Context Profiling in Mobile Devices

Aditi Gupta (Purdue University)
Markus Miettinen (Nokia Research Center)
Marcin Nagy (Aalto University)
N. Asokan (Nokia Research Center)

NOKIA
Connecting People

Automobile safety: early days

- UK Locomotives and Highways Act (1865) to assure safe driving
  - Man with a red flag or lantern 55 m in front of the car to warn
  - Max. speed in towns: 3 km/h
- Revised in 1878
  - Red flag man only 18 m in front
  - Widely ignored
- Repealed in 1896

Sources:
http://www.scienceandsociety.co.uk/results.asp?image=10326966&wwwflag=2&imagepos=4
http://en.wikipedia.org/wiki/Locomotives_and_Highways_Act

Automobile safety today

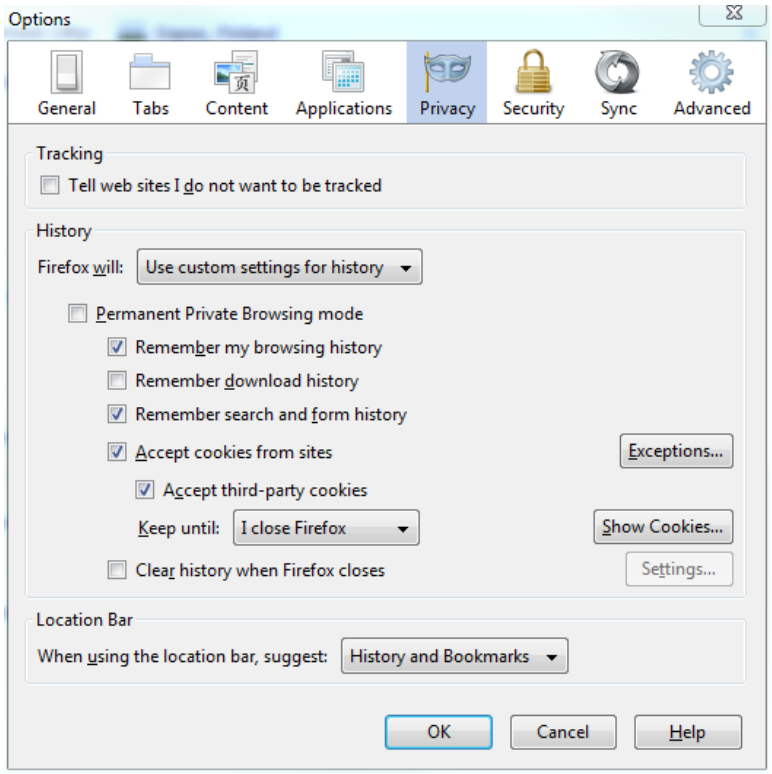
- The human is still in control
- Not just better “user interaction”
- But several underlying new technologies are in use
  - Traffic lights
  - Air bags
  - Anti-lock breaks

"People are still doing dumb things. But the fact is, the cars are now much safer and are more likely to save them. A crash that might have killed you 20 years ago is probably very survivable now."

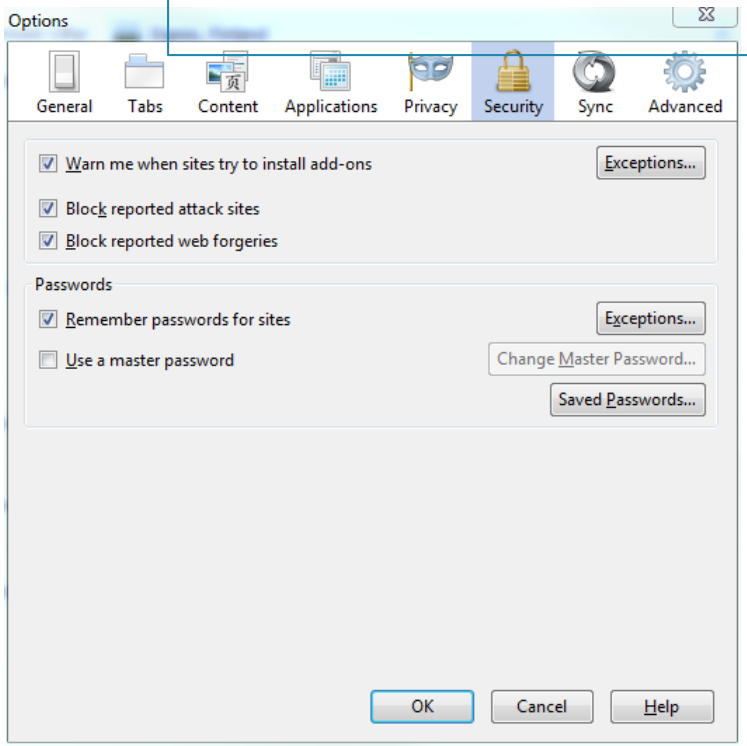
Sources:
http://research.cars.com/go/advice/Story.jsp?section=safe&subject=safe_tech&story=techIntro
http://research.cars.com/go/advice/Story.jsp?section=safe&subject=safe_tech&story=tech0ther&referer=advice&aff=national

Security policies for masses: early days

For ordinary users, on mass-market devices



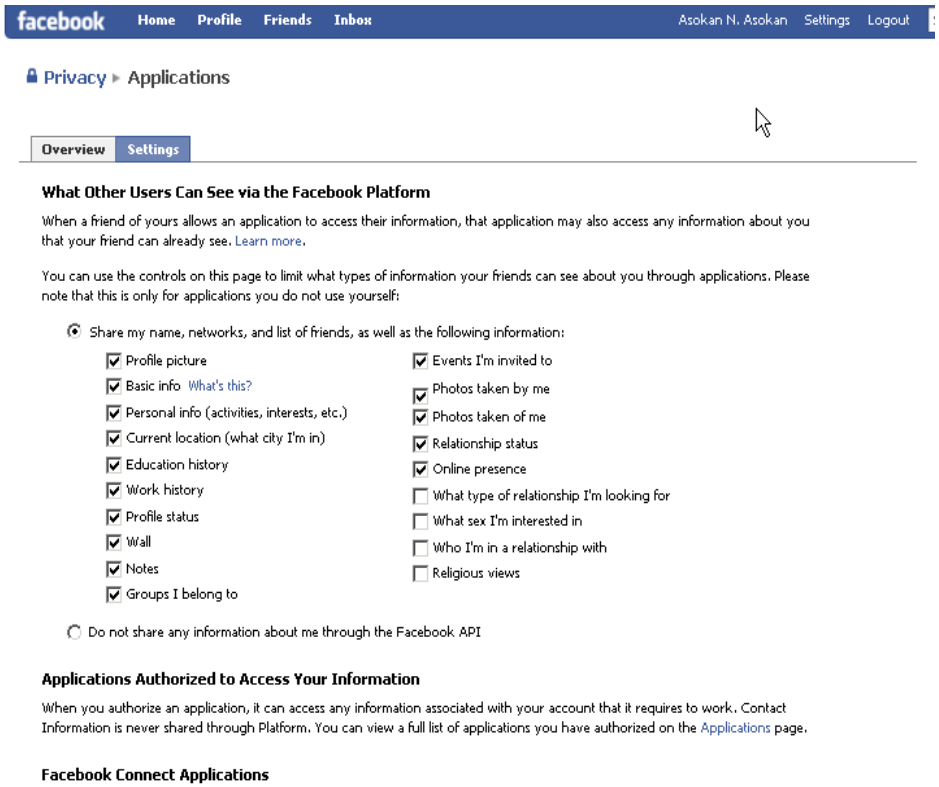
Firefox security/privacy settings Today!



