreen"					Ransomware	False Positive
"It popped up and stated that I had to pay to gain access back to my computer and I was unable to do anything."	•				Ransomware	False Positive
"the screen was flashing call this number immediately to get your computer repaired. I was gullible and scared so I called. the guy got a hold of my computer and then told me I had to pay $300 for him to fix it. I told him I didn't have that kind of money and he hung up on me. I then went and changed all my passwords and prayed he didn't get any important info from me."		•			False positive	False positive
"I don't recall exactly, However when I called to find out what was going on, I was told that I would have to pay to get what ever was holding up my computer off, I said, You put it on, just take it off. My computer was older, I just went and bought another computer, I decided not to be an ATM for criminals."					False positive	False positive
"Was told to send $1000.00 dollars to clean up computer."					False positive	False positive

Table 1: *Sample descriptions and reported characteristics of the attack, and their corresponding classification under the conservative and inclusive classification regimes. Responses classified as false positives under the conservative regime, but not the inclusive regime, are typically classified as such due to unclear or ambiguous descriptions, which cannot confirm that the attack was a ransomware attack. Responses classified as false positives under both regimes are typically classified as such because they include few ransomware characteristics (if any) and the descriptions provided typically describe other scams (technical support scams, scareware, etc.).*

their files. Figure 4.2 shows the distribution of attributes experienced by ransomware victims. Two observations stand out. First, a large proportion of victims reported experiencing strains that impersonate law enforcement agencies, typically the FBI (46%). These strains typically display a message claiming that the user's computer was locked because they engaged in illegal activities (e.g., browsed illegal pornographic websites), and a fine must be paid to regain access. Second, encryption does not appear to be commonly used in conjunction with law enforcement impersonation. Only 22% of victims reporting law enforcement strains also reported that their files were encrypted, while 43% of victims that did not experience law enforcement strains experienced encryption.

4.3 Ransom payment

A histogram of reported ransomware amounts demanded is shown in Figure 3. The average ransom is $530 (standard error $125), while the maximum amount reported reached

	All victims	Last 12 months
Self-reported	14%	5%
Re-classified (inclusive)	9%	3%
Re-classified (conservative)	6%	2%

Table 2: *Estimated proportions of ransomware victimization for the U.S. population under the conservative and inclusive classification schemes. Estimates are based on results from a probabilistic, census-representative sample of 1,180 U.S. respondents who completed an online survey between June and September, 2017. The "all victims" column includes all respondents who reported experiencing a ransomware attack at any time in the past, and the "last 12 months" column includes those who reported attacks within one year of the survey date.*

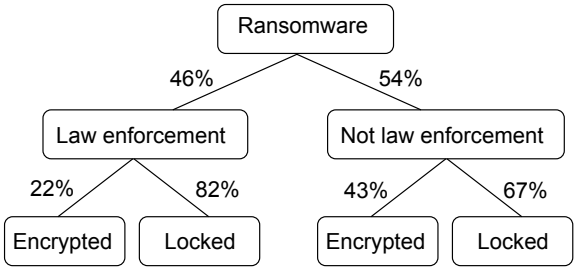

Figure 2: *Distribution of ransomware attributes (impersonation of law enforcement, screen lock, encryption). Encryption is not commonly used in conjunction with law enforcement impersonation.*

$8,000. The most common payment methods reported were wire transfers and payment voucher systems (e.g., Paysafecard, MoneyPak, CashU, MoneXy, prepaid Visa)[2], which together accounted for 56% of all reports. In contrast, only 12% of respondents reported being asked to pay only via cryptocurrency. Table 3 shows the distribution of payment method for all respondents reporting a ransomware attack. Results are qualitatively similar for victims within the last 12 months: 62% reported wire transfers or payment voucher systems, whereas only 2% reported cryptocurrencies. While recent work has focused on tracking Bitcoin payments as a means to estimate ransomware infections and quantify financial losses [16], this finding suggests that focusing solely on cryptocurrencies may underestimate losses. Secondly, it suggests that cryptocurrencies are not the driving force of the recent ransomware trend. Based on these results, a rough estimate places annual losses from ransomware to U.S. adults

[2]Respondents were presented with a multiple choice question and asked what payment method they were asked to pay the ransom in. As respondents could not select multiple payment methods, we are able to estimate a lower bound on the distribution of payment methods. Only respondents that were able to recall the ransom amount are included (n=66).

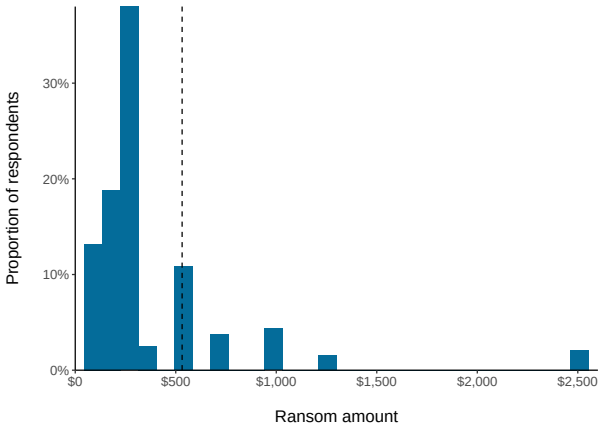

Figure 3: *Histogram of reported ransom amounts for respondents recalling an amount (n=66). The average ransom reported was $530 (standard error $125), and the maximum amount, $8,000.*

Method of payment	Proportion
Pre-paid cash voucher	41%
Wire transfer	15%
Cryptocurrency	12%
Premium-rate text message	7%
Not displayed	15%
Do not remember	10%

Table 3: *Distribution of payment methods for all respondents experiencing ransomware. Wire transfers and pre-paid cash vouchers predominate, whereas cryptocurrencies account for 12% of reported payment methods. The"not displayed" category includes cases where the payment method was not directly displayed. Respondents typically had to follow a link or call a number to find out the ransom amount (and did not do so).*

at over $100 million[3].

