Bad Domains: Exposure to Malicious Content Online*

Lucas Shen† Gaurav Sood‡

Feb 2025

Abstract

Concerns about the digital divide in the US have increasingly given way to fears about a new divide in online safety. By combining passively observed domain-level browsing data of a representative sample of over a thousand Americans with data on malicious domains, we assess if women, minorities, less educated, and older people are more exposed to malicious content than their respective counterparts. We start by looking at the aggregate. 51% of the respondents visited at least one malicious domain during the month-long observation period. However, the visits to malicious websites were highly skewed. The median user visited one malicious site, while the 95th percentile visited eight. Moving to questions about the digital divide, we find that men, African Americans, and individuals with lower levels of education are more exposed to malicious content. Exposure also varies by age, with those under 25 being the most exposed and those aged 35–49 being the least. This digital divide in exposure is driven by differences in internet usage, as all demographic differences at the median disappear once we account for the individual's degree of online presence.

Keywords: Digital divide, Cybersecurity, Malicious websites, YouGov, VirusTotal

*The replication materials are posted on http://github.com/themains/bad_domains.

†Institute for Human Development and Potential, Agency for Science, Technology, and Research, lucas@lucasshen.com.

‡Independent, gsood07@gmail.com.

1 Introduction

Concerns about the digital divide in the US have given way to fears of a new digital divide in online safety. In 2023, the Federal Bureau of Investigation's Internet Crime Complaint Center (IC3) received nearly 900,000 complaints with associated losses of over $12.5 billion. Worse, despite the increased use of automated detection tools (Aldwairi and Alsalman, 2012; Tanaka, Akiyama and Goto, 2017; Peng et al., 2019; Zhu et al., 2020; Baki and Verma, 2023; Choo et al., 2023), cybersecurity threats and associated losses have grown rapidly. The losses in 2023 were 22% higher than in 2022 (Federal Bureau of Investigation, 2023). When we consider that fewer than 15% of cybercrimes are reported (Federal Bureau of Investigation, 2016), the picture looks yet more concerning.

The risk, however, is not spread uniformly. People who are less digitally literate are liable to be more at risk. Part of the reason the less literate are more vulnerable is because they are targeted more aggressively. For instance, older people (who studies suggest are less digitally literate) are targets of more attacks (Federal Bureau of Investigation, 2023). Correspondingly, some research finds that older people are much more likely to be victims of ransomware (Simoiu et al., 2019; Whitty, 2019; Simoiu et al., 2020; Federal Bureau of Investigation, 2023). By the same token, some studies find that men are better at detecting phishing emails (Baki and Verma, 2023). And the implication is that they will be less exposed than women. But the overall risk of harm also depends on the extent to which you are online. For instance, someone who visits 100 'good' websites but has a false positive rate—how many 'good' websites are instead 'bad'—of 10 is less exposed than someone who visits 1000 but with a false positive rate of 5. Corresponding, some studies find that the kinds of people who are likelier to be online—those below 40 and the more educated—are more vulnerable (Hadlington and Chivers, 2018; Weems et al., 2018; Whitty, 2019; Diaz, Sherman and Joshi, 2020; Sood and Cor, 2019).

The two sources of risk—sophistication and online presence—have conflicting predictions about how exposed the traditionally disadvantaged groups are. Given that digital literacy is expected to follow the contours of social disadvantage, the first theory predicts that women, older people, racial minorities, and the less educated are more at risk. The predictions for the second theory are more equivocal and depend on the balance between sophistication and online presence. Combining passively observed browsing data from a representative sample of over a thousand Americans with data on malicious domains, our study sheds light on this question.

In using real-world browsing data, this study provides more valid estimates of the real-world quantities we care about. Much of the understanding of who is susceptible is based on self-reported surveys (Whitty, 2019; Hadlington and Chivers, 2018; Simoiu et al., 2019) or experiments where participants know they are being observed (Weems et al., 2018; Diaz, Sherman and Joshi, 2020) or actual crime reports (Federal Bureau of Investigation, 2016, 2023). All of these methods have serious shortcomings. Self-reported surveys only capture what people are willing to report, which is capped by what people are aware of. Self-reports also have noise stemming from failures in memory and satisficing. Experiments, where people are aware they are being watched, run the danger of artificially low estimates as people are liable to be more mindful of their activities. Switching surveys with actively reported crimes to understand the issue provides a skewed picture as well, as much of the crime goes unreported (Federal Bureau of Investigation, 2016, 2023). Our study circumvents these issues by combining passively observed browsing data with data on malicious websites. Passive tracking also enables us to objectively quantify individuals' level of online presence, which increases the likelihood of exposure (Simoiu et al., 2020; Whitty, 2019) but often goes unobserved in studies of susceptibility (Hadlington and Chivers, 2018; Weems et al., 2018; Sood and Cor, 2019; Baki and Verma, 2023; Federal Bureau of Investigation, 2016, 2023; Simoiu et al., 2019).

2 Data

2.1 Sample

We use data from YouGov to measure exposure to malicious content (Sood, 2022; Sood and Shen, 2024). YouGov maintains a large panel that it recruits using various methods. YouGov incentivizes panelists to respond to surveys using points that can be redeemed for various things. YouGov uses matched sampling to construct the survey sample. It draws a random sample from a large synthetic representative sampling frame, finds respondents matching the sampled individuals from its panel, and invites them to take the survey. Non-responders are substituted with similar others. For data on how well YouGov is able to approximate a random sample, see Rivers and Bailey (2009). More pertinently, our sample is broadly representative of the US population. Appendix SI 1 shows the comparison between our sample and the Current Population Survey (CPS) (Flood et al., 2024) on key demographic variables. Gender distributions are nearly identical, with less than a percentage point difference. Distributions of racial groups also correspond closely, with 63.5% Whites in the YouGov versus 67.3% in CPS, and Hispanics, African Americans, and "Other" differing by a few percentage points. The distribution of education in the sample closely corresponds with the population distribution, with differences of no more than two percentage points. We once again see minor differences in age, with the average age in the YouGov sample of 48.6 years vs. 49.8 in the CPS. The one major exception to these salutary patterns is geography. Geographically, YouGov underrepresents people in the West (20.2% vs. 27.4%) and overrepresents those in the South (42.1% vs. 37.1%).

For our broadly representative sample, we have de-identified web browsing data tracked via passive metering software, RealityMine, installed voluntarily on respondent computers. The software captures online visits independent of the browser type or browser-specific privacy settings.

In all, we have data on 1,200 respondents for June 2022. Of the 1,200 respondents, 66 did not have any browsing data. This may be because they have found a way to circumvent passive monitoring or were not online. We limit our analysis to 1,134 respondents who visited the Internet at least once over the month-long observation period. In all, we have 6.3 million visits to nearly 64,000 domains. For each visit, we have the domain name and category, the local time, and how long the person stayed on the domain.

2.2 Measuring Malicious Content

We measure exposure to malicious content by looking at engagement with websites flagged by major online services as hosting malicious content. On the assumption that what matters most is the total vectors of exposure, we opt for the number of websites with malicious content visited by a respondent as the primary measure of exposure to malicious content. We test the robustness of the patterns by also looking at the number of visits and total time spent. As we show in the Appendix (see SI 3), the major patterns that we highlight are largely similar, whatever measure we use.

We use VirusTotal, a Google subsidiary, to measure the presence of malicious content on a domain (Sood, 2023). VirusTotal is the largest online anti-malware scanning service. Security researchers widely use it for labeling malware (Aldwairi and Alsalman, 2012; Peng et al., 2019; Zhu et al., 2020). We feed the $\sim 64,000$ unique domains to VirusTotal and retrieve their classifications. A malicious domain is a site that carries exploits or other malicious artifacts. Each domain gets scanned by multiple security vendors (e.g., Forcepoint ThreatSeeker, Bitdefender).

4,185 domains (6.5% of the total unique domains in our data) are flagged as malicious by at least one security vendor. Most malicious websites are flagged by only a single vendor, with only 27% receiving malicious flags from more than one vendor. To use a measure with greater precision, our main results classify malicious sites as those with at least two vendors agreeing that the site is malicious (Zhu et al., 2020). This yields 1,128 malicious sites (1.8% of all observed domains).

