Prajit Kumar Das

Advisor: Anupam Joshi

Co-advisor: Tim Finin

Context-Dependent Privacy and Security Management on Mobile Devices

UMBC

AN HONORS UNIVERSITY IN MARYLAND

Outline

- Motivation
- Thesis statement and Contributions
- Related Work
- Approach: Mobile middleware and Application analytics
- Evaluations: Application analytics and Mobile middleware
- Extensions to previous work
- Conclusions and Future Work

Motivation

User data under threat

SECURITY

How many mobile apps collect data on users? Oh ... nearly all of them

Free or paid, Android or iOS, your apps are spying on YOU – report

By Neil McAllister, 21 Feb 2014 Follow 519 followers

Apple faces privacy breach charges with its secret user tracking file

By Monami Thakur
on April 21 2011 6:38 AM

f 0 0 in 0 g+ more

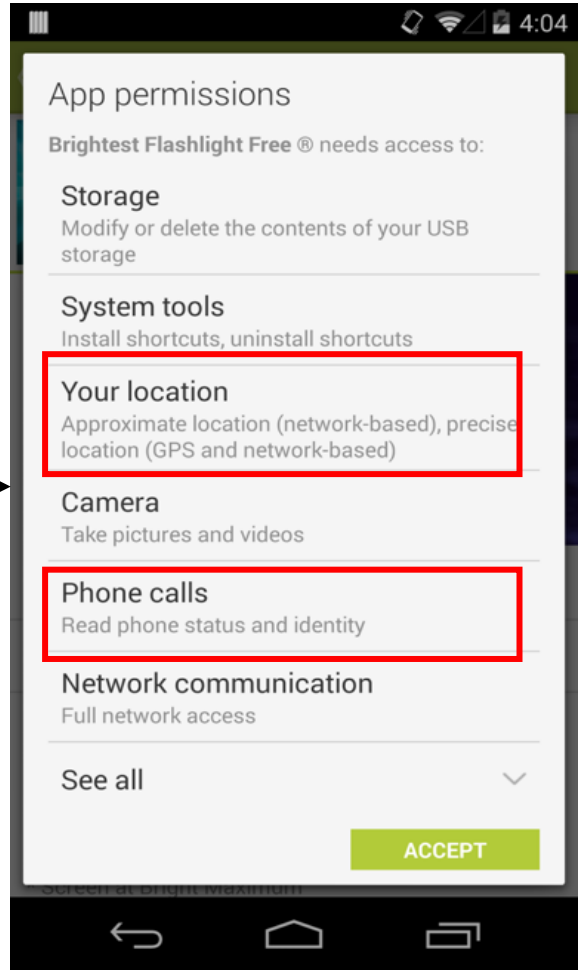
Researchers at a technology conference in San Francisco on Wednesday have accused Apple of breaching the privacy line of consumers by storing user's location and other details in a secret file. Reuters.

Excessive permission requests…

…can lead to security breach

Data haul by Android Flashlight app 'deceives' millions

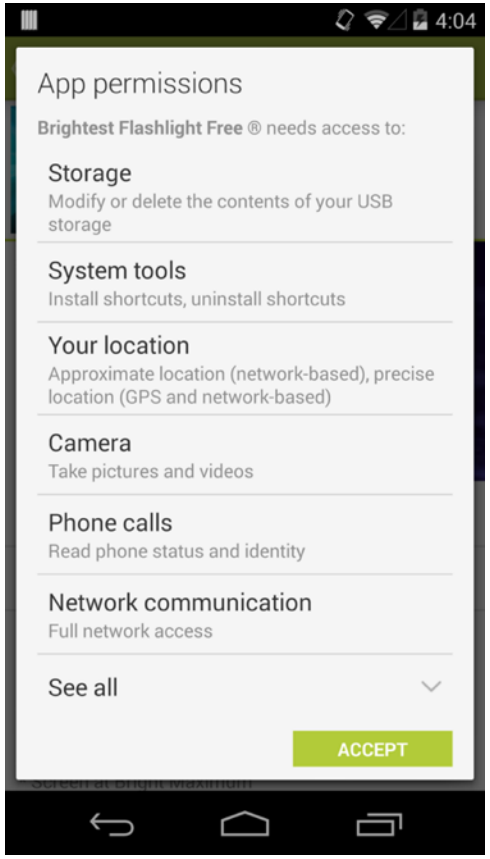
Mal-intent app downloaded by 50 million users sanctioned by United States Federal Trade Commission December 2013



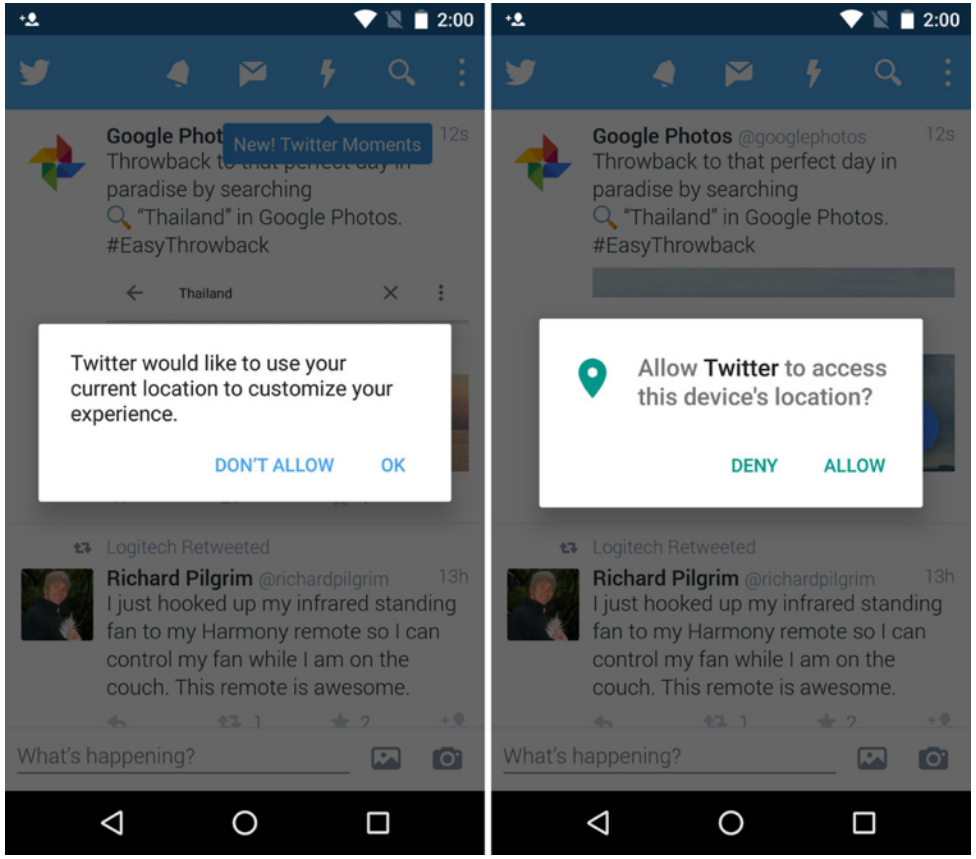
Source and Images courtesy: Google Play Store