4.4 Means of dealing with the attack

Respondents reported a wide range of methods for dealing with the ransomware infection, depending on the severity of the attack. The majority of victims were able to find means to remove the ransomware without seeking someone else's help (78%). The remaining 13% either paid for repairs at a computer shop or asked friends or family members for help in removing the ransomware (Table 4.4). Approximately a third of victims (30%) were able to circumvent the attack with a restart (either a normal restart or a restart in safe mode),

[3]We arrive at our estimate by multiplying the number of U.S. adults (approximately 252 million in 2017) by our estimated proportion of infection (a conservative estimate of 2%), the average ransom amount ($530), and the proportion of users that pay (4%).

Method	Proportion
Restarted computer	30%
Online tool	18%
Restored computer from backup	22%
Removed by someone else	13%
Reformatted computer	5%
Removed using AV software	5%
Paid ransom	4%
Other means	3%

Table 4: *Self-reported means of removing the ransomware. Few*

suggesting that these attacks relied on scare tactics and did not employ sophisticated locking or encryption mechanisms. None of these victims lost data as a result of the attack. 18% of victims used an online tool to remove the ransomware and/or decrypt their files, and an additional 5% of respondents used an antivirus software to remove the ransomware, in some instances purchasing the software for this purpose. Few victims lost data as a result of the attack. The majority of users that had to restore their computer had backups. Very few respondents paid the ransom (4%) or reported the attack to authorities (11%)

The self-reported reasons for paying the ransom focused on feelings of distress and aversion to loosing files, as well as lack of computer knowledge. Respondents' original reasons are given below:

- *I am computer illiterate. A little smarter now.*
- *We were very distressed and felt it was a legitimate request.*
- *I'm so scared*
- *Did not want to lose any files or programs on my system.*
- *I was a full time caregiver for my critically ill husband. He used the computer a great deal to maintain contact with friends and family. I did not want to take the computer somewhere to have the problem corrected at what likely would have been a more expensive cost. (Keep in mind this was around 10 years ago and the ransom was under $50.00 which my credit card company did not pay)*
- *The price was not that high.*

4.5 Behavioral changes post-attack

Victims were asked to indicate whether they changed any of their habits following the attack, if any. Approximately half of respondents reported changing at least 2 habits (Table 5). The top three changes reported were more careful browsing (65%), purchasing an antivirus software (44%), and updating their existing antivirus product (31%). Whether or not participants truly changed their browsing behavior following the attack, or if this is a form of social desirability bias, is difficult to know for sure. Nevertheless, this result suggests that the majority of victims attribute the cause of the attack, at least in part, to their own behaviors — as opposed to adopting the view they were merely unlucky in becoming infected.

Habit	Proportion
More careful browsing	65%
Purchased AV product	44%
Updated AV product	31%
Started to backup data	26%
Enable automatic updates	24%
Backup data more regularly	22%
Changed OS configurations	20%
Changed OS	10%
Changed default browser	12%
Encrypted hard drive	0%

Table 5: *Behavioral changes following the attack for ransomware victims. Multiple answers were permitted, and half of respondents reported changing at least two habits. The top three changes reported were more careful browsing, and purchasing or updating an antivirus product. "Enable automatic updates" refers to updates to the OS, browser, antivirus, and other programs. Examples of configuration changes are disabling Windows Script Host, restricting login access, enabling the "show file extension" feature in Windows, etc.)*

Few respondents reported changing their operating system (OS), although we find that victimization varies significantly with OS. The majority of respondents used Windows as their OS (82%)[4]. We find that 10% of Windows users were victims, whereas only 5% of non-Windows users were victims. This difference is statistically significant under both the conservative and inclusive classification schemes, using a two-proportion Z-test at the 5% significance level. We also find that only 36% of respondents began to backup their data *or* backed up their data more frequently following the attack. This is arguably the single most effective way to mitigate the effects of ransomware attacks, suggesting that more awareness is needed around the importance of this habit.

5 Risk assessment based on self-reported security habits

In decision-making scenarios, users and experts alike often make assessments based on experience and intuition rather than on statistical analysis [14]. Despite a large body of work showing that intuitive judgments are generally infe-

[4] 12% used a Mac, 4% used Chrome, while the remaining 2% used another OS.

rior to those based on statistical models [11, 20], decision-makers have consistently eschewed formal assessment models. This is in part because it has been difficult to create, understand, and apply them. Here we construct a statistically derived heuristic risk assessment of future ransomware infection (within the next 12 months), based on self-reported security habits. The result is a risk rubric that enables assessments to be made quickly in one's mind, without the aid of a computing tool, requires only limited information, and exposes the grounds on which assessments are made.

Following Jung et al. [17], we use the select-regress-and-round strategy [17] to construct a heuristic that performs on par with traditional machine learning algorithms. Given each respondent's profile, we first construct a model to predict infection status using two standard machine learning models, lasso and gradient boosted trees (GBM). To do so, we draw on several features extracted from the survey: demographics, socioeconomic status, the software used, level of computer knowledge (using an 8-question test), and general security habits. Table A2 in the Appendix includes a comprehensive record of features extracted from the survey questions, several of which have been inspired by previous work [12, 24, 23, 25, 5].