Table 1. Top domain categories of malicious websites across security vendors

	Forcepoint		alphaMountain		Sophos		Bitdefender		YouGov	
	Category	%	Category	%	Category	%	Category	%	Category	%
1	IT	20	Phishing	20	IT	17	Parked	25	Business	23
2	Search Engines & Portals	8	Malicious	17	Phishing & Fraud	16	Misc	20	Parked	10
3	Sex	6	Suspicious	10	Spyware & Malware	10	Business	10	Business, IT	8
4	Business & Economy	4	IT	2	Content Delivery	9	Porn	8	Adult	8
5	Hacking	3	Malicious, Phishing	2	Search Engines	8	Computers & Software	5	Entertainment	7
6	Malicious Web Sites	3	Unrated	2	General Business	6	Games	4	Business, Education	4
7	Suspicious Content	3	Search Engines/Portals	2	Sexually Explicit	5	Blogs	4	IT	4
8	Financial Data & Services	3	Entertainment	2	Video Hosting	4	Entertainment	3	Entertainment, Illegal Content	3
9	Web Infrastructure	3	Malicious, Parked Site	1	Parked Domains	4	Financial	2	IT, Media Sharing	2
10	Games	3	Malicious, Search Engines/Portals	1	Entertainment	4	Videos	2	Education	2
11	Compromised Websites	3	IT, Suspicious	1	Personal Network Storage	2	Hosting	2	Business, Economy & Finance	2
12	Shopping	3	Search Engines/Portals, Suspicious	1	Games	2	Filesharing	2	Dating & Personals	1
13	Entertainment	2	Content Servers, IoT, Suspicious	1	Spam URLs	1	Onlineshop	1	IT, Proxy & Filter Avoidance	1
14	Adult Content	2	Business/Economy, Suspicious	1	News	1	Education	1	Business, Shopping	1
15	Phishing & Other Frauds	1	Pornography	1	Dynamic DNS & ISP Sites	1	News	1	Adult, Entertainment	1

Table reports the top 15 domain categories of malicious sites (n = 1,128) from four security vendors (Forcepoint ThreatSeeker, alphaMountain, Sophos, and Bitdefender) and YouGov. Each column lists the categories and their corresponding percentage of malicious websites identified by the vendor. The percentage columns indicate the proportion of the 1,128 malicious websites classified into each category by the respective security vendor.

Table 1 summarizes the top 15 most common domain categories of malicious websites as identified by four security vendors. alphaMountain and Sophos have explicit categories for "Phishing", "phishing and fraud", and "spyware and malware" appearing as their top categories. The "information technology" category also appears frequently. Other categories commonly tied to malicious sites that are worth noting include: adult content (e.g., "sex", "sexually explicit", "porn"), "hacking", and "parked."[1]

3 Exposure to Malicious Content

Over the month-long observation period, 51% of the sample visited at least one malicious website. Moving to the number of malicious websites visited, the mean is 2 ($\hat{\sigma} = 5$) (Table 2). The mean, however, is a poor summary of the skewed data. The median user visited one

[1]The rationale behind "parked" is as follows: dormant sites can be revived as malware download sites (Tanaka, Akiyama and Goto, 2017). And some vendors, such as Bitdefender, flag not just active risks but also potential risks.

Table 2. Exposure to malicious and suspicious websites

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
				Percentiles								
	Mean	SD	Min	p5	p10	p25	p50	p75	p90	p95	p99	Max
# unique malicious sites	2	5	0	0	0	0	1	2	5	8	16	80
# visits to malicious sites	21	137	0	0	0	0	2	9	33	73	265	4,006
# minutes spent on malicious sites	14	115	0	0	0	0	0	2	12	37	206	2,879
# unique suspicious sites	3	4	0	0	0	0	2	4	7	10	18	58

Note: The table reports four measures of exposure to malicious content: (i) the number of unique malicious websites visited (the *primary measure*), (ii) the total number of visits to malicious websites visited, (iii) the total minutes of dwelling time on malicious websites, and (iv) the number of unique suspicious websites defined as those with 'suspicious' flag(s) over the month-long passive observation period for the $n = 1,134$ individuals.

unique website with malicious content during the month, the 75th percentile visited 2, the 95th percentile 8, the 99th percentile 16, and the maximum is 18.

30% of the visits to malicious sites lasted one second or less, and 54% lasted 5 seconds or less, compared to 18% and 43.5% for non-malicious visits (Appendix SI 2.1). This suggests some level of sophistication in recognizing a malicious website once on it. However, there is little correlation between time spent per visit and the number of vendors that flag a site as malicious (Appendix SI 2.2). More alarmingly, respondents visit the same malicious site repeatedly. 97% of the people who visited a malicious site visited it more than once (Appendix SI 2.3).

Moving to the total number of visits and the total time spent visiting all malicious sites, we see a large skew on both (Figure SI 4.1). On average, respondents visited malicious sites 21 times ($\hat{\sigma} = 137$), but the median user visited only twice. The 75th percentile is 9, the 95th percentile is 73, and the 99th percentile is 265 (more than 8 times per day, Table 2). Similarly, while the average time spent on malicious sites was 14 minutes ($\hat{\sigma} = 115$), the median was 0; the 99th percentile is 5.5 times the 95th percentile (37 minutes).

In addition to looking at engagement with websites flagged as malicious, we also examined engagement with suspicious websites. A small subset of 1,390 websites (2.2%)

are flagged as suspicious–a lower threat level than malicious. Suspicious sites are defined as those with at least one suspicious flag but no malicious flags. Users visit, on average, three ($\hat{\sigma} = 4$) different suspicious websites; the median is two, and the 75th percentile is four.

4 Exposure to Malicious Content by Sociodemographic Variables

4.1 Exposure to Malicious Content by Gender

The average number of malicious sites visited by women is 1.6 ($\hat{\sigma} = 4.3$) vs. 2.3 ($\hat{\sigma} = 5.0$) for men (Panel A, Table 3). Using robust statistics, we once again find that men visit more malicious sites than women. While the median number of malicious sites women visit is 0, the corresponding number for men is 1. The 95th percentile is 6.3 for women and 10 for men. We see a similar pattern for time spent on malicious sites (Table SI 3.1).

4.2 Exposure to Malicious Content Online by Race

African Americans, on average, visit more malicious sites ($\hat{mu} = 3.2, \hat{\sigma} = 8.8$) than other racial groups (a maximum mean of 1.9) (Panel B, Table 3). The median is 0 for Whites, Hispanics, and Asians, and 1 for African American and Others. At the 75th percentile, African Americans visit three different malicious sites compared to 2 for other races. The time spent on malicious websites exhibits similar differences. The median time spent on malicious sites is 0 minutes, but at the 75th percentile, African-Americans and Others spend more than 5 minutes, while it is 2 minutes or less for all other races Table SI 3.1). Overall, African Americans are the most exposed, while Asians are the least.

Table 3. Exposure to the number of unique malicious websites, by demographics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
					Percentiles							
	Count	Mean	SD	Min	p5	p10	p25	p50	p75	p90	p95	Max
Panel A. Gender												
Female	595 (52.5%)	1.6	4.3	0	0.0	0.0	0.0	0.0	2.0	4.0	6.3	67
Male	539 (47.5%)	2.3	5.0	0	0.0	0.0	0.0	1.0	3.0	6.2	10.0	80
Panel B. Race												
White	720 (63.5%)	1.8	3.8	0	0.0	0.0	0.0	0.0	2.0	5.0	7.0	67
Hispanic	168 (14.8%)	1.9	3.6	0	0.0	0.0	0.0	0.0	2.0	5.3	10.3	23
African American	144 (12.7%)	3.2	8.8	0	0.0	0.0	0.0	1.0	3.0	7.7	12.8	80
Other	56 (4.9%)	1.5	2.8	0	0.0	0.0	0.0	1.0	2.0	4.0	5.0	16
Asian	46 (4.1%)	1.5	2.6	0	0.0	0.0	0.0	0.0	2.0	4.5	6.8	12
Panel C. Education level												
HS or Below	411 (36.2%)	2.3	5.5	0	0.0	0.0	0.0	1.0	2.0	5.0	9.0	67
Some college	326 (28.7%)	2.1	5.4	0	0.0	0.0	0.0	0.0	2.0	5.0	9.8	80
College	255 (22.5%)	1.5	2.7	0	0.0	0.0	0.0	1.0	2.0	4.0	6.0	18
Postgrad	142 (12.5%)	1.4	2.5	0	0.0	0.0	0.0	0.0	1.0	4.0	7.0	12
Panel D. Age group												
< 25	93 (8.2%)	2.9	8.7	0	0.0	0.0	0.0	1.0	3.0	7.0	10.8	80
25–34	200 (17.6%)	1.9	3.6	0	0.0	0.0	0.0	0.5	2.0	5.0	9.0	23
35–49	285 (25.1%)	1.4	2.5	0	0.0	0.0	0.0	0.0	2.0	4.0	7.0	15
50–64	288 (25.4%)	2.4	6.2	0	0.0	0.0	0.0	1.0	2.0	6.0	10.0	67
> 65	268 (23.6%)	1.7	2.9	0	0.0	0.0	0.0	1.0	2.0	5.0	7.0	24

4.3 Exposure to Malicious Content by Education

The most discernible difference is between those with a postgraduate degree or more and others. Those with postgraduate degrees visited, on average, 1.4 different malicious sites ($\hat{\sigma} = 2.5$) compared to 1.5–2.3 for other education levels. The median number of malicious websites visited by postgraduate degree holders is zero compared to one for people with a college degree or high school diploma or below (Panel C, Table 3). At the 75th percentile, those with a postgraduate degree visit one unique malicious site compared to two for everyone else.

