Mining Mobile Security and Privacy Topics from Users' VPN App Reviews

Early stage paper

A J Burns
Louisiana State University
ajburns@lsu.edu

Clay Posey
Brigham Young University
clay.posey@byu.edu

ABSTRACT

As individuals' reliance on mobile devices has increased, mobile privacy-enhancing technologies (mPET) have grown in popularity. We contend that these apps provide an important new opportunity for security and privacy researchers to complement their more traditional study of security and privacy issues. To help shed light on individuals' privacy perceptions via their download and use of privacy-enhancing technologies, we suggest that researchers examine user-generated reviews of mPET apps such as VPNs. In this paper, we discuss an early stage research project that analyzes mPET reviews for security and privacy insights. Using a large set of VPN app reviews, we first generated a binary occurrence word matrix and manually evaluated the reviews for relevant insights. Next, to obtain a more complete understanding of contextual information within the VPN reviews, we turned to the Latent Dirichlet Allocation (LDA) algorithm. Based on our exploratory analyses, we believe that we have been able to uncover several important opportunities for security and privacy researchers which we discuss. Although we focus our textual analysis on a specific form of mPET, there are a variety of apps with millions of reviews that may contain novel insights for security and privacy researchers.

Keywords

VPNs, privacy, security, mobile app reviews, text-mining, privacy enhancing technologies.

INTRODUCTION

The security and privacy of individuals' sensitive information is an area of great concern for researchers, governments, businesses, and consumers (Posey et al., 2017; Xu & Dinev, 2022). Companies' and individuals' increased reliance on the cyber channel has only exacerbated the threats to personally identifiable information (Burns & Johnson, 2018). Despite the risks, individuals continue to engage in online activities at an increasing rate. In the U.S., as elsewhere, a primary driver of online engagement is mobile devices' pervasiveness. Mobile devices now account for more than half of all digital media consumption (comScore, 2017) and more than 30% of all eCommerce retail sales (Statista, 2019) in the U.S. The numbers are even more dramatic for younger technology users. For example, mobile applications (mApps), combined with smartphone web browsers, tablet apps, and tablet web browsers, account for more than 75% of the digital media consumption of 18-24-year-olds in the U.S. (comScore, 2017). The trend toward mobile is only accelerating with Americans consuming more than four hours of media per day on mobile devices in 2021 (Ali, 2021).

With such reliance on mobile devices, mobile privacy-enhancing technologies (mPET) that advertise their ability to provide increased security and privacy to individuals within the mobile environment have emerged. Often these mPET technologies come in the form of mApps that users can download onto their phones. One such mPET technology is virtual private networks (VPNs). As defined by Cisco, a VPN is a "private network constructed within a public network infrastructure, such as the global Internet" (Cisco, n.d.). Interestingly, VPN apps have become extremely popular in the iOS App Store in the productivity category. Figure 1 shows the top productivity apps on March 20, 2019, where three of the top ten 'free' apps were VPNs, and five of the top ten 'grossing' apps were also VPNs (AppAnnie, 2019). As of January 2023, an

astonishing 35 of the top 100 productivity apps in the iOS Appstore are advertised as VPNs (see Appendix). Despite VPNs' popularity and marketing promises that an individual can protect themselves "from cyber threats with a simple tap of the screen" (apple.com, 2019), relatively little research has examined users' reasons for downloading and experiences with such apps.

FIGURE 1. TOP PRODUCTIVITY APPS IN IOS (FROM 2019)

To help shed light on individuals' privacy perceptions via their downloading and use of privacy-based technologies, we propose that researchers examine the user-generated reviews of mPET apps such as VPNs. Analyzing user-generated content, such as user reviews, has emerged as a useful tool to help understand individuals' and firms' behaviors (Ghose et al., 2012; Hu et al., 2017). With the development of text-mining tools and techniques, extracting insights from large textual datasets has become more accessible. Related to information security and privacy, researchers have used text-mining techniques to examine corporate disclosures of security risk factors (Wang et al., 2013) and security and privacy violation disclosure notifications (Posey et al., 2017).

Textual app reviews are a valuable source of information about app usage, because they often contain diverse information (Genc-Nayebi & Abran, 2017). While the many studies focus on the positive or negative valence of reviews (i.e., sentiment analysis) on downloads and ratings (Genc-Nayebi & Abran, 2017; Wang et al., 2016), reviews also often contain other potentially

valuable information such as users' expectations of an app's features and its effectiveness. We contend that this diverse information is especially valuable for productivity/utility apps, such as VPNs, which can be useful in different ways across a variety of contexts. This is true because users will have disparate experiences and seek unique capabilities when using utility apps depending on their specific contexts, goals, and potential use-cases for the app.

DATA COLLECTION, SCREENING, AND PRELIMINARY ANALYSIS

To conduct our analyses, we first set out to discover the available VPN apps in the iOS Appstore in 2019. Using a third-party data provider (appfigures, 2019), we performed a search of all iOS apps containing the string "VPN." This search returned 527 unique apps in the Apple Appstore. Of these, we then chose for our analysis a subgroup of all VPN apps that were updated after January 1, 2018. This refinement resulted in a set of 271 apps with U.S. reviews from which we selected the top 25 apps in terms of *total U.S. ratings* since January 1, 2017. Table 1 exhibits the 25 apps selected for our study. Despite having many star ratings from the U.S., we excluded the app that included Chinese characters in the name (Transocks-VPN), because it had very few English language reviews. Thus, our final set included 24 popular VPN mApps.

