Who is targeted by email-based phishing and malware? Measuring factors that differentiate risk

Camelia Simoiu
csimoiu@stanford.edu
Stanford University

Ali Zand
zand@google.com
Google

Kurt Thomas
kurtthomas@google.com
Google

Elie Bursztein
elieb@google.com
Google

ABSTRACT

As technologies to defend against phishing and malware often impose an additional financial and usability cost on users (such as security keys), a question remains as to who should adopt these heightened protections. We measure over 1.2 billion email-based phishing and malware attacks against Gmail users to understand what factors place a person at heightened risk of attack. We find that attack campaigns are typically short-lived and at first glance indiscriminately target users on a global scale. However, by modeling the distribution of targeted users, we find that a person's demographics, location, email usage patterns, and security posture all significantly influence the likelihood of attack. Our findings represent a first step towards empirically identifying the most at-risk users.

ACM Reference Format:
Camelia Simoiu, Ali Zand, Kurt Thomas, and Elie Bursztein. 2020. Who is targeted by email-based phishing and malware? Measuring factors that differentiate risk. In *ACM Internet Measurement Conference (IMC '20), October 27–29, 2020, Virtual Event, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3419394.3423617

1 INTRODUCTION

Email-based phishing and malware persist as a major security threat, with hundreds of millions of attacks occurring daily [6]. The targets of these attacks are not uniformly distributed. Journalists, politicians, activists, and business owners alike face heightened levels of risk [26, 36, 38]. In response, tailored protections have started to emerge for "at-risk" individuals, including Google's Advanced Protection Program [22] and Microsoft's AccountGuard [40]. A key challenge in this space, however, is automatically identifying who is at-risk, as applying heightened protections to a broad user base incurs both a financial and usability cost [12, 14, 35].

Prior efforts to measure risk have focused largely on identifying which populations are most susceptible to deception. This includes understanding who falls for phishing or malware lures [15, 50] or what visible cues induce a person to click on a dangerous link or attachment [24, 29, 32, 50]. The ultimate goal of these studies has been to develop better education materials or to design better warnings for all users [1, 15] rather than to tailor protections to users based on their personal level of risk. Additionally, most of these studies rely on controlled lab simulations of attacks, as opposed to real-world threat data.

In this work, we present a measurement study of the factors that correlate with a higher likelihood of receiving email-based phishing and malware (i.e., being "targeted"). We derive our findings from 1.2 billion phishing and malware attacks sent to Gmail consumer users over a five month period. We first explore the anatomy of modern e-mail based phishing and malware campaigns in terms of their scale, duration, and reach. We then build on these lessons to infer which user attributes correlate with a higher risk of being targeted. Our model takes into consideration an individual's demographics, security posture, prior risk exposure, and various email engagement metrics.

During our measurement window, we find that attackers targeted, on average, 17.0 million users every week with hundreds of thousands of campaigns that last a median of just one day. These attacks follow a skewed distribution: 10% of phishing campaigns accounted for 76% of phishing attacks, and 10% of malware campaigns accounted for 61% of malicious attachments. Attackers broadly targeted users around the globe as part of their campaigns, with the majority of targets residing in North America and Europe. While 90% of attacks occurred in English, we show evidence that some attackers localize their efforts. For example, 34% of attacks that targeted users from France occurred in French, and 78% of attacks that targeted users from Japan occurred in Japanese.

Based on our model, we find that a gamut of factors correlate with a heightened risk, including where a person lives, their age, and their account hygiene. For example, the odds of receiving phishing or malware are over five times higher for users whose email address and other details were exposed in a third-party data breach, compared to users unaffected by data breaches. In some countries, the odds of attack is over twice that of the United States, such as the Democratic Republic of Congo or Australia. Likewise, the odds of being attacked increases with age, with people over 65 years old facing 1.50 times the risk compared to 18–24 year olds who face the lowest risk. Our measurements act as a first step towards understanding how to evaluate personal security risks. Ultimately, such estimates would enable automatically identifying, recommending, and tailoring protections to those users who need it most.

2 RELATED WORK

Empirical Studies of Attacks. Several researchers have developed frameworks to measure spam [31], phishing [45, 55], and malware distribution campaigns [8, 27]. In contrast to our work, these studies focus on understanding the inner-workings of the

IMC '20, October 27–29, 2020, Virtual Event, USA

ACM ISBN 978-1-4503-8138-3/20/10.
https://doi.org/10.1145/3419394.3423617

ecosystem from an attacker perspective, and do not provide insights about the impact these campaigns have on computer users. Only a few studies have questioned whether users who are targeted with spam have a different risk profile than those who are not. Two small-scale experiments (using under 300 email accounts) showed that spam is not randomly distributed, but targeted towards specific groups. For example, age, income, nationality, and the way the email account was exposed on the web (e.g., linking an account to a social network) were found to influence the amount of spam received [25, 39]. A separate study of email traffic logs received by a large UK ISP showed considerable disparity between the proportions of spam received by addresses with different first characters [11].

Some security entities such as Kaspersky, RSA, and the Anti-Phishing Work Group also routinely publish reports on the proportion of users who experience phishing and email malware attacks [4, 56]. While similar in spirit, such reports are not directly comparable to our study as they are based on the subsection of computer users that elect to purchase security software. These samples may thus be significantly biased towards the geographical areas where such products are marketed and sold, and a set of users which may have different levels of security awareness and risk profiles compared to the general computer user.

Lab-based Susceptibility Studies. More relevant to our work, are studies that aim to identify the vulnerable population and the characteristics that predispose users to various adverse security outcomes [16, 34, 41, 46, 51]. With respect to phishing, most research focuses on understanding the risk factors that predispose users to be *susceptible* to phishing. These studies are typically modestly sized in-lab user studies (ranging from 20 to 10,000 users) on specific populations (e.g., university staff or students) and cover a wide range of tasks and experimental designs. Several studies focus on identifying cues that individuals use to distinguish phishing emails from legitimate ones [13, 15, 16, 47, 58]. Others consider user attributes such as demographics, personality traits, habits, and situational factors as predictors of susceptibility.