One-time permissions – Inadequate

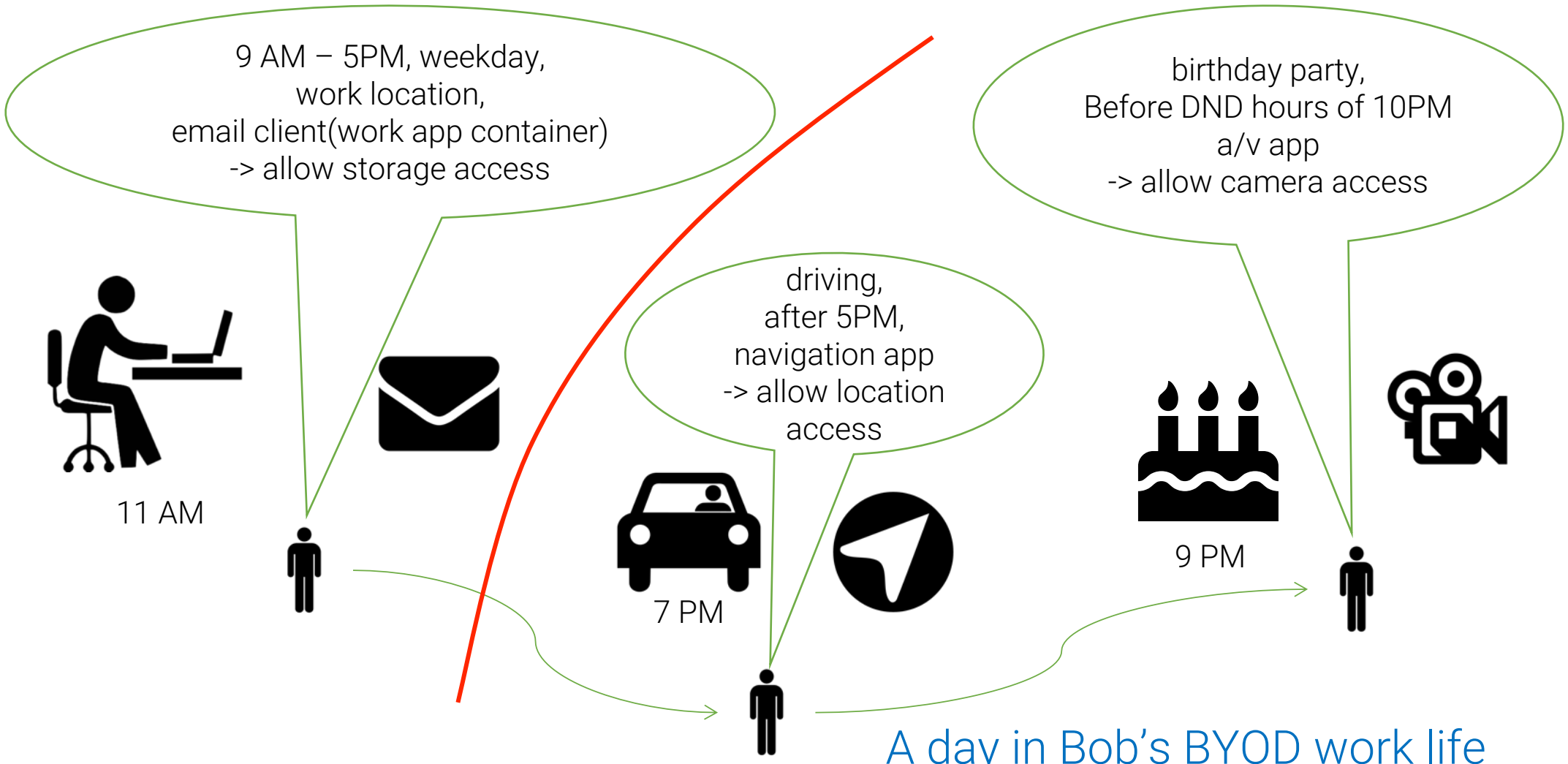
Pre-Marshmallow



Marshmallow



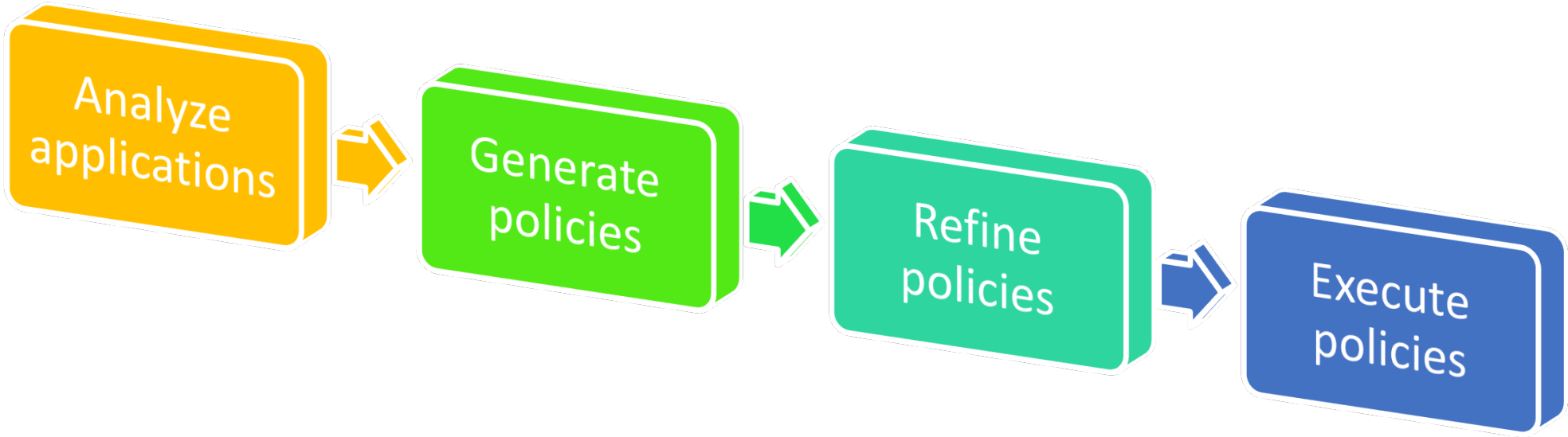
Solution: Context dependent permissions



A day in Bob's BYOD work life

How to create…

… context-dependent privacy and security policies?



Thesis Statement and Contributions

Thesis Statement

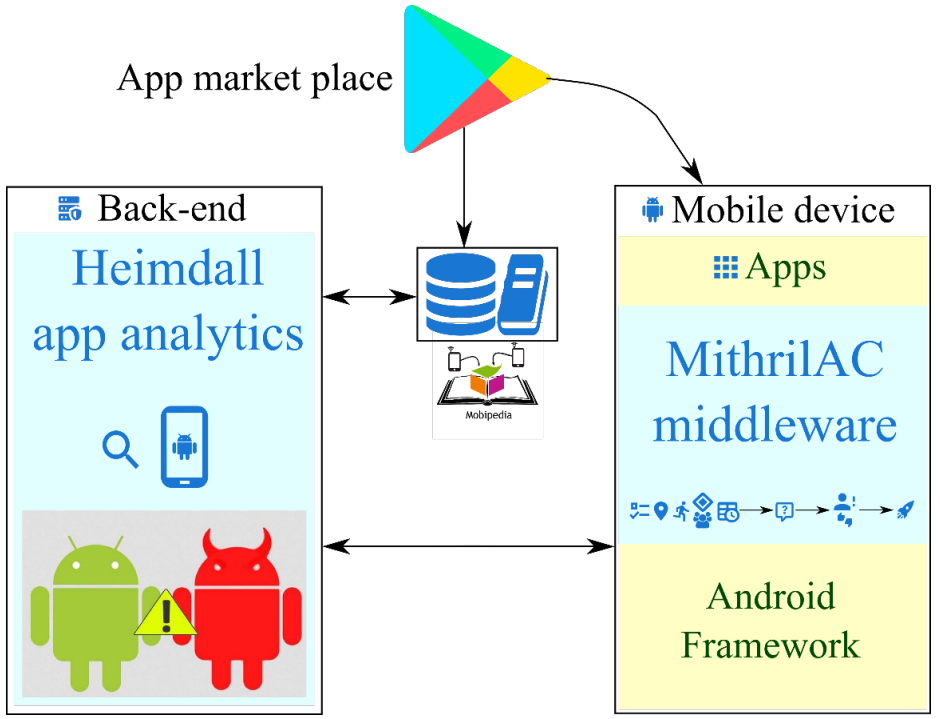
A semi-automated approach that combines *mobile application analysis* with *violation monitoring techniques* can *reduce* the amount of *user interaction required* in *capturing better access control policies* that are fine-grained and context-dependent.