Performance (average AUC score across K=10 folds) of the two models is given in Table 5. We find that models using only demographic and socioeconomic features achieve a maximum average AUC of 65%. Slightly higher performance is achieved using only features related to security habits (67% average AUC), suggesting that good security hygiene renders demographics and other socioeconomic differences, largely irrelevant. Previously experiencing an online scam proves to be highly predictive of ransomware infection, and the model including both security habits and past experience with an online scam (average AUC of 75%) achieves performance on par with the full model that includes all features.

Based on the performance of the two machine learning models, we construct a heuristic risk assessment rule based on self-reported security habits and previous experience with online scams. The heuristic outputs a risk score for each user, and is constructed by running the tuned lasso model on the full training set and rounding the coefficients. The heuristic is based on six factors: use of two-factor authentication, data backup habits, encryption of hard drive, frequency of using torrent services, password-protected computer for login, and previous experience with online scams. A comprehensive list of questions used to assess security habits and their corresponding score are included in Table 7. Higher scores correspond to a higher likelihood of infection.

We find that the heuristic performs on par with the more complex models, achieving an average AUC score of 78% across K=10 folds. We calibrate the model as follows: for each respondent, we calculate the risk score using the derived heuristic and predict ransomware status using the output risk score as the sole feature. The rule presented is predictive, in the sense that these factors are correlated with the risk of infection, not causally related. For example, not backing up your data is correlated with infection, although beginning to backup your data will not cause the likelihood of infection to decrease. In Figure 4, we show a calibration plot for this heuristic, which converts raw scores to a probability scale for easier interpretation.

Features	Lasso	GBM
Dem + SES	65	63
Dem, SES, tech, computer	61	65
Habits	66	67
Habits + scam	75	74
All features	76	76

Table 6: *Average AUC across K=10 folds for lasso and gradient boosting tree (GBM) models on survey responses. "Dem" refers to demographics and includes age, gender, "SES" includes all socioeconomic covariates, "Tech" refers to the technology used, "Computer" refers to computer knowledge, "Habits" includes all covariates related to security habits, "Scam" is a binary indicator of whether the individual has previously experienced an online scam (e.g., tech support scam). The optimal parameters are chosen via K=10 fold cross validation. Models based solely on self-reported security habits and previous experience with online scams performed on par with the full models including all covariates.*

6 Passive risk assessment based on browsing behavior

Previous work has shown that certain online activities present a higher risk than others [25, 8, 7], and whether users do so consciously or unconsciously, they must frequently trade off the risk of encountering malware with the benefits from carrying out their tasks (e.g., downloading a file, acquiring free software, streaming a video). Often, this risk is poorly understood and difficult to quantify. In the case of ransomware attacks, the primary attack vectors are thought to be e-mail attacks and web-based attacks, with some reports that web-based attacks are becoming increasingly common [3]. Here, we investigate whether web browsing behavior can be used to identify the vulnerable population and the behaviors that predispose users to such infections.

6.1 Data

We leverage several datasets to build our model. We consider a sample of users that have Symantec's antivirus product installed and are active users of the Safe Web toolbar—a free website reputation service offered by Norton. The Safe

Question	Points
How frequently do you download files from online torrent sites such as the Pirate Bay, ExtraTorrent, or TorrentZ2?	
• I frequently download files from torrent sites.	15
• I occasionally download files from torrent sites.	10
• I rarely download files from torrent sites.	5
• I never download files from torrent sites.	0
Do you backup your personal files to an external hard drive or a cloud-based storage service?	
• I do not have any of my files backed up.	8
• I backup my files once a year.	6
• I backup my files every couple of months.	4
• I backup my files every couple of weeks.	2
• I backup my files every day.	0
Is your hard drive encrypted?	
• Yes, my hard drive is encrypted.	0
• No, my hard drive is not encrypted.	1
Have you ever downloaded—or been asked to download—an application that you suspect was malicious, like fake anti-virus software?	
• Yes, I have.	10
• No, I haven't.	0
Do you use two-step authentication for at least one of your online personal accounts (i.e., not for a work-related account)?	
• Yes, I use two-step authentication.	0
• No, I don't use two-step authentication.	1
Is your computer password-protected for login?	
• Yes, my computer has a password.	0
• No, my computer doesn't have a password.	8

Table 7: *Questions included in the heuristic risk assessment based on self-reported security habits and previous experience with online scams.*

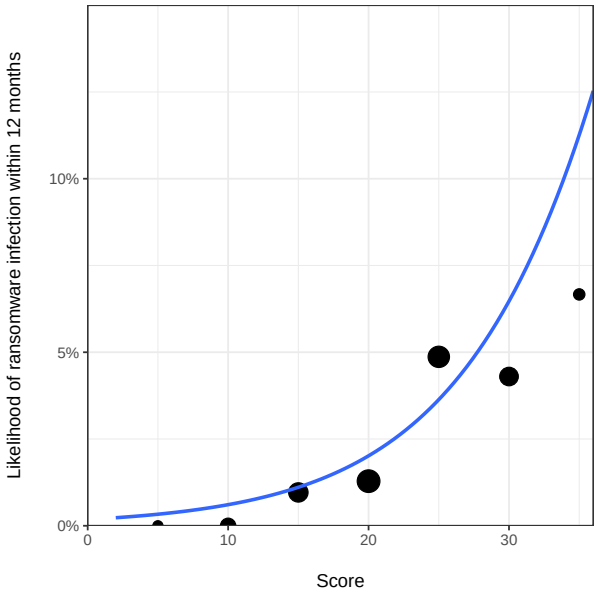

Figure 4: *Calibration plot for the heuristic showing the calculated score versus the empirical proportion of ransomware infections within 12 months, and the fitted logistic regression line. Scores are grouped into buckets of 5. Higher scores correspond to an increased likelihood of ransomware infection, giving us confidence that the heuristic is well-calibrated.*

Web toolbar is a free service that scans websites and displays a color-coded safety rating beside each search result. Green indicates a trusted website, whereas orange and red indicate that the site is suspicious or untrusted. For this set of users, Symantec passively logs a time-stamped history of certain URLs browsed, indexed by an anonymous and unique machine identifier. Other data sources include: system configurations for each machine (i.e., name, version, and build of the operating system, as well as any service packs available), and a listing of 30 threat attributes compiled by Norton for each domain. Threat attributes contain details regarding site certificates, as well as any threats detected on the website, such as Trojans, adware, viruses, etc. We leverage these logs in order to identify behavioral differences between clean machines and those that are likely to get infected by ransomware.