Demographics are a frequently-studied dimension which has generally been found to be an important predictor of susceptibility, however there is disagreement regarding the magnitude and directionality of the effects (e.g., which age groups are more at risk) [24, 28, 29, 32, 41, 46, 50]. A wide range of other characteristics have been put forth as increasing susceptibility. Among them, impulsivitiy in making decisions [7, 17, 30], receiving a large volume of emails [57], users' curiosity [5, 43], and risk propensity [43] were all found to increase susceptibility to phishing attacks. Given the small sample sizes, diverse subject pools and experimental designs, it remains difficult to draw general conclusions.

Security, Usability, and Personalization. Prior work has shown that phishing attacks often have a disparate impact on their targets, suggesting a need for personalizing defenses to protect the most at-risk users [9, 46, 51]. Such personalization becomes necessary since requiring additional defenses often means imposing additional usability costs on users [12, 14, 35]. Our study presents a first step towards designing a system that automatically identifies the subset of users requiring such heightened protections.

Attack type	Total emails	Breakdown
malware	679,835,204	56.1%
phishing	531,970,560	43.9%

Table 1: Summary of phishing and malware emails in our study, collected from April 7—August 31, 2020.

3 METHODOLOGY

Our measurement study relies on anonymized data collected from Gmail consumer users betwen April 7—August 31, 2020. In total, our dataset includes approximately 1.2 billion emails flagged as phishing and malware and, on average, 17.0 million weekly targeted users. In what follows, we describe the origin of the phishing and malware labels, the associated user and attack attributes we use to model risk, and the anonymity constraints that shape our methodology.

3.1 Detecting Attack Targets

Our dataset includes a weekly snapshot of users who received at least one email flagged by Gmail's automated classifiers as phishing or malware.

Phishing Detection. We define a phishing email to be one which contains a URL identified as phishing by Google Safe Browsing [21]. As detailed in Table 1, our dataset contains 531,970,560 phishing emails. We caution this is an underestimate of all email-based phishing, due to the possibility of evasion and detection latency [44], as well as non-URL phishing.

Malware Detection. Gmail applies a suite of proprietary anti-virus scanners to all email attachments [52]. As detailed in Table 1, our dataset contains 679,835,204 emails flagged as malicious by these scanners. We note that Gmail blocks the receipt of executable attachments by default [20], limiting attacks mainly to compressed files and documents (e.g., Office documents, PDFs). Attackers leverage the built-in code execution capabilities of these formats, such as VBA for Office documents, or JavaScript for PDFs.

3.2 User Annotations

The crux of our study is to understand what attributes or behaviors place an individual at higher risk of being targeted with email-based phishing or malware. We consider a variety of potential factors ranging from demographics, security experience, and email usage behaviors.

Demographics. We consider two coarse-grained demographic features for each user in our dataset: their age and the country from which a user accesses their account [19, 23]. Due to anonymization, the ages in our dataset are stratified into 10 year segments starting from "18 to 24" and ending with "65 or older". Previous studies have shown that age is a factor in susceptibility [37, 46], which we hypothesize may lead to higher rates of targeting by attackers. Likewise, age and country of access correlate with wealth, which we hypothesize may also be a factor in targeting due to financial incentives for attackers [3, 54].

Security Posture. We annotate each user with whether they have adopted some form of two-factor authentication (e.g., SMS, device

prompts) for their Google account, or established an account recovery mechanism via a secondary email account or phone number. These features allow us to examine whether users who are aware of their elevated risk status adopt critical account hygiene protections that would help protect against phishing attacks and some malware [14].

Prior Risk Exposure. We identify users who have personally suffered a data breach that exposed their email. Specifically, we flag email addresses in our dataset that also appear in a dataset shared from a previous analysis [55], which as of 2020 includes over 4 billion credentials. Understanding a user's prior risk exposure allows us to evaluate whether attackers harvest personal information from data breaches, in turn placing such users at elevated risk of future threats. Due to the lack of viable remediation options, data breaches may expose users to lasting harms beyond password exposure.

Devices & Engagement. Our final class of attributes includes a user's Gmail activity level (High, Medium, Low, or Inactive, based on quantiles of monthly usage statistics), whether a user relies solely on a mobile device, or solely on a computer. We rely on these features to understand whether certain classes of devices place users at higher risk of targeting, as well as to control for the possibility that a higher risk of attack may merely result from higher usage patterns. Finally, device usage may also be an indicator of socioeconomic status (SES), where mobile devices are more accessible than computers among low SES individuals [2].

3.3 Attack Annotations

We also leverage our dataset to understand how attackers tailor and distribute their phishing and malware lures.

Campaign Identifier. We use a cluster identifier produced by Gmail based on a similarity hash of the content [10]. We rely on this clustering to track the size and duration of phishing and malware campaigns across multiple weeks, if at least 50 users were the target of the campaign per week.

Language. For emails containing text, Gmail detects the language of the email. In total, roughly 89% of phishing and malware emails in our dataset have an associated language.

3.4 Data Anonymity Constraints & Ethics

All analysis was executed on Google's infrastructure and each result was aggregated and anonymized. The researchers involved in the study never had access to raw Gmail data. To this end, we could only examine a single week of data at a time to prevent tracking users over the course of our measurement. These constraints mean that a frequently targeted user may appear in multiple weekly samples.

Furthermore, when examining user attributes for modeling risk, we had to ensure that at least 50 users shared the same collection of attributes (e.g., country of access, device type, age) to avoid modeling or de-anonymizing any single user. This corresponds to a k-anonymity constraint of 50 over a Cartesian product of all features [53]. Due to these constraints, we had to limit our analysis to discretized or coarse-grained features in order to not fall below our anonymity threshold.

3.5 Limitations

Apart from the aforementioned limitations, we also note that our measurement period overlaps with the COVID-19 outbreak. This raises questions on the ecological validity of our results compared to prior time periods, as email providers like Gmail have noted a higher prevalence of spam and phishing during the outbreak [33]. We caution against interpreting the results of our analysis outside of this COVID-19 period. Our methodology, however, remains viable during any time period.