We see a similar pattern for time spent on malicious websites, with the 75 percentile for those with postgraduate education being 1 minute and 2–3 minutes for people with less education than that (Table SI 3.1). However, we note that those with "Some college" may be current college students, so a potential confound is age. We examine such potential confounds in Section 4.5.

4.4 Exposure to Malicious Content by Age

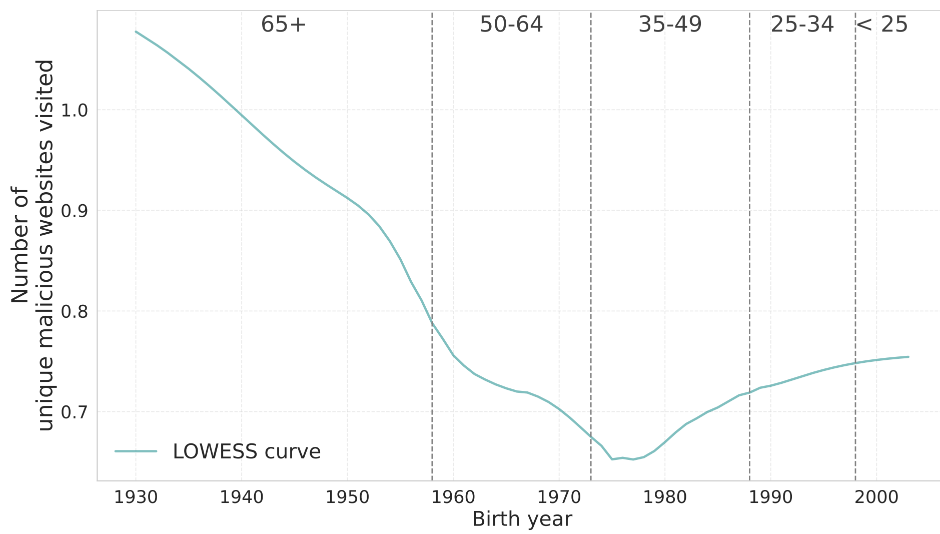


Figure 1. Relationship between birth year and the number of unique malicious websites visited using a LOWESS curve.

Panel D of Table 3 suggests that the number of malicious websites visited varies by age. Younger people are more exposed. For the under 25, the 75th and 90th percentiles are three and seven, respectively, higher than other age groups (25–34, 35–49, 50–65, > 65). The middle age group, those between 35–49, have the lowest exposure, with 75th and 90th percentiles at 2 and 7 visits.

To avoid artifacts from binning age groups, Figure 1 tracks the number of malicious websites visited by birth year (Baki and Verma, 2023). Earlier birth cohorts, particularly those that grew up before the early Internet or digital boom years, show peaks in exposure (Simoiu et al., 2019; Federal Bureau of Investigation, 2023). The steady decrease in exposure and birth year plateaus for those born around when the Internet and digital technology became more mainstream and rose again for the younger cohort who grew up with those technologies. Overall, the demographic most exposed to malicious sites are the very old and the very young, but not dramatically so (note the y-axis scale). A winsorized version of Figure 1 to reduce the influence of outliers yields similar conclusions.

In Appendix SI 3, we show that the broad patterns hold when we look at total duration instead of unique visits.

4.5 Interpreting Group Differences

The differences in exposure between groups are confounded by correlated demographic factors and the extent to which people are online. To better disentangle these confounds, we regress the number of unique malicious websites visited on group indicators and online presence (the total number of websites visited). We estimate quantile regression models for the median to account for the skewed nature of exposure (as seen in Table 3).

Table SI 3.4 reports the estimated differences in medians across demographics. These estimates mostly confirm the differences found in Table 3. As reported earlier, women, on average, visit fewer malicious sites than men. Relative to White Americans, African Americans are more exposed. Other racial groups have no detectable differences. As suggested in Table 3, those with postgraduate degrees are less exposed than those with high school diplomas or less. Relative to those aged 18–24, the 35–49 group is less exposed, although not statistically significant.

Adjusting for online presence (the even-numbered columns) eliminates all demographic differences at the median. This indicates that most observed demographic disparities in exposure are attributable to differences in overall browsing activity rather than differences in demographic characteristics alone. Adjusting for all demographic baselines at once (Columns 9–10) does not substantially change the estimates of group differences.

As anticipated, the number of total websites visited as a measure of online presence is a strong and consistent predictor of exposure–those who browse more websites encounter

Table 4. The number of unique malicious websites visited by demographic characteristics (median regression)

	Dependent variable is Number of unique malicious website visited									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Woman	-1.000^{a}	-0.036							-0.500	-0.078
	(0.292)	(0.102)							(0.379)	(0.078)
Race: African American			1.000^{b}	0.384					0.500	0.332
			(0.507)	(0.240)					(0.373)	(0.212)
Race: Asian			-0.000	-0.016					-0.000	-0.063
			(0.695)	(0.105)					(0.257)	(0.148)
Race: Hispanic			-0.000	0.007					-0.000	-0.003
			(0.693)	(0.061)					(0.216)	(0.068)
Race: Other			1.000	0.017					0.500	0.030
			(0.642)	(0.182)					(0.378)	(0.185)
Educ: Some college					-1.000	-0.021			-0.500	-0.106
					(0.510)	(0.229)			(0.347)	(0.150)
Educ: College					0.000	-0.076			-0.000	-0.111
					(0.471)	(0.226)			(0.323)	(0.148)
Educ: Postgraduate					-1.000^{a}	-0.072			-0.500	-0.119
					(0.318)	(0.232)			(0.357)	(0.156)
Age: 25–34							0.000	-0.400	-0.000	-0.268
							(0.641)	(0.343)	(0.382)	(0.265)
Age: 35–49							-1.000	-0.418	-0.000	-0.298
							(0.607)	(0.325)	(0.371)	(0.251)
Age: 50–64							0.000	-0.418	0.000	-0.276
							(0.605)	(0.325)	(0.388)	(0.256)
Age: 65+							0.000	-0.426	0.000	-0.324
							(0.492)	(0.325)	(0.413)	(0.261)
Total visits (scaled)		13.526^{a}		12.875^{a}		13.674^{a}		13.388^{a}		13.325^{a}
		(1.718)		(1.670)		(1.775)		(1.680)		(1.655)
Total visits^2 (scaled)		-11.413^{b}		-10.299^{b}		-11.603^{b}		-11.141^{b}		-11.077^{b}
		(4.733)		(4.493)		(4.679)		(4.917)		(4.723)
Constant	1.000^{a}	-0.002	0.000	-0.023^{a}	1.000^{a}	-0.001	1.000^{b}	0.398	1.000^{b}	0.377
	(0.055)	(0.102)	(0.499)	(0.008)	(0.133)	(0.230)	(0.428)	(0.325)	(0.419)	(0.301)
Observations	1,134	1,134	1,134	1,134	1,134	1,134	1,134	1,134	1,134	1,134

Note: All models are quantile regression models for medians ($q = .5$). The dependent variable is the total number of unique malicious websites visited over the month-long period. Even-numbered columns adjust for the total number of website visits (*Total visits*). Total visits are scaled to 0–1 so that all variables are between 0–1. The baseline categories are man for gender, White for race/ethnicity, high school or below for education, and 18–24 for age. Bootstrapped standard errors ($n = 1000$) are reported in parentheses. Significance levels: ^{c} 0.1 ^{b} 0.05 ^{a} 0.01.

more malicious sites. Its squared term is negative, indicating diminishing marginal effects as the total number of visits increases.[2]

[2]We also examine differences in means exposure between groups using OLS models Table SI 3.4. Unlike the median estimates, adjusting for online presence does not attenuate demographic differences at the means. Except for age, adjusting for online presence (the even-numbered columns) does not substantially affect the estimates, suggesting that the

5 Discussion

Combining passively observed browsing data of a large representative random sample of adult Americans with data from widely used online services that analyze URLs for malicious content, we find that most people are exposed to malicious content, though only a few people are exposed from multiple sources.