We consider an observation period of 10 months from August 2016 to May 2017. For each month, "eligible" users are those registered in the U.S. that have browsed at least 100 URLs per month in the current and previous month. We hence discard users who are likely to have uninstalled the toolbar shortly after their first use. In total, our dataset includes 4.1 million distinct machines that typically visit a total of over 25 million distinct domains each month.

Figure 6 shows the distribution of ransomware strains in our sample. A total of 366 distinct ransomware strains (including different versions of the same family) were identified. We find that the top three strains – CryptoLocker, Locky, and Cerber, together account for over 76% of all infections and are all primarily spread via spam emails.

CryptoLocker is typically propagated as an attachment to a seemingly innocuous e-mail message, which appears to have been sent by a legitimate company. A ZIP file attached to an email message contains an executable file with the filename and the icon disguised as a PDF file, taking advantage of Windows' default behavior of hiding the extension from file names to disguise the real .exe extension. Locky is typically delivered via an invoice requiring payment with an attached Microsoft Word document that contains malicious macros. When the user opens the document, it displays random characters and uses social engineering techniques to convince the user to activate macros to "correct" the encoding. If the user enables macros, they save and run a binary

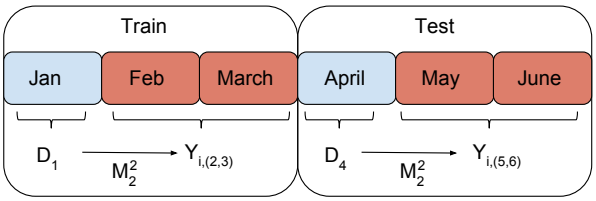

Figure 5: *Example of the prediction problem setup. The training set consists of features derived from data in January, D_{t-1}, and infection status during February and March, $Y_{i,(t,t+1)}$. The test set consists of features derived from data in April, and infection status during May and June.*

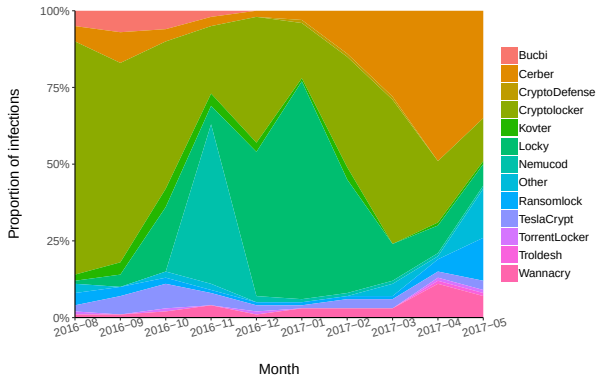

Figure 6: *Distribution of ransomware strains across the ten month observation period. Three families of ransomware predominate (CryptoLocker, Locky, and Cerber), together accounting for over 76% of all infections.*

file that downloads the actual encryption Trojan, which encrypts all files that match particular extensions. Lastly, Cerber is a ransomware Trojan targeting Microsoft Windows that uses a similar technique. A .docx file arrives attached to an email message displaying corrupted encoding and relies on social engineering to convince the user to activate macros.

6.2 Defining the prediction problem

For each month t, we construct features using web browsing history during month $t-1$, and predict $Y_{i,(t,t+j)}$, the infection status for each user i between months t and $t+j$, where $j = \{1,...,5\}$ months. We evaluate the fitted model, M_j^t on an out of sample test set as shown in Figure 5. Th figure shows the model setup for an arbitrary month, t, and infection status labels between months t and $t+1$. We also contrast our results for predicting ransomware infections to predicting any type of malware infection, more generally. To do so, for each month t, we re-define the set of infected users to be those that have at least one file or instance of malicious traffic blocked during between months t and $t+j$. Conversely, we define clean users to be those that have no blocks between months t and $t+j$.

Our ground truth is based on the observation of malicious activity by the end point protection software. To determine infection status, we use telemetry from two types of defense systems: the antivirus system (AV) and intrusion prevention (IPS) system. The AV system is a disk-level defense that blocks malicious executable files already on the hard drive via signatures, machine learning, and heuristic-based methods. In contrast, the IPS system does not detect specific files, but monitors and blocks both incoming and outgoing network traffic that is identified to be malicious. For each machine, we make use of the following fields: time stamp, malware family name, file name (if available), and action taken. "Infected" users are defined to be those that have at least one ransomware block generated by either AV or IPS during the month $t+j$. Ransomware-specific blocks are found by string matching "ransom" on the malware family name. We define a machine as clean if it has no ransomware blocks between months t and $t+j$.

6.3 Feature construction

We construct a number of features that capture the content, volume, and temporal behavior of users over the one-month browsing period. Volume-based features include the number of days active (i.e., browsing at least one URL), the number of distinct domains, and the total number of URLs visited. We also aim to capture temporal differences in browsing behavior, in order to understand whether browsing at different times during the day or on weekends is correlated with facing higher risk of infection. Our hypothesis is that weekend and evening browsing is more likely to be varied, exploratory, and focused on entertainment rather than work-related purposes – factors that could be correlated with riskier activities and increased likelihood of infections. The features we extract are the proportion of URLs and domains visited in each hour and during weekends (the times tamps used correspond to local time, therefore timezone differences do not affect the accuracy of these features).