Key Contributions

Mithril: Semi-automated access control approach, that combines

- Mobile application analytics back-end
- Application monitoring mobile middleware

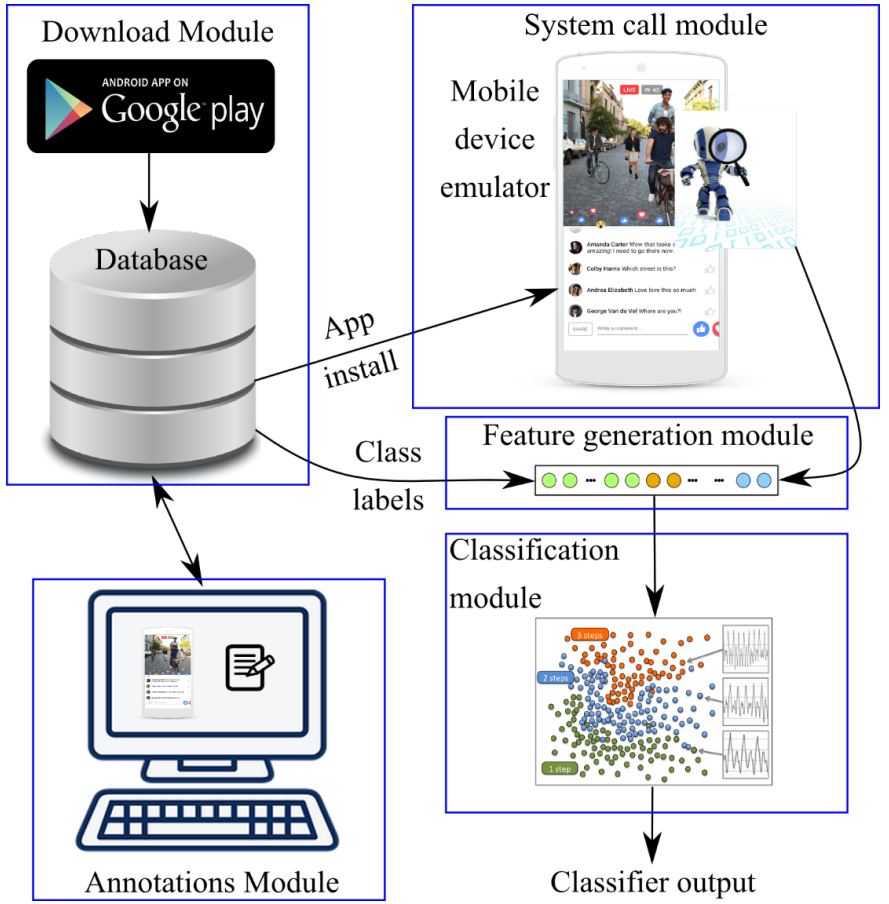
...to capture better access control policies that are fine-grained and context-dependent.



Mithril's access control approach

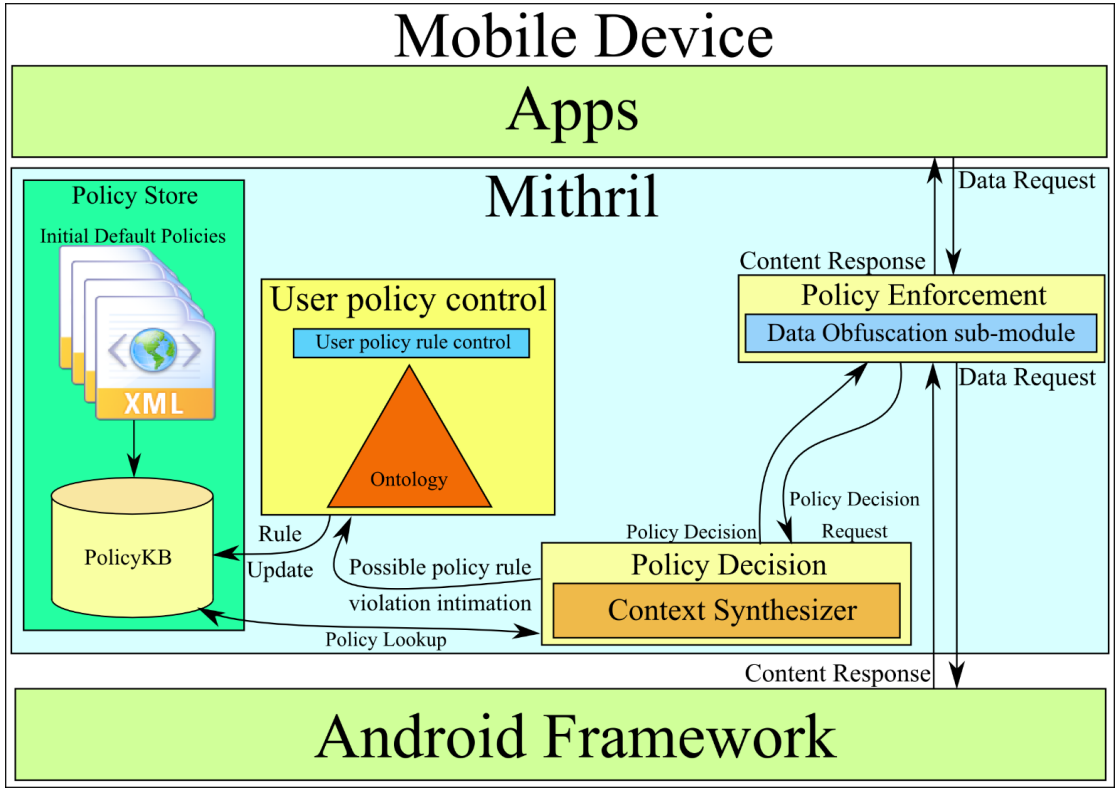
Contributions: App analytics

- Defined “application behavior” for creating curated policies
  - Used static and dynamic features to determine classes like Alarm Clock, File Explorer etc.
  - Used features important to malware detection for computing application risks
  - Generated initial default policy through crowdsourced data and behavior classification



Contributions: Mobile Middleware

- Defined “violation metric” as a way to determine completion of policy capture
- Reduced user interaction required – using output curated policy
  - User study performed to show feasibility of using violation metric
- Custom ROM built for executing context-dependent policies



Contributions: Extending earlier work

- Added behavior knowledge to Mobipedia (MoDeST'15)
- Enhanced presence context generation using nearby messages and low energy beacons (Jagtap'11, Das'16)

Related Work

Usable privacy and security

- Convergence of personal and enterprise application usage in BYOD scenario leads to new threat models (Kodeswaran et. al.'13)
- People are privacy pragmatists (Kumaraguru'12)
- User permission decision simplified through profiles (Liu et. al.'14)
- Contextual policies on Mobile: CRêPE (Conti'11)
  - Focus on user context capture for which policy variations exist

Context generation

- Dey'99: created the popular definition of primary user context as:
  - Identity (we use presence) Location, Activity, and time
- Wibisono et.al.'13 defined context in a peer-to-peer environment by reasoning about situations in presence of uncertain context information
- Jagtap et.al.'11 created the Platys ontology to model high-level notion of "context" using collaborative information sharing
  - Mithril uses Platys ontology and extends presence context using nearby messaging and beacons