Somewhat alarmingly, 97% of the visitors to malicious sites end up on the same malicious sites again. Malicious visits tend to occur during private hours, indicating a shift in online use, reduced supervision, and greater privacy.

Exposure to malicious content is highly skewed. The median individual visits one unique malicious site. The 95th percentile visits 8.

The exposure also varies considerably by sociodemographics but not in ways that always align with classic digital divides. For instance, even though men are better at detecting phishing (and hence plausibly more digitally literate) Baki and Verma (2023), they are more exposed than women. Adjusting for online presence and other sociodemographic factors doesn't alter the result.

Racial differences are also notable. African Americans are more exposed than other racial groups. African Americans, on average, visit 3.2 malicious sites, at least 1.3 more sites, on average, than other racial groups. Moving to education, people with high school diplomas or less education are the most exposed, visiting, on average, 2.3 different malicious websites, while postgraduates are the least exposed, with a mean of 1.4. Our findings align with Hadlington and Chivers (2018), who find that students and the unemployed, who are

mean differences are not founded in the extent to which different groups spend time online. The quadratic pattern observed with total visits remains consistent with that in Table 4, indicating a concave relationship between total visits and exposure. The low R^2 values (< 0.12) indicate that the variation in exposure is not well explained by the basic demographics.

typically less educated, are more vulnerable. However, our finding on education contradicts studies finding that the more educated are more vulnerable (Sood and Cor, 2019), possibly because of a belief in vulnerability (Weems et al., 2018; Whitty, 2019; Diaz, Sherman and Joshi, 2020). It may also be that disaggregating education gives a clearer picture. As Diaz, Sherman and Joshi (2020) find, STEM students are less susceptible, likely because of their greater technical literacy.

Lastly, we find a nuanced U-shaped relationship for age. The youngest and oldest are the most exposed, with the latter more so than the former (Baki and Verma, 2023; Federal Bureau of Investigation, 2023). This finding is consistent with studies observing that the older demographic is more aggressively targeted and vulnerable, perhaps because of lower digital literacy and higher financial resources (Simoiu et al., 2019; Whitty, 2019; Simoiu et al., 2020; Federal Bureau of Investigation, 2023). The youngest demographics' high exposure is also consistent with their more frequent online presence and impulsive behavior (Hadlington and Chivers, 2018).

5.1 Limitations

Our study has two main limitations. The first is that even though the browsing data is passively collected, it isn't collected without the respondent's knowledge (even though YouGov clarifies that the data is de-identified and the measurement is unobtrusive). If respondents change their online behavior because they know that their data is being collected, they may modify their behavior or figure out ways to evade detection, which may bias our results. In fact, we think it is likely that people would be less likely to search for risky content, like pornographic content, which is associated with a greater chance of carrying malicious content. If that is so, our estimates are a lower bound of the exposure to malicious content. If this bias varies by the attributes we split on, our estimates of differences across groups will also be biased.

The second concern with our measurement is that we code content at a domain level. This runs the risk of incurring some ecological fallacy, where the classification of an entire domain may not reflect the risks of its subdomains. For example, certain domains, like a file-sharing platform, may generally be innocuous, but certain shared files or user-uploaded content may contain malware, phishing links, or other harmful content. The associated concern is that we only have measures for potential exposure but not actual exposure.

6 Conclusion

Our study leverages unique data to shed light on an important concern. Over half the participants are exposed to malicious content during the observation period. The exposure, however, varies dramatically across people, with very little of its variation explained by sociodemographic variables.

References

Aldwairi, Monther and Rami Alsalman. 2012. "Malurls: A lightweight malicious website classification based on url features." *Journal of Emerging Technologies in Web Intelligence* 4(2):128–133. DOI: 10.4304/JETWI.4.2.128-133.

Baki, Shahryar and Rakesh M. Verma. 2023. "Sixteen Years of Phishing User Studies: What Have We Learned?" *IEEE Transactions on Dependable and Secure Computing* 20(02):1200–1212. DOI: 10.1109/TDSC.2022.3151103.

Choo, Euijin, Mohamed Nabeel, Doowon Kim, Ravindu De Silva, Ting Yu and Issa Khalil. 2023. "A Large Scale Study and Classification of VirusTotal Reports on Phishing and Malware URLs." *Proc. ACM Meas. Anal. Comput. Syst.* 7(3). DOI: 10.1145/3626790.

Diaz, Alejandra, Alan T. Sherman and Anupam Joshi. 2020. "Phishing in an academic community: A study of user susceptibility and behavior." *Cryptologia* 44(1):53–67. DOI: 10.1080/01611194.2019.1623343.

Federal Bureau of Investigation. 2016. "Internet Crime Report.". Available at https://www.ic3.gov/AnnualReport/Reports/2016_IC3Report.pdf.

Federal Bureau of Investigation. 2023. "Internet Crime Report.". Available at https://www.ic3.gov/AnnualReport/Reports/2023_IC3Report.pdf.

Flood, Sarah, Miriam King, Renae Rodgers, Steven Ruggles, J. Robert Warren, Daniel Backman, Annie Chen, Grace Cooper, Stephanie Richards, Megan Schouweiler and Michael Westberry. 2024. "IPUMS CPS: Version 12.0 [dataset].". DOI: https://doi.org/10.18128/D030.V12.0.

Hadlington, Lee and Sally Chivers. 2018. "Segmentation Analysis of Susceptibility to Cybercrime: Exploring Individual Differences in Information Security Awareness and Personality Factors." *Policing: A Journal of Policy and Practice* 14(2):479–492. DOI: 10.1093/police/pay027.

Peng, Peng, Limin Yang, Linhai Song and Gang Wang. 2019. Opening the Blackbox of VirusTotal: Analyzing Online Phishing Scan Engines. In *Proceedings of the Internet Measurement Conference*. IMC '19 New York, NY, USA: Association for Computing Machinery p. 478–485. DOI: 10.1145/3355369.3355585.

Rivers, Douglas and Delia Bailey. 2009. Inference from matched samples in the 2008 US national elections. In *Proceedings of the Joint Statistical Meetings*. pp. 627–639. www.asasrms.org/Proceedings/y2009/Files/303309.pdf.

Simoiu, Camelia, Ali Zand, Kurt Thomas and Elie Bursztein. 2020. Who is targeted by email-based phishing and malware? Measuring factors that differentiate risk. In *Proceedings of the ACM Internet Measurement Conference*. IMC '20 New York, NY, USA: Association for Computing Machinery p. 567–576. DOI: 10.1145/3419394.3423617.

Simoiu, Camelia, Christopher Gates, Joseph Bonneau and Sharad Goel. 2019. "I was told to buy a software or lose my computer. i ignored it": a study of ransomware. In *Proceedings of the Fifteenth USENIX Conference on Usable Privacy and Security*. SOUPS'19 USA: USENIX Association p. 155–174. DOI: 10.5555/3361476.3361487.

Sood, Gaurav. 2022. "YouGov Pulse Data for 1200 people for June 2022.". DOI: 10.7910/DVN/VIV4TS.

Sood, Gaurav. 2023. “Virustotal Data (2023 July) for 64k Domains in YG File.”. DOI: 10.7910/DVN/GMNP04.

Sood, Gaurav and Ken Cor. 2019. Pwned: The Risk of Exposure From Data Breaches. In *Proceedings of the 10th ACM Conference on Web Science*. WebSci ’19 New York, NY, USA: Association for Computing Machinery p. 289–292. DOI: 10.1145/3292522.3326046. **URL:** *https://doi.org/10.1145/3292522.3326046*

Sood, Gaurav and Lucas Shen. 2024. “Holier Than Thou? No Large Partisan Gaps in the Consumption of Pornography Online.” *Journal of Quantitative Description: Digital Media* 4. DOI: 10.51685/jqd.2024.011.

Tanaka, Yasuyuki, Mitsuaki Akiyama and Atsuhiro Goto. 2017. “Analysis of malware download sites by focusing on time series variation of malware.” *Journal of Computational Science* 22:301–313. DOI: 10.1016/j.jocs.2017.05.027.

Weems, Carl F., Irfan Ahmed, Golden G. Richard, III, Justin D. Russell and Erin L. Neill. 2018. “Susceptibility and resilience to cyber threat: Findings from a scenario decision program to measure secure and insecure computing behavior.” *PLOS ONE* 13(12):1–18. DOI: 10.1371/journal.pone.0207408.