Table 1. VPN Apps

Ranks	App Name	Developer	Total Star Ratings since 2017	Reviews in Analysis	Date of First Review Included	Date of Last Review Included
1	VPN 360 – Unlimited VPN Proxy	TouchVPN	228,728	3,640	Feb 2017	Feb 2019
2	X-VPN Unlimited VPN Proxy	Free Connected Limited	138,346	1,168	July 2017	Feb 2019
3	Best VPN Proxy Betternet	BetterNet LLC	123,503	53,687	Dec 2014	Feb 2019
4	VPN Master Unlimited proxy	ALL Connected Co.,Ltd	108,407	17,394	Sept 2015	Feb 2019
5	HotspotShield VPN & Wifi Proxy	AnchorFree Inc.	99,598	24,664	Nov 2011	Feb 2019
6	VPN – Super Unlimited Proxy	Mobile Jump	61,399	7,649	May 2018	Feb 2019

7	Free VPN by Free VPN .org	Free VPN LLC	50,208	2,486	Nov 2015	Feb 2019
8	Turbo VPN Private Browser	ALL Connected Co.,Ltd	40,060	1,093	April 2018	Feb 2019
9	Hexatech Unlimited VPN Proxy	BetterNet LLC	39,951	10,239	Jan 2016	Feb 2019
10	VPN Security & Master Defender	Ever Fun Apps LLC	39,231	448	Mar 2018	Feb 2019
11	OrNET – Private Onion Browser	Free VPN Fast VPN	36,887	23,125	Mar 2018	Feb 2019
12	TunnelBear VPN & Wifi Proxy	TunnelBear, LLC	33,969	4,290	Nov 2012	Mar 2019
13	IPVanish VPN: The Fastest VPN	Mudhook Marketing	26,026	1,728	March 2014	Jan 2019
14	ExpressVPN - #1 Trusted VPN	ExpressVPN	25,153	1,816	Aug 2014	Feb 2019
15	VPN by Private Internet Access	London Trust Media, Inc.	24,427	2,554	May 2015	March 2019
16	NordVPN: VPN Fast & Unlimited	Tefinkom & C.O. S.A.	21,569	2,903	Dec 2014	Feb 2019
17	VPN Proxy by Seed4.Me VPN	S4M Tech, Inc.	14,480	9,942	June 2014	Jan 2019
18	Transocks-VPN	Chengdu Fobwifi Networks	14,438	n/a	n/a	n/a
19	VPN for iPhone – Proxy Server	Brain Craft Ltd	14,373	124	Sept 2017	Feb 2019
20	HMA! Hotspot VPN & Proxy – PRO	Privax	13,734	425	Oct 2013	Mar 2019
21	#VPN – Wi-Fi Hotspot Security	Apalon Apps	13,154	282	Sept 2016	Feb 2019
22	SkyVPN – Best VPN Proxy Shield	Sentry Secure Communication	12,770	512	April 2017	Feb 2019
23	Onion TOR Browser + VPN	Art Fusion	11,895	8,144	April 2017	Feb 2019
24	SpeedVPN-VPN Speed vpn Master	Speed VPN	11,456	9,623	Dec 2018	Jan 2019
25	SurfEasy VPN	SurfEasy	10,603	24,038	May 2013	March 2019

Across the 24 apps we identified, we collected a total 211,975 app text reviews. Using Rapidminer (rapidminer.com) we extracted 40,848 unique ‘terms’ or ‘tokens’ that are strings found in the corpus (i.e., entire set of reviews). Of these terms, over half (24,067) appeared only once in the corpus. Many of these were non-English words, nonsense strings (e.g., aawwweehhfhryyyeeaaaah), misspellings (e.g., aceptable, acesss), or unusual slang (e.g., omgggg, omgosh, oooook).

Binary Word-Occurrence Matrix

As a first step to explore the data, we generated a binary occurrence word matrix in Rapidminer. A binary word-occurrence matrix is simply a word matrix with a 1 or 0 for each review depending on whether the review includes the corresponding term. This matrix can be used to conduct an analysis similar to a latent semantic analysis (LSA) (e.g., Posey et al., 2017; Sidorova et al., 2008). As described, "the main idea behind LSA is to collect all of the contexts within which words appear, and to establish common factors that represent underlying concepts" (Sidorova et al., 2008, p. 471).

To make the assessment more manageable, we generated the binary word-occurrence matrix using a 20% random sample of the original dataset and pruned the word list to include terms that were found in at least 5% of the sampled documents and in no more than 30% of the sampled documents. We also filtered out English language stopwords (e.g., "a," "an," "the"), and applied the Porter stemming algorithm (removing word endings such as -ing, -ed). Because we were interested in understanding users' experiences with the apps rather than their opinion about the app, we also filtered out positive and negative terms using the Loughran and McDonald (2016) positive and negative sentiment word lists. Lastly, we filtered the tokens by their length (min = 3, max = 25 characters). The resulting term list included 533 words. Thus, with our initial analysis procedures we were able to narrow our range of analysis to 1.3% of the original terms. We then manually screened the terms and retained 120 terms of potential interest. Interestingly, two or more of these 120 terms appeared in 7,392 reviews.