Application analytics

- CHABADA (Gorla'14) studied application descriptions for malware detection (56% precision in malware detection)
- CrowDroid (Burguera'11) project analyzed system calls for self-written malwares (92.5% precision in malware detection)
- State-of-the art for mobile malware detection: 96.76%
  - Using deep learning, static and dynamic features (Yuan'16)
  - Ebiquity research used deep learning to improve malware detection
- System calls used for software analysis (Kosoresow'97)
  - System calls for mobile application behavior classification

Approach: Mithril Mobile Middleware

Capturing policies for fine-grained access control on mobile devices*; IEEE CIC'16*
Refine policies ⇒ Execute policies

Access control model

- Traditional access control models
  - Discretionary Access Control (DAC), Mandatory Access Control (MAC), Role-Based Access Control (RBAC) – Sandhu et.al.’96
  - Do not define context-awareness
- Attribute-based Access Control (ABAC) – Draft NIST security standard (Hu et.al.’13)
  - Attributes in form of tuple (R, C, Q)
  - R: Represents the requester's context
  - C: Represents the user's context
  - Q: Represents the query received

Policy representation

- Semantic Web Rule Language (SWRL, Horrocks'04)
  - Antecedent ⇒ Consequent
  - Antecedents represent context attributes

    Requester(?app) ^ hasCategory(?app, "Social_Media") ^

    User(?u) ^ hasLocation(?u, "School") ^

    Request(?app, "$Camera") ⇒ denyAccess("Camera")
- Policy consequent: Both deny/allow
  - Things that are allowed by policy but blocked by OS – critical to system
  - Things that are blocked by policy but allowed by OS – critical to user policy

Relevant Terminologies

- **Observer**: Observe system; captures rules
- **Enforcer:** Enforce captured rules
- **Violation**: App and resource usage contradicts known policy in contextual situation(s)
  - At location or during event or in presence of or At time
- True Violation – **TV**: User votes - Policy is good
- False Violation – **FV**: User votes - Policy needs revision

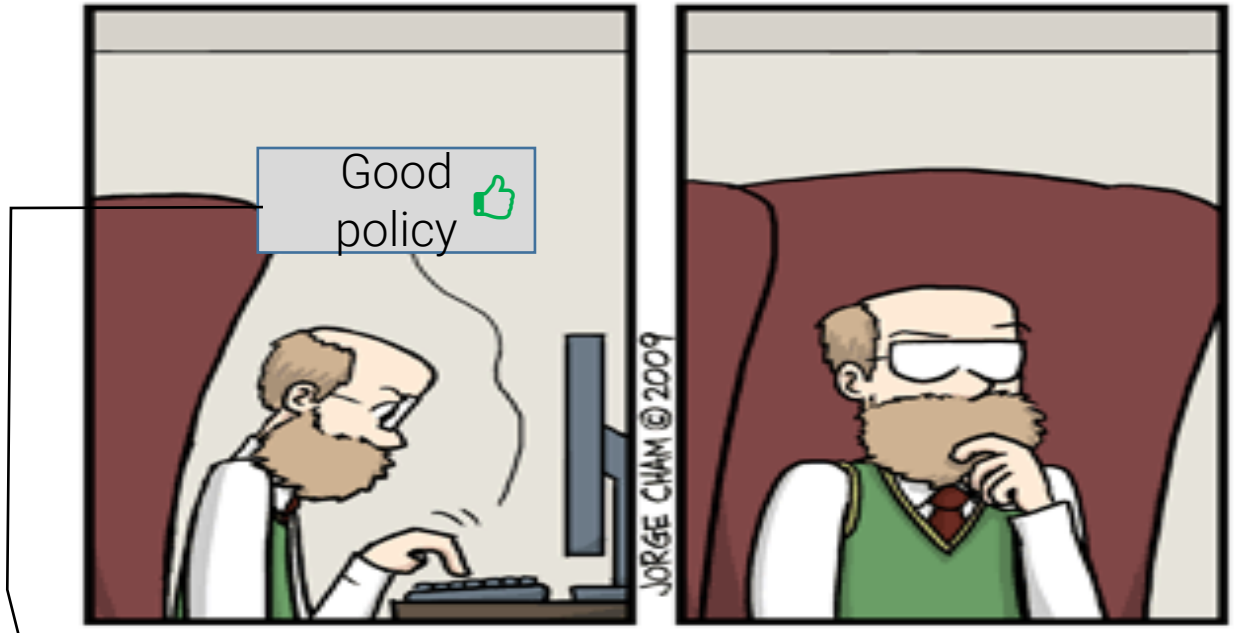
User feedback algorithm: Violation detection

- Step 1: Detect app launch
- Step 2: Detect resource usage
- Step 3: Detect context: Using Platys ontology and mobile APIs
- Step 4: Look up existing policies and detect violations

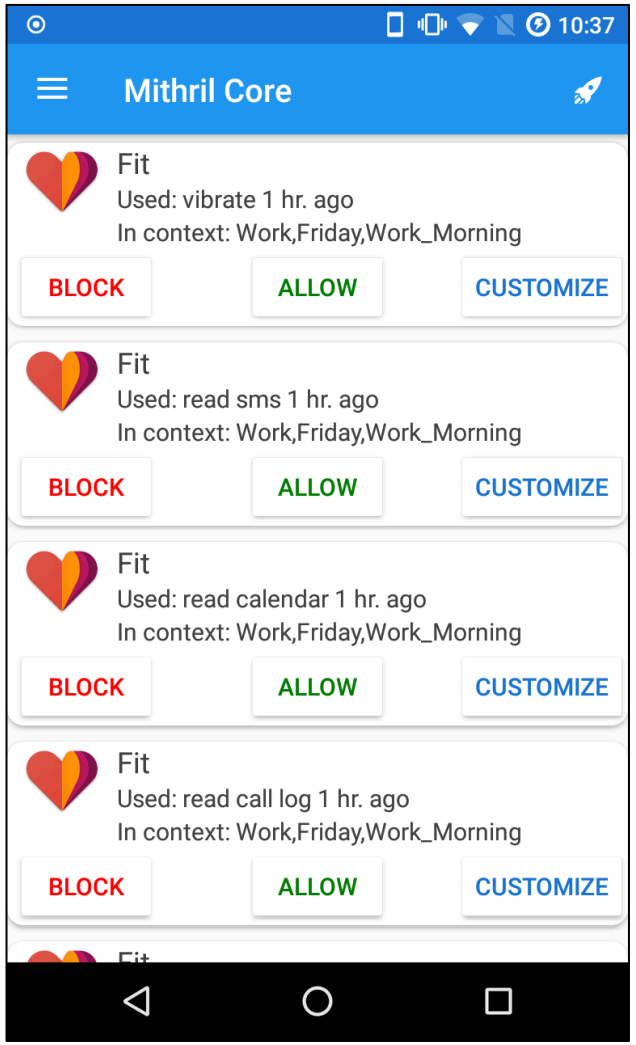
User feedback algorithm: Policy modification

Step 5: User votes on policy

Step 6: True violations: good!