Whitty, Monica T. 2019. “Predicting susceptibility to cyber-fraud victimhood.” *Journal of Financial Crime* 26(1):277–292. DOI: 10.1108/JFC-10-2017-0095.

Zhu, Shuofei, Jianjun Shi, Limin Yang, Boqin Qin, Ziyi Zhang, Linhai Song and Gang Wang. 2020. Measuring and modeling the label dynamics of online anti-malware engines. In *Proceedings of the 29th USENIX Conference on Security Symposium*. SEC’20 USA: USENIX Association. DOI: 10.5555/3489212.3489345.

Supporting Information

SI 1 Representativeness of the Sample

Table SI 1.1. Comparison of YouGov sample to the Current Population Survey

	YouGov sample (1)	Current Population Survey (1)
Female	0.525	0.52
Male	0.475	0.48
White	0.635	0.673
Hispanic	0.148	0.141
African American	0.127	0.099
Other	0.049	0.023
Asian	0.041	0.064
Age (mean)	48.6	49.8
18–24	0.094	0.112
25–34	0.177	0.143
35–49	0.257	0.240
50–64	0.247	0.248
65+	0.226	0.257
High school or below	0.362	0.382
Some college	0.287	0.267
College degree	0.225	0.219
Postgraduate degree	0.125	0.132
West region	0.202	0.274
Midwest region	0.200	0.193
Northeast region	0.178	0.161
South region	0.421	0.371

Column (1) is this study's YouGov sample in June 2022. Column (2) is the Current Population Survey for all months in 2022. All figures in the table are proportions, except for *Age (mean)*.

SI 2 Understanding Visits to Malicious Websites

In this appendix, we leverage the 6.3 million browsing-level data to further understand online user behavior around malicious versus non-malicious content.

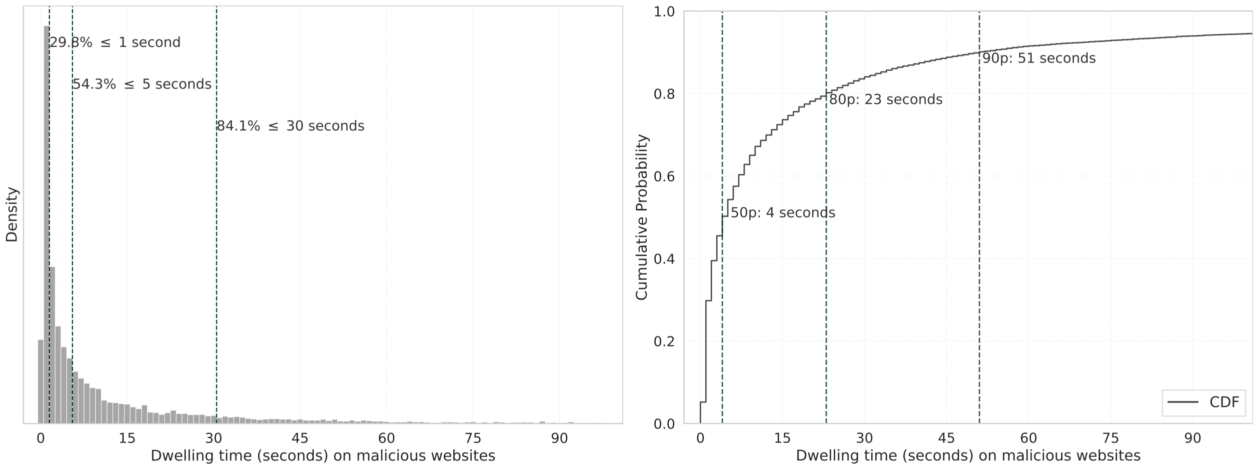
People are generally good at detecting potentially malicious content once they encounter one, given the relative speed of egress on such sites relative to non-malicious sites. 30% of visits to malicious sites end within a second, compared to 18% for visits to non-malicious sites (Figure SI 2.1). Unfortunately, this pattern does not correspond with the level of maliciousness measured by the number of malicious flags. Dwelling time appears to decrease when going from two to five flags, but the pattern reverses as the number of flags increases (Appendix SI 2.2). Unfortunately, we also fail to observe that most visits to malicious sites are singleton or one-off visits–only 17 individuals have all their exposure to malicious content made up of singleton visits.

Finally, we examine the timing of visits and find that people are likelier to visit malicious sites after office hours (Appendix SI 2.4). While this behavior suggests that individuals primarily bear the risk, it also points to the potential exposure of organizations to security vulnerabilities through the personal use of work devices during private hours.

SI 2.1 How Good Are People At Detecting a Malicious Site?

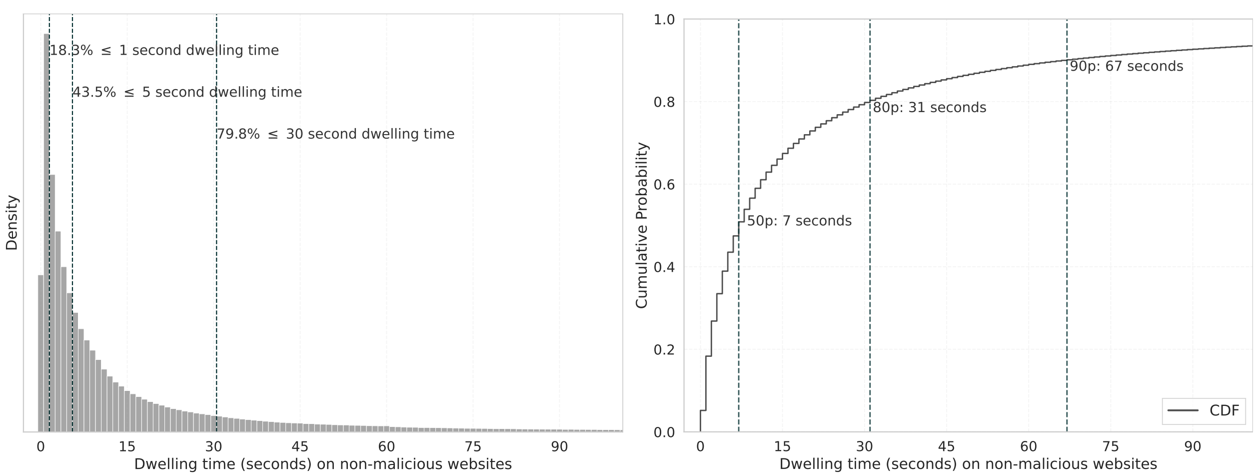
If people are good at detecting a malicious site once they are on it, the average and modal dwelling time per visit should be short. To test this, we analyze the length of visits to malicious websites (n = 23,677) (Figure SI 2.1).

Visits of a second or less constitute more than 30% of the visits to malicious sites. More than half, about 54%, are five seconds or less. The 90th percentile of the duration of a visit to a malicious site is 51 seconds (Figure SI 2.1). In comparison, visits to non-malicious sites last longer. Only 18% of the visits to non-malicious sites are one second or less (Figure SI 2.2). About 43.5% of visits to non-malicious sites are five seconds or less. At the 90th percentile, the dwelling time is 67 seconds. Overall, dwelling times on malicious websites suggest that people are fairly good at detecting malicious content once on it.



(a) Histogram **(b)** Cumulative distribution

Figure SI 2.1. Cumulative distribution of dwelling time (in seconds) on malicious websites, based on 23,677 visits from the individual-browsing level data. Malicious websites are those with at least two malicious flags.



(a) Histogram **(b)** Cumulative distribution

Figure SI 2.2. Cumulative distribution of dwelling time (in seconds) on non-malicious websites, based on 6,002,879 visits from the individual-browsing level data. Here, we define non-malicious websites as those with zero malicious flags and zero suspicious flags.

SI 2.2 Dwelling Time Does Not Depend on Maliciousness

If people are adept at recognizing malicious content, especially on websites where multiple vendors agree on its maliciousness, they should disengage quicker from these sites, as captured by dwelling times. Unfortunately, Figure SI 2.3 suggests the contrary.

