"I just hated it and I want my money back":

Data-driven Understanding of Mobile VPN Service Switching Preferences in The Wild

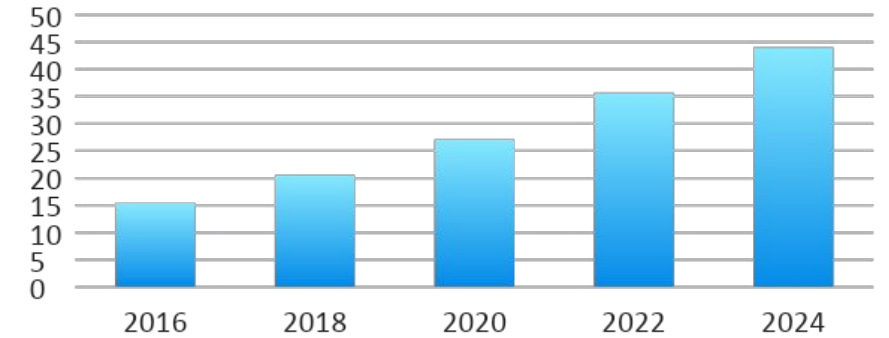Rohit Raj, Mridul Newar, **Mainack Mondal**

IIT Kharagpur

VPN services are becoming popular

VPN market share is increasing

44 B+ USD and **1.5 B+** users in 2024

69% used VPN apps on mobile device

VPN services are becoming popular

VPN market share is increasing

44 B+ USD and **1.5 B+** users in 2024

69% used VPN apps on mobile device

Great market share comes with great number of VPN apps

Applications / Virtual private network
From sources across the web

ExpressVPN