Table 2. Terms of Interest in the Sample of 7,392 Reviews

Term	Occurrence	Term	Occurrence	Term	Occurrence	Term	Occurrence	Term	Occurrence	Term	Occurrence
vpn	2686	devic	293	media	156	facebook	115	trust	68	music	50
connect	1875	site	292	appl	146	advertis	113	track	66	oversea	50
school	1381	brows	291	privat	142	comput	112	bought	64	viru	50
wifi	1303	privaci	286	youtub	142	googl	112	auto	62	govern	49
block	807	web	271	annoi	141	stream	112	amazon	59	net	48
internet	672	instal	269	anonym	140	address	110	appreci	59	view	48
phone	595	network	252	mobil	140	place	105	passwor	59	appear	47
download	594	countri	234	unblock	140	read	105	onion	56	answer	46
try	575	ipad	220	restrict	139	bypass	104	safari	56	filter	46
secur	552	china	209	video	138	mac	94	amount	55	region	46
browser	549	safe	198	log	133	page	92	iran	55	android	45
access	468	locat	197	game	131	laptop	90	usa	55	pandora	45
data	455	ask	189	instagra	130	absolut	84	firewal	53	content	44
tor	442	email	187	compani	128	fact	81	listen	52	tool	42
iphon	438	netflix	179	surf	128	anywher	78	attempt	51	itun	39
server	398	social	174	inform	125	softwar	74	dark	51	bandwidt	37
websit	333	snapchat	173	onlin	124	info	70	hide	51	freedom	35
allow	332	account	169	avail	122	desktop	69	twitter	51	safer	35
watch	327	travel	165	applic	121	hack	69	unabl	51	platform	34
protect	311	proxi	157	automat	119	configur	68	encrypt	50	fortnit	29

Manually Exploring Term Co-occurrences

We were surprised to see that the term "school" was one of the most frequent terms in our review set, occurring 1,381 times in only 7,392 reviews. To better understand the context of this term within our data, we examined terms that co-occurred with "school" in the reviews. Interestingly, we found that 59% of the reviews containing the term "school" also included the term "block." Upon further examination, we found that multiple reviewers mention "school," "block," and at least one other term of interest (including "*snapchat*," "*facebook*," "*fortnite*," "*game*," "*instragram*," and "*youtube"*). See table 3 for these co-occurrences.

Table 3. Term Co-occurrences with "School" and "Block"

Term	**Total**	**School +**	**%age**	**School + Block +**	**%age**
school	1,381				
block	807	447	59%		
snapchat	173	143	83%	66	38%
facebook	115	50	43%	21	18%
fortnit	29	13	45%	3	10%
game	131	47	36%	19	15%
instagram	130	91	70%	42	32%
youtub	142	58	41%	22	15%

> Example Review containing "school" + "block" + "snapchap" + "instagram": *I use this app for **school** because in study halls I'm bored and our **school** WiFi **blocks Snapchat** and **instagram**, the old app I had would give u 30 mins of **vpn** time and make u watch ad after ad after ad , but this stays **connected** with no interruptions. Very good!!*

Leveraging the binary word-occurrence matrix, we next examined other examples of reviews containing a combination of the terms of interest. Again, as in LSA, this process serves to help identify the various contexts terms to help identify underlying concepts. For example, we noticed reviewers mentioned the term "government" in 49 reviews. Exploring the co-occurrence of terms we found themes around users' desire to protect their privacy from governments for various purposes. One was concerned about government tracking, while others were concerned with the privacy of their online data, still others wanted to use the VPN to bypass governmental restrictions.

> Example Review containing "privaci" + "track" + "govern" + "vpn": *As we know the **government** is **tracking** us we need to be more careful with our **privacy**. **VPN** is a great form of this hopefully this app delivers.*

> Example Review containing "govern" + "privat" + "brows" + "data" + "online" + "vpn": *The best **VPN** I've used to keep the **government** or anyone from getting my **private browsing data** and personal **online data**. Great app!*

> Example Review containing "govern" + "restrict" + "connect": *I studied abroad and was unable to use any form of internet calling due to **restrictions** placed by the **government**. Using this app I was able to make voice/video calls as if I were in the United States, and it barely slowed down my **connection**.*

As a final example, we explored users' feedback regarding their desire for anonymity. In our subsample, 40 reviews included the terms "brows" and "anonym." Meanwhile, 69 reviews mentioned the term "hack." We show examples of each below.

> Example Review containing "anonym" + "brows": *It's a great app and it lets you **anonymously browse** the web without big brother watching you.*

> Example Review containing "hack" + "secur" + "inform" + "online": *I love this! I feel much more **secure** when I go **online** now out in public say at the library or the coffee shop. No*

> *more worrying if the guy behind me if stealing my **information** or if my ex is **hacking** into my account. It's nice to be able to feel **secure** again! Thanks Betternet!*

We contend that these examples demonstrate how insightful the reviews can be once the noise is filtered out of the corpus. However, with such a large universe of reviews, researchers may be interested in more automated processes to better understand the most important contexts for mobile VPN users. For this purpose, we explored the corpus using an automated topic modelling algorithm.

Latent Dirichlet Allocation (LDA)

To obtain a more complete understanding of contextual information within the VPN reviews in an automated fashion, we turned to the Latent Dirichlet Allocation (LDA) algorithm. LDA is one of the most commonly adopted topic modelling tools (Palese & Piccoli, 2020) and is available in Rapidminer. Because LDA is an unsupervised process, we found that the results were more sensitive to data quality than was our manual process. This is true because as experts, we were able to ignore irrelevant reviews and terms even when they were not excluded by the screening process. Thus, in keeping with our exploratory lens, we returned to our original set of reviews and re-engaged in screening processes. We manually removed all reviews containing only non-letters, we manually removed reviews containing non-English characters, and we filtered out any duplicate reviews. These efforts left us with 98,047 reviews.