Policy-by-drudgery: set precise and detailed policies manually

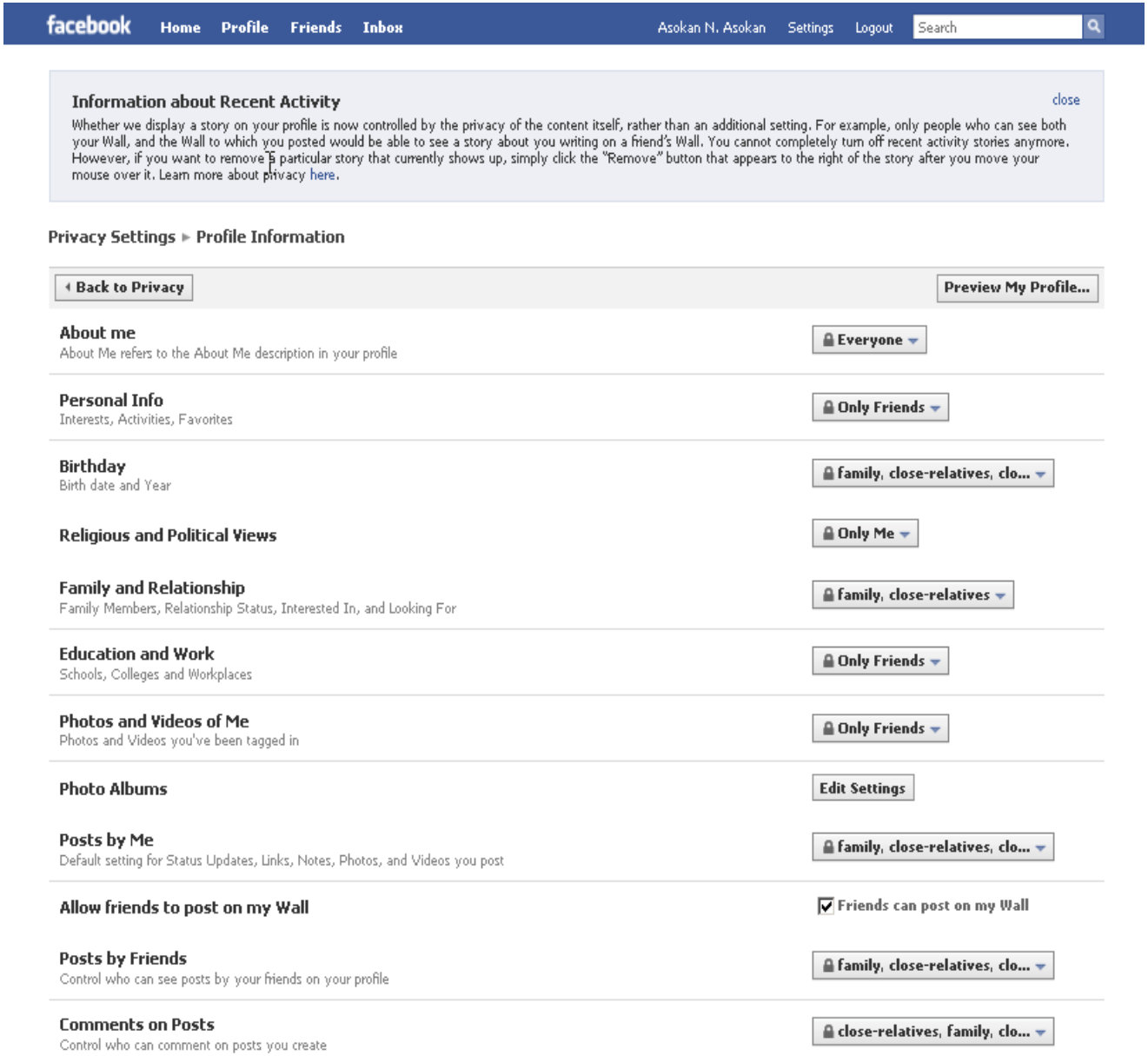
Security policies for masses: early days

For ordinary users, on mass-market devices



Facebook privacy settings Circa 2010

Policy-by-drudgery: set precise and detailed policies manually



Facebook privacy settings Circa 2010

Policy-by-drudgery: set precise and detailed policies manually

Default policies are not always right

One-size does not always fit all

Policy-by-fiat: No choice - defaults specified by developer/administrator

Current state of access control policies

Today the choice for ordinary users is *between* "sensible" and "intuitive"

"sensible" = personalized, appropriate
"intuitive" = easy to use

Problem:
How can ordinary users set
and manage access control policies?

Objective:
Intuitive means to set/manage
sensible access control policies

Idea:
use *context* and *history* to infer sensible policies

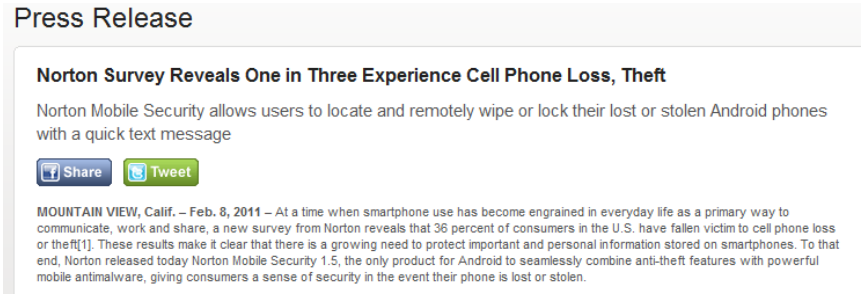
How do users set access control policies?

Eventually…
Policy-by-inference: trusted assistant

Tomorrow
Policy-by-imitation: "do what he/she does"
(E.g., see "Privacy Suites" by Bonneau et al, SOUPS 2009)

Today
Policy-by-fiat: developer/administrator-set defaults
Policy-by-drudgery: user suffers through fine-grained policies

An example: Device Lock



Press Release

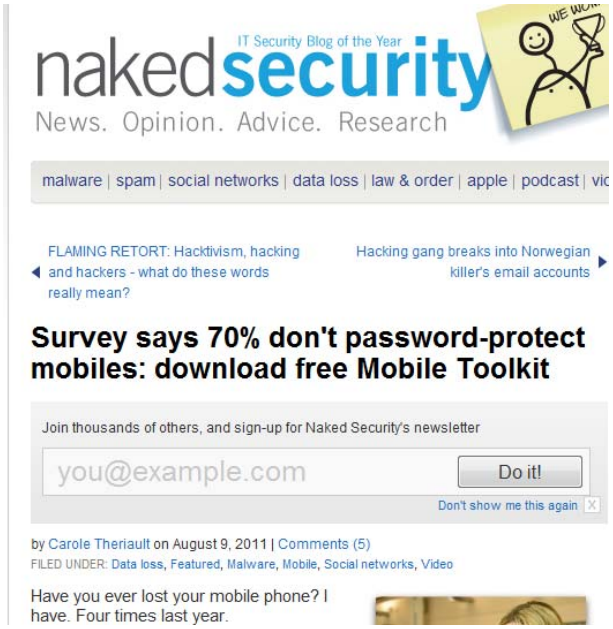
Norton Survey Reveals One in Three Experience Cell Phone Loss, Theft

Norton Mobile Security allows users to locate and remotely wipe or lock their lost or stolen Android phones with a quick text message

Share Tweet

MOUNTAIN VIEW, Calif. – Feb. 8, 2011 – At a time when smartphone use has become engrained in everyday life as a primary way to communicate, work and share, a new survey from Norton reveals that 36 percent of consumers in the U.S. have fallen victim to cell phone loss or theft[1]. These results make it clear that there is a growing need to protect important and personal information stored on smartphones. To that end, Norton released today Norton Mobile Security 1.5, the only product for Android to seamlessly combine anti-theft features with powerful mobile antimalware, giving consumers a sense of security in the event their phone is lost or stolen.

http://www.symantec.com/about/news/release/article.jsp?prid=20110208_01



nakedsecurity IT Security Blog of the Year

News. Opinion. Advice. Research

malware | spam | social networks | data loss | law & order | apple | podcast | vic

FLAMING RETORT: Hacktivism, hacking and hackers - what do these words really mean?

Hacking gang breaks into Norwegian killer's email accounts

Survey says 70% don't password-protect mobiles: download free Mobile Toolkit

Join thousands of others, and sign-up for Naked Security's newsletter

you@example.com Do it!