In order to generate content-related features, we first preprocess the URL to extract the domain, the top level domain, and file extension, if available. As expected, browsing behavior is heavy-tailed, with popular domains accounting for the majority of the volume. We extract the 10,000 most visited domains (i.e., the 10,000 visited by the most users in each training month), and calculate the proportion of URLs each user visited at each domain[5]. We transform this domain feature through one-hot encoding, expanding it to a set of 10,000 binary features. For each user, we construct

[5]Various experiments were tried using the top 25,000 and 100,000 domains, however these were not found to improve the performance of the model. In a separate experiment, 372 Alexa categories and sub categories were mapped to the domains to be used as features instead of the raw domains, however this also failed to improve model performance.

a 1x10,000 frequency vector with each element representing the proportion of URLs visited at the particular domain. We also include several features that capture the tail of the browsing distribution for each user, as well as their browsing consistency and diversity. These include: the proportion of URLs accounted for by the top N visited domains, and the proportion of domains visited less than N times, where N ranged from 1 to 100 (1, 2, 3, 5, 10, 25, 50, 100). Our goal is not to provide a comprehensive statistical account of each and every domain browsed, but to capture a snapshot the varied interests and skill sets of individuals. For example, a user who frequently visits stack overflow and Wikipedia is likely to have a different skill set, set of interests, and risk profile, than a user who frequents popular culture, celebrity websites and encounters many ads.

In a similar vein, we hypothesize that accessing a greater and more diverse set of files and top level domains is correlated with an increase in risk. We follow a similar approach for top level domains and file extensions[6], transforming them through one-hot encoding, where each element represents (1) the proportion of domains visited with that top level domain / file extension, and (2) the proportion of URLs visited with that top level domain / file extension, for each user.

The safety posture of the domains visited is captured by two features, (1) the proportion of URLs with each threat attribute (n=30 features), and (2) the proportion of URLs with each Safe Web safety rating (e.g., recommended, warning, bad, unrated) that is displayed to the user; We also include the overall proportion of referrer URLs from major search engines (e.g. Google, Safe Web, Yahoo, Bing, Ask), the proportion of URLs that: contain ads, are associated with adult websites, are email clients, and are associated with account logins.[7] Lastly, we included several features related to the operating system characteristics: type (e.g., Windows, Mac, Linux, etc.), name (e.g. Home, XP), version, build, and service pack. In total, an average of 136 distinct operating systems configurations were observed during each month (name, version). Browser information was only available for 10% of users and was not used.

6.4 Model

In order to distinguish machines infected with ransomware from clean ones, a supervised learning model was constructed using gradient boosted trees[8] [9]. A separate model was run for each training month t, and tuned via K=5 fold cross validation on the training set[9]. We examine the effect of forecasting individual-level infection status at the end of the following 1-6 months.

A common concern in designing machine learning models for security is related to concept drift [30], the risk that the statistical properties of the behavior being modeled changes over time, rendering the models less and less effective. For example, one might worry that as the threat vectors for ransomware evolve over time (e.g. from email, exploit kits, or malvertising campaigns, to a new attack vector) or target new categories of domains, the effectiveness of the algorithm will decrease. We address this problem by limiting the training period to one month, so that if different behaviors become "risky" (e.g. malvertisement campaigns begin targeting sports or streaming domains), the training data will reflect this updated pattern and – provided the behavior continues for at least another month, maintain their accuracy over time.

6.5 Results

An average of 4,700 (0.3%) of users experienced a ransomware attack each month during our observation period, 8% of which experienced more than one attack. This is approximately in line with the 2%-3% annual estimate from our representative survey. By comparison, an average of 245,000 (13%) users experienced a malware attack each month. In predicting ransomware infection at the end of the following month $t = 1$, we achieve an average of 73% AUC across the ten test months. Although the distribution of ransomware families changes considerably over time, we observe a low standard error (2%) in model performance across the ten months, suggesting that the accuracy of the model is stable over time. We achieve similar performance predicting malware infections at the end of the following month $t = 1$, obtaining an average AUC of 74% (standard error of 4%) across the ten test months.

Next, we consider predictions at longer time frames in the future. Figure 7 shows the AUC for predicting ransomware infections at the end of the next j months, where $j = 1, 2, ...5$ for the first training month in our dataset, August 2016. We observe that performance drops from 74% to 68% AUC as we consider longer prediction windows. A similar trend can be seen for the malware prediction problem.

Two factors are likely limiting the performance of our passive risk-assessment model. First, web browsing behavior likely captures exposure to certain threat vectors better than others. For example, we expect that web-based attacks that

[6] A total of 284 distinct top level domains and 173 distinct file extensions were extracted.

[7] These five features were all generated using string matching on the url for associated terms. Ads were identified via string matching on popular ad terms (e.g. doubleclick, utm_campaign, utm_medium, adclick, adserver, ad_keyword, etc.). Account logins were identified via string matching on common login terms (e.g., mail, inbox, account, sign_in, etc.

[8] We used Python's implementation of xgboost, an interpretable algorithm that has been shown to consistently achieve top performance on categorical data. Two other implementations were tried, but did not improve performance: Vowpal Wabbit, a fast online machine learning library was used, and a neural network model implemented in PyTorch.

[9] Tuning was performed in a two-step process, first tuning parameters that control model complexity, and then adjusting parameters that add randomness to make training robust to noise. Observations weighted inversely proportional to their class frequency.