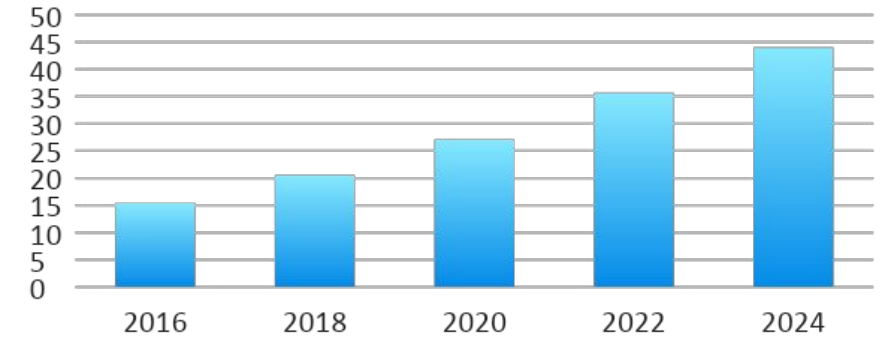
NordVPN

Proton VPN

CyberGhost VPN

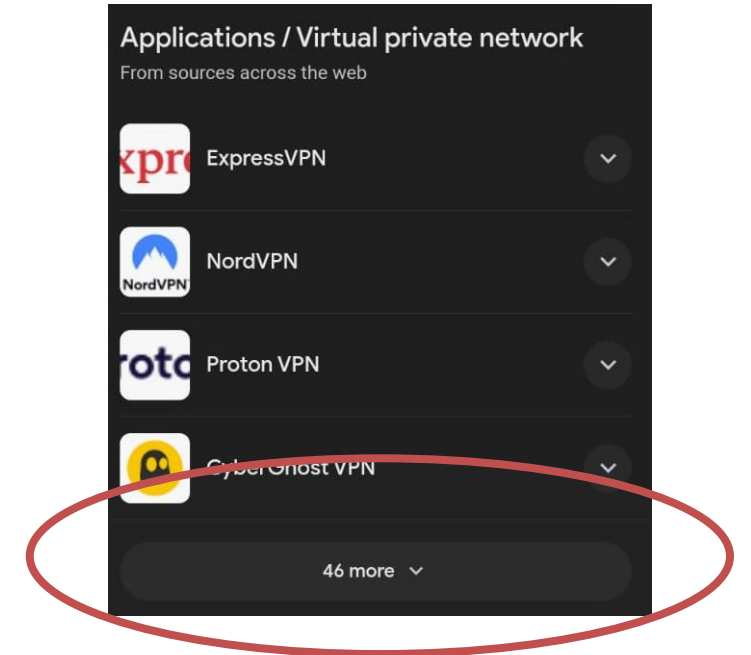46 more

VPN services are becoming popular

VPN market share is increasing

44 B+ USD and **1.5 B+** users in 2024

69% used VPN apps on mobile device

Great market share comes with great

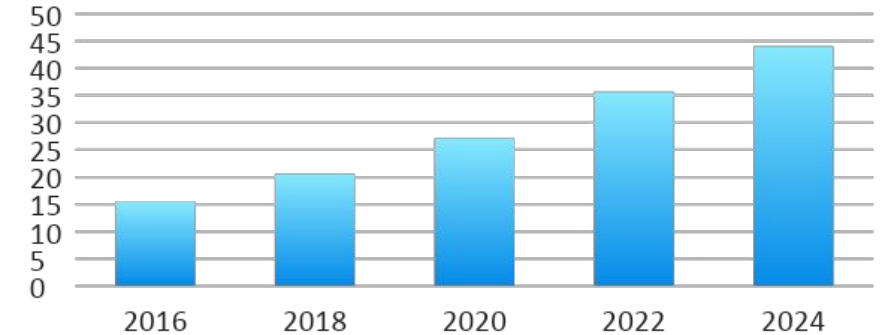number of VPN apps

… and **great confusion**

VPN services are becoming popular
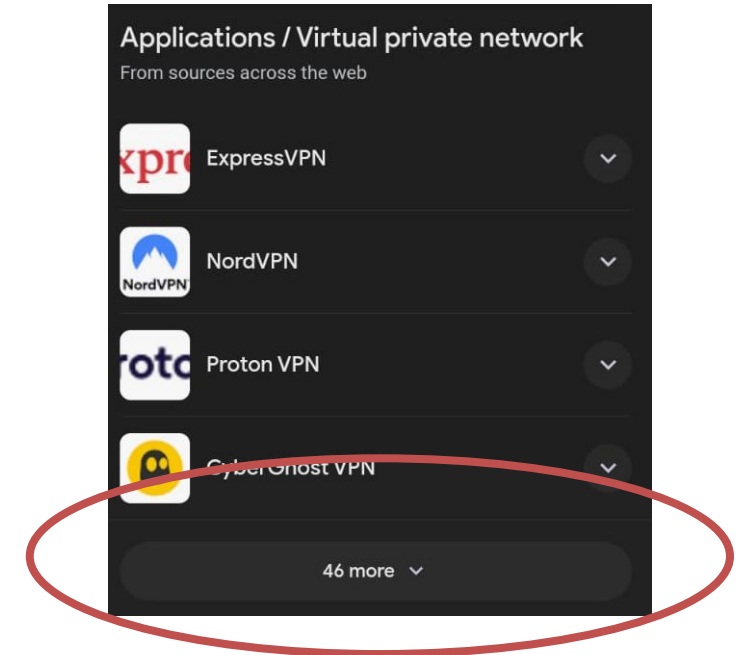
VPN market share is increasing

44 B+ USD and **1.5 B+** users in 2024

69% used VPN apps on mobile device
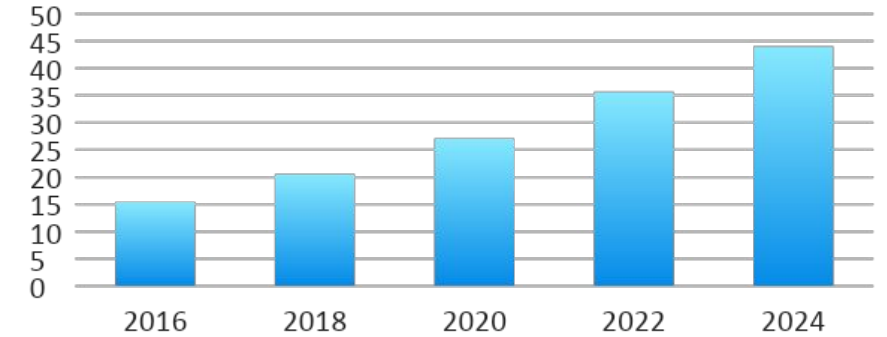
Great market share comes with great