Don't show me this again

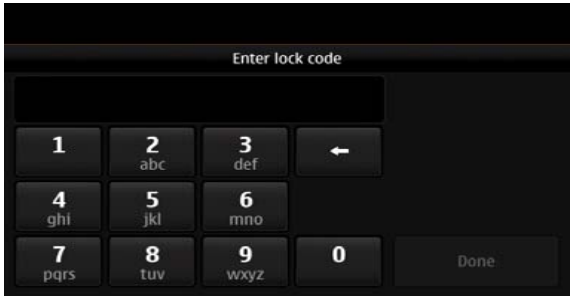
by Carole Theriault on August 9, 2011 | Comments (5)

FILED UNDER: Data loss, Featured, Malware, Mobile, Social networks, Video

Have you ever lost your mobile phone? I have. Four times last year.

http://nakedsecurity.sophos.com/2011/08/09/free-sophos-mobile-security-toolkit/

- Intended for theft protection
- Example of one-size-fits-all
  - device lock always kicks in
- Can be annoying in
  - freezing weather
  - groggy mornings
  - …

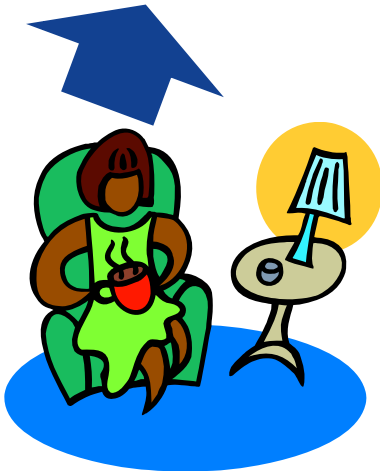


NOKIA

Device lock: desired user experience

Timeout and unlocking method adjusted based on context

Long timeout



Home

Medium timeout

Work Cafeteria

Short timeout

Unknown

NOKIA

How to estimate safety of a place?

Identify places of interest and profile them over time

A place may not be always safe (or unsafe)

1. Identify places (generally ”contexts”) of interest: CoIs
2. Profile CoIs by keeping track of what is seen there
3. Estimate **familiarity of a device** in a CoI
4. Estimate **familiarity of CoI** based on devices present
5. Estimate **safety** based on current/historical familiarity

Context Profiling: the intuition

Find and profile *contexts of interest (CoI)*

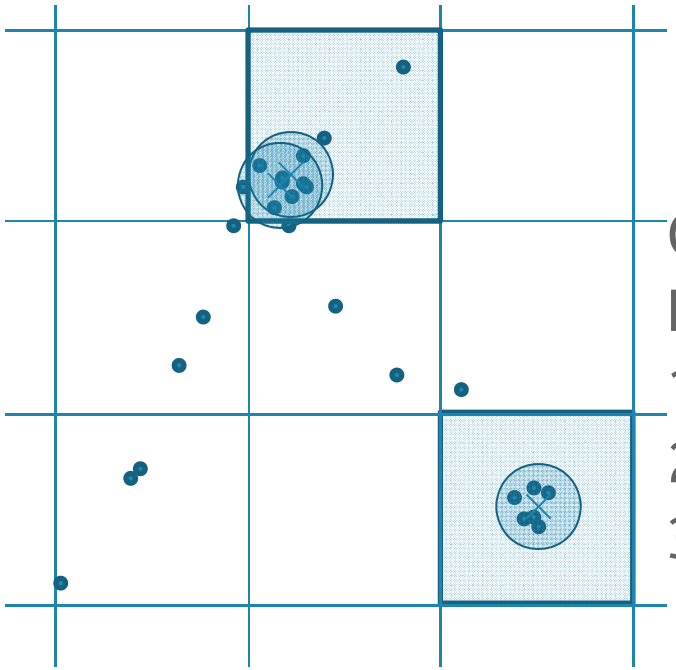
- For now context = place
- Finding places of interest is easy
- Profile a CoI by keeping track of what is seen in that CoI
  - Now - radio environment: Bluetooth devices, WiFi APs
  - Later – ambient noise, light, …

Data collection

Observe current context

- Periodically scan the environment (every 5 minutes)
- In each observation, record
  - GPS co-ordinates
  - Bluetooth devices
  - WiFi access points
- When there is no GPS fix, attempt to guess (more later)

Contexts of Interest (CoI) detection



Grid-based clustering

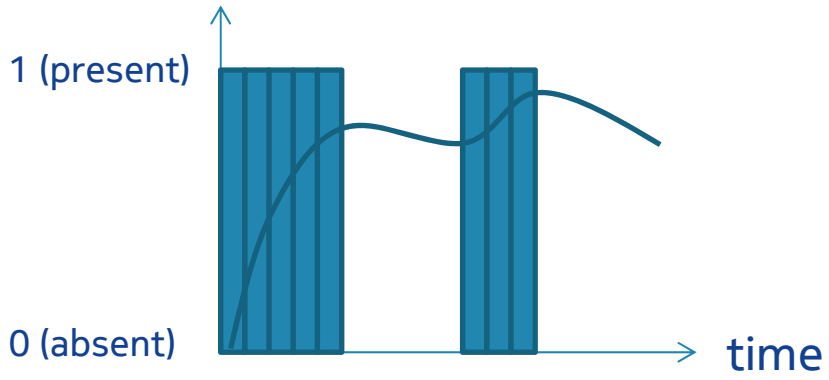
Count observations in grid cells (d = 250m)

Periodically (every 6 hours)

1. identify cells that cross count threshold
2. compute cluster centroid for cell
3. observations within distance r of centroid belong to CoI (r = 100m)
4. recompute cluster centroid of CoI
5. repeat steps 3 & 4 until stable (max 10 iterations)

From now on update CoI centroid whenever a new observation falls with distance r

Profiling CoIs: Device Familiarity



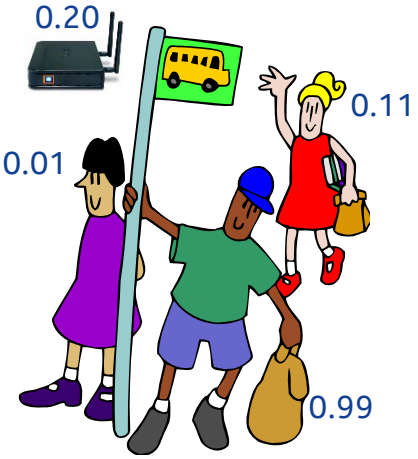
$$devFam(d, C, n) = \alpha_D \times occ(d, C, n) + (1 - \alpha_D) \times devFam(d, C, n-1)$$

Aggressive growth, but conservative decay - penalize only if absence is prolonged

$$occ(d, C, n) = \begin{cases} 1, \text{if } d \text{ seen in } n^{th} \text{ sample in } C \\ 0, \text{if } d \textbf{ not} \text{ seen in sample in } C \text{ and } n - N_{last} \bmod N_0 = 0 \\ devFam(d, C, n-1) \text{ otherwise} \end{cases}$$

where N_{last} is the last observation in which d was seen in in C
N_0 is the length of absence after which *devFam* is penalized

