Predicting Impending Exposure to Malicious Content by Learning User Behavior

Ayumu Kubota, KDDI Research Inc.

KDDI Research

content

- Brief summary of AI/ML related work in KDDI’s cyber security research

- Recent result from user behavior analysis

AI/ML related work in KDDI's cyber security research

- Analysis of various security logs (IDS, firewalls etc.)
  - For supporting cyber security operations
- Spam e-mail detection in mobile network
  - Spam e-mails specifically targeting mobile users are different from other spam e-mails and existing spam filters do not work well
  - By analyzing a large collection of spam e-mails we collect, we have developed and deployed additional spam filters to our service
- Android application inspection for KDDI's app market
  - We have been inspecting all the applications on the KDDI's app market (au Market) before their release
  - Applying AI for improving accuracy and reducing the need for analysis conducted by human experts

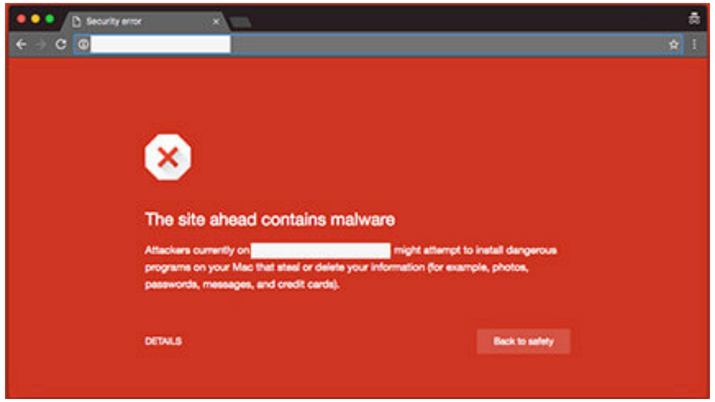
PREDICTING IMPENDING EXPOSURE TO MALICIOUS CONTENT FROM USER BEHAVIOR

Mahmood Sharif*, Jumpei Urakawa†, Nicolas Christin*, Ayumu Kubota†, Akira Yamada†

*Carnegie Mellon University †KDDI Research, Inc

Presented at ACM CCS 2018 (15-19 Oct. 2018, Toronto, Canada)

Traditional defenses are reactive



Blacklists react to prevent users' visits to malicious websites

Anti-viruses react at the time of download or execution of malware

By the time they react, it might be too late

Proactive defenses work over long periods

For example:

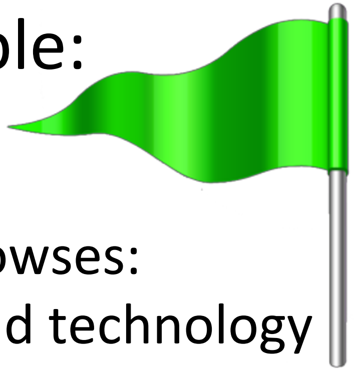
- Forecast whether users will visit malicious websites within **3 months** [Canali et al., AsiaCCS '14]
- Predict whether websites will be compromised within **1 year** [Soska and Christin, USENIX Security '14]

Limited sets of interventions can be taken

Our work

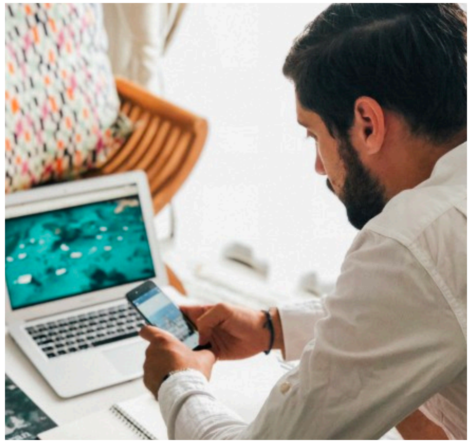
Predict exposure to malicious pages shortly before occurrence (e.g., milliseconds, 5 seconds, 30 seconds) using network traffic

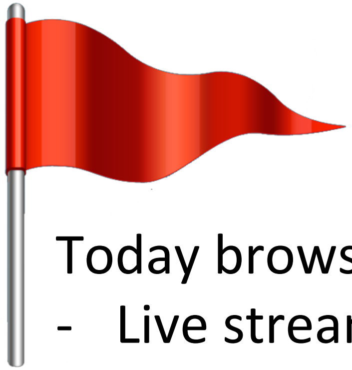
For example:



Usually browses:
- News and technology
- Popular websites
- Avg. 3MB per session

John





Today browses:
- Live streaming and ads
- Unpopular websites
- 30MB in 1 minute

Exposed to a malicious page

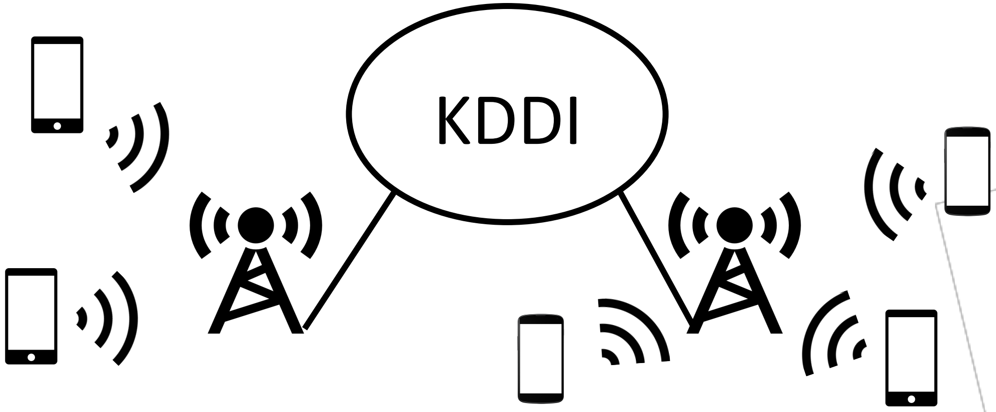
Various interventions can be enabled

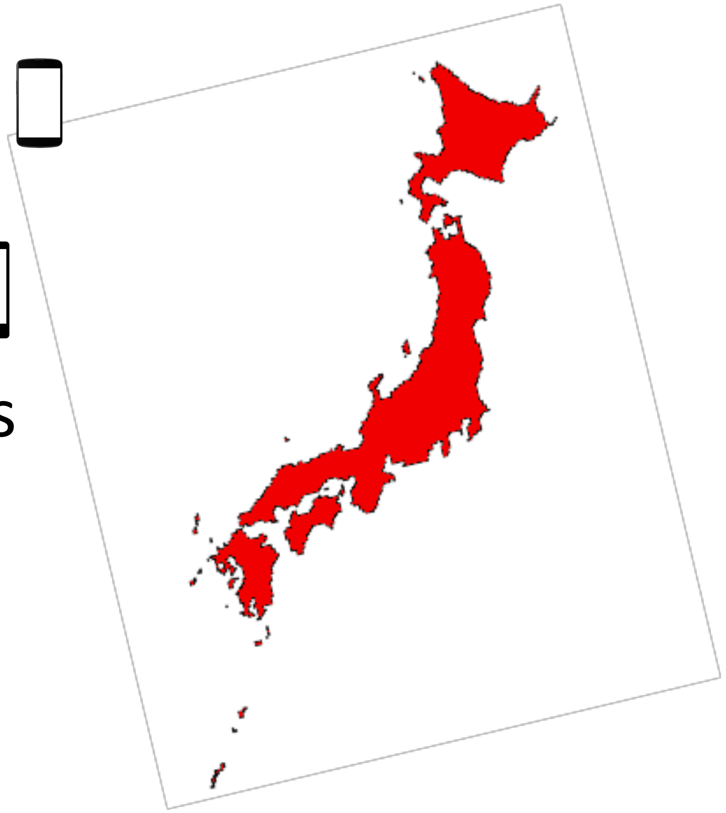
- Alerting users about potential exposure
- Prioritizing traffic for expensive inspection
- Block downloads of 3rd party apps
- Terminate Internet connections

Less aggressive

More aggressive

Our data (1/3): HTTP requests





- HTTP requests (*text/html* only) of 20,645 customers of KDDI
- Fields: consistent user ID, timestamp, URL, # bytes up/down, …
- Spanning 3 months: April to June, 2017
- Collected and used securely with user consent

Our data (2/3): Online survey

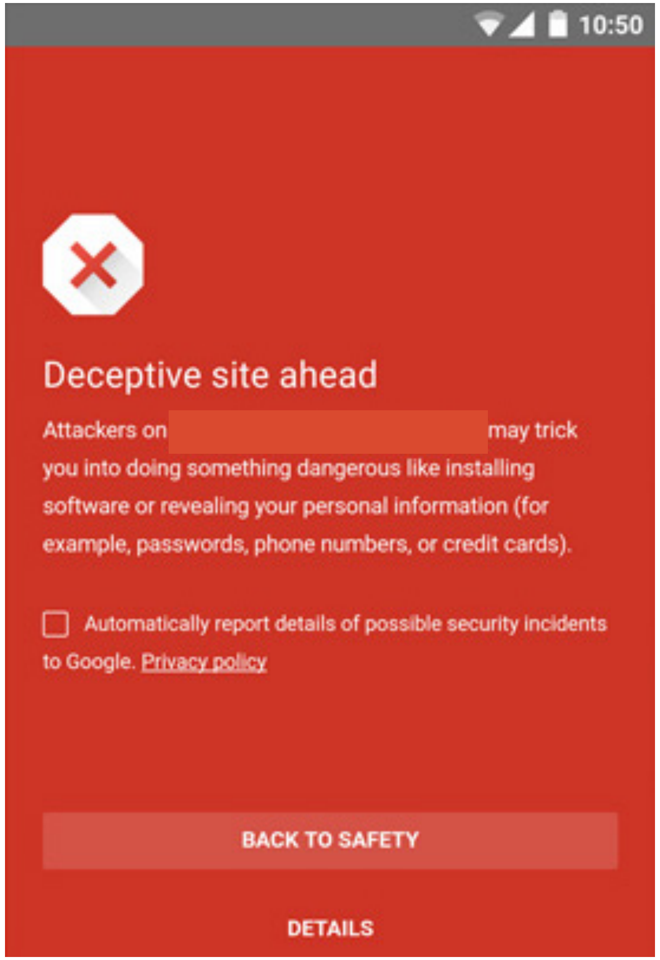
Answered by the 20,645 customers. Asked about:

1. Prior security incidents (e.g., account breaches)
2. Whether the customer runs an anti-virus
3. Types of App marketplaces used (official/unofficial)
4. Whether the customer proceeds on browser warnings
5. Standard security-behavior questions
   (from the Security Behavior Intentions scale*)
6. Self-confidence in security knowledge

* Egelman and Peer. "Scaling the security wall: Developing a security behavior intentions scale." CHI, 2015.

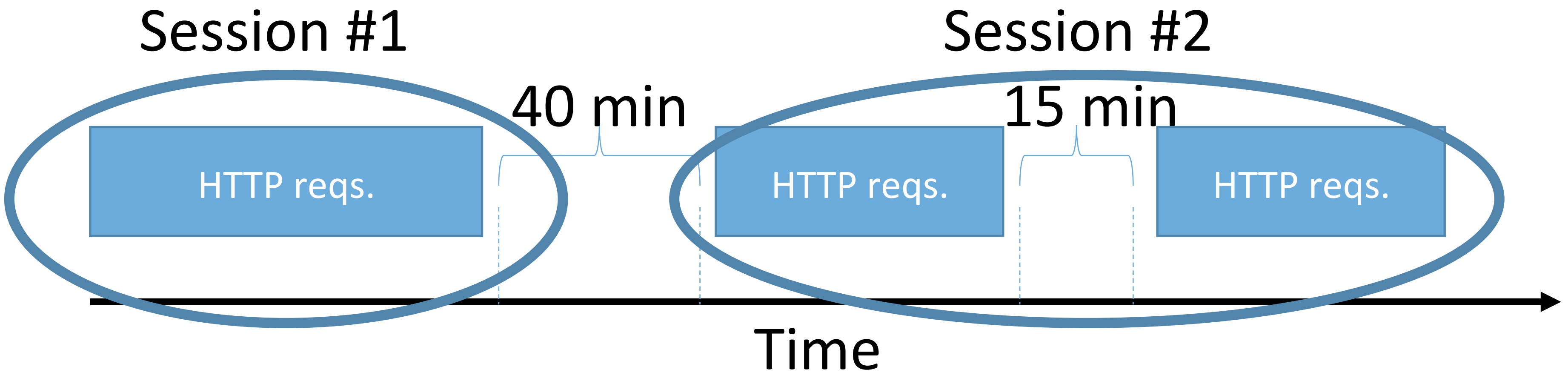
Our data (3/3): Google Safe Browsing (GSB)