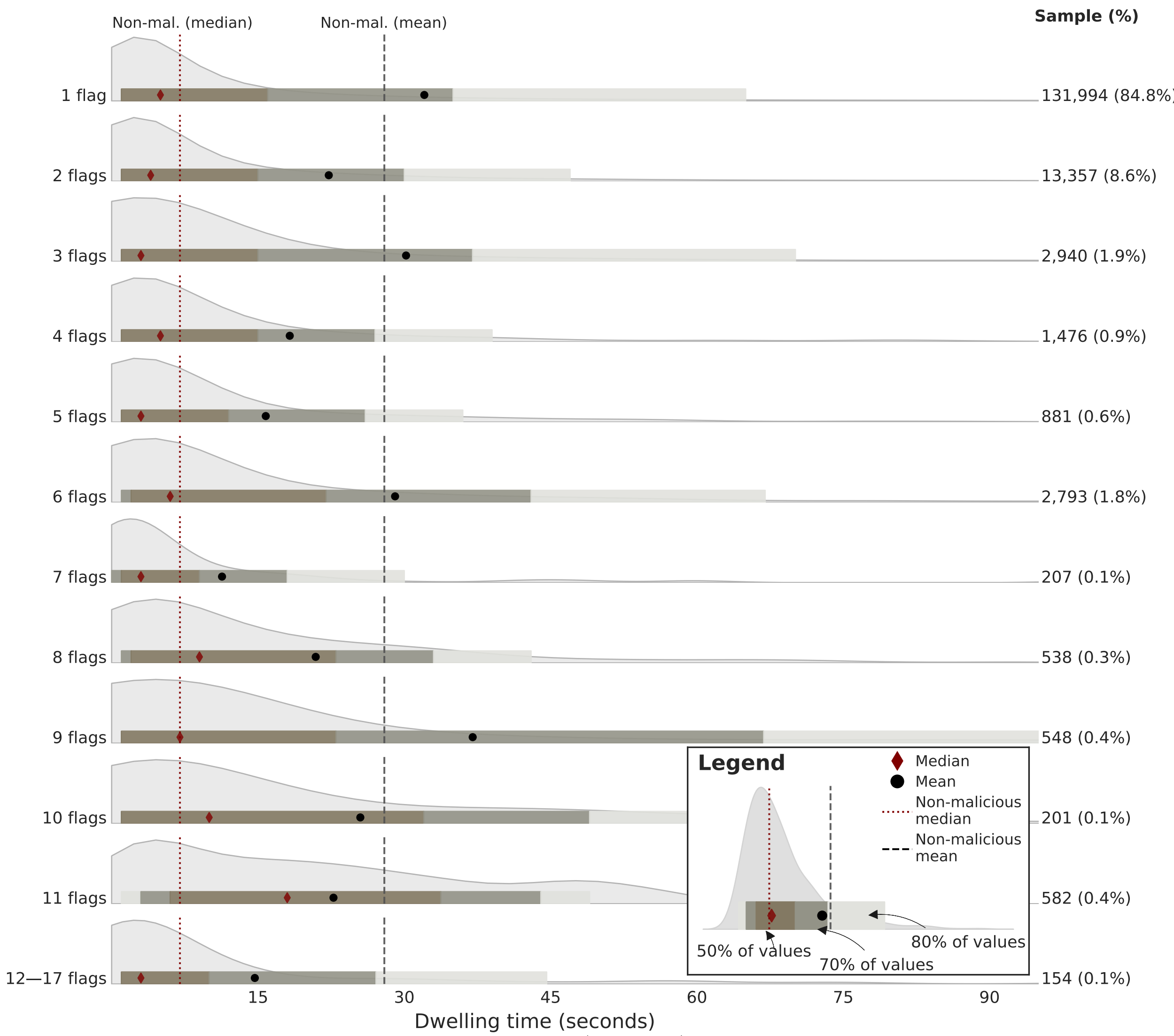


Figure SI 2.3. Distribution of dwelling time (seconds) on websites with at least one malicious flag (n = 155,671 visits). The dotted (dashed) vertical line indicates the median (mean) duration per visit on non-malicious websites (zero malicious and zero suspicious flags; n ≈ 6 m visits). The graph soft censors at the right tail (at approximately the 95th percentile).

Figure SI 2.3 summarizes distributions of dwelling time on visits by maliciousness, revealing a fairly nuanced pattern. While websites with 1-5 flags exhibit progressively shorter median and 90th percentile dwelling times, the trend reverses for websites with six or more

Table SI 2.1. Duration on websites flagged as malicious

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
					Percentiles							
# flags	Count	Mean	SD	Min	p5	p10	p25	p50	p75	p90	p95	Max
1	131,994 (84.8%)	67	880	0	1	1	1	5	16	65	197	85,621
2	13,357 (8.6%)	46	438	0	0	1	1	4	15	47	101	21,601
3	2,940 (1.9%)	35	121	0	0	1	1	3	15	70	178	1,812
4	1,476 (0.9%)	28	194	0	0	1	1	5	15	39	81	3,601
5	881 (0.6%)	25	165	0	0	1	1	3	12	36	55	2,986
6	2,793 (1.8%)	30	77	0	1	1	2	6	22	67	152	999
7–17	2,230 (1.4%)	28	96	0	1	1	2	8	26	51	91	2,313

The table reports the distribution of dwelling time on the 155,671 visits to malicious websites by the number of security vendors flagging the website as malicious.

flags. For example, websites flagged by a single vendor—which we do not classify as malicious in the main text—have a median dwelling time of 5 seconds and a 90th percentile of 65 seconds (Table SI 2.1). For websites with two flags (the threshold for classifying as malicious), these metrics decrease to 4 and 47 seconds, respectively. For sites with five flags, the median drops to 3 seconds and the 90th percentile to 36 seconds. However, websites flagged by more than five vendors see these values increase again. For example, sites with 6–8 flags have median dwelling times of 6–8 seconds and 90th percentiles of 51–67 seconds. This reversal complicates the narrative that users are quicker to disengage as maliciousness increases.

Generally, Figure SI 2.3 supports this conclusion, showing that although the medians (red diamonds) and means (black circles) shift downward for moderately flagged websites, they rise again for highly flagged sites. The distributions also remain right-skewed across all groups, indicating that some users lingered significantly on flagged content. Moreover, highly flagged websites can have longer mean or median dwelling times compared to non-malicious websites as a baseline (indicated by the vertical lines), challenging the notion that users are quicker to egress from more dangerous sites.

SI 2.3 (Lack of) One-Off Visits

Next, we look for evidence of adaptive behavior, or lack thereof, by examining if visits to malicious sites are predominantly "singletons" or "one-off" visits—websites visited only once—or repeated visits to the same malicious websites within our one-month sample period. 17 individuals have all their visits to malicious sites as singleton visits, never visiting a malicious site more than once in our sample period. Notably, this behavior is absent for non-malicious websites–no one has singleton visits for non-malicious sites (Figure SI 2.4).

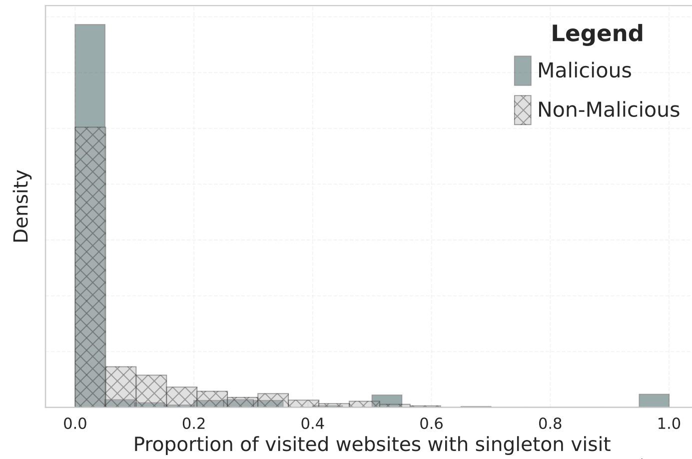


Figure SI 2.4. Individuals visiting at least one malicious website (n = 582) and individuals visiting non-malicious websites (n = 1134). The proportion of singleton visits (websites with one-off visits by the individual) is calculated by grouping the 6.3 million visits by individual and domains, computing visits per domain by the individual, and then computing the proportion of domains with only one visit.

However, Figure SI 2.4 shows that the proportion of such individuals with only singleton visits to malicious sites is very small. Repeated visits to malicious sites are common for most individuals, as with non-malicious sites. Overall, while singleton behavior exists where all visits to malicious sites are one-off, it is exceedingly rare.