Profiling CoIs: Context Familiarity



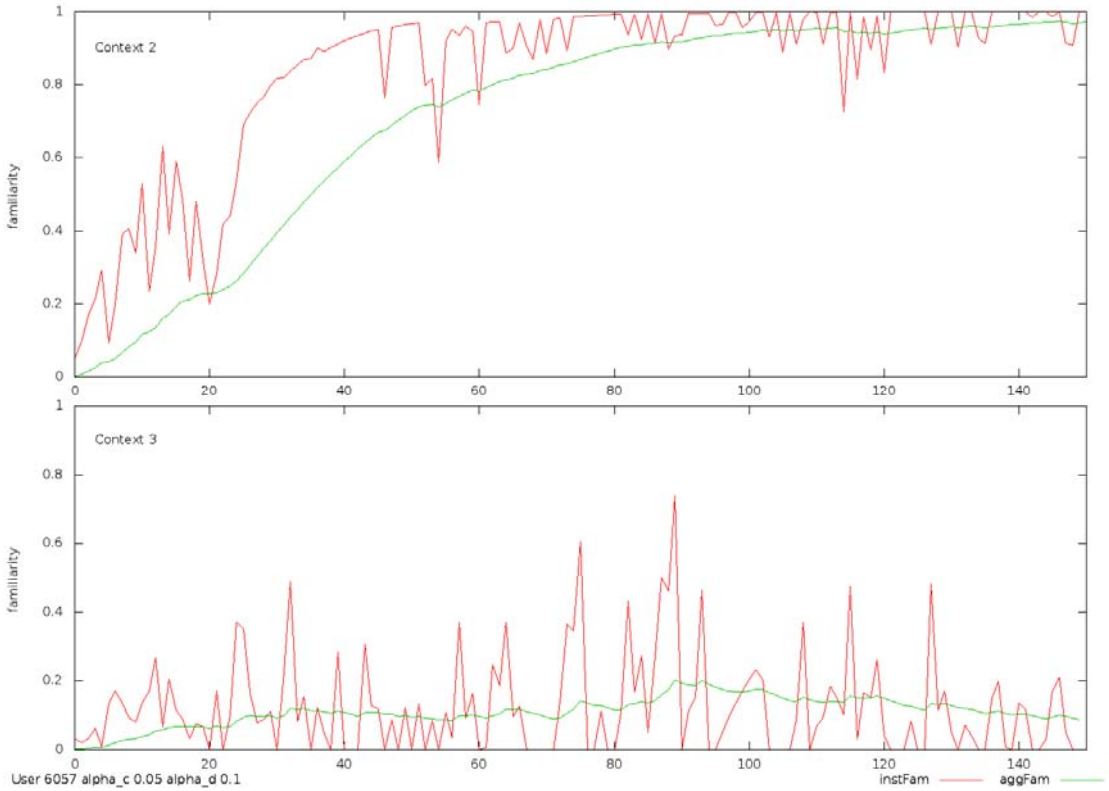
Instant familiarity of a context w.r.t to a type of devices:

$$instFam_t(C,n) = \frac{1}{|D_{C,t,n}|} \sum_{d \in D_{C,t,n}} DFam(d,C,n)$$

Instant familiarity of the context:

$$instFam(C,n) = \frac{1}{t} \sum_{i=1}^{t} instFam_t(C,n)$$

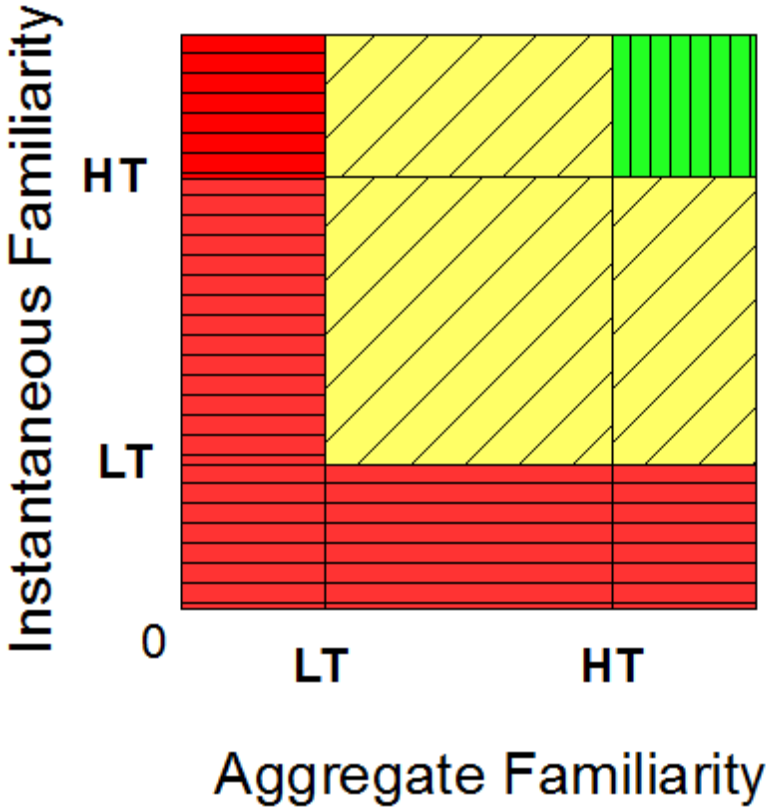
Profiling Cols: Aggregate Familiarity



Smoothing instant familiarity scores of a context to get aggregate familiarity:

$$aggFam(C,n) = \alpha_C \times instFam(C,n) + (1 - \alpha_C) \times aggFam(C,n-1)$$