Prior to running our LDA assessment, we processed these reviews to remove English stop words, positive and negative words, and pruned the reviews to include only words that occurred 5% of the time and no more than 30% of the time. This process removed an additional 1,774 reviews which contained only words that were screened out. This left 96,273 reviews to use for the LDA assessment. Again to help manage the size of our sample, we used a 20% subsample,

resulting in 19,609 reviews for the LDA assessment. Since LDAs require the researcher to use their expertise in determining the number of topics to extract, we took an exploratory approach to determining how the LDA would perform under different conditions (i.e., 10 topics, 20 topics, and 30 topics).

First, we ran the LDA with 10 topics. As expected, the majority of review topics seem to focus on providing feedback of the app experience (e.g., updates, payment, star rating, speed, service). However, three of ten topics, representing 25.8% of the corpus (or 5,058 reviews), reference potential topics of interest. These topics include travelling in countries like China and accessing Netflix in other countries (Topic 2), security and privacy, (Topic 5) and avoiding school Wi-Fi blocks (Topic 7).

Next, we ran the LDA with 20 topics. Again, the majority of review topics seem to focus on providing feedback of the app experience (e.g., updates, payment, star rating, speed, service). However, six of twenty topics, representing 24.7% of the corpus (or 4,850 reviews), reference potential topics of interest. These topics include safety, security, and privacy (Topic 1), user information privacy (Topic 9), web browsing security (Topic 16), school Wi-Fi blocks (Topic 17), server locations (Topic 18), and Netflix and video countries (Topic 19).

Finally, we ran the LDA with 30 topics. Again, the majority of review topics seem to focus on providing feedback of the app experience (e.g., updates, payment, star rating, speed, service). As with the previous LDA, six of thirty topics reference potential topics of interest. These topics are like those identified in the 20-topic LDA. Thus, we concluded that 20 topics were sufficient for adequately understanding the contexts within the VPN reviews. Table 4

shows the results of the 20-topic LDA along with the top five terms from each topic. In the following section we delve into the topics of interest from these results.

Table 4. LDA Results – 20 Topics

Topic #	Top five terms	Reviews	%	Human topic interpretation
Topic 0	connect time speed internet load	654	3.3%	Connection speed
Topic 1	secur protect keep privaci safe	700	3.6%	Safety, security, & privacy
Topic 2	trial subscript monei pai charg	846	4.3%	Payment
Topic 3	connect updat fix time issu	5189	26.5%	Updates
Topic 4	connect time wifi disconnect turn	427	2.2%	Disconnections
Topic 5	sai know see download get	162	0.8%	Download
Topic 6	thank help lot make gui	493	2.5%	Review
Topic 7	iphon servic ipad devic version	234	1.2%	Devices
Topic 8	data nice get pai limit	1006	5.1%	Data limits
Topic 9	user privaci inform pia secur	2153	11.0%	User information privacy
Topic 10	review upgrad star get give	459	2.3%	Star rating
Topic 11	easi fast recommend connect simpl	349	1.8%	Recommendation
Topic 12	support servic email custom account	1005	5.1%	Customer support
Topic 13	que mui con esta funciona	960	4.9%	Spanish language reviews
Topic 14	phone get turn instal delet	1439	7.3%	Delete
Topic 15	get game plai dai time	1162	5.9%	Get and play games
Topic 16	browser tor web brows secur	1165	5.9%	Web browsing security
Topic 17	school wifi block get help	443	2.3%	School Wi-Fi blocks
Topic 18	countri server china locat connect	580	3.0%	Server locations
Topic 19	watch netflix give video countri	183	0.9%	Netflix & video countries
Total		**19,609**	**100%**	

Topic 1: Safety, Security, and Privacy

To better understand safety, security, and privacy topics that reviewers mentioned in their reviews, we sorted the reviews labeled as Topic 1 and examined the contexts of the terms safety, security, and privacy. First, we looked at reviews that included each of the terms "*safe*", "*privaci*," and "*secur*." As shown in table 5, the contexts for these terms are varied, including airports, hotels, Wi-Fi networks, social media, travelling, hacking, tracking, governments, IP address, and banking.

Table 5. Topic 1 Context: Safety, Security, and Privacy

Term	Review
Safe	I really feel **safe** having this.
	I like it it makes me feel **safe**!
	Effective, practical and **safe** when connecting on airports and hotels
	Great stuff to keep you **safe**. Take that Facebook.
	It works great and it helps me stay **safe** online so keep up the good work
	Just the fact it let's you use free WiFi is perfect and knowing your **safe** and it's good is the best thing I have found
	In today's tech world, you can never be too careful. Nord gives you superior protection, super easy to use, a large selection of servers around the world, you just sign in, pick the server you want to use, and surf the net **safely**. It is really that easy.
	I love my PIA VPN service. It does exactly what it's supposed to do, route all your Internet traffic through remote locations, encrypted and **safe**. I feel **safer** knowing that when I have my VPN turned on, all of my activities are shielded from prying eyes. I use public WIFI often and with PIA you could rest assured no one will be grabbing your log in information or private conversations. You can't beat the price and in today's world we need all the protection we could get. If you are on the fence about purchasing a VPN service, you have just found the most affordable and user friendly program out there.
Privaci	This app protect your **privacy**.
	If you value **privacy** this app is for you.
	I love the protection on Wi-Fi networks and the overall **privacy** of my data online
	I love this app. The **privacy** is a real concern with all of the hacking out there. Can't go wrong.
	Having **privacy** while online is a priority, you don't want strangers taking your information or knowing what you're doing.
	Using this app will allow you to have **privacy** and reduce large corporations from collecting your moves at every moment. It's scary enough that the government can track your actions, now it's corporations that are trying to collect intel on you. This would help reduce that!
Secur	Great app for bank **security**
	I'm happy to know my phone is **secure**.
	Never felt more **secure** while using unsecured wireless networks.
	I really like it but it keeps my information **secure** and keeps my IP address stealth.
	I have to use Wi-Fi networks that are not **secure** at times and I trust Tor to protect me from the nasty people.
	I bought a subscription for traveling. Currently in Portugal. App works great. I feel **secure** when transmitting info via hotel internet connections.
	I have been using this VPN in N. America, India, and Southeast Asia. No problems anywhere and now I know that when I access banking, credit card, and brokerage accounts I have a much higher level of **security** especially with public WIFI locations.