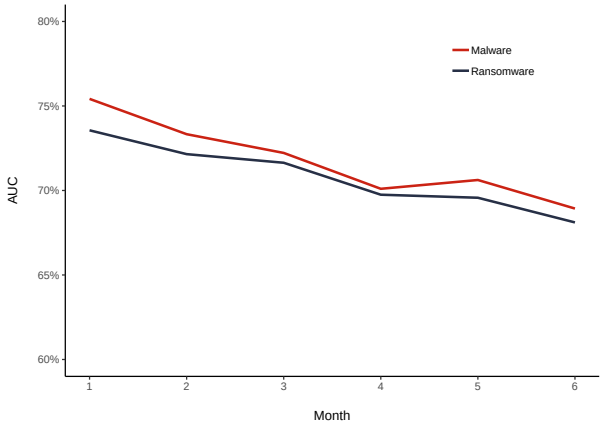

Figure 7: *AUC for predicting ransomware and malware infections at the end of the next i months, where $i = 1, 2, ...6$ for the first training month in our dataset, August 2016. Performance decays gradually for longer-term predictions, losing approximately 6% over the six month period.*

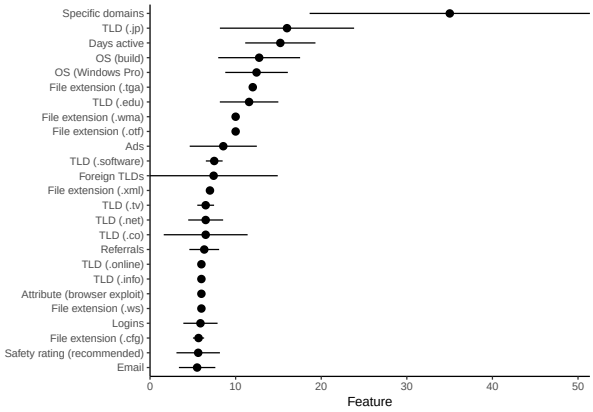

Figure 8: *Average feature importance across the ten months for predicting ransomware. Importance is defined as the number of times a feature is used to split the data across all trees. The "specific domain" feature includes domains with an importance score greater than or equal to 3. Confidence intervals represent two standard errors from the mean.*

are spread via exploit kits and malvertisement campaigns should be easier to predict using web browsing behavior, than attacks spread via spam emails, which rely on social engineering techniques to trick users into downloading the payload from an email. Examining the distribution of ransowmare strains in our sample, we find that the majority of infections (at least 76%) are due to strains that typically spread via spam emails. Unfortunately we have no visibility into user email behaviors. The hope, in this case, is that web browsing behavior is correlated with susceptibility to social engineering attacks and can be used as an informative proxy.

Second, our labels define clean users as those not infected with ransomware, indifferent of whether they are infected with other types of malware. If users who are at risk of other types of malware share similar behaviors to those who are at risk of ransomware, this increases the difficulty in distinguishing between the two groups.

6.6 Feature significance

We report feature importances resulting from each of the models, where importance is defined as the number of times a feature is used to split the data across all trees [9]. The distribution of feature importances for the 10,000 most popular domains exhibits a heavy tail, with approximately 29% of domains having an importance greater than or equal to 1. For ease of interpretation, we group all domains with importance score of at least 3 into one category (n=100), and all others into a second category. Each of the ten separate models produces a slightly different feature importance list. In order to determine which features have the highest impact on distinguishing clean user profiles from risky ones, we calculate the average importance score across the ten months for each feature.

The resulting top 25 features are shown in Figure 8. The list is dominated by: specific domains, level of activity (number of days browsing at least one URL), the operating system build and name (and specifically, Windows Pro), visiting a large and diverse number of top level domains (e.g., .jp, .tv, .software), and accessing a diverse set of files extensions. Upon inspecting the list of most discriminant domains, we find that, in contrast to some previous work (e.g., [7]), there is no one specific domain category that is particularly dominant. Domains range from online shopping sites, a blog hosting platform, gaming, and more.

We contrast these observations with the the top 25 most discriminant features for the more general task of predicting any malware infection (Figure 9). Several differences stand out. First, feature importance is distributed more evenly among the top features, and specific domains have less discriminant power. Search engine referral behavior (i.e., proportion of referrals from Yahoo, Bing, etc.) appears to be much more important than for ransomware attacks, suggesting that default search engine choice may be an informative indicator. Being exposed to ads, the proportion of weekend activity, visiting adult sites (as measured by the proportion of days and URLs browsing adult content), are all more discriminant predictors for malware than for ransomware. There is also some indication that a large volume and diversity of top level domains and file extensions are discriminant, however there does not seem to be as much dependence on these features as there was for ransomware.

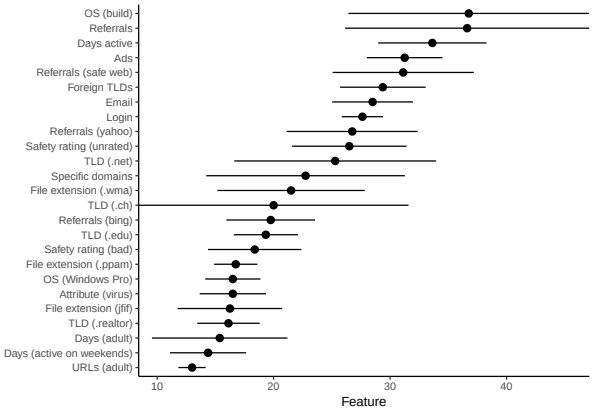

Figure 9: *Average feature importance (the number of times a feature is used to split the data across all trees) across the ten months for predicting any malware infection. The "specific domain" feature includes domains with an importance score greater than or equal to 3.*

7 Conclusions and future work

Our survey results shed new light on the scale of ransomware in the general population. While our estimated victimization rate of 2–3% of the population per year indicates the vast majority of Americans are not directly affected, this still suggests millions of cases per year. Additionally, even with a small payment rate (4%), the large average payment value ($530) suggests that ransomware may be yielding over $100 million in illicit revenue.