- The most deployed blacklist (used by the major browsers)
- We collected daily snapshots
- Used to detect users' accesses to malicious pages



Processing data into sessions

Session: set of contiguous requests made by the same user, which terminates when the user is idle for more ≥20 minutes*



This work: From early observations in the session, predict whether the user will get exposed to malicious pages later in the session

* Wang, Gang, et al. "You Are How You Click: Clickstream Analysis for Sybil Detection." USENIX Security, 2013.

Next

1. Window of exposure to malicious pages
2. Behavioral differences between exposed and unexposed users
3. Short-term prediction: methodology and evaluation

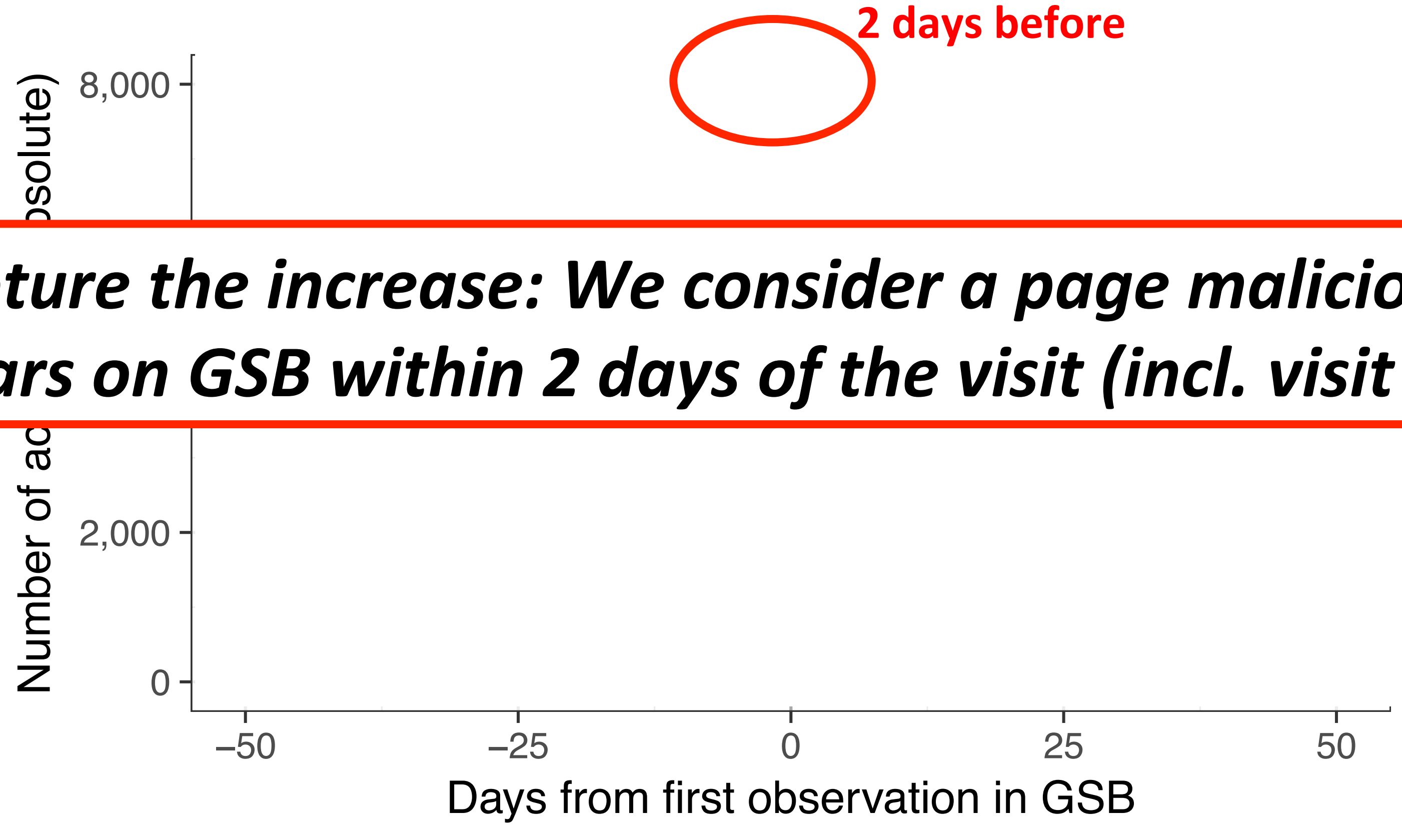
User exposure

- About one session per 1,000 sessions is exposed
- 2,172 users (~11%) exposed to pages on GSB

The blacklisting approach used by major browsers is not enough!