Lastly, we make extensive use of Gmail's proprietary detection and labeling systems throughout this study. We caution this may introduce biases due to false positives and negatives in detection, or more subtly, variations in detection accuracy across regions or campaign structure. However, with billions of attacks in our corpus, our analysis can at least shed light on the targeting strategies of known threats today.

4 ATTACK ANALYSIS

We examine how attackers orchestrate phishing and malware campaigns, including the scale, duration, and reach of attacks. We find that a majority of campaigns span less than one day, with attackers predominantly focusing their attacks on North America and Europe. These measurements lay the foundation for our risk modeling.

4.1 Volume of attacks

We provide a weekly breakdown of the phishing and malware attacks covered in our study in Figure 1 and Figure 2 respectively. In both cases, attacks are bursty, with the volume of attacks increasing by 500% at times from week to week. At its peak, we observed 117 million phishing emails targeting 41 million distinct users during the week of May 11, 2020. Across every week, the median targeted user received just one phishing email and 10% of targeted users were attacked at least five times. By comparison, malware campaigns were largely absent prior to July 20, 2020, totaling fewer than 500,000 emails per week. However, following the return of the Emotet botnet [18], we observed a peak of 224 million malware emails targeting 46 million distinct users. Emotet is a malware family that serves as a multi-stage dropper for other malware families. The botnet's orchestrators rely on Office and PDF documents as the initial infection vector, with multiple different variants emerging and disappearing since 2014. Repeated attacks with malware were more frequent, with 25% of targeted users receiving five or more emails with malicious attachments in a given week. Based on our results for both phishing and malware, we opted to model all users who received at least a single attack, rather than a smaller population of higher-risk users (5+ attacks).

4.2 Campaign Size and Duration

Over the course of our measurement period, we observed a total of 406,002 distinct phishing campaigns and 1,724,160 malware campaigns. Both classes of attacks exhibit a highly skewed distribution. The top 10% of phishing campaigns account for 76% of phishing emails, while the top 10% of malware campaigns account for 61% of emails with malicious attachments. As shown in Figure 3, phishing and malware campaigns vary widely in size. We find that 91% of phishing campaigns distribute fewer than 1,000 emails, and 99.9%

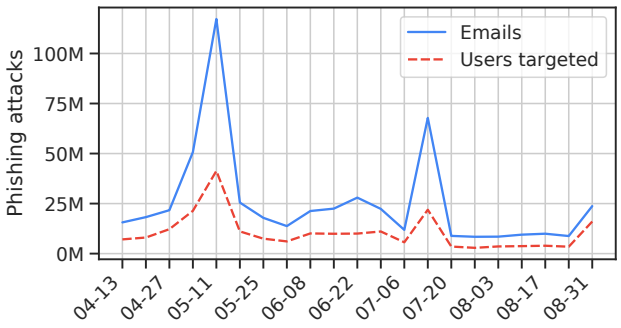

Figure 1: Weekly volume of phishing attacks, in terms of emails sent and number of distinct users targeted.

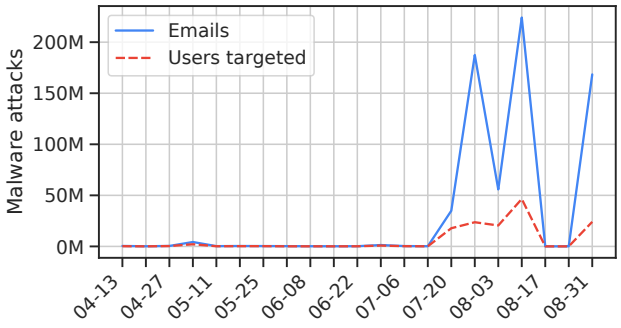

Figure 2: Weekly volume of malware attacks, in terms of emails sent and number of distinct users targeted.

of campaigns fewer than 10,000 emails. Malware campaigns on the other hand appear tactically smaller, with 99% of campaigns generating less than 1,000 emails.

The majority of campaigns are brief, as shown in Figure 4. We observe that 89% of malware campaigns last just one day, whereas the median phishing campaign lasts three days or less; 80% of phishing campaigns last less than one week. This short duration is likely a direct response to attackers attempting to re-configure campaigns to evade detection. Absent cycling to new campaigns, traffic to phishing pages has been found to disappear within a few hours after detection [45].

4.3 Reach and Localization

Attackers largely focus their phishing and malware attacks on North America and Europe, as shown in Table 2 and Table 3 respectively. [1] For both classes of attacks, the United States receives the highest attack email volume. The top targeted countries identified in our five month observation window closely match victims of phishing kits and keyloggers from 2016–2017 [55], as well as the phishing victims identified by a large financial institution from 2019 [45]. This suggests that the dynamics behind targeting have remained stable over time.

[1]Due to our anonymity constraints, we are unable to normalize attacks per capita, as we cannot count unique users across our weekly analysis windows.

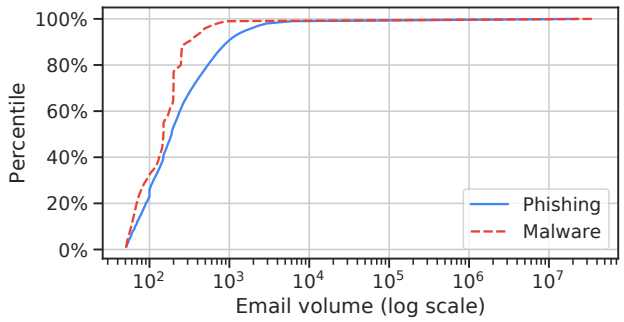

Figure 3: CDF of the volume of emails observed over the course of campaigns.

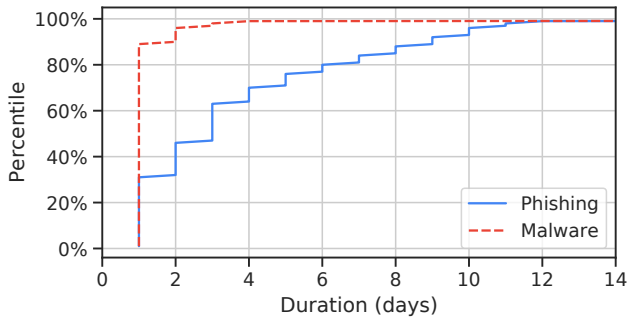

Figure 4: CDF of campaign durations, measured in terms of the number of days each campaign was active (x-axis truncated from maximum of 146 days).