Topic 9: User Information Privacy

Next, we explored the reviews associated with topic 9. Based on the LDA results, these reviews are associated with the terms "*user*," *privaci*," "*inform*," and "*secur*." As shown in table 6, there again were numerous contexts for these terms. Interestingly, as opposed to the contexts

for security and privacy in topic one which tended to focus on privacy enhancements from mPET, many reviews in topic 9 appear to be referring to the security and privacy risks to information that result from using the VPN. For example, there are several reviews mentioning that the VPN app *is not secure* or *does not respect users' privacy*. While there certainly were reviews that mentioned such vulnerabilities in topic 1, it appears to be a more common theme in topic 9.

Table 6. Topic 9 Context: User Information Privacy

Term	Review
Inform	i love this app for hiding my ip from intruders and people willing to go to prison for my personal **information**
	this is the worst vpn ever,dont consider it a vpn,for one thing it doesnt change your ip,so you are not anonymous,it probably just collects your data and your **information**.
	I like how you can change the VPN protocol, speeds are good, works good, ads are fine. BUT clearly no one is reading what **information** they're collecting from you. Yes personal **information** isn't collected BUT how they define non-personal **information** that IS collected is quite concerning. This app collects IP address, device ID number, apps you have on your device and how you use them, etc. Clearly some of this could borderline personal **information** especially IP address. If your looking to remain truly anonymous this app is a JOKE!!!
Secur	This VPN leaks. Do a DNS leak test and you will be disappointed if you thought your connection was **secure**. Shame on you SurfEasy.
	I used to LOVE this app. But Opera just sold this app to a company based in China, so I am no longer confident this app will remain as **secure** in the future. China doesn't exactly embrace VPN. That's a shame because it was very good. (albeit a little pricy)
	The app works great, and I greatly appreciated the addition of the option to use Openvpn; however the authentication being used could be stronger. SHA-1 authentication has been broken, and should be replaced with SHA-2 (Ex:SHA-256). Also it would be much better for **security** if AES-256 was being used for encryption.
Secur + Privaci	I recommend this tool. The only drawback I have noticed is that many of the terms of the company's **privacy** policy make the use of their product much less **secure** and anonymous. But it's better than nothing.
	Doesn't work. Also there are multiple articles out there on **security** blogs showing that Surfeasy is extremely bad for actual **privacy**. They keep logs and lie about it, etc.
Inform + Privaci	In these days of **information** theft, **privacy** is more important than ever!
	Have tried a few VPNs and this beats them hands down. They have taken steps in 2017 to assure **privacy** and to not sell/share any location **information**.
	I love the fact that they are offsite. I chose them because they believe in **privacy** and NOT surrendering **information**. Connection is always fast.
User + Privaci + Inform	Multiple servers in a handful of good locations!! Awesome! I can use the same service on my computer and phone!! Awesome. Good **privacy**: TunnelBear does NOT log IP addresses, NOR do we monitor, collect, analyze or store any information about the applications, services or websites our **users** use while connected to our Services. (From their privacy policy)

	This app allows **users** to seamlessly surf the web and I feel much better knowing that my information is not being collected by malicious individuals. I really gained a lot of respect for the developers when reading the **privacy** policy. They clearly state that though there are some instances that they may collect data (basically if you're doing something illegal, which you shouldn't be doing anyway), they provide a truly free betternet to people around the world. I respect their dedication to my **privacy**.

Topic 16: Web Browsing Security

Next, we explored the reviews associated with topic 16. Based on the 20-topic LDA, these reviews are associated with terms like "*web*, *"brows*," "*tor*," "*browser*," and "*secur*." As shown in table 7, there are again numerous contexts for these terms. Unlike topic 1, which focused on safety, security, and privacy of using VPNs in general. Several reviews in this topic focus on specific VPN technology of onion routing or the use of Tor technology.

Table 7. Topic 16 Context: Web Browsing Security

Brows + Secur	Love the ability to **browse securely** and discreetly with The Onion
Brows + secure + tor	I feel **Tor** gives me **security** and safety when **browsing**.
	I love **browsing** using **Tor**. Not only do I feel **secured** and safe but speed is always fast and efficient.
	Use **Tor browser** for safe and **secure browsing**, my biz ness stays my biz ness. Highly recommend!

Topic 17: School Wi-Fi Block

Looking at topic 17, the theme of using VPNs to bypass URL-filtering school Wi-Fi networks is prominent. The terms are "*school*," "*wifi*," "*block*," "*get*" and "*help.*" This topic appears to be one of the most internally consistent in the LDA results. As shown in table 8, these reviews almost exclusively focus on using VPN to bypass firewall restrictions and access Internet resources like social media and video streaming services.