Window of exposure

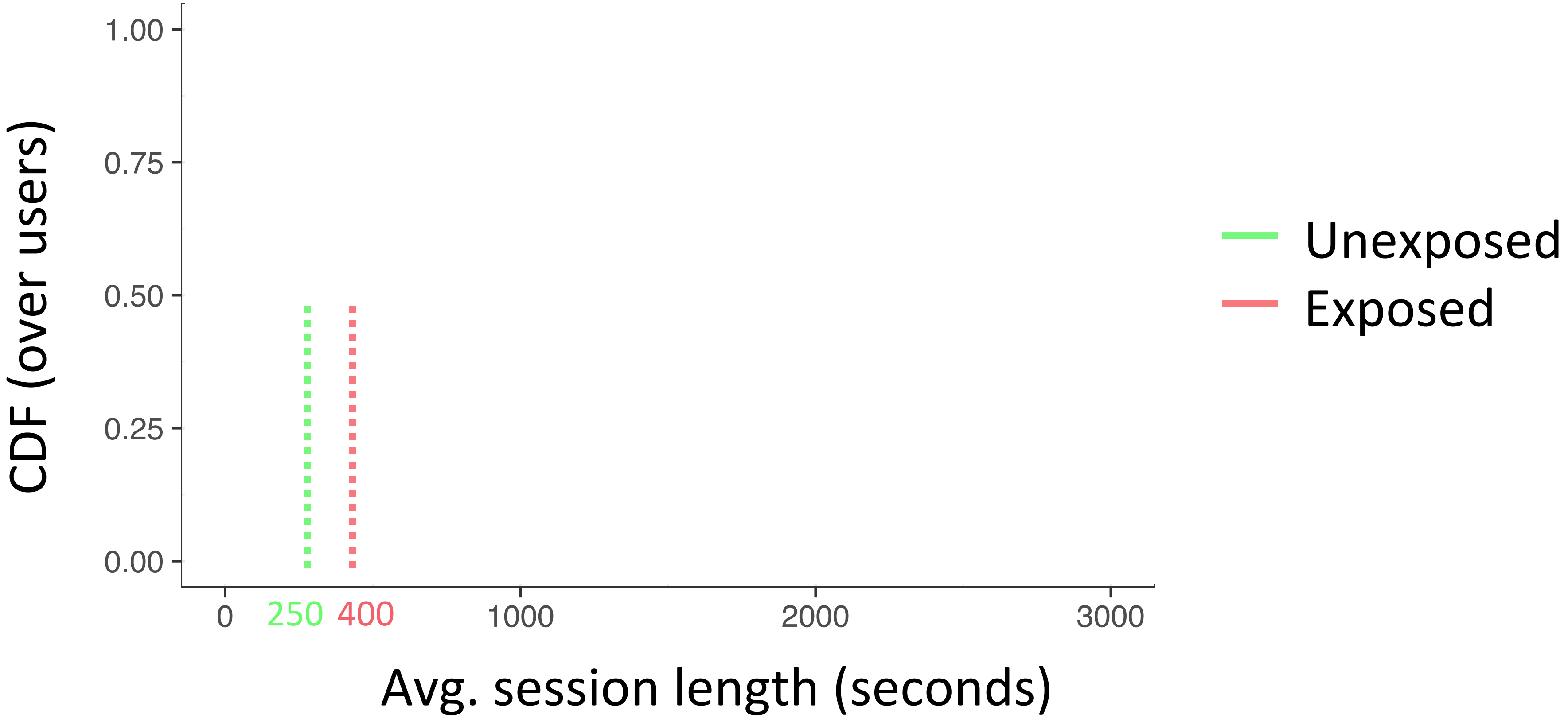
Visits to malicious pages increase days before they appear on GSB



To capture the increase: We consider a page malicious if it appears on GSB within 2 days of the visit (incl. visit time)

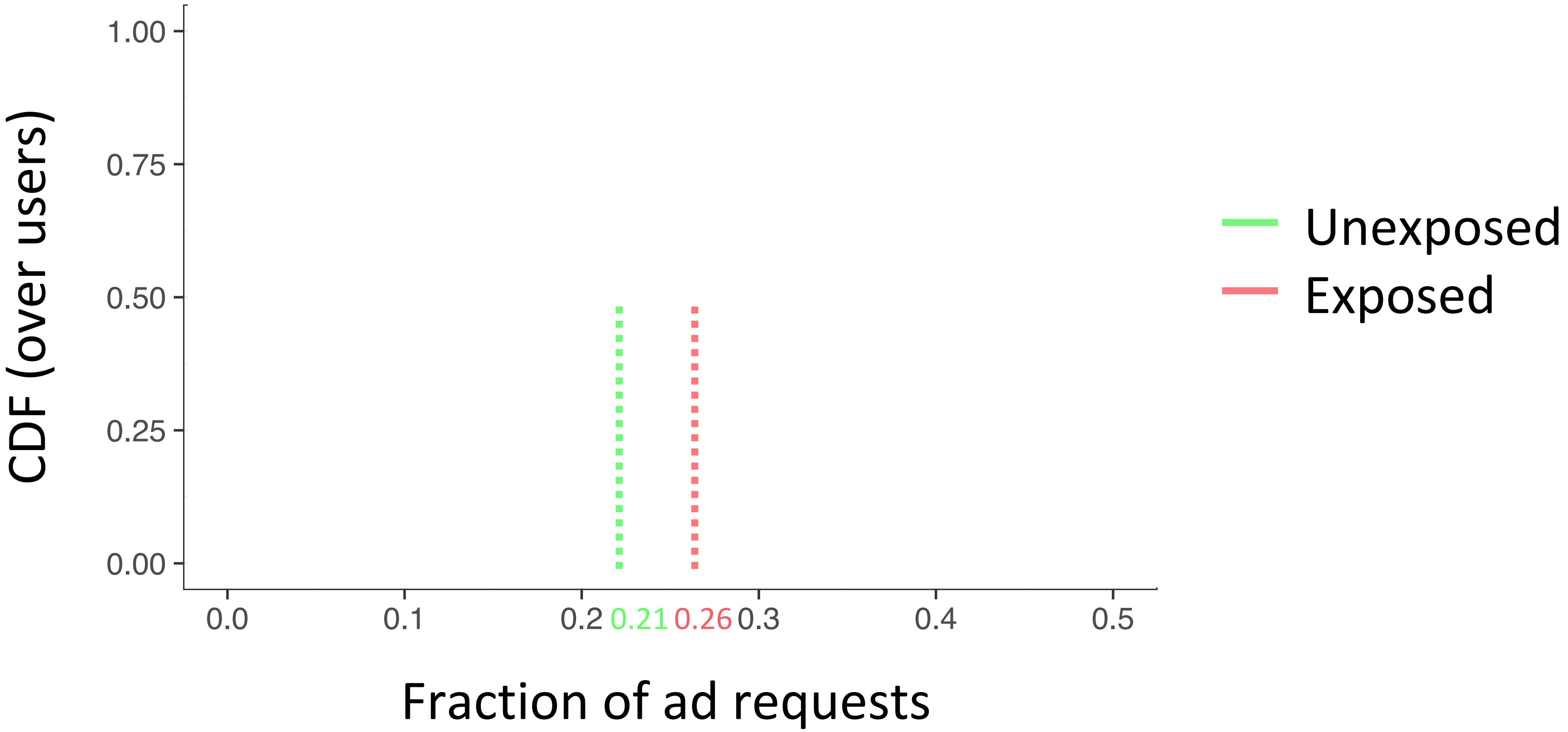
Behavioral differences between users (1/3)