When examining the localization of attacks, we find that attackers compose 83% of phishing emails and 97% of malware emails in English. However, we do observe some attempts at localization as shown in Table 4. In particular, 34% of attacks targeting users in France involve emails composed in French. Similarly, 78% of attacks targeting users in Japan involve emails composed in Japanese. These results are primarily driven by phishing attacks, in which targeting appears to be more prominent than in malware attacks.[2] This suggests a bifurcation of targeting strategies: broad-based campaigns where attackers target users globally, as well as localized campaigns where attackers focus on specific countries.

5 EXPLORING INFLUENTIAL RISK FACTORS

Despite the dynamic nature and reach of phishing and malware campaigns, we nevertheless are able to identify several stable attributes that correlate with a higher personal risk of attack. We discuss each risk factor in detail and explore potential explanations.

5.1 Model Generation

We model the risk that a user is targeted by either phishing or malware on a given week as a binomial distribution $Y_i \sim bin(n_i, \pi_i)$

[2]See Appendix, Table 6 and Table 7 for localization data separated by phishing and malware, where localization is largely absent from malware.

Country	Phishing emails	Breakdown
United States	127,852,455	28.4%
Japan	49,315,177	10.9%
India	30,454,888	6.8%
United Kingdom	20,555,605	4.6%
Brazil	16,481,347	3.7%
Spain	16,355,335	3.6%
France	12,310,909	2.7%
Canada	9,168,380	2.0%
Australia	8,833,911	2.0%
Indonesia	8,431,852	1.9%
Other	150,879,314	33.5%

Table 2: Top 10 countries by volume of phishing.

Country	Malware emails	Breakdown
United States	320,628,618	51.2%
United Kingdom	82,779,864	13.2%
Australia	40,707,850	6.5%
Netherlands	39,566,141	6.3%
France	28,976,576	4.6%
Spain	25,888,990	4.1%
Belgium	11,756,854	1.9%
India	4,747,387	0.8%
Sweden	4,251,462	0.7%
Finland	4,240,147	0.7%
Other	62,132,420	9.9%

Table 3: Top 10 countries by volume of malware.

Country	English	Dutch	French	Japanese	Portuguese	Spanish	Other
Australia	91.8%	0.0%	6.6%	0.3%	0.1%	0.2%	0.9%
Canada	94.0%	0.0%	3.1%	0.7%	0.3%	0.5%	1.5%
India	97.6%	0.1%	0.2%	0.4%	0.1%	0.1%	1.5%
United Kingdom	96.9%	0.1%	1.0%	0.3%	0.1%	0.1%	1.4%
United States	95.1%	0.2%	1.0%	0.5%	0.1%	0.2%	2.9%
Netherlands	83.1%	6.7%	8.6%	0.5%	0.1%	0.1%	0.9%
France	65.1%	0.0%	33.8%	0.3%	0.1%	0.2%	0.6%
Japan	20.5%	0.0%	0.1%	78.2%	0.0%	0.0%	1.2%
Brazil	32.2%	0.0%	0.2%	0.2%	66.8%	0.3%	0.3%
Spain	73.5%	0.1%	17.7%	0.2%	0.1%	7.4%	0.9%
Other	75.9%	0.2%	3.8%	1.1%	0.4%	5.2%	13.3%

Table 4: Localization of all phishing and malware attacks combined in terms of the proportion of emails sent to each country.

using a logarithmic link function using R's GLM library [49]. We consider each week to be an independent sample—as most attack campaigns last only a short period (as we show in Section 4) and due to our anonymity constraints—and run a separate logistic regression for each week. We define a *targeted* user to be one who receives at least one malware or phishing email in a given week, and treat all users who did not receive any phishing or malicious emails as baseline, low-risk users. We note that false negatives will result in some targeted users being included in the baseline group. We caution that this will cause our model to potentially underestimate effect sizes for attributes that place users at higher risk of attack. Due to processing constraints, we rely on a 10% random sample of baseline users. We note that a logistic regression is robust to this sampling without any need to re-weight [48]. The model's covariates consist of the user attributes detailed in Section 3.2. Our model allows us to assess the influence of a user attribute while controlling for other potential explanatory factors.

Given multiple weeks and variation between weeks, we average the results across our entire analysis period and report the weekly average (μ) and standard deviation (σ) of the odds ratio for each covariate in Table 5.[3] We note that all reported values are significant with $p < 0.0001$. Our model includes over 200 countries, however we report coefficients for only the top 10 countries with the highest odds that met our significance threshold for at least 14 weeks of our study.

[3]Figure 5 in the Appendix shows the complete distribution of odds ratios resulting from each of the weekly models summarized in Figure 5.

5.2 Exploring Risk Correlations

Country of Access. We find that the country where a user accesses Gmail represents a considerable risk factor. The highest risk countries are concentrated in Europe and Africa, with average weekly odds ranging from 1.14 to 2.64, however the specific countries and exact odds vary substantially week by week, as detailed by the high standard deviation in relative odds. Overall, 16 countries exhibited a higher risk on average than the United States, even though the United States is the largest target by volume of emails.

Age. We find that the odds of being targeted increase slightly with each subsequent age group ($\mu = [1.29, 1.64]$). For example, the odds of someone 55-64 experiencing an attack is, on average, 1.64 times that of an 18–24 year old. One possible explanation is that attackers specifically target older users, potentially due to their reported higher susceptibility to deception and coercion [37]. Alternatively, these older users may have larger online footprints, thus making discovery of their accounts easier.

Security Posture. We find only a nominal difference ($\mu = 1.34$) between the odds that someone with two-factor authentication enabled will experience an attack versus password-only authentication. This suggests that many users who are at risk of attack have yet to enable additional protections. At the same time, we find that users who have proactively established a recovery mechanism face a higher odds of attack ($\mu = 2.34$). These users would likely be better protected by strict two-factor authentication.