SI 2.4 Visits to Malicious Websites Are Likelier Outside of Office Hours

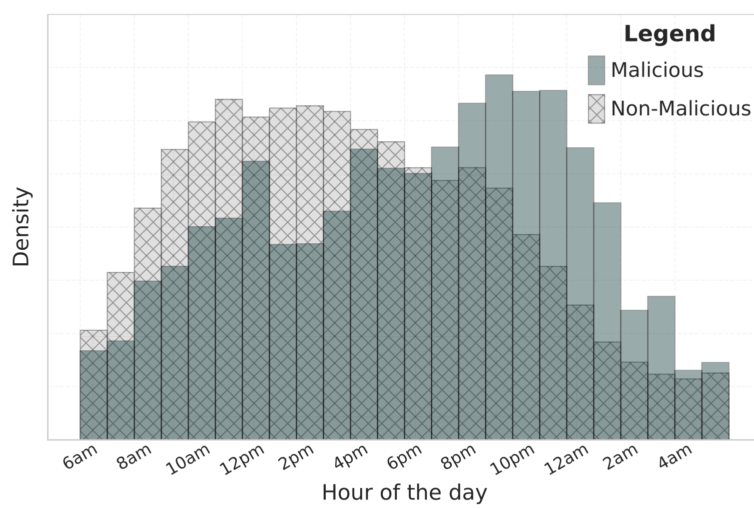


Figure SI 2.5. Time of day during visits to malicious and non-malicious websites based on the 6.3 million visits.

Lastly, we examine the time of day at which individuals visit malicious websites. Figure SI 2.5 presents a clear pattern where, relative to non-malicious visits, proportionally more visits to malicious websites occur during "private hours" (7 pm–3 am) or periods outside regular office hours. This pattern, where more risky visits happen during the private and late hours, indicates a shift in internet use behavior, perhaps encouraged by reduced supervision and greater privacy. Overall, this finding suggests that targeted interventions to curb risky online behavior should occur beyond the workplace, aligning with the observed peaks in risky visits (Baki and Verma, 2023). Likewise, this pattern also raises concerns that individuals in the broader population using company-assigned work machines for personal use during private hours may inadvertently expose their organizations to security risks.

SI 3 Alternate Measures of Exposure to Malicious Content

SI 3.1 Exposure by Duration

Table SI 3.1. Exposure by time (minutes) spent on unique malicious websites, by demographics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
					Percentiles							
	Count	Mean	SD	Min	p5	p10	p25	p50	p75	p90	p95	Max
Panel A. Gender												
Female	595 (52.5%)	15.8	150.8	0	0.0	0.0	0.0	0.0	1.0	8.6	31.3	2879
Male	539 (47.5%)	10.8	50.5	0	0.0	0.0	0.0	0.0	4.0	18.0	50.0	890
Panel B. Race												
White	720 (63.5%)	14.5	140.1	0	0.0	0.0	0.0	0.0	2.0	9.0	30.1	2879
Hispanic	168 (14.8%)	6.5	28.0	0	0.0	0.0	0.0	0.0	1.0	9.0	21.3	250
African American	144 (12.7%)	19.8	62.3	0	0.0	0.0	0.0	0.0	5.2	43.8	128.5	407
Other	56 (4.9%)	10.0	33.3	0	0.0	0.0	0.0	0.0	6.0	21.0	39.8	229
Asian	46 (4.1%)	5.8	15.5	0	0.0	0.0	0.0	0.0	1.0	17.0	49.8	64
Panel C. Education level												
HS or Below	411 (36.2%)	20.9	178.3	0	0.0	0.0	0.0	0.0	3.0	10.0	41.0	2879
Some college	326 (28.7%)	11.7	67.4	0	0.0	0.0	0.0	0.0	2.0	14.5	40.8	890
College	255 (22.5%)	9.2	35.3	0	0.0	0.0	0.0	0.0	2.0	11.0	51.8	321
Postgrad	142 (12.5%)	3.0	8.2	0	0.0	0.0	0.0	0.0	1.0	9.0	18.0	49
Panel D. Age group												
< 25	93 (8.2%)	13.6	44.6	0	0.0	0.0	0.0	0.0	5.0	31.8	63.2	329
25–34	200 (17.6%)	28.7	209.1	0	0.0	0.0	0.0	0.0	2.0	29.4	85.9	2879
35–49	285 (25.1%)	5.3	21.7	0	0.0	0.0	0.0	0.0	1.0	8.6	27.8	192
50–64	288 (25.4%)	17.0	135.3	0	0.0	0.0	0.0	0.0	2.0	13.0	34.6	2067
> 65	268 (23.6%)	6.6	45.0	0	0.0	0.0	0.0	0.0	3.0	9.0	19.6	711

SI 3.2 Exposure to Suspicious Websites

Table SI 3.2. Exposure by number of unique suspicious websites, by demographics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
					Percentiles							
	Count	Mean	SD	Min	p5	p10	p25	p50	p75	p90	p95	Max
Panel A. Gender												
Female	595 (52.5%)	2.7	3.9	0	0.0	0.0	1.0	2.0	4.0	6.0	9.3	45
Male	539 (47.5%)	3.2	4.6	0	0.0	0.0	0.0	2.0	5.0	8.0	10.0	58
Panel B. Race												
White	720 (63.5%)	3.0	3.8	0	0.0	0.0	0.0	2.0	4.0	7.0	10.0	33
Hispanic	168 (14.8%)	2.8	3.7	0	0.0	0.0	0.0	2.0	3.0	7.3	10.0	17
African American	144 (12.7%)	3.5	6.9	0	0.0	0.0	0.0	1.0	4.0	6.7	12.0	58
Other	56 (4.9%)	2.6	2.5	0	0.0	0.0	1.0	2.0	4.0	6.0	7.2	11
Asian	46 (4.1%)	2.8	3.7	0	0.0	0.0	0.0	1.5	4.0	6.5	10.5	15
Panel C. Education level												
HS or Below	411 (36.2%)	3.0	4.3	0	0.0	0.0	0.0	2.0	4.0	7.0	9.5	45
Some college	326 (28.7%)	3.1	5.0	0	0.0	0.0	0.0	2.0	4.0	8.0	11.0	58
College	255 (22.5%)	3.0	3.6	0	0.0	0.0	1.0	2.0	4.0	7.0	9.0	23
Postgrad	142 (12.5%)	2.7	3.5	0	0.0	0.0	0.0	1.0	4.0	6.0	9.9	18
Panel D. Age group												
< 25	93 (8.2%)	3.4	6.8	0	0.0	0.0	0.0	1.0	5.0	7.0	9.4	58
25–34	200 (17.6%)	2.8	3.8	0	0.0	0.0	0.0	1.0	4.0	8.0	10.1	20
35–49	285 (25.1%)	2.4	2.9	0	0.0	0.0	0.0	1.0	4.0	6.0	8.0	18
50–64	288 (25.4%)	3.2	4.8	0	0.0	0.0	1.0	2.0	4.0	7.0	10.0	45
> 65	268 (23.6%)	3.3	3.9	0	0.0	0.0	1.0	2.0	5.0	8.0	11.6	25

SI 3.3 Probability of Visiting Malicious Websites (Individual-Level)

Table SI 3.3. Probability of exposure to malicious websites by demographic characteristics

	Dependent variable is 1(Malicious website visitor)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Woman	-0.083^{a}	-0.070^{b}							-0.085^{a}	-0.072^{a}
	(0.030)	(0.028)							(0.030)	(0.028)
Race: African American			0.106^{b}	0.134^{a}					0.115^{b}	0.130^{a}
			(0.045)	(0.043)					(0.045)	(0.042)
Race: Asian			−0.020	−0.053					−0.007	−0.054
			(0.076)	(0.069)					(0.077)	(0.070)
Race: Hispanic			−0.005	0.038					0.007	0.029
			(0.043)	(0.041)					(0.044)	(0.042)
Race: Other			0.055	0.066					0.073	0.083
			(0.069)	(0.063)					(0.070)	(0.064)
Educ: Some college					−0.056	-0.070^{b}			−0.052	-0.071^{b}
					(0.037)	(0.035)			(0.038)	(0.035)
Educ: College					−0.032	-0.079^{b}			−0.023	-0.071^{c}
					(0.040)	(0.037)			(0.040)	(0.037)
Educ: Postgraduate					-0.106^{b}	-0.152^{a}			-0.099^{b}	-0.139^{a}
					(0.048)	(0.045)			(0.048)	(0.045)
Age: 25–34							−0.038	−0.052	−0.027	−0.032
							(0.063)	(0.060)	(0.062)	(0.060)
Age: 35–49							−0.057	−0.082	−0.035	−0.051
							(0.060)	(0.057)	(0.060)	(0.057)
Age: 50–64							−0.020	−0.084	−0.002	−0.056
							(0.060)	(0.058)	(0.059)	(0.058)
Age: 65+							0.007	−0.087	0.029	−0.062
							(0.060)	(0.058)	(0.060)	(0.058)
Total visits (scaled)		3.01^{a}		3.08^{a}		3.10^{a}		3.08^{a}		3.20^{a}
		(0.258)		(0.262)		(0.261)		(0.266)		(0.266)
Total visits^2 (scaled)		-2.96^{a}		-3.02^{a}		-3.03^{a}		-3.03^{a}		-3.20^{a}
		(0.489)		(0.505)		(0.500)		(0.506)		(0.516)
Constant	0.557^{a}	0.398^{a}	0.499^{a}	0.333^{a}	0.550^{a}	0.413^{a}	0.538^{a}	0.429^{a}	0.578^{a}	0.469^{a}
	(0.021)	(0.024)	(0.019)	(0.022)	(0.025)	(0.026)	(0.052)	(0.051)	(0.057)	(0.056)
R^2	0.007	0.127	0.005	0.131	0.005	0.132	0.002	0.125	0.020	0.147
Observations	1,134	1,134	1,134	1,134	1,134	1,134	1,134	1,134	1,134	1,134