Exposed users browse the web more than unexposed users



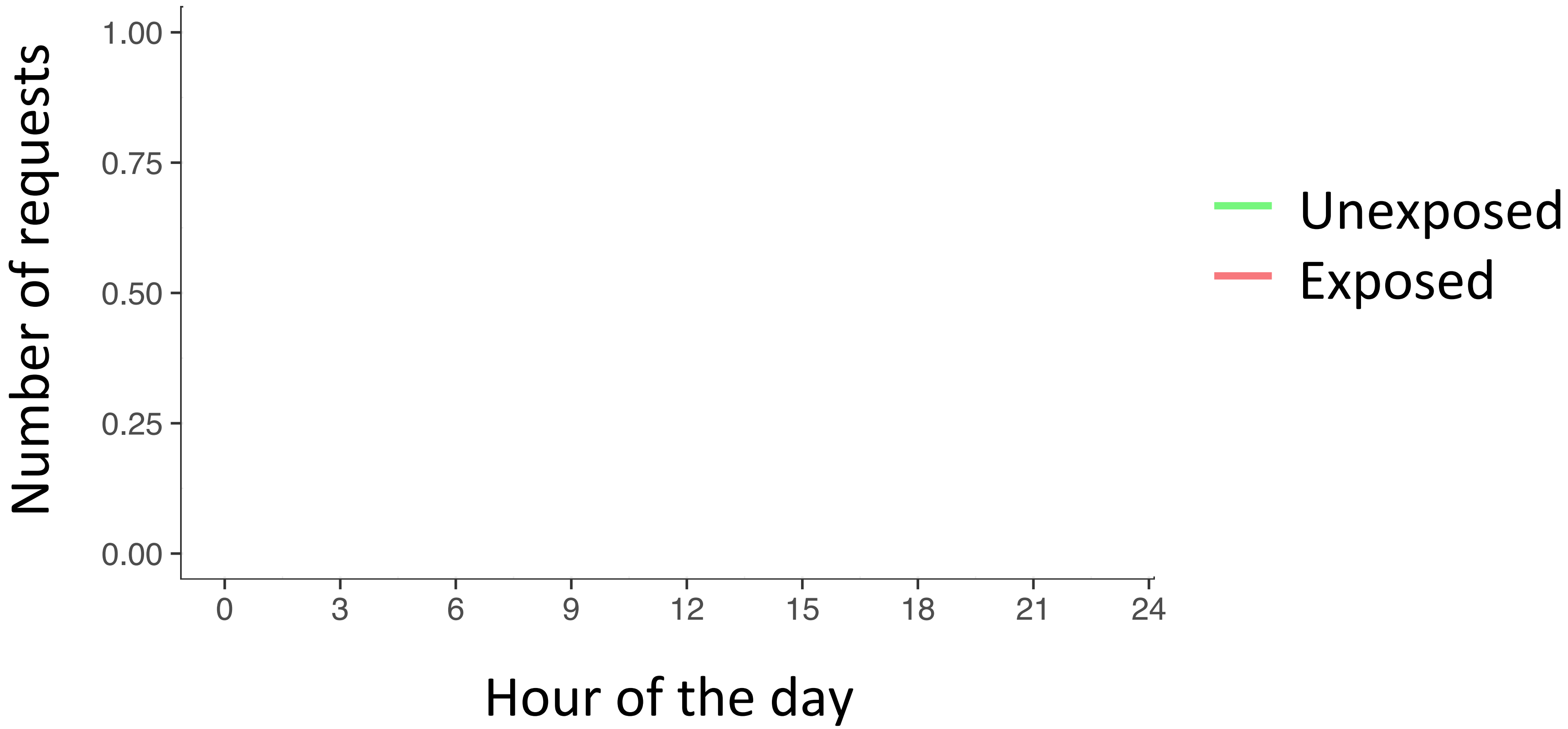
Behavioral differences between users (2/3)

Exposed users request pages of certain topics at different rates than unexposed users (e.g., they request more ads)



Behavioral differences between users (3/3)

Exposed users browse the Internet more frequently at night and outside of working hours



Survey responses and exposure

- Built a logistic regression model to understand correlation
- Dependent variable: whether the user gets exposed
- Independent variables: survey responses
- Some results:
  - Men are ~1.9 times more likely than women to get exposed
  - Users who run anti-virus are ~2.5 times more likely to get exposed

But, model explains only 5% of variance in data.
I.e., self-reported data may not be sufficient on its own.

Exposure prediction: Methodology (1/2)

Based on findings, we developed 3 types of features for prediction:

Contextual
(Updated *during* session)

- # requests
- Session length
- Distribution of topics
- Time of day/week
- …

Past behavior
(Updated *after* session)

- Avg. # requests per session
- Avg. session length
- Past exposures?
- …

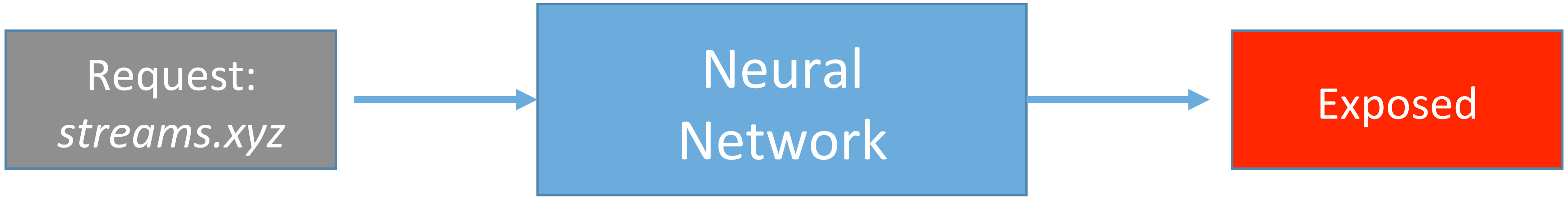
Self reported
(Collected via survey)