Prior Risk Exposure. Users with personal data exposed by third-party breaches face far higher average odds of attack ($\mu = 5.20$). This suggests that attackers actively harvest data breach information, both for enumerating email addresses, but also potentially for demographic information in order to identify a user's age or country of access. As such, our results suggest that data breaches expose users to lasting harms due to the lack of viable remediation options.

Type of Device. Compared to users owning multiple types of devices, we find that users who own only a personal computer face slightly lower odds of targeting (0.90) and mobile-only users face even lower risks of attack (0.80). This may be due to the socio-economic (SES) factors affecting device ownership (i.e., lower SES groups more likely to own only mobile or only desktop devices) and attackers targeting wealthier groups. Device ownership may also be correlated with technical savviness and online footprint; users that only sign in from one type of device may sign up for less online services and accounts, further reducing their likelihood of being targeted.

Email activity. The odds of being targeted increase with the level of engagement with Gmail. Active users face higher likelihoods of being targeted, with those most frequently interacting with Gmail being, on average, 5.18 times more likely to be targeted than an inactive user. We speculate that this could be due to inactive accounts being set up with limited scope and/or having smaller online footprints.

User attribute	Treatment	Ref	Weekly Odds μ	Weekly Odds σ
Country of access	DR Congo	USA	2.64	3.87
	Australia	USA	2.20	3.12
	Netherlands	USA	1.97	2.95
	United Kingdom	USA	1.65	1.65
	Belgium	USA	1.58	2.86
	Finland	USA	1.57	2.93
	Japan	USA	1.32	1.19
	Lesotho	USA	1.22	1.50
	Spain	USA	1.15	1.12
	Denmark	USA	1.14	0.77
Age	25-34	18–24	1.29	0.15
	35-44	18–24	1.50	0.29
	45-54	18–24	1.63	0.38
	55-64	18–24	1.64	0.43
	65+	18–24	1.50	0.44
Two factor auth?	Yes	No	1.34	0.12
Recovery setup?	Yes	No	2.34	0.32
In data breach?	Yes	No	5.20	4.43
Desktop only?	Yes	No	0.95	0.16
Mobile only?	Yes	No	0.80	0.12
Email activity	Low	Inactive	1.54	0.23
	Medium	Inactive	2.84	0.49
	High	Inactive	5.18	1.68

Table 5: Odds of being targeted by phishing or malware according to a logistic regression model. We report the average (μ) and standard deviation (σ) of the odds ratios, aggregated over each week of our analysis.

6 DISCUSSION AND FUTURE WORK

Our modelling has revealed several factors that correlate with the likelihood of being targeted with email-based phishing and malware attacks for Gmail users. The list identified is by no means exhaustive. An important area for future work will be to use individual-level data in order to identify more precise risk factors, develop predictive models to identify targeted users, and evaluate their accuracy.

Secondly, given the shift in the type of attack over time, with phishing attacks predominating from April 13 to July 20, and malware attacks predominating from July 20 to August 31, there is a possibility that Emotet's re-gained prominence could represent a change in attack strategy, and hence, a change in the risk profile of targeted users. Inspecting individual weekly odds ratios for each covariate, one notable observation is that for users whose credentials were compromised, the weekly odds ratios of being targeted pre 07/20—when phishing attacks predominate—are higher than post 07/20, when malware attacks predominate (see Appendix, Figure 5). This suggests that attackers may make use of email accounts and/or other identifying information from data breaches in phishing attacks more so than in malware attacks. No such separation is observed among any of the other covariates. Although we cannot directly compare effect sizes between weekly models [42], an interesting line of inquiry for future work would be to investigate whether the effect of our covariates differs by attack type, or more subtly, per campaign.

7 CONCLUSION

We presented a global measurement study of Gmail consumer users who are presently most at-risk of email-based phishing and malware attacks. We shed new light on the scale and reach of attack campaigns, and identify several factors that correlate with a higher risk of being targeted by attackers. We identified several stable factors that have a bearing on an invididual's risk level, including age, locality, device classes, and even prior security incidents. Our results represent a first step towards empirically identifying at-risk user populations and the promise of tailoring protections to those users that need it most. We hope that future work will build on these insights to add a richer understanding of which factors influence risk, as well as to establish a minimum threshold for who needs high-friction protections.

REFERENCES

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

[2] Monica Anderson and Madhumitha Kumar. Digital divide persists even as lower-income americans make gains in tech adoption. https://www.pewresearch.org/fact-tank/2019/05/07/digital-divide-persists-even-as-lower-income-americans-make-gains-in-tech-adoption/, 2019.

[3] Ross Anderson, Chris Barton, Rainer Boehme, Richard Clayton, Michel J.G. van Eeten, Michael Levi, Tyler Moore, and Stefan Savage. Measuring the cost of cybercrime. In *Proceedings of the Workshop on Economics of Information Security*, 2012.

[4] Anti-Phishing Working Group. APWG Trends Report Q1 2019. https://docs.apwg.org/reports/apwg_trends_report_q1_2019.pdf.

[5] Zinaida Benenson, Freya Gassmann, and Robert Landwirth. Unpacking spear phishing susceptibility. In *International Conference on Financial Cryptography and Data Security*, pages 610–627. Springer, 2017.

[6] Elie Bursztein and Daniela Oliveira. Understanding why phishing attacks are so effective and how to mitigate them. https://security.googleblog.com/2019/08/understanding-why-phishing-attacks-are.html, 2019.

[7] Marcus Butavicius, Kathryn Parsons, Malcolm Pattinson, and Agata McCormac. Breaching the human firewall: Social engineering in phishing and spear-phishing emails. *arXiv preprint arXiv:1606.00887*, 2016.

[8] Juan Caballero, Chris Grier, Christian Kreibich, and Vern Paxson. Measuring pay-per-install: The commoditization of malware distribution. In *Proceedings of the USENIX Security Symposium*, 2011.

[9] Davide Canali, Leyla Bilge, and Davide Balzarotti. On the effectiveness of risk prediction based on users browsing behavior. In *Proceedings of the 9th ACM symposium on Information, computer and communications security*, pages 171–182, 2014.