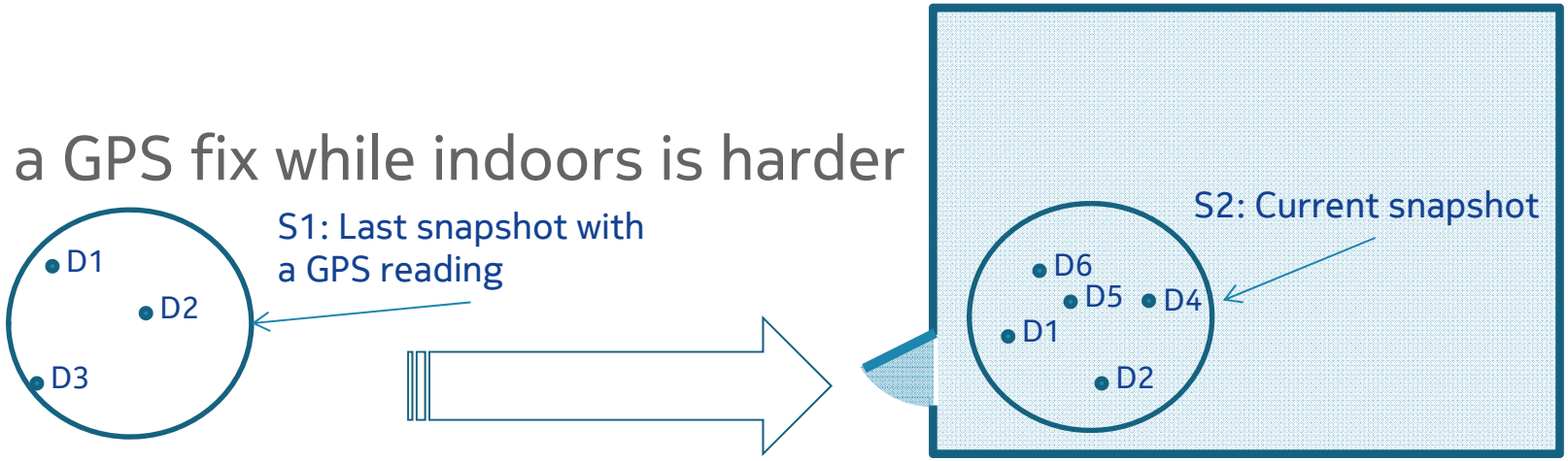
From Familiarity to Safety



LT: Low threshold for familiarity
HT: High threshold for familiarity

Guessing Geolocation

Getting a GPS fix while indoors is harder



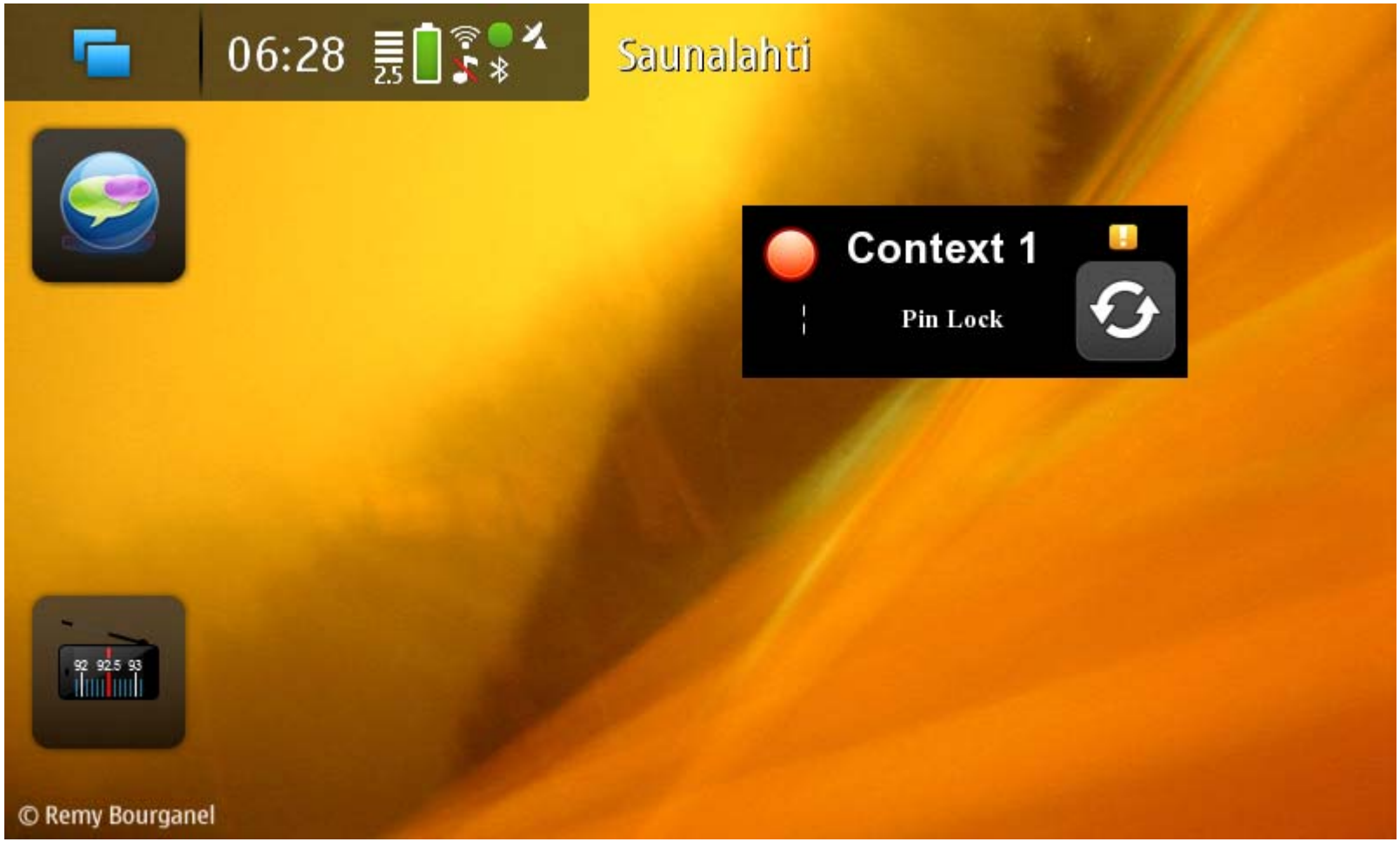
- WiFi- and Cell-tower localization via reverse lookup
  - Needs access to a database
- WiFi Similarity
  - If Jaccard distance (S1, S2) < 0.5, use the location of S1
- Best matching CoI
  - Calculate $instFam_{WiFi}(C_i, n)$ for S2 w.r.t all known CoIs C_i
  - Use centroid of CoI X with the highest $instFam_{WiFi}(X, n)$
    - (if it is also above a threshold, 0.75)

Null device

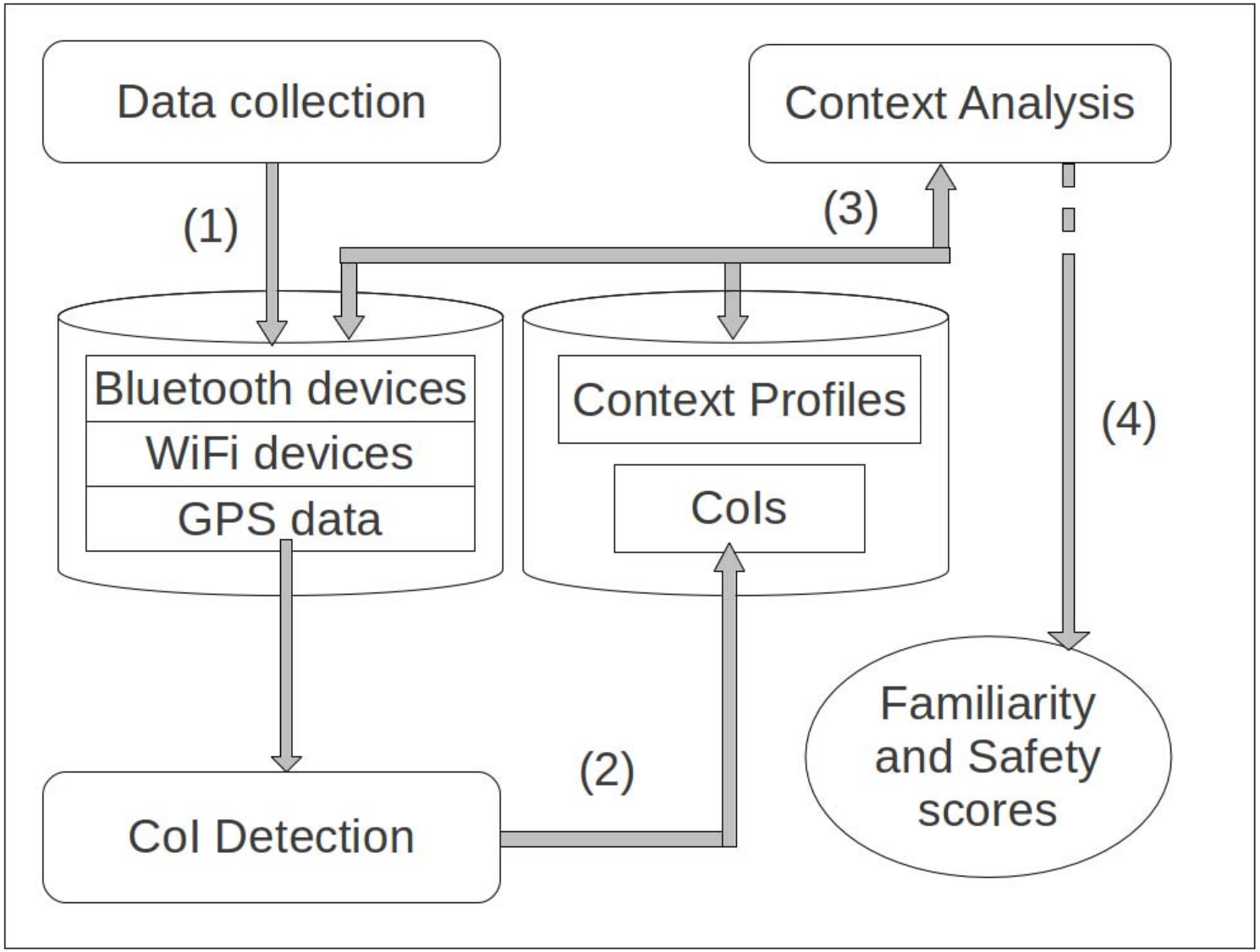
Model the absence of any devices as a ”null device”

- Each type has its own NULL device
- $devFam(NULL, C, n)$ is computed as for other devices
  - High in CoIs where absence of devices is typical
  - Low in CoIs where absence if devices is untypical

Incorporating user feedback



Context Profiler Architecture



OKIA

Choosing Context Profiler Parameters

Parameters identified so far: α_D, N_0, α_C, LT and HT

Datasets from Lausanne Data Collection Campaign

- GPS, WiFi and Bluetooth observations at different rates
- Mapped to our needs: observations with all three types
- Experiments using data of 37 users (led to 167 CoIs)

Frequent CoIs: set of CoIs containing two CoIs with the highest number of observations for each user