Table 8. Topic 17 Context: School Wi-Fi Block

School	My **school** has a lot of restrictions and it allows me to go through the firewall, nuff said.
	Works well at **school**
	Now I can Instagram at **school** ;)
	Let's me use social media in **school**

	It's good for getting pass **school** filters
	I go to a **school** with very few allowed sites and apps so I got through quite a few vpns so far this one works the best
	I turn it on at **school** and it lets me use snapchat or Instagram which is good, but it can be a spotty and sometimes it can take a while to turn on.
School + block + wifi	This app let's me use my phone at my **school** that has no service and their **wifi blocks** everything. I love this app.
	My **school blocked** Facebook, snapchat and Instagram on their **wifi**, but this app bypasses all the blocked applications; and now I can fail in peace ??
	This app is so awesome. I honestly don't ever write reviews but I wrote this one so another person in high school can see this and benefit from it! My **school blocked** a bunch of things on their **wifi** including YouTube. This app consistently helped me get through to the real Internet and let me watch YouTube videos to help me study! It made a huge difference. The app is super easy to use too, another benefit.

Topic 18: Server Locations

The reviews in topic 18 focus on the ability to access resources that are restricted to one country. The top terms for this topic are "*countri*," "*server*," "*china*," "*locat*," "*connect*." For example, some reviewers use VPNs to route traffic to a specific country to access geo-restricted content (i.e., route traffic into a geographic area). Others mention the ability to access blocked resources in a specific country (i.e., route traffic out of a geographic area). For instance, several reviews mention countries with strict internet restrictions. Individuals residing in or visiting these countries might use VPNs to access news, social media, and communicate with the outside world.

Table 9. Topic 18 Context: Server Locations

Countri	I just go this to change my Netflix **country** and it works flawlessly
	Allow you to subscribe and use services not yet available in your **country**
	Being deployed to a foreign **country** where porn is banned this is a true life saving app
	I love that I can finally connect to international apps which I can't access in my home country!
	When you live a in a oppressive government **country** like mine (Ecuador) where Twitter is sometimes censored, this is the best vpn free to overcome restrictions and learn what is really going on!
Server	This app works great when you can connect to the **server** you want. US and Canada servers seem to go down a lot.
	Currently in Europe and the Nordvpn allows me to communicate safely in hotels with open wi fi networks. With the new location identified **server** selection by city am able to watch most of my favorites.
	This is an excellent VPN, the only one that made me switch from PIA after 2 years. It has many

	servers to choose, the speeds are good and the apps are easy to use. Their privacy policy is very good and they are outside 14 eyes jurisdiction. The service works very well for unblocking sites as well (other VPNs are blocked). Highly recommended.
Countri + Server	You cannot pick the **server** sometime USA or other **country**. :(
	Works very well. I would like more **servers** around other **countries** like in South America.
	When I'm in another **country** and I can't access US **Servers**, I just turn this on and it's like I'm there.
	A reasonably priced service with good speeds and tons of **servers** around the globe. The app is very easy to setup and use and the service stays connected when needed. It is very easy to change servers and/or **countries** as required to bypass restrictions.
China	Good app! Especially when you are traveling in mainland China.
	Works well in China most of the time. More and more VPN'S are blocked, especially in the west.
	Although it's a little bit slow, the app is still capable to get through the Great Firewall of China.
	I got to China and nothing was working until I downloaded this app! Saved my life because i could communicate with everyone from the USA!

Topic 19: Watch Netflix and Videos

Finally, topic 19 focuses on the use of VPNs specifically to watch videos that are geographically restricted. The top terms for this topic are "*watch*," "*netflix*," "*give*," "*video*," "*countri*," "*connect*." In some cases, these restrictions are geography-based, such as the ability to watch UK-based shows by routing traffic through the UK. In other cases, the restrictions are content based, such as using VPNs to watch adult videos online where they censored.

Table 10. Topic 19 Context: Watch Netflix and Videos

Watch	Now I can **watch** anime in Mexico thx
	Now i can **watch** tv from any country!
	I am able to **watch** a series from Spain with no hassle.
	Effective as well as a proxy server. I can **watch** BBC iPlayer in the US.
	In Saudi Arabia moviebox and a lot of things don't come up finally I can **watch** movies for free on moviebox
Watch + Netflix	Great way to **watch** shows on **Netflix** that other wise you would only be able to watch on other countries.
	The only thing this is good for is **watching** Italian **Netflix** from the U.S., and it isn't able to do that.
	It was just what I needed here in India to **watch Netflix** so it was just perfect I can' find any glitches so far!

RESEARCH OPPORTUNITIES

Based on these exploratory analyses, we believe that we have been able to uncover several important opportunities for security and privacy researchers. From a methodological perspective, while we focused our textual analysis on a specific form of mPET, there are a variety of apps with millions of reviews that are relevant for security and privacy researchers who could also benefit from the LDA approach. For example, recent research investigates the adoption of password managers (Mattson et al., 2022). A quick search of mApps for password managers reveals dozens of password managers with millions of cumulative reviews. These reviews reveal reasons why the individual might adopt a password manager in their own words. For example, the following review points to prior experience with having a password stolen as a potential motivating factor for future research: "… *For someone (like me) that's paranoid (having an account hacked before) and don't ever use the same passwords, it's a great app.*"

As another example, security and privacy research has examined the use of anti-malware and -spyware applications (Johnston & Warkentin, 2010; Lee & Kozar, 2005; Lee & Larsen, 2009). There are many mApps that advertise anti-malware and anti-spyware capabilities. We contend that these apps' reviews likely contain useful context for users' intentions to install these applications. As noted in the following two reviews, some individuals reported downloading these applications only after encountering a specific incident:

> "…a*fter downloading some sketchy modified apps however, I had a pop-up problem. I went through and cleaned a bunch of stuff up, but couldnt find the root. I decided to give this a try and it found the culprit (mobidash) very quickly and I was able to remove it.*"
>
> "*I downloaded some photo editing apps. Right after, my phone screen froze and then stopped working properly... I immediately ran Malwarebytes for a virus scan. Sure enough, one of the photo editing apps had a trojan attached to it. Malwarebytes found it, removed it from my phone and uninstalled the malware app...*"

Thus, our first recommendation for future research is to use app reviews to help contextualize and supplement traditional research methods like surveys. This will allow researchers to map their survey-based quantitative findings back to contextualized reviews to explain their significance. We believe this opportunity will be especially useful when researchers have specific terms in mind that they can use to mine spontaneous reviews and find user comments that help contextualize or explain their findings.