- Runs anti-virus?
- Prior security incidents?
- …

Exposure prediction: Methodology (2/2)

- Train neural networks to predict exposure after each request
- Session is exposed if neural network predicts exposure after a request
- E.g., user browses:

reddit.com → *streams.xyz* → *malicious.com*

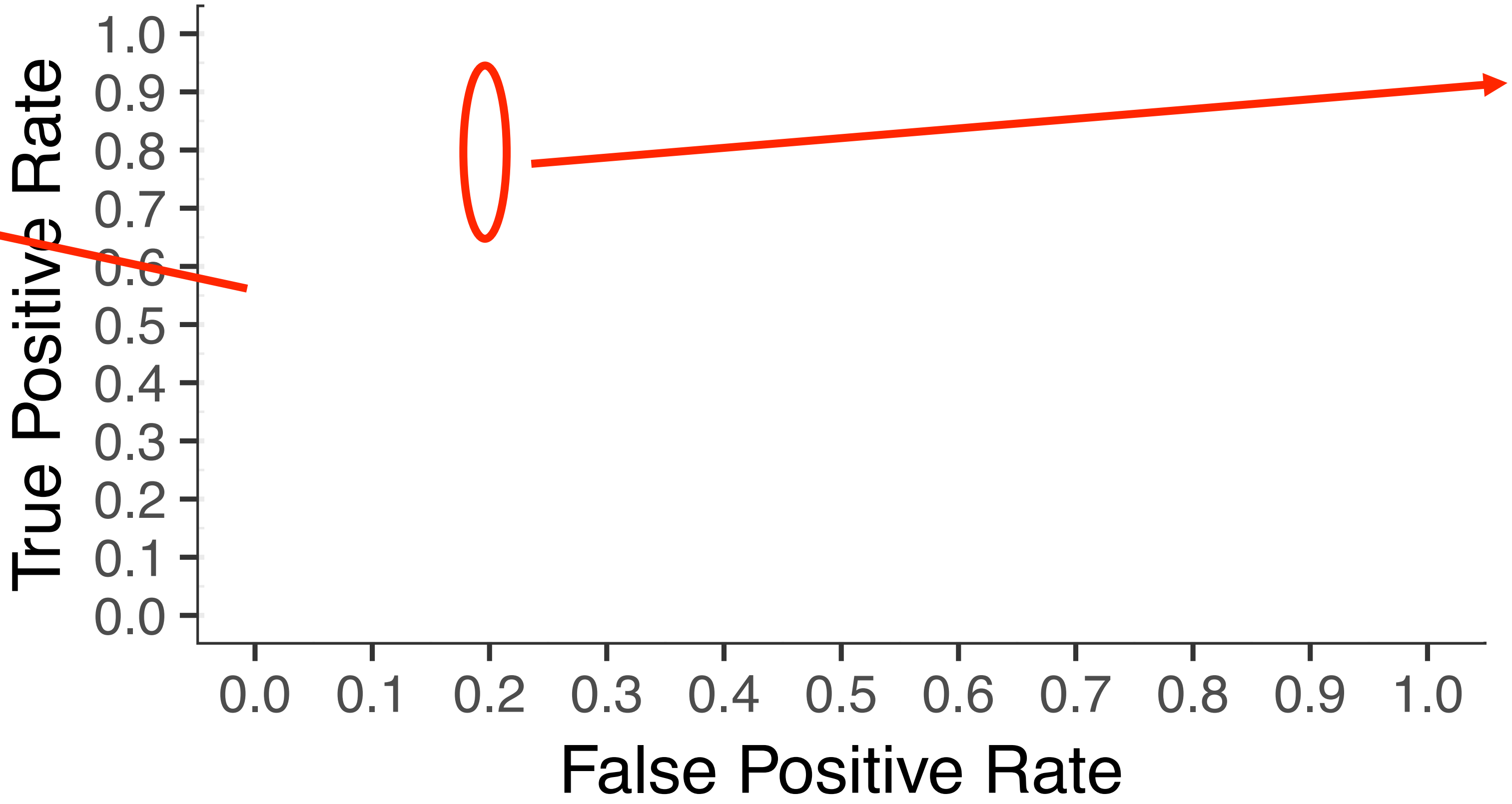


- Evaluate using five 20-day periods: 15 days to train, 5 to test

Exposure prediction: Results (1/2)