[10] Moses S Charikar. Similarity estimation techniques from rounding algorithms. In *Proceedings of the ACM Symposium on Theory of Computing*, 2002.

[11] Richard Clayton. Do zebras get more spam than aardvarks? *ratio*, 20:40, 2008.

[12] Sanchari Das, Andrew Dingman, and L Jean Camp. Why johnny doesn't use two factor a two-phase usability study of the fido u2f security key. In *Proceedings of the International Conference on Financial Cryptography and Data Security*, 2018.

[13] Rachna Dhamija, J. D. Tygar, and Marti Hearst. Why phishing works. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pages 581–590, New York, NY, USA, 2006. ACM.

[14] Periwinkle Doerfler, Maija Marincenko, Juri Ranieri, Angelika Moscicki Yu Jiang, Damon McCoy, and Kurt Thomas. Evaluating login challenges as a defense against account takeover. In *Proceedings of the Web Conference*, 2019.

[15] Julie S Downs, Mandy Holbrook, and Lorrie Faith Cranor. Behavioral response to phishing risk. In *Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit*, pages 37–44. ACM, 2007.

[16] Julie S. Downs, Mandy B. Holbrook, and Lorrie Faith Cranor. Decision strategies and susceptibility to phishing. In *Proceedings of the Second Symposium on Usable Privacy and Security*, SOUPS '06, pages 79–90, New York, NY, USA, 2006. ACM.

[17] Waldo Rocha Flores and Mathias Ekstedt. Shaping intention to resist social engineering through transformational leadership, information security culture and awareness. *computers & security*, 59:26–44, 2016.

[18] Dan Goodin. There's a reason your inbox has more malicious spam–emotet is back. https://arstechnica.com/information-technology/2020/07/destructive-emotet-botnet-returns-with-250k-strong-blast-of-toxic-email/, 2020.

[19] Google. About targeting geographic locations. https://support.google.com/google-ads/answer/2453995?visit_id=637363906136362321-1839693281&rd=1, 2020.

[20] Google. File types blocked in Gmail. https://support.google.com/mail/answer/6590?hl=en, 2020.

[21] Google. Google Safe Browsing. https://https://safebrowsing.google.com/, 2020.

[22] Google. Google's strongest security for those who need it most. https://landing.google.com/advancedprotection/, 2020.

[23] Google. How Google infers interest and demographic categories. https://support.google.com/google-ads/answer/2580383?hl=en, 2020.

[24] Tzipora Halevi, Nasir Memon, and Oded Nov. Spear-phishing in the wild: A real-world study of personality, phishing self-efficacy and vulnerability to spear-phishing attacks. *Phishing Self-Efficacy and Vulnerability to Spear-Phishing Attacks (January 2, 2015)*, 2015.

[25] Il-Horn Hann, Kai-Lung Hui, Yee-Lin Lai, Sang-Yong Tom Lee, and Ivan PL Png. Who gets spammed? *Communications of the ACM*, 49(10):83–87, 2006.

[26] Seth Hardy, Masashi Crete-Nishihata, Katharine Kleemola, Adam Senft, Byron Sonne, Greg Wiseman, Phillipa Gill, and Ronald J Deibert. Targeted Threat Index: Characterizing and Quantifying Politically-Motivated Targeted Malware. In *Proceedings of the USENIX Security Symposium*, 2014.

[27] Luca Invernizzi, Stanislav Miskovic, Ruben Torres, Christopher Kruegel, Sabyasachi Saha, Giovanni Vigna, Sung-Ju Lee, and Marco Mellia. Nazca: Detecting malware distribution in large-scale networks. In *NDSS*, volume 14, pages 23–26. Citeseer, 2014.

[28] Cristian Iuga, Jason RC Nurse, and Arnau Erola. Baiting the hook: factors impacting susceptibility to phishing attacks. *Human-centric Computing and Information Sciences*, 6(1):8, 2016.

[29] Tom N Jagatic, Nathaniel A Johnson, Markus Jakobsson, and Filippo Menczer. Social phishing. *Communications of the ACM*, 2007.

[30] Helen S Jones, John N Towse, Nicholas Race, and Timothy Harrison. Email fraud: The search for psychological predictors of susceptibility. *PloS one*, 14(1):e0209684, 2019.

[31] Christian Kreibich, Chris Kanich, Kirill Levchenko, Brandon Enright, Geoffrey M. Voelker, Vern Paxson, and Stefan Savage. On the spam campaign trail. In *Proceedings of the USENIX Workshop on Large-Scale Exploits and Emergent Threats*, 2008.

[32] Ponnurangam Kumaraguru, Steve Sheng, Alessandro Acquisti, Lorrie Faith Cranor, and Jason Hong. Teaching johnny not to fall for phish. *ACM Transactions on Internet Technology (TOIT)*, 10(2):7, 2010.

[33] Neil Kumaran and Sam Lugani. Protecting businesses against cyber threats during COVID-19 and beyond. https://cloud.google.com/blog/products/identity-security/protecting-against-cyber-threats-during-covid-19-and-beyond, 2020.

[34] Fanny Lalonde Levesque, Jude Nsiempba, José M Fernandez, Sonia Chiasson, and Anil Somayaji. A clinical study of risk factors related to malware infections. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*, pages 97–108, 2013.

[35] Juan Lang, Alexei Czeskis, Dirk Balfanz, Marius Schilder, and Sampath Srinivas. Security keys: Practical cryptographic second factors for the modern web. In *Proceedings of the International Conference on Financial Cryptography and Data Security*, 2016.

[36] Stevens Le Blond, Adina Uritesc, Cédric Gilbert, Zheng Leong Chua, Prateek Saxena, and Engin Kirda. A Look at Targeted Attacks Through the Lense of an NGO. In *Proceedings of the USENIX Security Symposium*, 2014.

[37] Tian Lin, Daniel E Capecci, Donovan M Ellis, Harold A Rocha, Sandeep Dommaraju, Daniela S Oliveira, and Natalie C Ebner. Susceptibility to spear-phishing emails: Effects of internet user demographics and email content. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 2019.