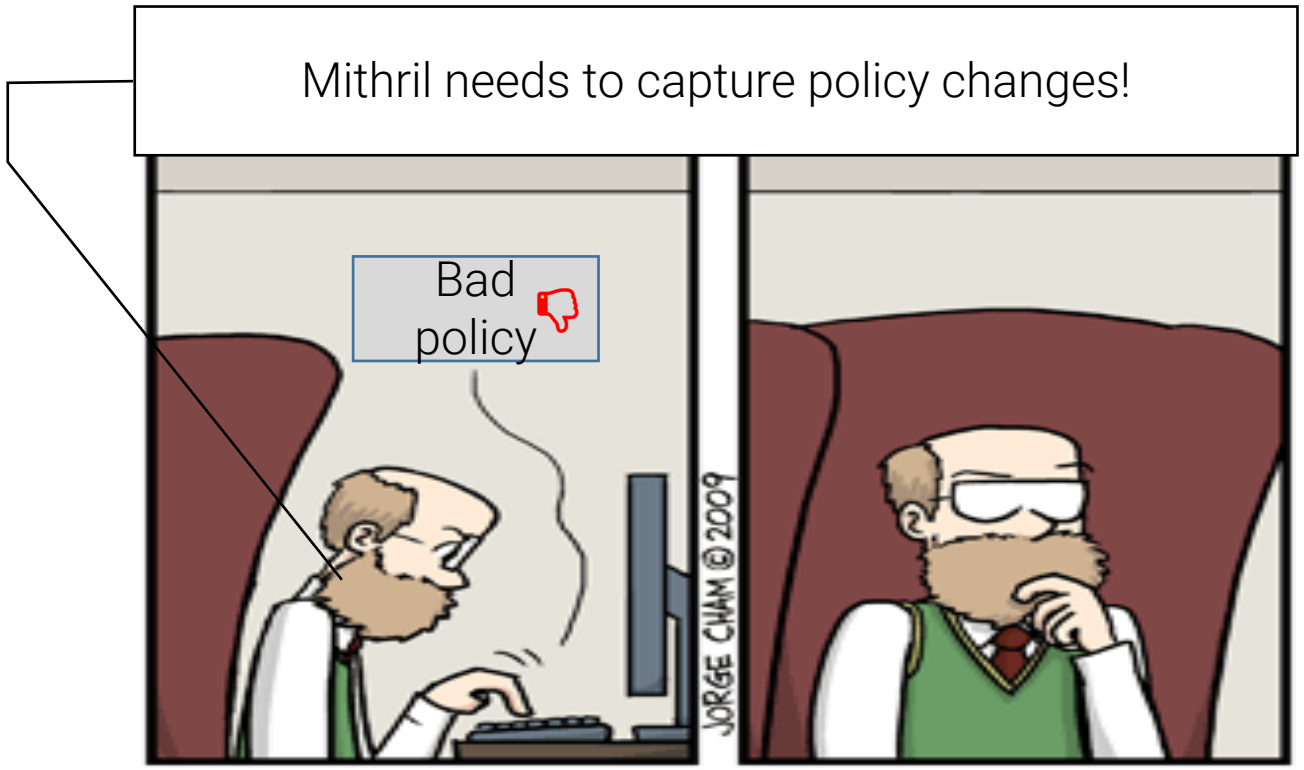


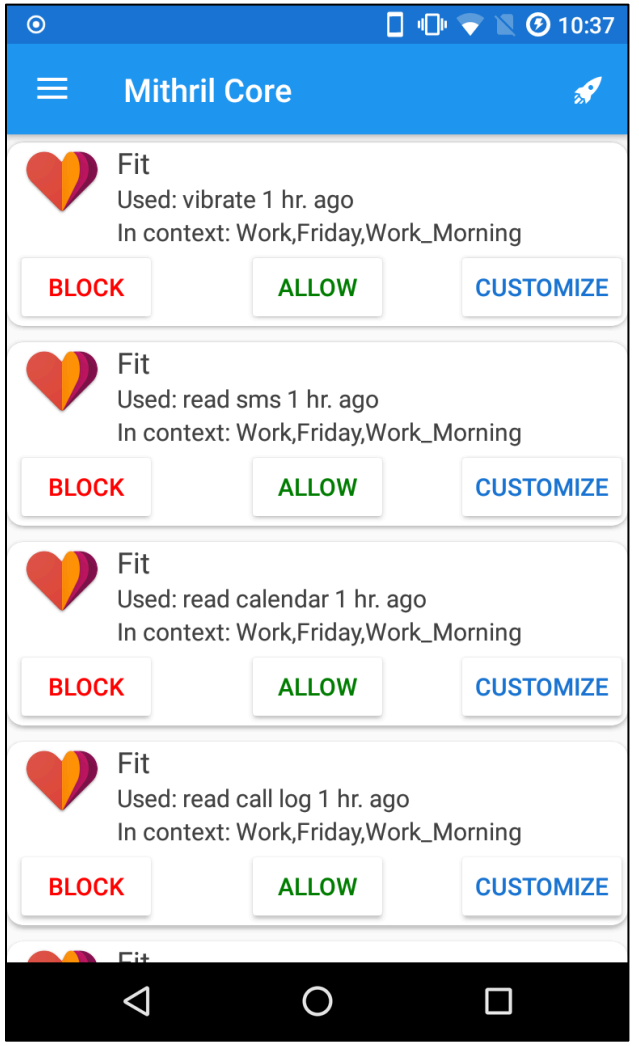
Mithril knows user's policies, recording true violations!



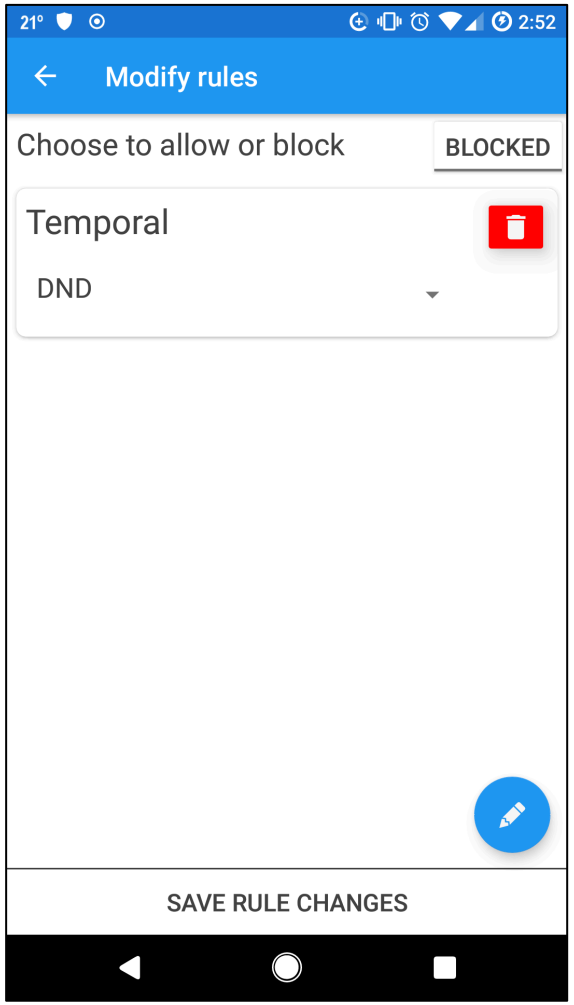
User feedback algorithm: Policy modification