number of VPN apps

… and **great confusion**

Prior work: How users adopt VPNs
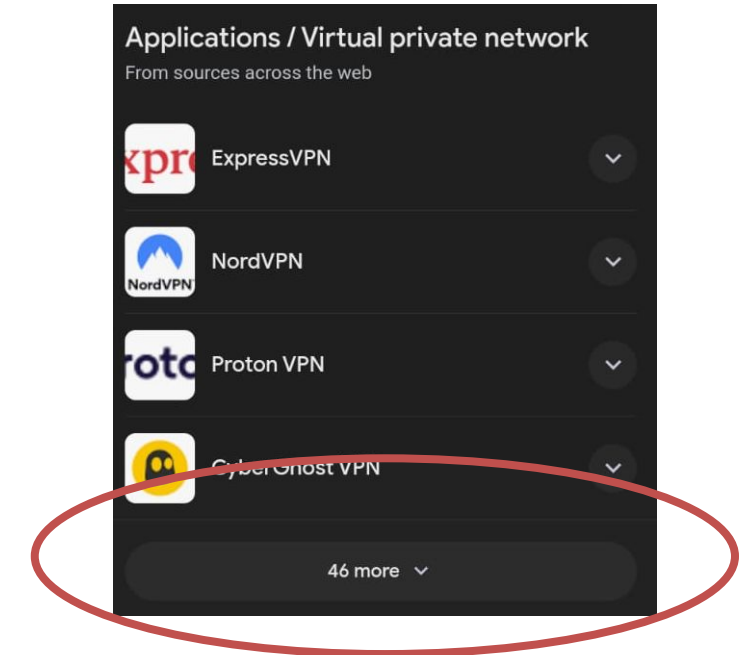
Prior work on VPN awareness and adoption

VPNs for security, accessing restricted content
    Recommendation of online reviews and ratings
    Wrong mental model of users

[Namara et al. PETS'20]

[Sombatruang et al., ICISS'20]

[Dutkowska-Zuk et al. USENIX'22]

[Ramesh et al., USENIX'23]

Prior work on VPN awareness and adoption

VPNs for security, accessing restricted content

  Recommendation of online reviews and ratings

  Wrong mental model of users

Prior work on VPN awareness and adoption

VPNs for security, accessing restricted content

    Recommendation of online reviews and ratings

    Wrong mental model of users

They mostly considered first-time VPN adoption

    Users can **change VPN app** based on requirement

Prior work on VPN awareness and adoption

VPNs for security, accessing restricted content
   Recommendation of online reviews and ratings
   Wrong mental model of users

They mostly considered first-time VPN adoption
   Users can **change VPN app** based on requirement

VPN switching

Prior work on VPN awareness and adoption

VPNs for security, accessing restricted content
    Recommendation of online reviews and ratings
    Wrong mental model of users

They mostly considered first-time VPN adoption
    Users can **change VPN app** based on requirement

VPN switching

No investigation on user requirements that lead to VPN-switching

Our goal

Data-driven **detection and analysis** of the **underlying VPN features** that lead to **VPN-switching** among **mobile VPN app users**

Rest of the talk

How to collect user experiences at scale about VPN-switching?

What is the prevalence and reasons for VPN-switching?

Are desired features of VPNs correlated?

Does information from VPN review websites capture desired features?

Rest of the talk

How to collect user experiences at scale about VPN-switching?

What is the prevalence and reasons for VPN-switching?

Are desired features of VPNs correlated?

Does information from VPN review websites capture desired features?

Detecting VPN-switching experiences in the wild

Earlier work used user surveys: Difficult to scale ❌

Idea: Use publicly posted VPN-app reviews from App store ✓

Detecting VPN-switching experiences in the wild

Earlier work used user surveys: Difficult to scale ❌

Idea: Use publicly posted VPN-app reviews from App store ✓

Challenge 1: How to select the VPN apps?

 Search "best vpn 2023" in app store

 Then randomly sample 20 apps from popularity buckets

Detecting VPN-switching experiences in the wild

Earlier work used user surveys: Difficult to scale ❌

Idea: Use publicly posted VPN-app reviews from App store ✓

Challenge 1: How to select the VPN apps?

    Search "best vpn 2023" in app store

    Then randomly sample 20 apps from popularity buckets

Challenge 2: How to select candidate reviews where users might mention VPN-switching?

Intuition: Reviews where VPN-switch can happen

I left [VPN1] for these guys! [VPN1] was slow.. and i mean slow... tried [VPN2]...and never looked back. love the speed and security!!