An important future research question is whether these figures are growing (and at what rate), which will require longitudinal follow-up studies. Conventional wisdom has held that cryptocurrency will fuel growth in ransomware, but our results suggest most cases in 2016–2017 were not reliant solely on cryptocurrency for payment. Another open question for future research is if payment rates will increase or decrease as more individuals affected have either been previously victimized themselves or have heard more about ransomware from affected friends and family.

Our risk assessment models also suggest that vulnerability can be estimated from self-reported security habits and observed web browsing behavior. Importantly, our model suggests that ransomware vulnerability is distinct from general vulnerability to malware. Our model is fast, frugal in terms of features used, and transparent. While prior research suggests these qualities make a risk-assessment heuristic more acceptable to users [14], future research is required to gauge user reaction to our model. Future work might also leverage privacy-preserving data collection techniques to incorporate potentially sensitive historical web browsing data into a risk assessment model. Finally, an important open question is how effective different interventions might be for limiting ransomware exposure to users identified as being high-risk.

8 Acknowledgments

The authors would like to thank Leyla Bilge, Petros Efstathopoulos, Zhiyuan Jerry Lin, Dan Boneh, Darren Shou, Ansh Shukla for helpful comments and feedback.

References

[1] Lucrative ransomware attacks: Analysis of the cryptowall version 3 threat. Tech. rep., 2015.

[2] 2016 internet crime report. Tech. rep., 2016.

[3] 2016 internet security threat report. Tech. rep., 2016.

[4] BILGE, L., HAN, Y., AND DELL'AMICO, M. Riskteller: Predicting the risk of cyber incidents. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security* (2017), ACM, pp. 1299–1311.

[5] BOSSLER, A. M., AND HOLT, T. J. On-line activities, guardianship, and malware infection: An examination of routine activities theory. *International Journal of Cyber Criminology 3*, 1 (2009).

[6] CABAJ, K., AND MAZURCZYK, W. Using software-defined networking for ransomware mitigation: the case of cryptowall. *IEEE Network 30*, 6 (2016), 14–20.

[7] CANALI, D., BILGE, L., AND BALZAROTTI, D. On the effectiveness of risk prediction based on users browsing behavior. In *Proceedings of the 9th ACM symposium on Information, computer and communications security* (2014), ACM, pp. 171–182.

[8] CARLINET, Y., MÉ, L., DEBAR, H., AND GOURHANT, Y. Analysis of computer infection risk factors based on customer network usage. In *Emerging Security Information, Systems and Technologies, 2008. SECURWARE'08. Second International Conference on* (2008), IEEE, pp. 317–325.

[9] CHEN, T., AND GUESTRIN, C. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (2016), ACM, pp. 785–794.

[10] CHOI, K.-S. Computer crime victimization and integrated theory: An empirical assessment. *International Journal of Cyber Criminology 2*, 1 (2008).

[11] DAWES, R. M. The robust beauty of improper linear models in decision making. *American psychologist 34*, 7 (1979), 571.

[12] EGELMAN, S., AND PEER, E. Scaling the security wall: Developing a security behavior intentions scale (sebis). In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (2015), ACM, pp. 2873–2882.

[13] GAZET, A. Comparative analysis of various ransomware virii. *Journal in computer virology 6*, 1 (2010), 77–90.

[14] GIGERENZER, G., HERTWIG, R., AND PACHUR, T. *Heuristics: The Foundations of Adaptive Behavior*. OUP USA, 2011.

[15] HALAWA, H., BEZNOSOV, K., BOSHMAF, Y., COSKUN, B., RIPEANU, M., AND SANTOS-NETO, E. Harvesting the low-hanging fruits: defending against automated large-scale cyber-intrusions by focusing on the vulnerable population. In *Proceedings of the 2016 New Security Paradigms Workshop* (2016), ACM, pp. 11–22.

[16] HUANG, D. Y., MCCOY, D., ALIAPOULIOS, M. M., LI, V. G., INVERNIZZI, L., BURSZTEIN, E., MCROBERTS, K., LEVIN, J., LEVCHENKO, K., AND SNOEREN, A. C. Tracking ransomware end-to-end. In *Tracking Ransomware End-to-end*, IEEE, p. 0.

[17] JUNG, J., CONCANNON, C., SHROFF, R., GOEL, S., AND GOLDSTEIN, D. G. Simple rules for complex decisions.

[18] KHARRAZ, A., ARSHAD, S., MULLINER, C., ROBERTSON, W. K., AND KIRDA, E. Unveil: A large-scale, automated approach to detecting ransomware. In *USENIX Security Symposium* (2016), pp. 757–772.

[19] KHARRAZ, A., ROBERTSON, W., BALZAROTTI, D., BILGE, L., AND KIRDA, E. Cutting the gordian knot: A look under the hood of ransomware attacks. In *International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment* (2015), Springer, pp. 3–24.

[20] KLEINBERG, J., LAKKARAJU, H., LESKOVEC, J., LUDWIG, J., AND MULLAINATHAN, S. Human decisions and machine predictions. *The quarterly journal of economics 133*, 1 (2017), 237–293.