[38] William R Marczak, John Scott-Railton, Morgan Marquis-Boire, and Vern Paxson. When governments hack opponents: a look at actors and technology. In *Proceedings of the USENIX Security Symposium*, 2014.

[39] Rodrigo Sanches Miani, Danielle Oliveira, Kil Jin Brandini Park, and Bruno Bogaz Zarpelao. An empirical study of factors affecting the rate of spam. In *Anais Principais do XXXVI Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*. SBC, 2018.

[40] Microsft. Microsoft AccountGuard. https://www.microsoftaccountguard.com/en-us/, 2020.

[41] Jamshaid G Mohebzada, Ahmed El Zarka, Arsalan H BHojani, and Ali Darwish. Phishing in a university community: Two large scale phishing experiments. In *2012 International Conference on Innovations in Information Technology (IIT)*, pages 249–254. IEEE, 2012.

[42] Carina Mood. Logistic regression: Why we cannot do what we think we can do, and what we can do about it. *European sociological review*, 26(1):67–82, 2010.

[43] Gregory D Moody, Dennis F Galletta, and Brian Kimball Dunn. Which phish get caught? an exploratory study of individuals' susceptibility to phishing. *European Journal of Information Systems*, 26(6):564–584, 2017.

[44] Adam Oest, Yeganeh Safaei, Adam Doupé, Gail-Joon Ahn, Brad Wardman, and Kevin Tyers. Phishfarm: A scalable framework for measuring the effectiveness of evasion techniques against browser phishing blacklists. In *Proceedings of the IEEE Symposium on Security and Privacy*, 2019.

[45] Adam Oest, Penghui Zhang, Brad Wardman, Eric Nunes, Jakub Burgis, Ali Zand, Kurt Thomas, Adam Doupé, and Gail-Joon Ahn. Sunrise to sunset: Analyzing the end-to-end life cycle and effectiveness of phishing attacks at scale. In *Proceedings of the USENIX Security Symposium*, 2020.

[46] Daniela Oliveira, Harold Rocha, Huizi Yang, Donovan Ellis, Sandeep Dommaraju, Melis Muradoglu, Devon Weir, Adam Soliman, Tian Lin, and Natalie Ebner. Dissecting spear phishing emails for older vs young adults: On the interplay of weapons of influence and life domains in predicting susceptibility to phishing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 6412–6424. ACM, 2017.

[47] Kathryn Parsons, Marcus Butavicius, Malcolm Pattinson, Dragana Calic, Agata Mccormac, and Cate Jerram. Do users focus on the correct cues to differentiate between phishing and genuine emails? In *Australasian Conference on Information Systems*, 2016.

[48] Ross L Prentice and Ronald Pyke. Logistic disease incidence models and case-control studies. *Biometrika*, 66(3):403–411, 1979.

[49] rdocumentation. Fitting generalized linear models. https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/glm, 2020.

[50] Steve Sheng, Mandy Holbrook, Ponnurangam Kumaraguru, Lorrie Faith Cranor, and Julie Downs. Who falls for phish?: A demographic analysis of phishing susceptibility and effectiveness of interventions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 373–382, New York, NY, USA, 2010. ACM.

[51] 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)*, 2019.

[52] Sri Somanchi. New built-in gmail protections to combat malware in attachments. https://gsuiteupdates.googleblog.com/2017/05/new-built-in-gmail-protections-to.html, 2017.

[53] Latanya Sweeney. Achieving k-anonymity privacy protection using generalization and suppression. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05):571–588, 2002.

[54] Kurt Thomas, Danny Yuxing Huang, David Wang, Elie Bursztein, Chris Grier, Tom Holt, Christopher Kruegel, Damon McCoy, Stefan Savage, and Giovanni Vigna. Framing Dependencies Introduced by Underground Commoditization. In *Proceedings of the Workshop on the Economics of Information Security*, 2015.

[55] Kurt Thomas, Frank Li, Ali Zand, Jacob Barrett, Juri Ranieri, Luca Invernizzi, Yarik Markov, Oxana Comanescu, Vijay Eranti, Angelika Moscicki, et al. Data breaches, phishing, or malware?: Understanding the risks of stolen credentials. In *Proceedings of the ACM Conference on Computer and Communications Security*, 2017.

[56] Maria Vergelis, Tatyana Shcherbakova, and Tatyana Sidorina. Spam and phishing in Q2 2019. https://securelist.com/spam-and-phishing-in-q2-2019/92379/, 2019.

[57] Arun Vishwanath, Tejaswini Herath, Rui Chen, Jingguo Wang, and H Raghav Rao. Why do people get phished? testing individual differences in phishing vulnerability within an integrated, information processing model. *Decision Support Systems*, 51(3):576–586, 2011.

[58] Jingguo Wang, Tejaswini Herath, Rui Chen, Arun Vishwanath, and H Raghav Rao. Research article phishing susceptibility: An investigation into the processing of a targeted spear phishing email. *IEEE transactions on professional communication*, 55(4):345–362, 2012.

APPENDIX

Localization, Phishing vs. Malware

We present the localization of phishing and malware attacks in Table 6 and Table 7 respectively, limited to the top 10 regions targeted by each class of attack. While we find evidence of localization for phishing, such as in France, Japan, and Brazil, malware appears more indiscriminate in their targeting.

Country	English	Dutch	French	Japanese	Portuguese	Spanish	Other
Australia	98.0%	0.0%	0.3%	0.4%	0.1%	0.3%	0.8%
Canada	94.0%	0.0%	3.1%	0.8%	0.3%	0.5%	1.3%
India	97.6%	0.0%	0.1%	0.5%	0.1%	0.1%	1.6%
United Kingdom	97.4%	0.0%	0.4%	0.3%	0.1%	0.2%	1.4%
United States	97.4%	0.1%	0.3%	0.5%	0.1%	0.3%	1.3%
Netherlands	89.6%	8.6%	0.4%	0.3%	0.1%	0.2%	0.8%
France	36.0%	0.1%	62.2%	0.4%	0.1%	0.3%	0.9%
Japan	19.5%	0.0%	0.0%	78.9%	0.0%	0.0%	1.5%
Brazil	29.1%	0.0%	0.1%	0.2%	70.1%	0.2%	0.3%
Spain	80.2%	0.2%	1.4%	0.2%	0.3%	15.9%	1.7%
Other	73.5%	0.2%	2.3%	1.2%	0.5%	6.4%	15.8%

Table 6: Localization of phishing attacks against the top 10 countries (ranked by phishing email volume, in terms of the proportion of emails sent to each country).