Intuition: Reviews where VPN-switch can happen

I left [VPN1] for these guys! [VPN1] was slow.. and i mean slow... tried [VPN2]...and never looked back. love the speed and security!!

Focus on reviews where multiple VPN names occur

Intuition: Reviews where VPN-switch can happen

I left [VPN1] for these guys! [VPN1] was slow.. and i mean slow... tried [VPN2]...and never looked back. love the speed and security!!

Focus on reviews where multiple VPN names occur

I have used this [VPN1] for a Long time.....We usually run ... through it. Then personal devices are on the [VPN2]!.....

Issue: Might not always work

Our data collection and scaling pipeline

Collected 1.3 million
English reviews from
   20 VPN apps
  (Android + iOS)

Our data collection and scaling pipeline

Collected 1.3 million English reviews from 20 VPN apps (Android + iOS)

Manual analysis to identify VPN-switching (with reason) at small scale

Annotating reviews to find VPN-switching

Filtered reviews with > 5 words

Identified reviews with names of multiple VPN apps

Annotated 1600 such reviews
Actual, potential, irrelevant
(inter-rater agreement: 0.74)

Example

I hate I can't have [VPN1] on all the time…I just hated it and I want my money back... Don't install this crap, going for [VPN2]

Actual switch

….My only drawback is the parent companys are subject to U S laws. anything else [VPN1] or [VPN2].

potential switch

Annotating reviews to find VPN-switching

Filtered reviews with > 5 words

Identified reviews with names of multiple VPN apps

Annotated 1600 such reviews Actual, potential, irrelevant (inter-rater agreement: 0.74)

Open coding + affinity diagramming on 1,305 Reviews which indicate actual / potential VPN switching

4-level hierarchy of reasons for switching

Example

[VPN1] was no longer able to support my needs and looking for another solution [VPN2] has come to my rescue and I'm loving the price, I just wish there was more pricing options…

Example

[VPN1] was no longer able to support my needs and looking for another solution [VPN2] has come to my rescue and I'm loving the price, I just wish there was more pricing options…

Level 1 — Financial aspects

Level 2 — Payment

Level 3 — Subscription options

Level 4 — Less flexible plans

Example

[VPN1] was no longer able to support my needs and looking for another solution [VPN2] has come to my rescue and I'm loving the price, I just wish there was more pricing options…

Most abstract

Level 1	Financial aspects
Level 2	Payment
Level 3	Subscription options
Level 4	Less flexible plans

Most specific

Example

[VPN1] was no longer able to support my needs and looking for another solution [VPN2] has come to my rescue and I'm loving the price, I just wish there was more pricing options…

Most abstract

Level 1	Financial aspects
Level 2	Payment
Level 3	Subscription options
Level 4	Less flexible plans

Most specific

These reasons also provide **desired features of VPNs**

Our data collection and scaling pipeline

Collected 1.3 million English reviews from 20 VPN apps (Android + iOS)

Preprocessing and manual coding to identify VPN-switching (with reason) at small scale

Scale up "VPN switching" dataset using Machine learning

Machine learning for identifying VPN-switching

Viewed the problem as two classification tasks

1. Classify a review into "actual", "potential" or "irrelevant"

2. Classify a review into Level-2 reasons

> **80% accuracy for both tasks**

Machine learning for identifying VPN-switching

Viewed the problem as two classification tasks

    1. Classify a review into "actual", "potential" or "irrelevant"

    2. Classify a review into Level-2 reasons

    > 80% accuracy for both tasks

Details, accuracy, code, dataset, models?

Machine learning for identifying VPN-switching

Viewed the problem as two classification tasks

1. Classify a review into "actual", "potential" or "irrelevant"

2. Classify a review into Level-2 reasons

> **80% accuracy for both tasks**

Details, accuracy, code, dataset, models?

Check the repository and artifact

https://github.com/Mainack/switch-vpn-datacode-sec24

Rest of the talk

How to collect user experiences at scale about VPN-switching?

What is the prevalence and reasons for VPN-switching?

Are desirable features of VPNs correlated during VPN-switching?

Does information from VPN review websites capture desired features?

Do users switch VPN apps?

We checked reviews with #words > 5

Do users switch VPN apps?

We checked reviews with #words > 5

	#reviews	% reviews
Reviews with actual switch	90,305	22
Reviews with potential switch	95,094	23.1
Irrelevant	286,352	54.9

Do users switch VPN apps?