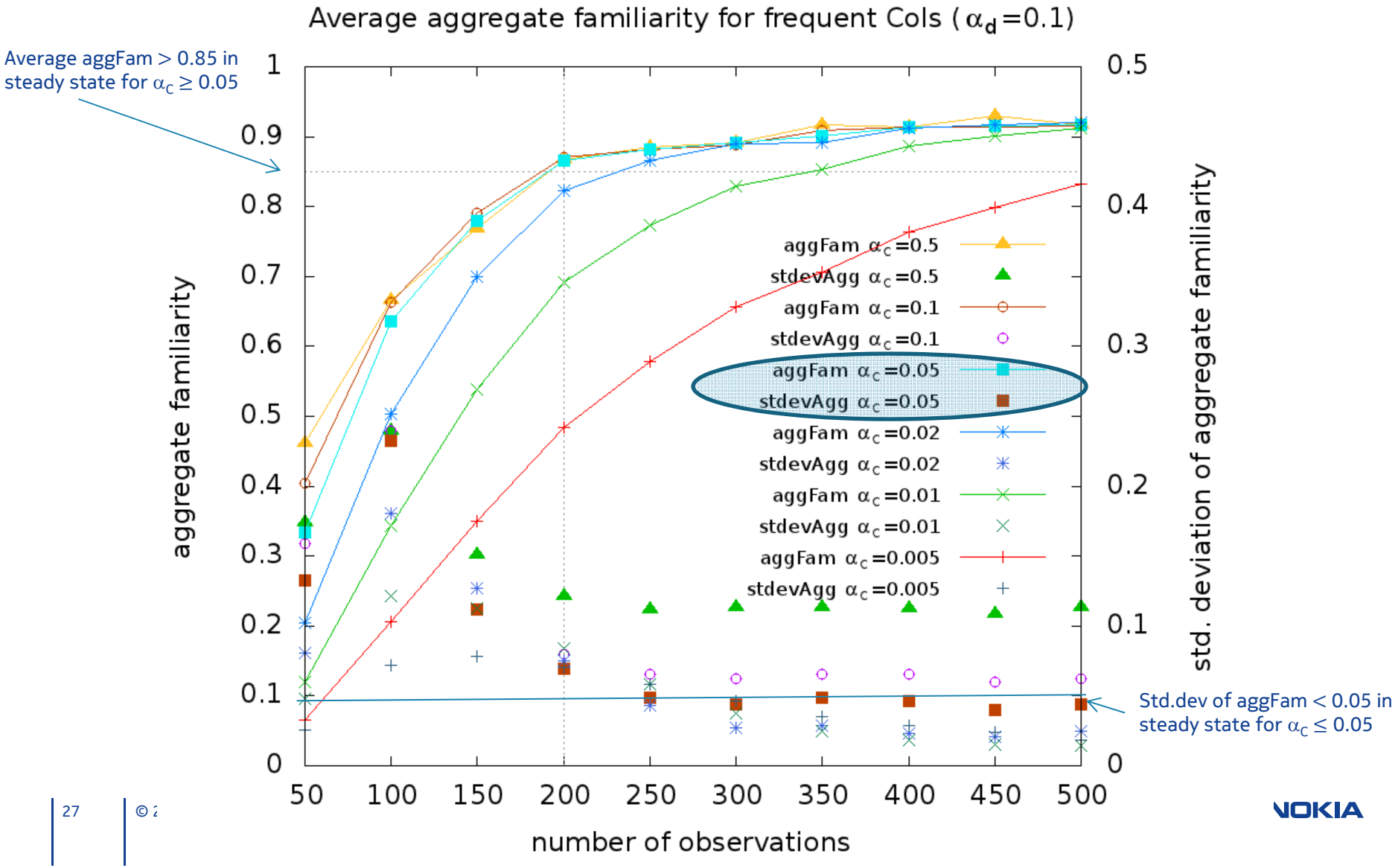
Choosing Context Profiler Parameters

Rough targets and assumptions

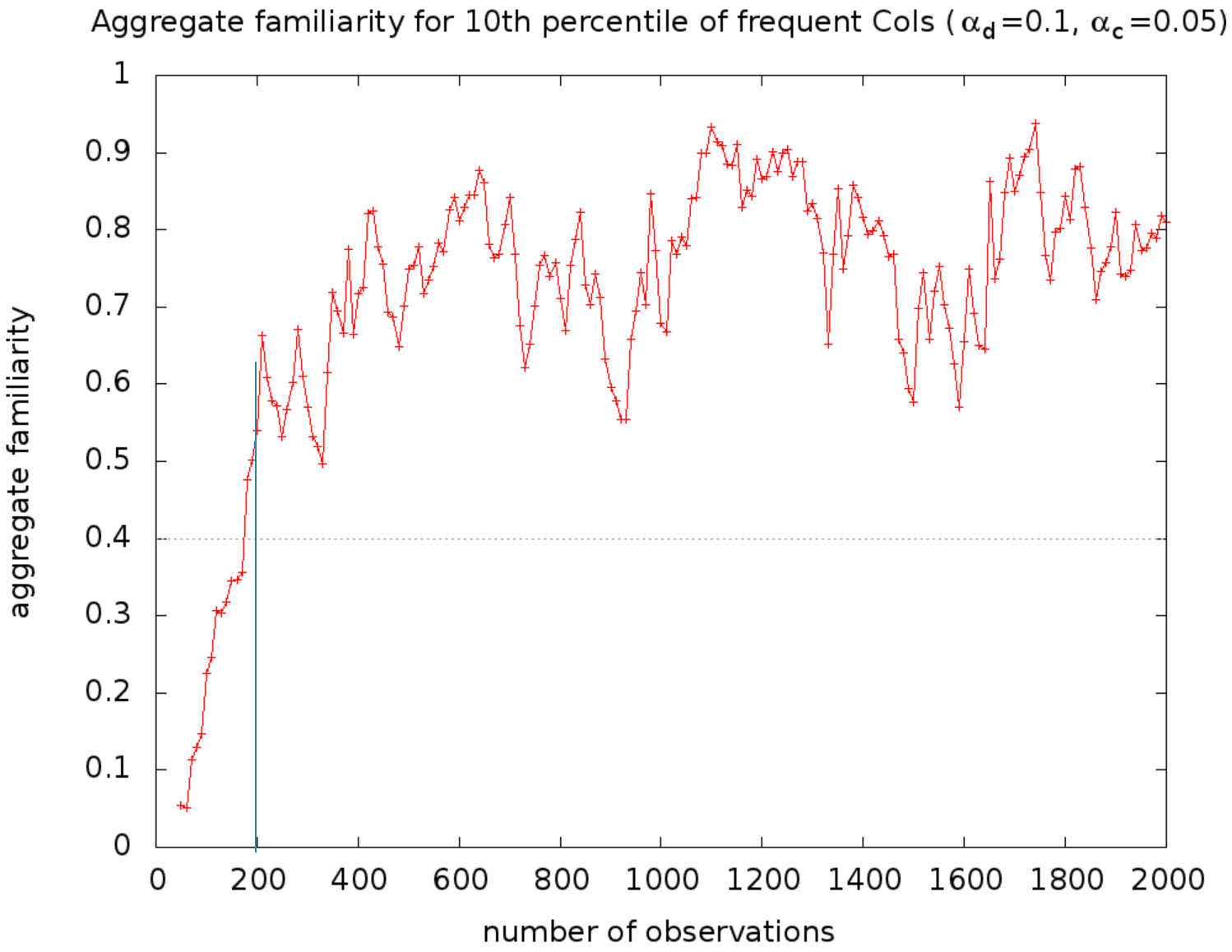
- Assume most people have two safe CoIs (work/home)
  - Identify the set of *"frequent CoIs"* in the data
- Context Profiler should recognize safe CoIs within 2 days
- A third of a day is probably spent in a given safe CoI (~100 observations)
  - Set N_0 100 ("prolonged absence")
- Device familiarity is smoothing of a locally constant step function
  - Set α_D to 0.1*

* following guidance in "Smoothing, Forecasting and Prediction of Discrete Time Series", R. G. Brown, Dover Phoenix Edition, 2004.