Note: The dependent variable is a dummy for whether the individual ever visited a malicious website during the month-long period. All models are linear probability models. Even-numbered columns adjust for the total number of website visits (*Total visits*). Total visits are scaled to 0–1 so that all variables are between 0–1. The baseline categories are man for gender, White for race/ethnicity, high school or below for education, and 18–24 for age. Standard errors are reported in parentheses. Significance levels: ^{c} 0.1 ^{b} 0.05 ^{a} 0.01.

SI 3.4 Differences in Means

Table SI 3.4. The number of unique malicious websites visited by demographic characteristics

	Dependent variable is Number of unique malicious website visited									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Woman	-0.748^a	-0.639^b							-0.735^b	-0.624^b
	(0.279)	(0.254)							(0.286)	(0.260)
Race: African American			1.47^b	1.67^b					1.45^b	1.56^b
			(0.750)	(0.735)					(0.734)	(0.703)
Race: Asian			−0.272	−0.523					−0.177	−0.535
			(0.402)	(0.410)					(0.424)	(0.442)
Race: Hispanic			0.101	0.402					0.106	0.257
			(0.311)	(0.302)					(0.315)	(0.302)
Race: Other			−0.250	−0.158					−0.117	−0.033
			(0.392)	(0.382)					(0.388)	(0.376)
Educ: Some college					−0.199	−0.301			−0.242	−0.385
					(0.404)	(0.382)			(0.427)	(0.401)
Educ: College					-0.715^b	-1.08^a			-0.594^c	-0.968^a
					(0.319)	(0.327)			(0.352)	(0.339)
Educ: Postgraduate					-0.908^a	-1.25^a			-0.736^b	-1.04^a
					(0.341)	(0.347)			(0.373)	(0.364)
Age: 25–34							−1.03	−1.18	−0.834	−0.913
							(0.936)	(0.922)	(0.921)	(0.895)
Age: 35–49							-1.54^c	-1.73^c	−1.24	−1.35
							(0.913)	(0.883)	(0.866)	(0.826)
Age: 50–64							−0.553	−1.02	−0.297	−0.705
							(0.972)	(0.974)	(0.966)	(0.954)
Age: 65+							−1.21	-1.91^b	−0.913	-1.60^c
							(0.918)	(0.962)	(0.855)	(0.895)
Total visits (scaled)		20.4^a		21.3^a		21.4^a		21.5^a		22.8^a
		(3.48)		(3.60)		(3.64)		(3.76)		(3.90)
Total visits^2 (scaled)		-15.9^a		-16.8^a		-16.5^a		-17.1^a		-18.8^a
		(4.68)		(4.53)		(4.99)		(4.59)		(4.79)
Constant	2.32^a	1.18^a	1.75^a	0.554^a	2.26^a	1.27^a	2.92^a	2.14^a	3.17^a	2.36^a
	(0.215)	(0.168)	(0.142)	(0.209)	(0.270)	(0.263)	(0.901)	(0.745)	(0.804)	(0.655)
R^2	0.006	0.084	0.011	0.095	0.005	0.091	0.010	0.093	0.030	0.119
Observations	1,134	1,134	1,134	1,134	1,134	1,134	1,134	1,134	1,134	1,134

Note: Similar to Table 4, except using OLS regression models. The dependent variable is the total number of unique malicious websites visited over the month-long period. Even-numbered columns adjust for the total number of website visits (*Total visits*). Total visits are scaled to 0–1 so that all variables are between 0–1. The baseline categories are man for gender, White for race/ethnicity, high school or below for education, and 18–24 for age. Standard errors are reported in parentheses. Significance levels: ^{c} 0.1 ^{b} 0.05 ^{a} 0.01.

SI 4 Skewness in Alternate Measures

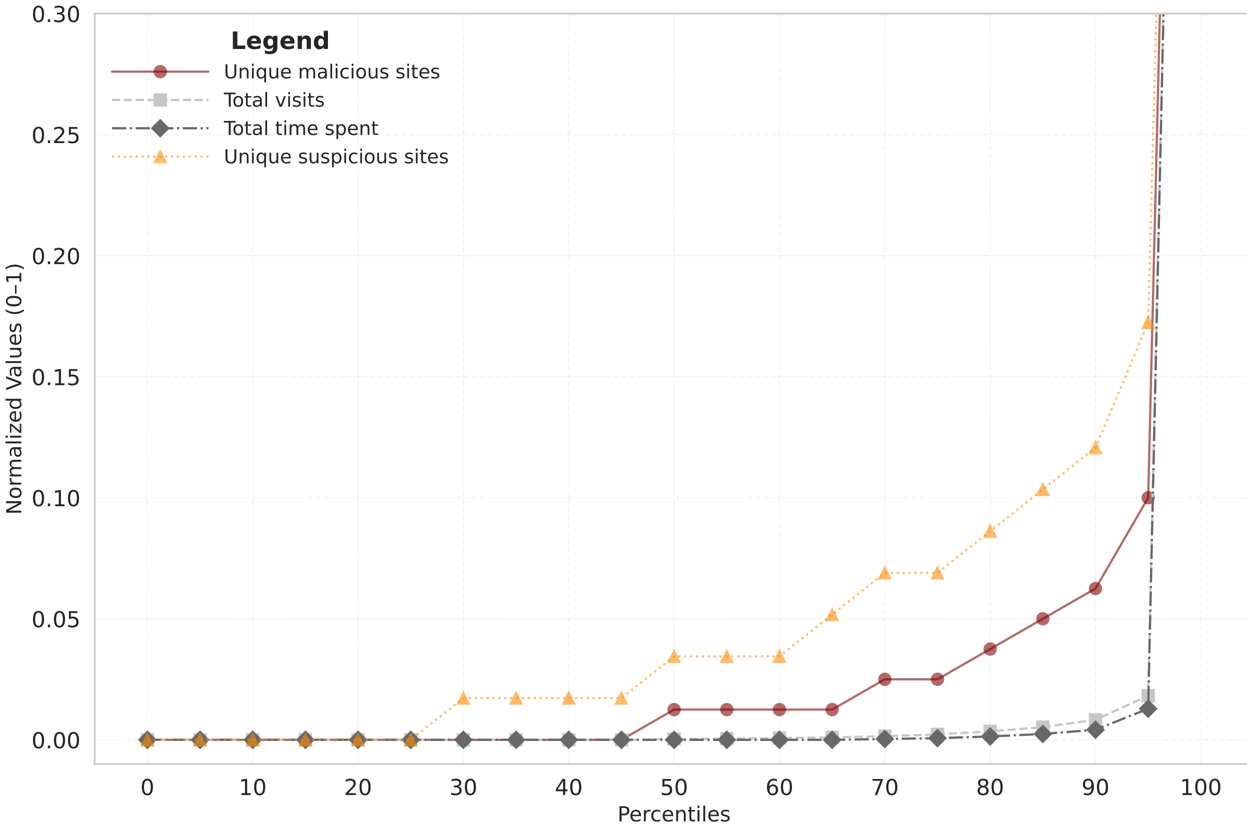


Figure SI 4.1. Distribution of Exposure to Malicious and Suspicious Websites. This graph presents the exposure to malicious and suspicious websites across percentiles for four metrics: "Unique malicious sites," "Total visits," "Total time spent," and "Unique suspicious sites." Values are rescaled to lie between 0 and 1. The plot "soft-censors" the upper portion for visual articulation–all points converge to 1 at the 100th percentile.