Accurate within-session exposure prediction is possible

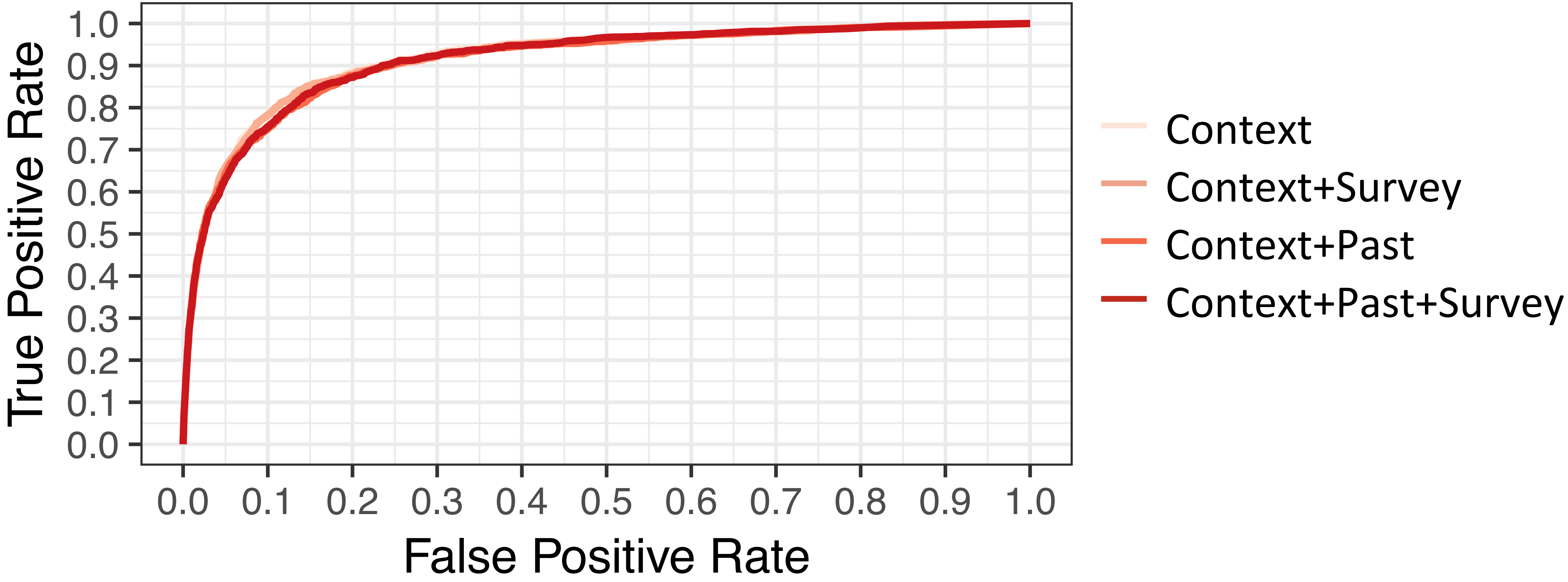
Less false-alarms:
56% TPR at
3% FPR



≥74% TPR
at 20% FPR

Short-term prediction: Results (2/2)

Contextual features are sufficient to predict exposure



Base-rate effect (1/2)

Exposure rate is low (~1000 unexposed sessions per exposed session)

$\Downarrow$

Potentially high number of false detections

For example, at 56% TPR and 3% FPR:

56 true detections and ~3000 false detections per 100K sessions

Is the system not useful?

Base-rate effect (2/2)

In reality, most of the false detections may be true detections

Checking against VirusTotal's (more inclusive) blacklists, we found:

- Exposure rate: 24 exposed sessions per 976 unexposed
- TPR=56% FPR=3% corresponds to TPR=96% FPR=1%

⇒ The system was actually achieving 2,186 true detections and 870 false detections per 100K sessions

Wrap up

- Proposed short-term prediction to enable proactive defenses
- Explored the behavioral differences between unexposed and exposed users to devise useful features
- Showed that short-term prediction can be done accurately

PREDICTING IMPENDING EXPOSURE TO MALICIOUS CONTENT FROM USER BEHAVIOR

Mahmood Sharif, Jumpei Urakawa, Nicolas Christin, Ayumu Kubota, Akira Yamada

E-mail: mahmoods@cmu.edu

Carnegie Mellon University

KDDI Research

Defining malice

τ**-malicious page:** a page visited at time t is τ-malicious ($\tau \geq 0$) if it appears on GSB within τ days from the visit (i.e., before $t+\tau$)

- τ=0: page has to be on GSB to be considered malicious
- We set $\tau=2$ to capture the spike
- Larger τ leads to higher coverage, but might decrease soundness

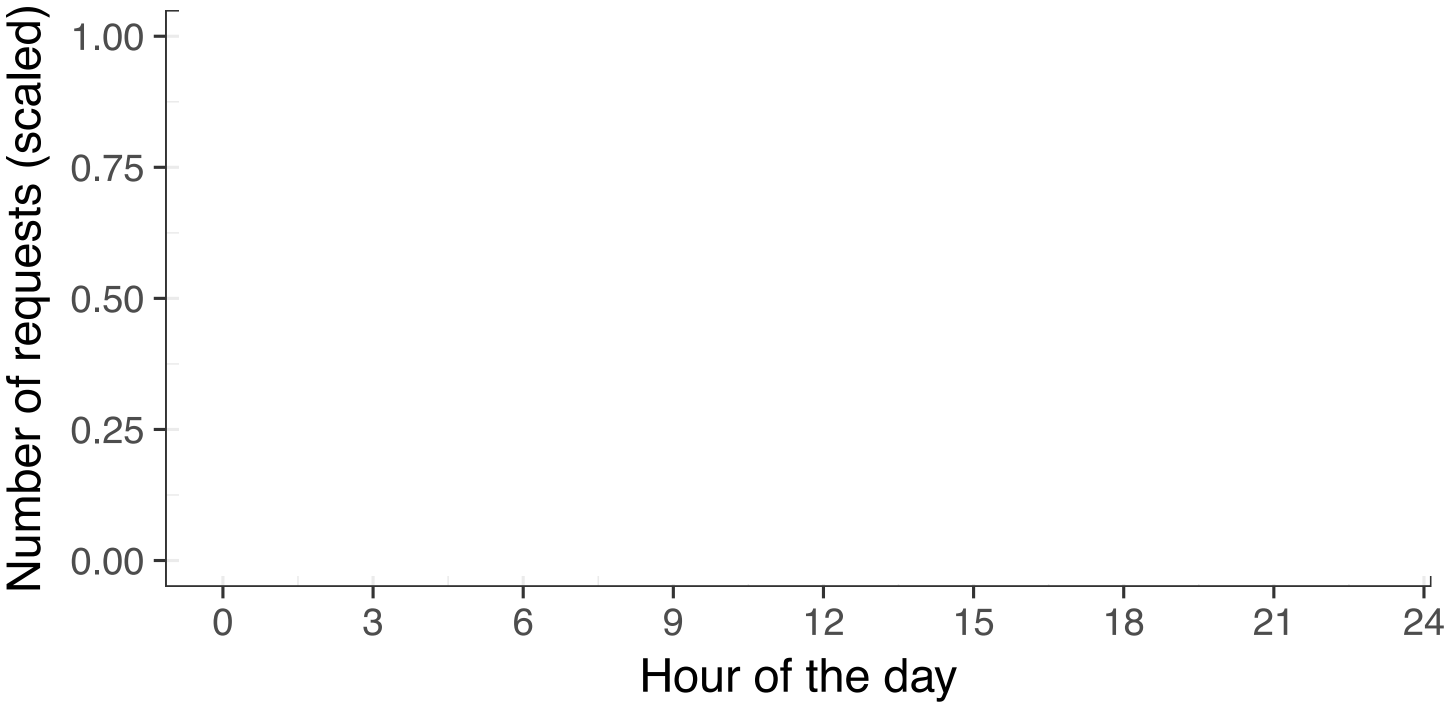
Limitation

Incomplete picture of users' browsing behavior:

- Only *text/html* content (no scripts, images, …)
- No HTTPS traffic
- No Wi-Fi traffic

Behavioral differences between users (2)

Exposed users tend to browse the web more frequently at night and outside of working hours



Survey responses and exposure

- Used logistic regression to understand correlation
- Dependent variable: user exposure
- Independent variables: survey responses

Variable	Odds	p-value
Is female?	0.54	<0.01
RSeBIS score	0.82	<0.01
Proceeds on warning?	1.26	<0.01
Suffered from compromised	1.67	<0.01
Uses anti-virus?	2.51	<0.01
Uses unofficial App market?	1.17	<0.01

RSeBIS scale is a good predictor of user exposure

Users who report to have anti-virus are more likely to get exposed!