Step 7: False violations: trigger policy modification steps





User feedback algorithm: Policy modification



Step 7: False violations

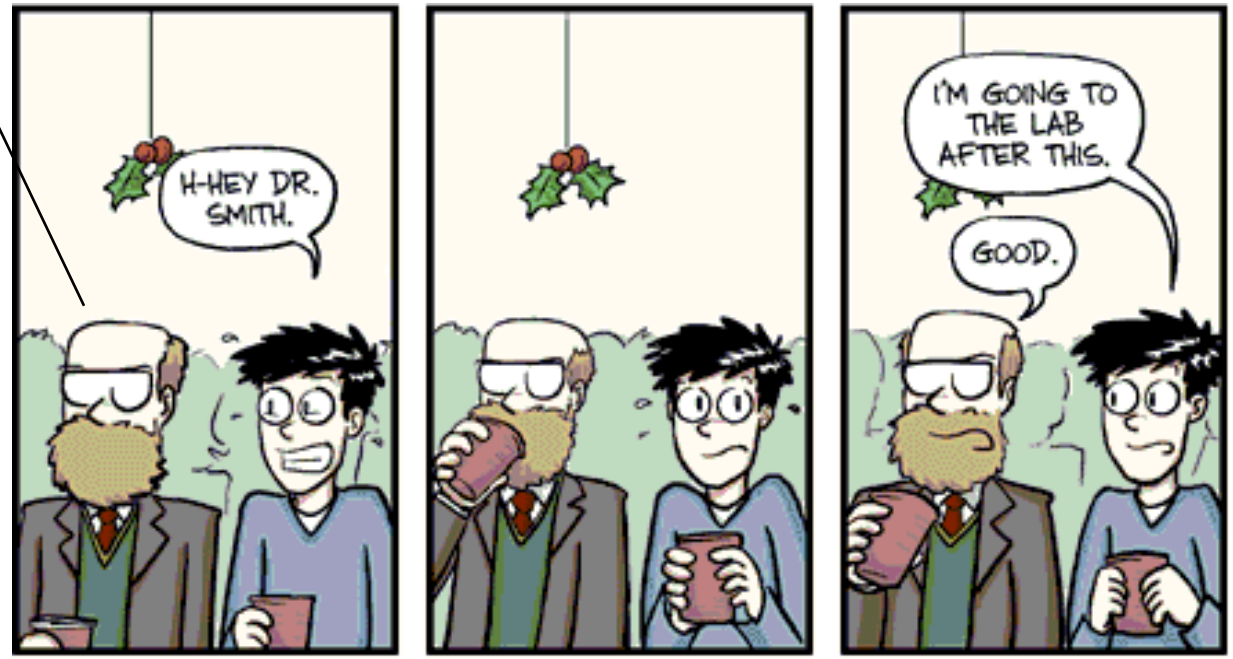
- Condition generalization
- Condition specialization
- Delete conditions (Policy deletion)
- Add conditions (Policy creation)

Why capture policy modifications?

Generic Rule: When at work Professors do not open location sharing apps

This is Prof. Smith.
He sometimes has lunch with his students.

When out to lunch, Prof. Smith used a location sharing app to coordinate with students.

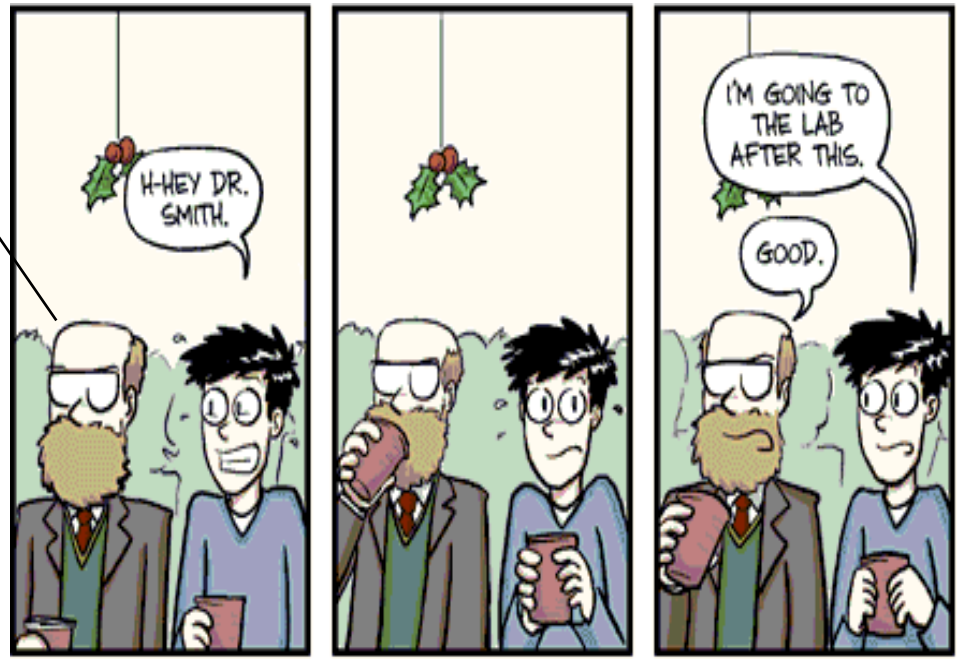


Sadeh et.al'09 has shown asking too many questions leads to user confusion and fatigue

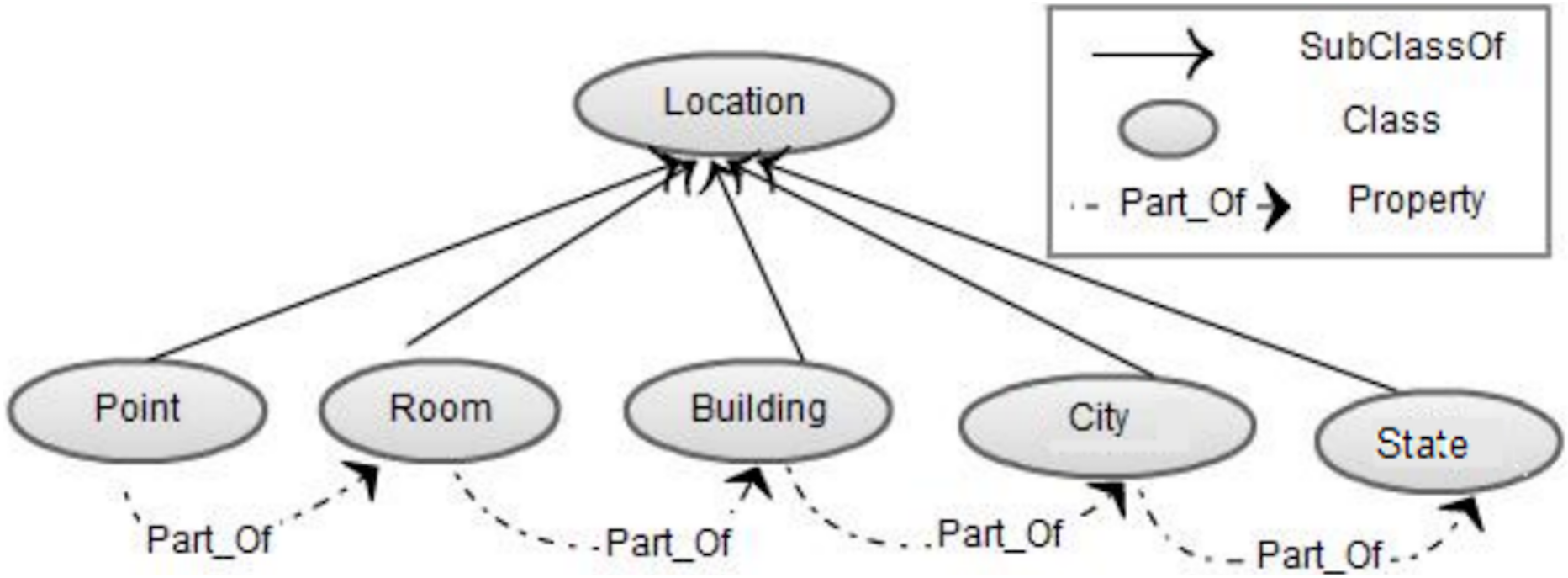
Context definition: Platys Ontology

- Activity
  - Personal_Activity
  - Professional_Activity
    - Attending_Lecture
    - Delivering_Lecture
    - Meeting
      - Lab_Meeting
      - Lunch_Meeting
      - Professor_Meeting
      - Project_Meeting
    - Studying
      - Course_Work
      - Project_Work