Choosing α_C and High Threshold HT

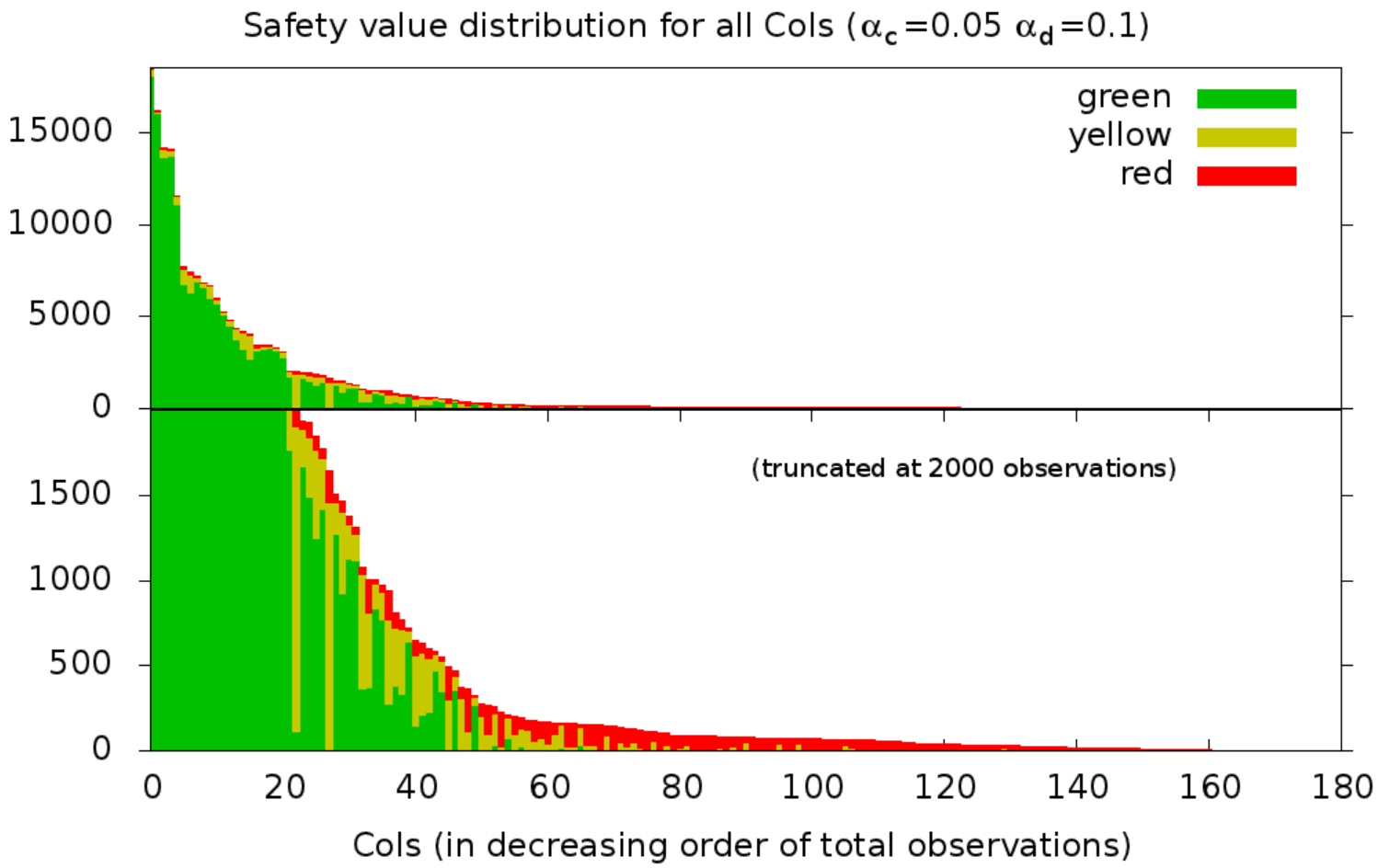


Choosing Low Threshold LT



NOKIA

Distribution of safety scores



Comparing with ”Ground Truth”

No ground truth regarding specific observations. But users labelled places:

safe

My home
My freetime home
My main workplace

Home of a friend
My main school or college place
My other work place
Other
I don’t know

unclassified

Holiday resort or vacation spot
Shop or shopping center
Location related to transportation (bus stop)
Place for indoor sports (e.g., gym)
Place for outdoor sports (e.g., walking)

unsafe

Assume all observations in safe places must be safe (similarly for unsafe)

Comparing with ”Ground Truth”

Recognizing safe situations		
Precision	$\frac{\lvert G_{safe} \cap C_{GREEN} \rvert}{\lvert C_{GREEN} \rvert}$	0.854
Recall	$\frac{\lvert G_{safe} \cap C_{GREEN} \rvert}{\lvert G_{safe} \rvert}$	0.917
Fallout w.r.t ”unsafe”	$\frac{\lvert G_{unsafe} \cap C_{GREEN} \rvert}{\lvert G_{unsafe} \rvert}$	0.152
Fallout w.r.t ”unclassified”	$\frac{\lvert G_{unclassfied} \cap C_{GREEN} \rvert}{\lvert G_{unclassified} \rvert}$	0.755

Recognizing unsafe situations		
Precision	$\frac{\lvert G_{unsafe} \cap C_{RED} \rvert}{\lvert C_{RED} \rvert}$	0.311
Recall	$\frac{\lvert G_{unsafe} \cap C_{RED} \rvert}{\lvert G_{unsafe} \rvert}$	0.341
Fallout w.r.t ”safe”	$\frac{\lvert G_{safe} \cap C_{RED} \rvert}{\lvert G_{safe} \rvert}$	0.019
Fallout w.r.t ”unclassified”	$\frac{\lvert G_{unclassfied} \cap C_{RED} \rvert}{\lvert G_{unclassified} \rvert}$	0.096

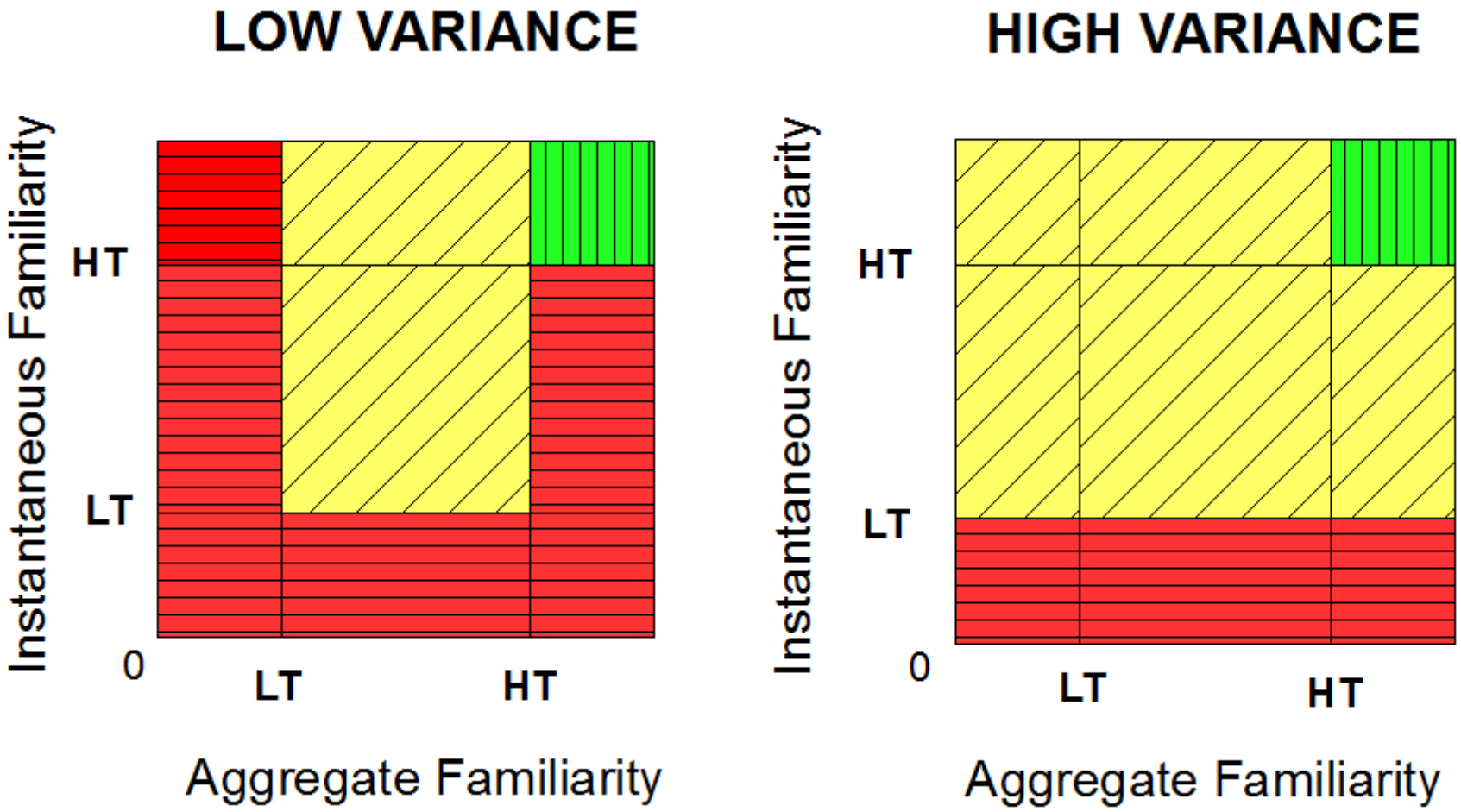
G: set of ground truth ”observations”
C: set of context profiler observations