But, model explains only 5% of variance in data. I.e., self-reported data may not be sufficient on its own.

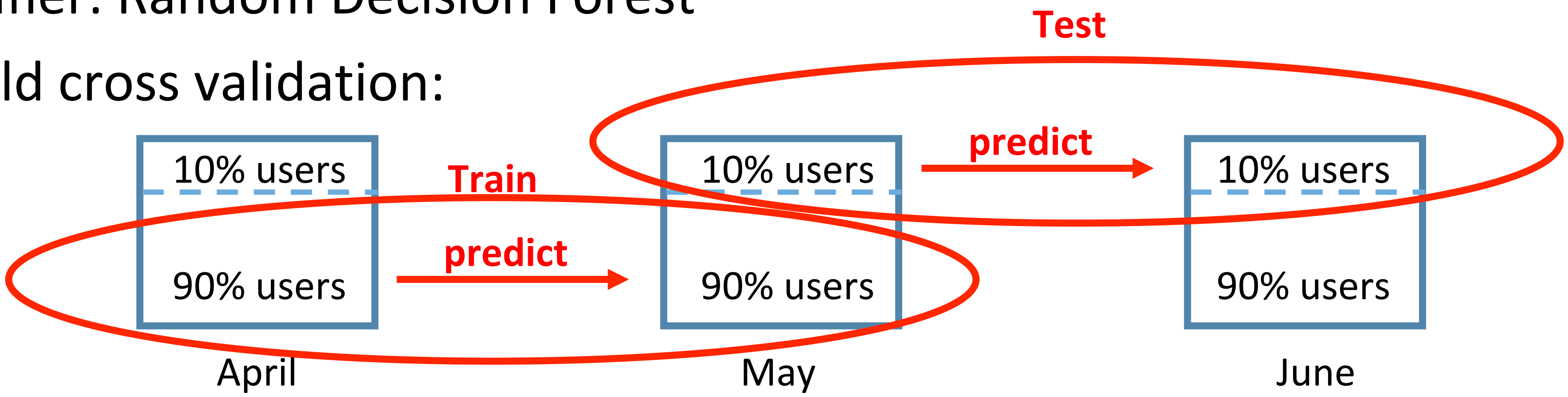
Long-term prediction: Methodology

- Rely on two sets of features: Past behavior (*P*) and Self reported (*S*)

Past behavior features: motivated by behavioral differences, efficiently computable

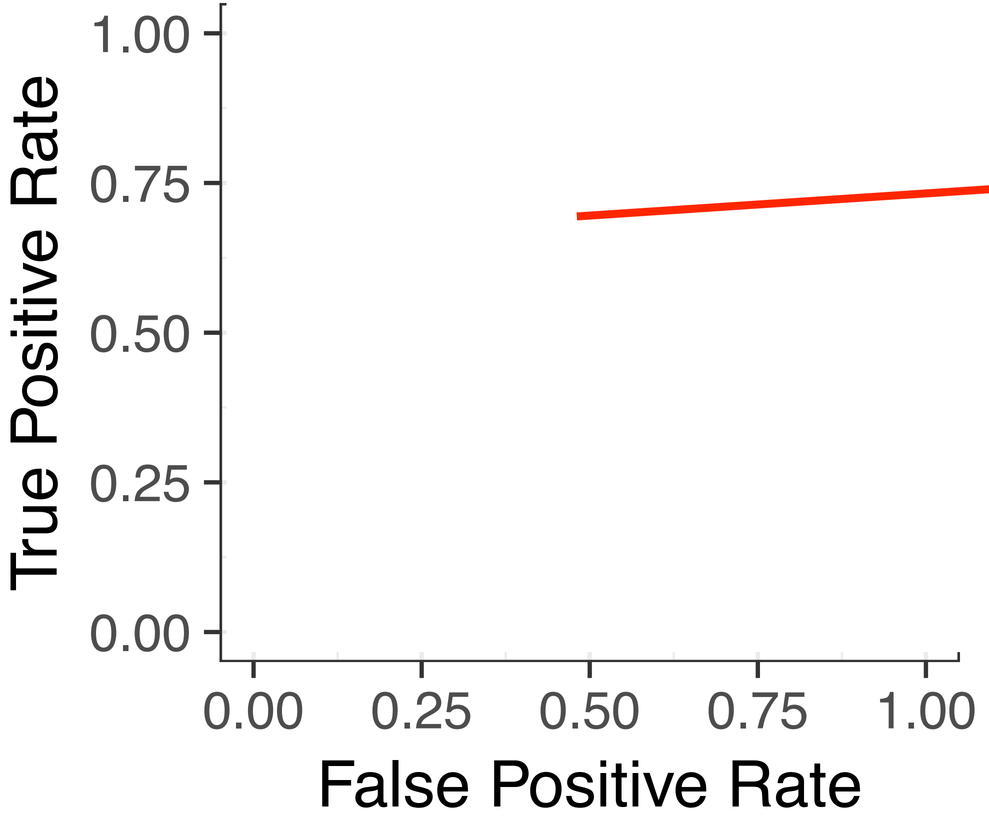
- Avg. # daily sessions and requests
- Prior exposure?
- Fraction of top Alexa websites
- Activity in different times of day/week
- Distribution of URL topics
- …

- Classifier: Random Decision Forest
- 10-fold cross validation:



Long-term prediction: Results

Comparable to prior work [Canali et al., ’14], while less intrusive and using more efficiently computable features (e.g., require no history)



Self-reported data is insufficient on its own

Behavioral differences between users (3/3)

Exposed users browse the Internet more frequently at night and outside of working hours