[21] LALONDE LEVESQUE, F., NSIEMPBA, J., FERNANDEZ, J. M., CHIASSON, S., AND SOMAYAJI, A. A clinical study of risk factors related to malware infections. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security* (2013), ACM, pp. 97–108.

[22] LÉVESQUE, F. L., FERNANDEZ, J. M., AND SOMAYAJI, A. Risk prediction of malware victimization based on user behavior. In *Malicious and Unwanted Software: The Americas (MALWARE), 2014 9th International Conference on* (2014), IEEE, pp. 128–134.

[23] MAIER, G., FELDMANN, A., PAXSON, V., SOMMER, R., AND VALLENTIN, M. An assessment of overt malicious activity manifest in residential networks. In *International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment* (2011), Springer, pp. 144–163.

[24] MILNE, G. R., LABRECQUE, L. I., AND CROMER, C. Toward an understanding of the online consumer's risky behavior and protection practices. *Journal of Consumer Affairs 43*, 3 (2009), 449–473.

[25] NGO, F. T., AND PATERNOSTER, R. Cybercrime victimization: An examination of individual and situational level factors. *International Journal of Cyber Criminology 5*, 1 (2011).

[26] OVELGÖNNE, M., DUMITRAŞ, T., PRAKASH, B. A., SUBRAHMANIAN, V., AND WANG, B. Understanding the relationship between human behavior and susceptibility to cyber attacks: a data-driven approach. *ACM Transactions on Intelligent Systems and Technology (TIST) 8*, 4 (2017), 51.

[27] REDMILES, E. M., KROSS, S., PRADHAN, A., AND MAZUREK, M. L. How well do my results generalize? comparing security and privacy survey results from mturk and web panels to the us. Tech. rep., 2017.

[28] SCAIFE, N., CARTER, H., TRAYNOR, P., AND BUTLER, K. R. Cryptolock (and drop it): stopping ransomware attacks on user data. In *Distributed Computing Systems (ICDCS), 2016 IEEE 36th International Conference on* (2016), IEEE, pp. 303–312.

[29] SHARIF, M., URAKAWA, J., CHRISTIN, N., KUBOTA, A., AND YAMADA, A. Predicting impending exposure to malicious content from user behavior. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security* (2018), ACM, pp. 1487–1501.

[30] TSYMBAL, A. The problem of concept drift: definitions and related work. *Computer Science Department, Trinity College Dublin 106*, 2 (2004).

[31] YEN, T.-F., HEORHIADI, V., OPREA, A., REITER, M. K., AND JUELS, A. An epidemiological study of malware encounters in a large enterprise. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security* (2014), ACM, pp. 1117–1130.

[32] ZAVARSKY, P., LINDSKOG, D., ET AL. Experimental analysis of ransomware on windows and android platforms: Evolution and characterization. *Procedia Computer Science 94* (2016), 465–472.

Category	Raw proportion	Weighted proportion
Female	55%	54%
Male	45%	46%
White	81%	75%
Black or African American	8%	11%
Hispanic or Latino	5%	8%
Asian	2%	2%
Mixed	3%	2%
Native American	1%	1%
Other	1%	1%
Middle Eastern	0%	0%
Age (19 – 30)	11%	16%
Age (30 – 45)	19%	24%
Age (45 – 60)	28%	27%
Age (over 60)	42%	32%
No High school	1%	2%
High school	20%	32%
Some college	22%	22%
2-year grad	15%	13%
4-year grad	24%	20%
Post-grad	17%	12%
Full-time	38%	40 %
Retired	28%	22%
Part-time	11%	10%
Permanently disabled	9%	8%
Student	4%	7%
Unemployed	3%	5%
Homemaker	5%	5%
Temporarily laid off	1%	1%
Other	1%	1%
Married	51%	49%
Never married	26%	31%
Divorced	13%	11%
Widowed	6%	6%
Domestic / civil partnership	3%	2%
Separated	1%	0%
Child under 18 in household - yes	19%	25%
Child under 18 in household - no	81%	75%
Less than $10,000	3%	4%
$10,000 - $19,999	8%	8%
$20,000 - $29,999	10%	10%
$30,000 - $39,999	11%	12%
$40,000 - $49,999	9%	9%
$50,000 - $59,999	10%	11%
$60,000 - $69,999	7%	6%
$70,000 - $79,999	8%	7%
$80,000 - $99,999	9%	8%
$100,000 - $119,999	6%	6%
$120,000 - $149,999	5%	5%
$150,000 - $199,999	4%	4%
$200,000 - $249,999	2%	1%
$250,000 - $349,999	1%	1%
Prefer not to say	8%	9%

Table A1: *Demographics and socioeconomic characteristics of respondents, n=1,180. The raw proportion represents the fraction of respondents out of n=1,180 having a particular characteristic, and the weighted proportion represents the post-stratified proportion.*

Category	Features
Demographics	Gender, race, age
Socioeconomic (SES)	Highest level of education completed, household income, employment status, marital status, field of work or study, child under 18 in household
Computer knowledge	8 question multiple choice test
Security habits	Time spent on the computer each day, number of emails opened per day, frequency of downloading files from online torrent sites, data backup habits (on external hard drive or cloud-based storage device), storage strategy for sensitive information on personal computer (e.g., use of password-protected computer or folder), has encrypted hard drive, credential saving habits in browser, software updating habits (e.g., postpone, install immediately, etc.), own a blog or website, use two-factor authentication (if yes, for which services), password creation habits (e.g., use the same password for all sites), use of computer at work (if yes, for which tasks)
Software used	Operating system (name and version), most commonly used browser (name and version), list of plugins installed

Table A2: *Survey features. Software used were collected passively, and the name of operating system and browser currently used was also asked as a survey question.*