Comparing with ”Ground Truth”

Recognizing unsafe situations		
Recall (RED + YELLOW)	$\frac{\lvert G_{unsafe} \cap (C_{RED} \cup C_{YELLOW}) \rvert}{\lvert G_{unsafe} \rvert}$	0.848

YELLOW safety level is not significantly safer than RED

Revisiting the safety algorithm



What about security and privacy?

- Yes, WiFi and Bluetooth addresses can be easily faked
  - Our goal is to make device lock acceptable for those who do not use it now
  - *Any increase in security is a win*
- Yes, this data collection is like iPhone tracking; but
  - data never leaves the device
  - data can be encrypted using device-specific hardware key
  - feature can be opt-in

Context profiler status (with internal links)

ISAC (Intuitive and Sensible Access Control) is an exploratory research activity in NRC (withstudents from Purdue University, EPFL and Aalto University)

- Context profiler prototyped on N900: Demo video
- Draft paper describing the analysis of the algorithms using dataset from Lausanne data collection campaign
- Planned work
  - Port to other platforms
  - User study
  - Extensions:
    - other context variables,
    - benefits of sharing context profiles vs. privacy issues

Context profiler status

Part of *Intuitive and Sensible Access Control (ISAC)* exploratory research activity in NRC (with students from Purdue University, EPFL and Aalto University)

- Context profiler prototype: Demo video, demo paper
- Draft paper describing analysis of algorithms using dataset from Lausanne data collection campaign
- Current work
  - Porting to Qt/QML (and possibly other platforms)
  - User study
  - Extensions: other context variables, benefits of sharing context profiles vs. privacy issues, distinguishing verifiable device addresses (without wasting battery)

Open issues

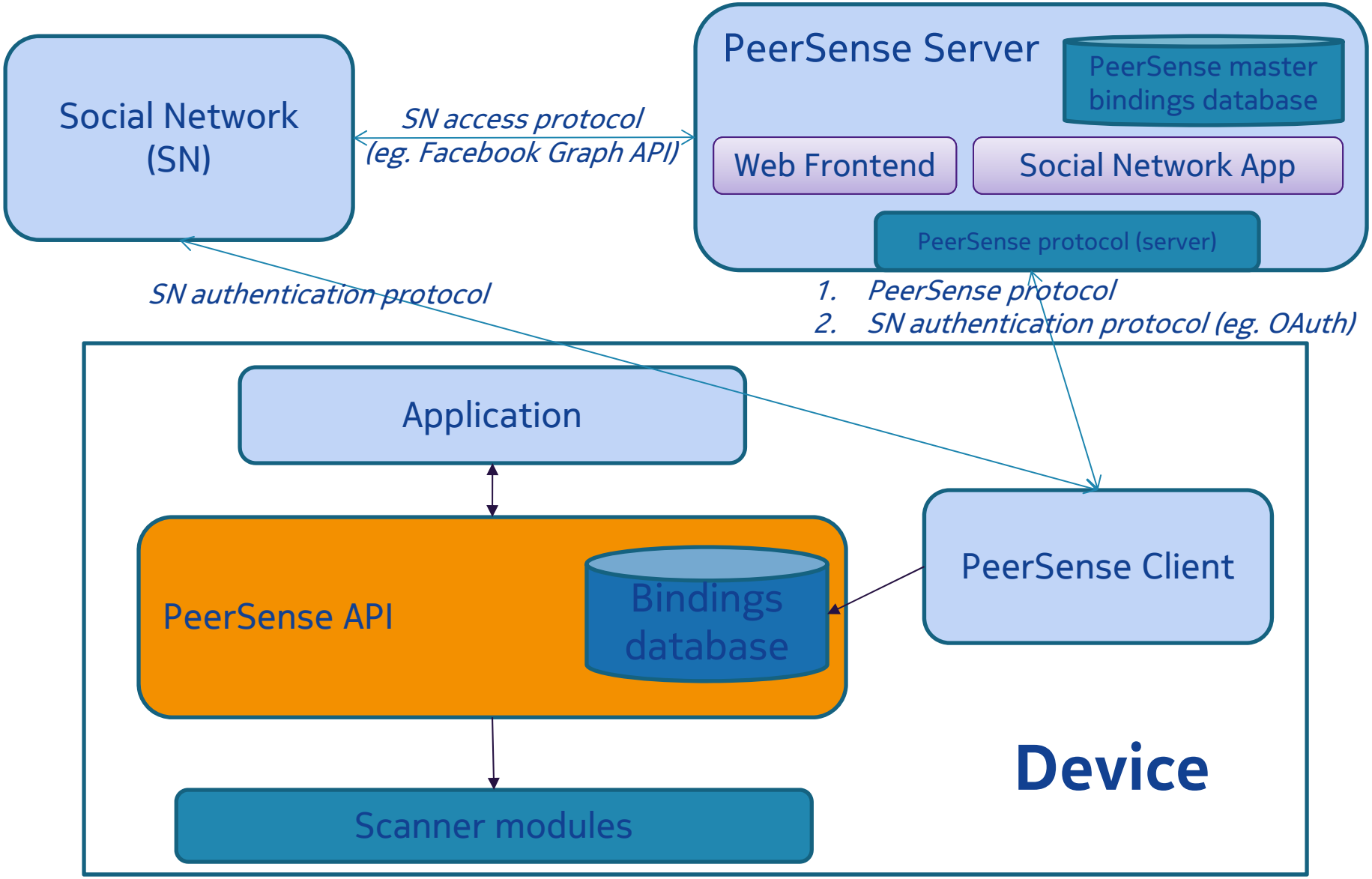
- Scanning without running the battery down
- Finding the notion of safety right for the application
- Making context profiler inferences intelligible to the user
- Sharing context profiler data without damaging privacy
- Distinguishing verifiable device addresses
- Dealing with more general types of contexts
  - Family car, group of colleagues at a bar, …
- Finding other applications besides device lock
- …

Another ISAC example

PeerSense: recognizing nearby friends (work in progress)

- **How can your device recognize your friends' devices?**
  - **intuitive**: one-time simple user action to get started; user need not manually bind friends' names to device addresses
  - **private**: eavesdroppers do not learn names; servers do not learn location or co-location of devices/users
- PeerSense API allows an application to find information about nearby "friends"
  - Example: camera recording nearby friends as photo metadata(as in TagSense); use to infer likely sharing targets
- Status: Demo (to be shown at Percom 2012)

PeerSense architecture



Summary

An open problem: Ordinary users need intuitive means to set sensible security and privacy policies

Cues in context and history can help in solving this problem

- Two instances explored
  - Context-profiling and its application to device lock policy
  - PeerSense Person-to-Device binding and its application for photo sharing policies
- Several open issues
  - What other applications can use familiarity and safety estimates? Can we generalize beyond geolocational contexts?