Second, mApp reviews could also provide a source of data to develop taxonomies of interest. For example, Posey et al. (2017) categorized textual breach data to develop a taxonomy of privacy threats. In similar fashion, researchers might be able to categorize textual mApp reviews to develop security- or privacy-related taxonomies. For example, our preliminary analysis of VPN reviews might serve as a foundation for developing a taxonomy of mobile VPN use-cases.

Third, mApp reviews might be useful in direct theory building. For example, as reviews uncover novel motives and experiences, new theoretical frameworks can be developed that account for these real-world behaviors and experiences. These theories can then be tested through more traditional research methods such as qualitative interviews or quantitative surveys. In this way, when security or privacy technology adoption is of interest, mApp reviews can assist researchers in triangulating more impactful research models and theoretical explanations.

Fourth, in our LDA assessment, one of our topics (topic 9) uncovered users' concerns about their own security and privacy when using mPET applications such as VPNs. This is an important point and uncovers another potentially fruitful opportunity for future research involving VPNs. Specifically, when users route their traffic through a VPN provider, they are by

definition making themselves vulnerable to the security and privacy characteristics of the mApp itself. Future research might use the mApp reviews to better understand how trust is established with these VPN providers.

In all, we recommend security and privacy researchers consider the rich datasets available in mApp reviews.

REFERENCES

Ali, A. (2021). The rise and rise of media on your mobile phone - in one chart. Retrieved January 10, 2023, from https://www.weforum.org/agenda/2021/05/rise-of-media-on-mobile-phone-chart/

AppAnnie. (2019). *Ios top app charts*, from https://www.appannie.com

appfigures. (2019). Retrieved March 9, 2019, from appfigures.com

apple.com. (2019). *App store preview for hotspotshield vpn & wifi proxy*. Retrieved March 20, 2019, from https://itunes.apple.com/app/id443369807

Burns, A. J. & Johnson, E. (2018). The evolving cyberthreat to privacy. *IEEE IT Professional*, 20(3), 64-72.

Cisco. (n.d.). *Virtual private network (vpn)*. Retrieved March 20, 2019, from https://www.cisco.com/c/en_uk/solutions/security/vpn.html

comScore. (2017). The 2017 u.S. Mobile app report. Retrieved March 20, 2019, from https://www.comscore.com/Insights/Presentations-and-Whitepapers/2017/The-2017-US-Mobile-App-Report

Genc-Nayebi, N. & Abran, A. (2017). A systematic literature review: Opinion mining studies from mobile app store user reviews. *Journal of Systems and Software*, 125, 207-219.

Ghose, A., Ipeirotis, P. G., & Li, B. (2012). Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content. *Marketing Science*, 31(3), 493-520.

Hu, Y.-H., Chen, Y.-L., & Chou, H.-L. (2017). Opinion mining from online hotel reviews – a text summarization approach. *Information Processing & Management*, 53(2), 436-449.

Johnston, A. C. & Warkentin, M. E. (2010). Fear appeals and information security behaviors: An empirical study. *MIS Quarterly*, 34(2), 549-566.

Lee, Y. & Kozar, K. A. (2005). Investigating factors affecting the adoption of anti-spyware systems. *Communications of the ACM*, 48(8), 72-77.

Lee, Y. & Larsen, K. R. (2009). Threat or coping appraisal: Determinants of smb executives' decision to adopt anti-malware software. *European Journal of Information Systems*, 18(2), 177-187.

Loughran, T. & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, 54(4), 1187-1230.

Mattson, T., Aurigemma, S., & Ren, J. (2022). Positively fearful: Activating an individual's hero within to explain volitional security technology adoption. *Journal of the Association for Information Systems*, JAIS Preprints(Forthcoming).

Palese, B. & Piccoli, G. (2020). Evaluating topic modeling interpretability using topic labeled gold-standard sets. *Communications of the Association for Information Systems*, 47(1), 54.

Posey, C., Raja, U., Crossler, R. E., & Burns, A. J. (2017). Taking stock of organisations' protection of privacy: Categorising and assessing threats to personally identifiable information in the USA. *European Journal of Information Systems*, 26, 585–604.

rapidminer.com.

Sidorova, A., Evangelopoulos, N., Valacich, J. S., & Ramakrishnan, T. (2008). Uncovering the intellectual core of the information systems discipline. *MIS Quarterly*, 32(3), 467-482.