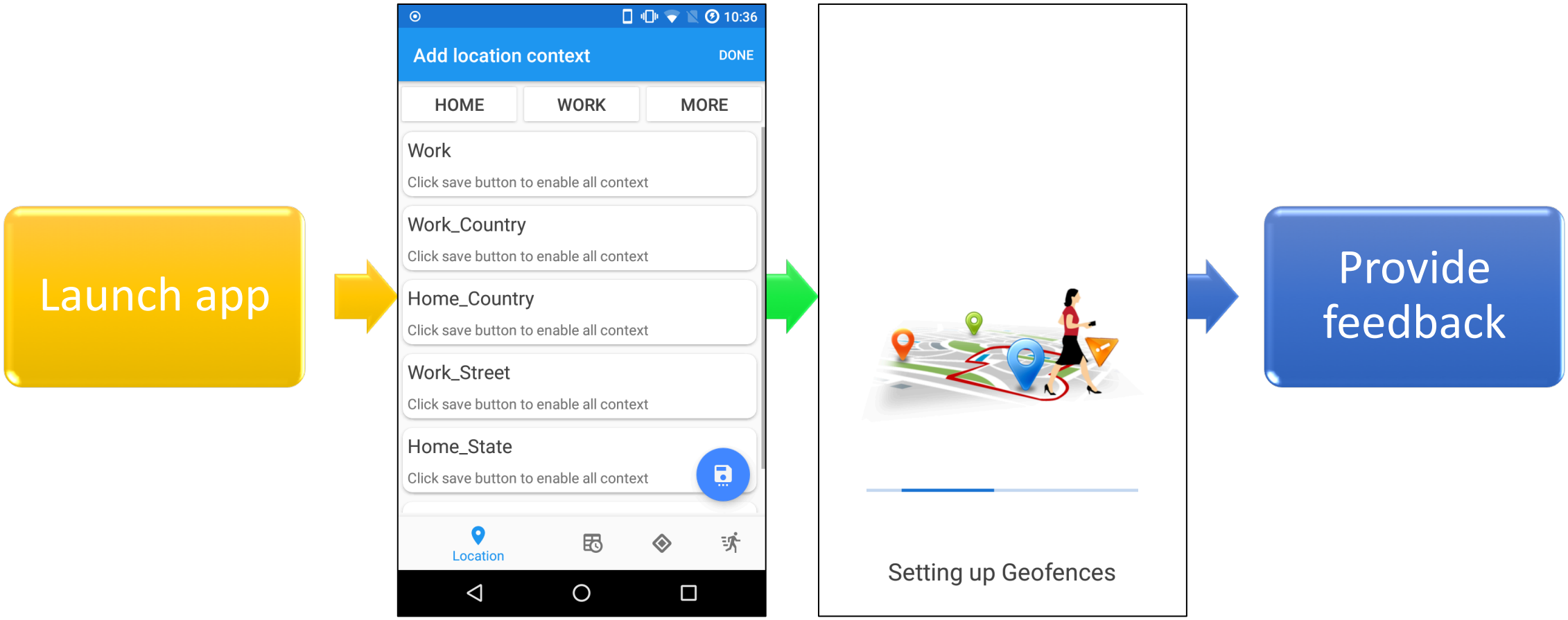
When out to lunch, Prof. Smith used a location sharing app to coordinate with students.



Context definition: Location hierarchy



User study: Creating context instances



Android Fence API: Location, activity and proximity fences – Energy efficient

User study: Challenges and solutions

- Application operations API
- Inaccessible to user land app and hidden since Android 4.3
- Middleware requires “privileged status”, even rooted phones won’t work
- ***Solution***: Created custom Android ROM; added Middleware with OEM privileges

Policy Enforcement

Built Android (LineageOS) – MithrilAC as a *priv-app*

Approach: Heimdall App Analytics

App behavioral analysis using system calls*; MobiSec'17*
Analyze apps ⇒ Generate policies