We checked reviews with #words > 5

	#reviews	% reviews	
Reviews with actual switch	90,305	22	
Reviews with potential switch	95,094	23.1	
Irrelevant	286,352	54.9	

Close to **45% of reviews** mentioned switching!

185,399 VPN-switching reviews **posted by 169,201 unique usernames**

Why do users switch VPN apps?

Our qualitative analysis identified 4-level hierarchical themes

Eight level – 1 (most abstract) themes and 35 level - 2 themes

Recall: These themes capture **desired features of VPN**

Why do users switch VPN apps?

Our qualitative analysis identified 4-level hierarchical themes
  Eight level – 1 (most abstract) themes and 35 level - 2 themes
  Recall: These themes capture **desired features of VPN**

Network	User Interaction	Connection	Geography
Security & Privacy	Financial Aspects	Dynamic Data Services	Software

Are some desired features more important?

Top reasons
Connection
Network speed
User-friendliness
Service provided
UI/UX

Privacy and Security is NOT in top 5

Are some desired features more important?

Top reasons
Connection
Network speed
User-friendliness
Service provided
UI/UX

Privacy and Security is NOT in top 5

But usability is …

Rest of the talk

How to collect user experiences at scale about VPN-switching?

What is the prevalence and reasons for VPN-switching?

Are desired features of VPNs correlated?

Does information from VPN review websites capture desired features?

Finding correlation of desired features

We created a weighted **desired-feature co-occurance** network

Nodes are features

Weighted edges indicate #reviews where two features co-occur

Communities in the graph capture **correlated set of features**

Are desired features correlated?

There were ten communities of desired features, e.g.,...

Connection, UI/UX, Network speed, Security, Resource consumption, Geography, Service provided, User-friendliness, Not Self Sufficient, Reliability
Network Traffic, Ads, OTT, Privacy
Security, Privacy, Switching server, Geography, Premium
Protocol, Bypassing capability, Network, Updates, Privacy

Are desired features correlated?

There were ten communities of desired features, e.g.,...

Connection, UI/UX, Network speed, **Security,** Resource consumption, Geography, Service provided, User-friendliness, Not Self Sufficient, Reliability
Network Traffic, Ads, OTT, **Privacy**
Security, Privacy, Switching server, Geography, Premium
Protocol, Bypassing capability, Network, Updates, **Privacy**

Security and Privacy is strongly present in set of desired features!

Rest of the talk

How to collect user experiences at scale about VPN-switching?

What is the prevalence and reasons for VPN-switching?

Are desired features of VPNs correlated?

Does information from VPN review websites capture desired features?

Do blogs mention desired features of VPN?

Collected 583 blogs from 6 websites

176 were VPN review blogs

200 were VPN recommendation blogs

Do blogs mention desired features of VPN?

Collected 583 blogs from 6 websites

    176 were VPN review blogs

    200 were VPN recommendation blogs

LDA Topic modelling

Do blogs mention desired features of VPN?

Collected 583 blogs from 6 websites

176 were VPN review blogs

200 were VPN recommendation blogs

LDA Topic modelling

BART based feature classification

Do blogs mention desired features of VPN?

Collected 583 blogs from 6 websites

176 were VPN review blogs

200 were VPN recommendation blogs

LDA Topic modelling

BART based feature classification

Named entity recognition

Do blogs mention desired features of VPN?

Collected 583 blogs from 6 websites

176 were VPN review blogs

200 were VPN recommendation blogs

LDA Topic modelling

BART based text classification

Named entity recognition

Word association analysis

Do blogs mention desired features of VPN?

Collected 583 blogs from 6 websites

176 were VPN review blogs

200 were VPN recommendation blogs

LDA Topic modelling

Named entity recognition

BART based for classification

Word association analysis

NO! **important features** are **not covered**

Significant bias towards specific VPN app names for specific features

Summary

Close to 45% reviews of mobile VPN services indicate VPN switching

VPN switching happens due to set of specific desired features

These features are not captured by VPN review blogs

Summary

Close to 45% reviews of mobile VPN services indicate VPN switching

VPN switching happens due to set of specific desired features

These features are not captured by VPN review blogs

**"I just hated it and I want my money back":
Data-driven Understanding of Mobile VPN
Service Switching Preferences in The Wild**

Rohit Raj, Mridul Newar, Mainack Mondal

https://github.com/Mainack/switch-vpn-data
code-sec24