Statista. (2019). *U.S. Mobile retail commerce sales as percentage of retail e-commerce sales from 2017 to 2021*. Retrieved March 20, 2019, from https://www.statista.com/statistics/249863/us-mobile-retail-commerce-sales-as-percentage-of-e-commerce-sales/

U.S. Department of Commerce. (2022). Quarterly retail e-commrce sales: 3rd quarter 2022. *U.S. Census Bureau News*. Retrieved January 10, 2023, from https://www2.census.gov/retail/releases/historical/ecomm/22q3.pdf

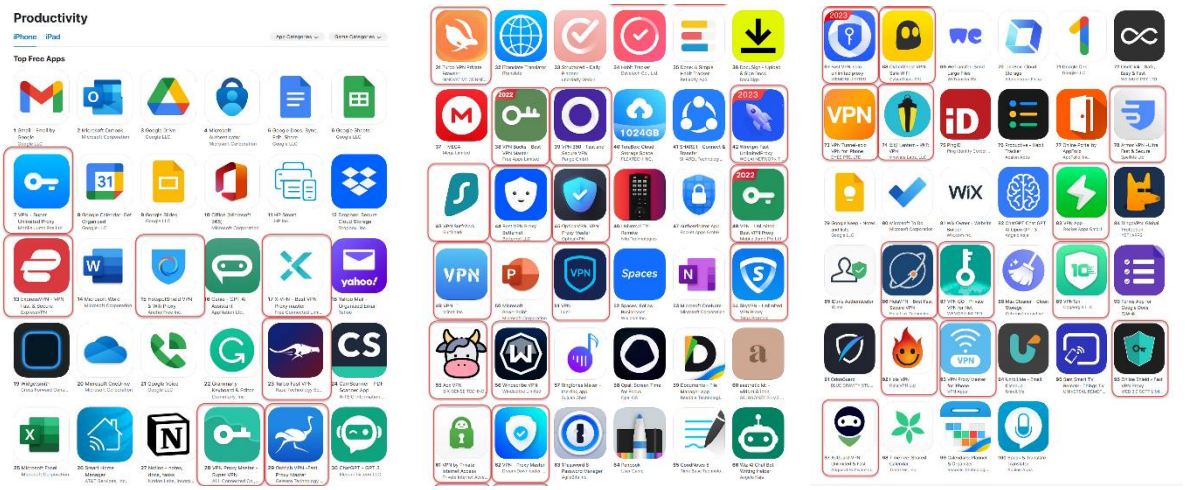
Wang, T., Kannan, K. N., & Ulmer, J. R. (2013). The association between the disclosure and the realization of information security risk factors. *Information Systems Research*, 24(2), 201-218.

Wang, Y., Miguel, A.-U., & Song, J. (2016). Investigating the value of information in mobile commerce : A text mining approach. 26(4), 577-592.

Xu, H. & Dinev, T. (2022). Guest editorial: Reflections on the 2021 impact award: Why privacy still matters. *MIS Quarterly*, 46(4), xx-xxxii.

APPENDIX

Top Free iOS Productivity apps as of January 13, 2023*



*VPNs circled in red

10-Topic LDA

Topic #	Top five terms	Reviews	%	Human topic interpretation
Topic 0	monei subscript trial month pai	1,212	6.2%	Payment
Topic 1	connect time update turn fix	3,437	17.5%	Updates
Topic 2	watch country china netflix travel	819	4.2%	Netflix, Countries like China, and Travel
Topic 3	servic support email custom get	453	2.3%	Customer service
Topic 4	connect service server speed iphon	1,074	5.5%	Connection speed
Topic 5	browser tor secur privacy web	1,918	9.8%	Security and Privacy
Topic 6	review give upgrad star write	847	4.3%	Star rating
Topic 7	school wifi block get help	2,321	11.8%	School wifi blocks
Topic 8	easi fast recommend nice connect	5,342	27.2%	Recommendation
Topic 9	get thank thing help data	2,186	11.1%	Service
Total		**19,609**	**(100%)**	

30-topic LDA

Topic #	Top five terms	Reviews	%	Human topic interpretation
Topic 0	recommend highli definit amaz fast	868	4.4%	Recommendation
Topic 1	que mui con esta funciona	142	0.7%	Spanish language reviews
Topic 2	batteri password account son log	293	1.5%	Battery performance
Topic 3	browser tor web brows secur	711	3.6%	Secure web browsing
Topic 4	turn get run look softwar	1,810	9.2%	Software performance
Topic 5	watch netflix video countri stream	306	1.6%	Watch video countries
Topic 6	sai pop button page click	1,453	7.4%	App performance
Topic 7	thank help awesom make lot	120	0.6%	Recommendation
Topic 8	download get try sai see	993	5.1%	Download issues
Topic 9	secur protect privaci keep inform	146	0.7%	Information security & privacy
Topic 10	iphon ipad devic phone mac	85	0.4%	Devices
Topic 11	review upgrad get write give	571	2.9%	Upgrades
Topic 12	connect time disconnect turn wifi	81	0.4%	Disconnect issues
Topic 13	support servic custom email get	416	2.1%	Customer support
Topic 14	connect get pretti help internet	189	1.0%	Internet connection
Topic 15	speed test server mbp devic	3,488	17.8%	Speed
Topic 16	monei trial subscript pai charg	291	1.5%	Payments
Topic 17	know thing want sai get	138	0.7%	Perspective
Topic 18	give star toe thoe rate	485	2.5%	Star rating
Topic 19	updat connect fix version iphon	245	1.2%	Updates
Topic 20	servic year price speed tri	26	0.1%	Price
Topic 21	game plai get dai time	1,943	9.9%	Games
Topic 22	easi fast simpl connect reliabl	322	1.6%	Ease of use
Topic 23	connect server pia issu set	1,088	5.5%	Connection issues
Topic 24	server locat countri usa address	93	0.5%	Server locations countries
Topic 25	connect china access internet travel	444	2.3%	Internet access travel
Topic 26	phone delet instal turn download	248	1.3%	Deletions
Topic 27	nice derp mama diddili ooh	1,250	6.4%	Nonsense
Topic 28	school wifi block get help	767	3.9%	School wifi block
Topic 29	data unlimit limit get pai	597	3.0%	Data limits
Total		**19,609**	**100%**	