Country	English	Dutch	French	Italian	Spanish	Swedish	Other
Australia	90.0%	0.0%	8.9%	0.1%	0.0%	0.0%	1.0%
India	97.9%	0.2%	1.1%	0.0%	0.1%	0.0%	0.6%
United Kingdom	96.8%	0.1%	1.3%	0.1%	0.0%	0.1%	1.6%
United States	92.4%	0.3%	1.8%	0.4%	0.1%	0.2%	4.8%
Belgium	97.1%	0.1%	2.0%	0.0%	0.1%	0.0%	0.7%
Netherlands	80.8%	4.2%	13.2%	0.1%	0.0%	0.0%	1.6%
France	98.9%	0.0%	0.5%	0.0%	0.0%	0.0%	0.5%
Italy	91.9%	0.0%	0.8%	0.7%	0.2%	0.0%	6.3%
Spain	70.6%	0.0%	27.3%	0.0%	1.0%	0.0%	1.1%
Sweden	67.3%	0.0%	21.1%	0.0%	0.1%	9.5%	1.9%
Other	86.6%	0.2%	7.4%	0.2%	1.0%	0.2%	4.4%

Table 7: Localization of malware attacks against the top 10 countries (ranked by malware email volume, in terms of the proportion of emails sent to each country).

Weekly Model Odds

We present the distribution of odds ratios in Figure 5, where each week represents a single point. We distinguish two periods: "Predominantly phishing" covering 04/07/2020 - 07/20/2020 shown in blue, and "Predominantly malware" covering 07/27/2020 - 08/31/2020 shown in red (dates are inclusive). The plot is annotated with significance ($p < 0.01$) and boxplots to emphasize the distribution. For readability, we separate the y-axis ranges; note the different scaling factors.

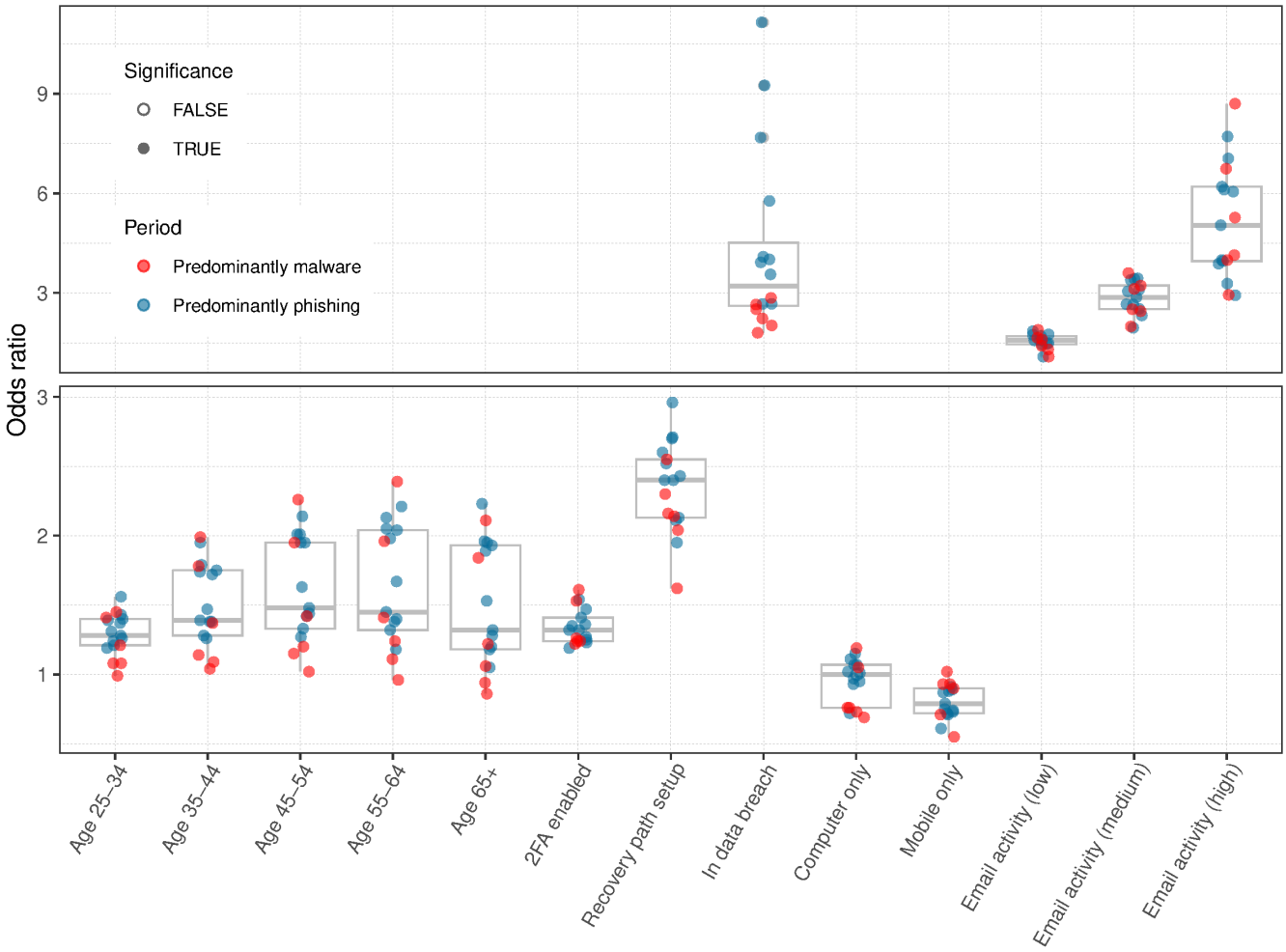

Figure 5: Odds ratio for weekly logistic regression models, with box plots denoting the median and first and third quantiles.