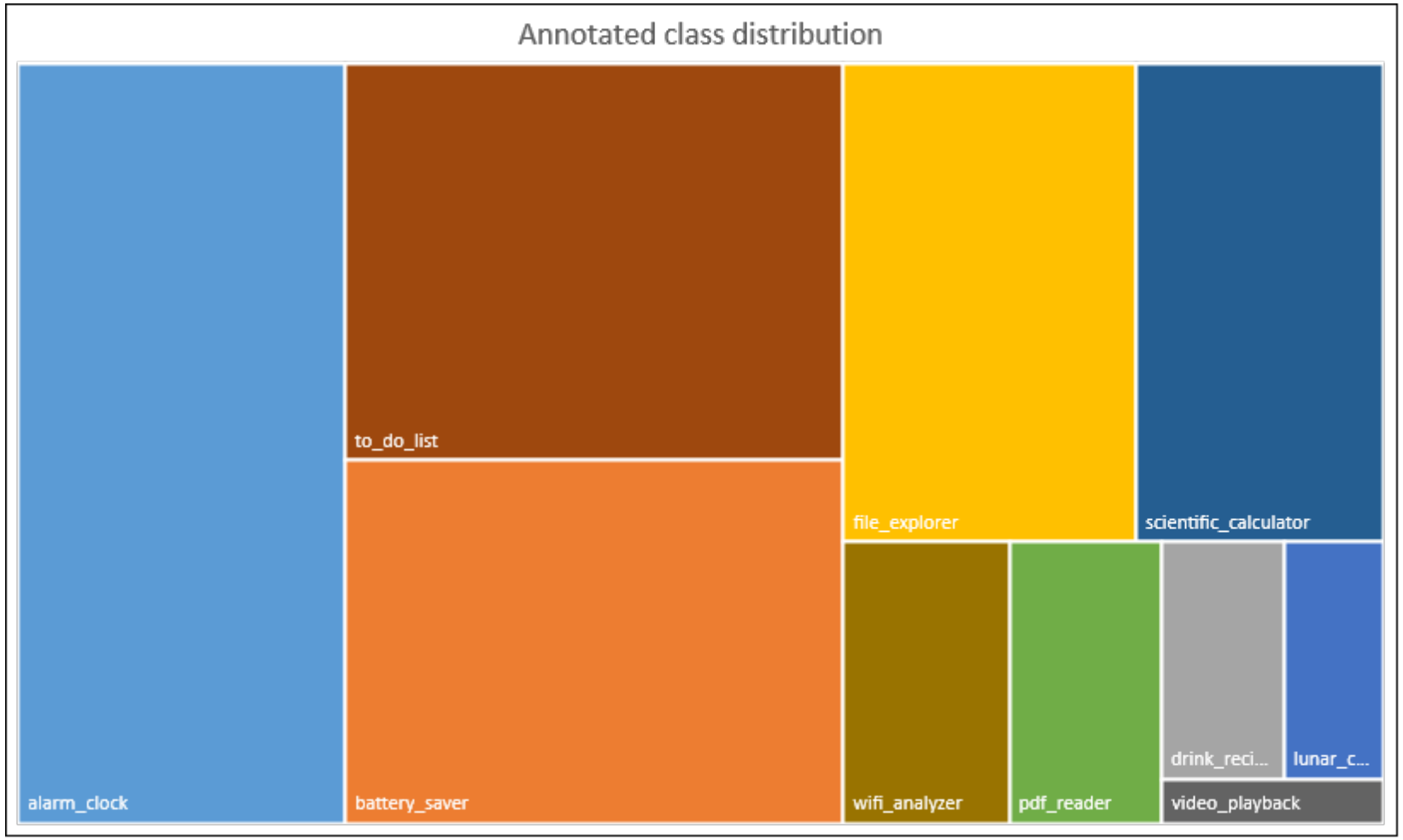
Classification tasks and features

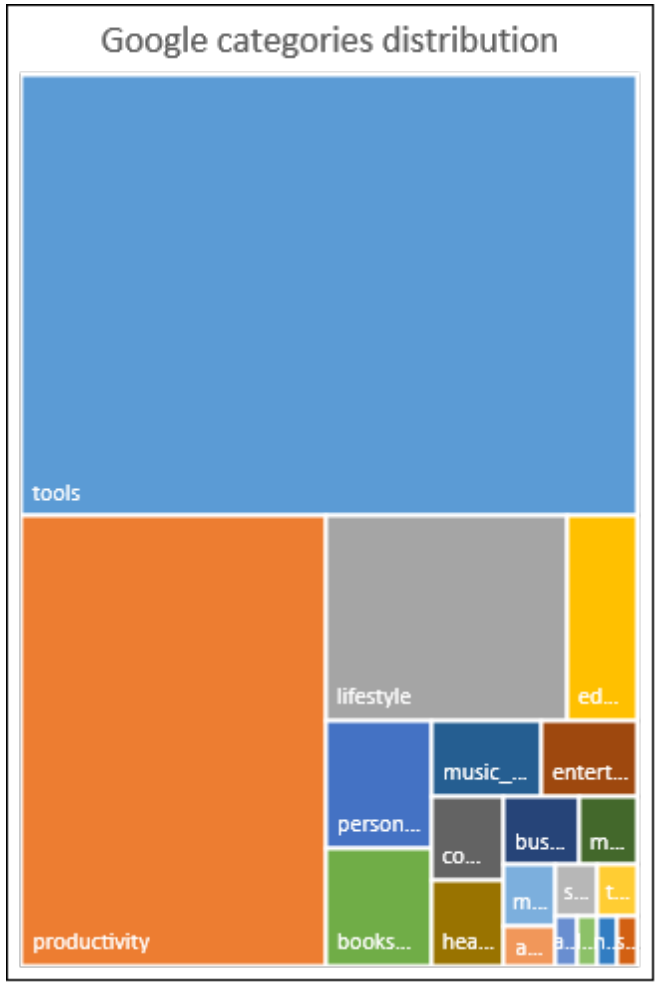
- Android application behavior
- Google Play category
- Malware detection (just static features)
- Features – System calls:
  - Round 1: Unigram model
  - Round 2: N-Gram language models
- Features – Static permissions: Just 1-hot vectors

Meaning of application behavior

- Problem: Non-granular app categorization (only 50+ categories)
  - Too coarse
- Application behavior: Human perception of what an app does
- Classes defined: Alarm_clock, Battery_saver, Drink_recipes, File_explorer, Lunar_calendar, Pdf_reader, Scientific_calculator, To_do_list, Video_playback, Wifi_analyzer
- Example policy:
  - alarm_clock **deny contacts** access **anytime**
  - scientific_calculator **deny location** access **anytime**
  - to_do_list **allow location** access at **grocery stores**
  - file_explorer **deny storage** access at **home**

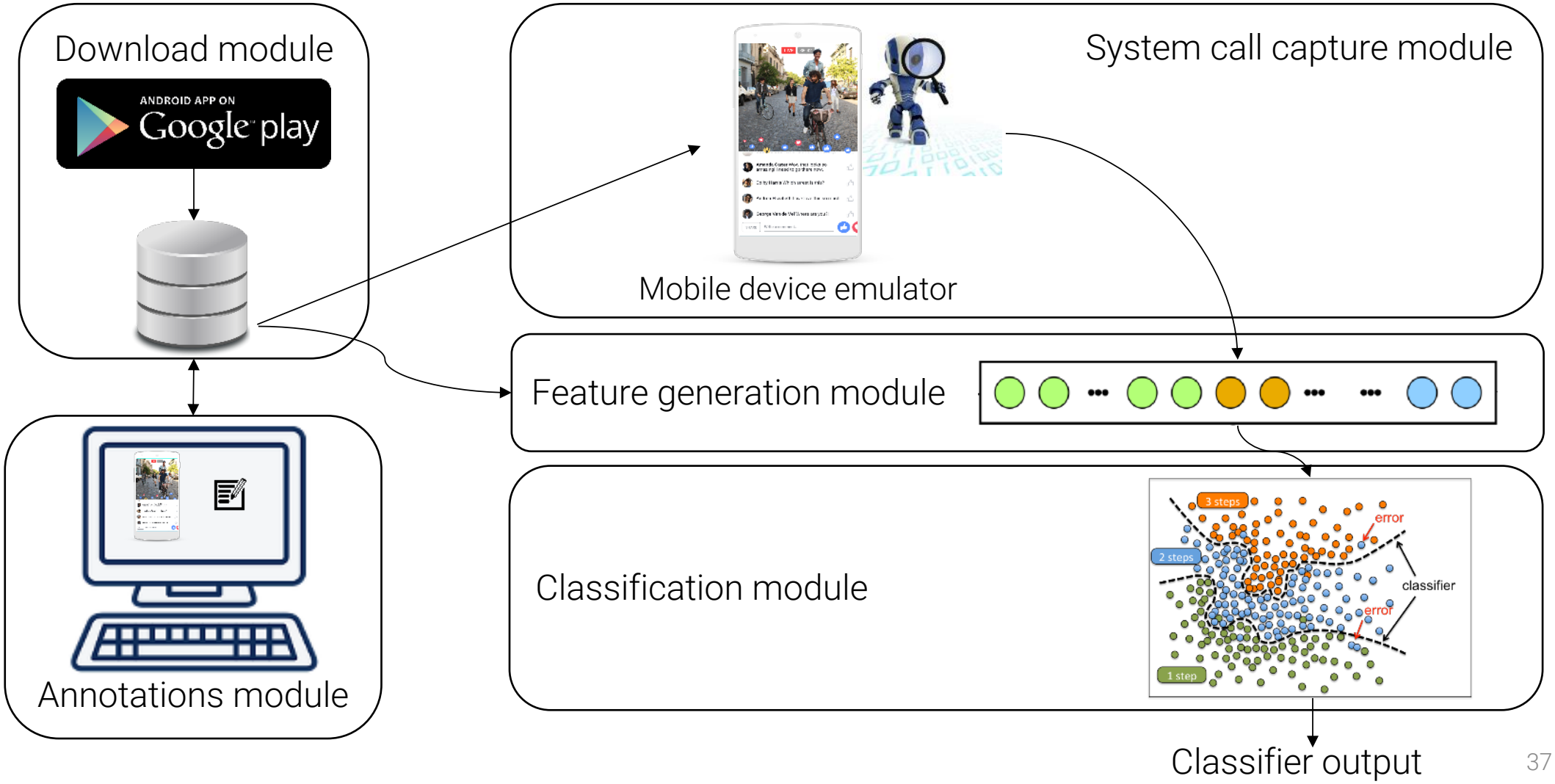
Dataset distribution





10 annotated categories, 20 Google Play categories – 75% tool and productivity

Pipeline for app classification



Crowdsourced policy generation

- Xprivacy – Privacy application on Android
- Collects crowd feedback on app permission
- 21 million rules for 17k apps
- Majority voting to generate policies as per app classification

Experimental setup

- 1560 apps
- 534 successfully executed
- **strace** used to capture system calls
  - can only be used on emulator
- Monkeyrunner UI/Application exerciser tool
- Android 6.0.1 December 2015 build
- Round 1: 1-hot and TF-IDF weight vectors
  - Best F1-score with MLP at 0.44
- Round 2: Uni, Bi, Tri, Quad and All gram sequences generated
  - Best F1-score with bigram models and KNN 0.50

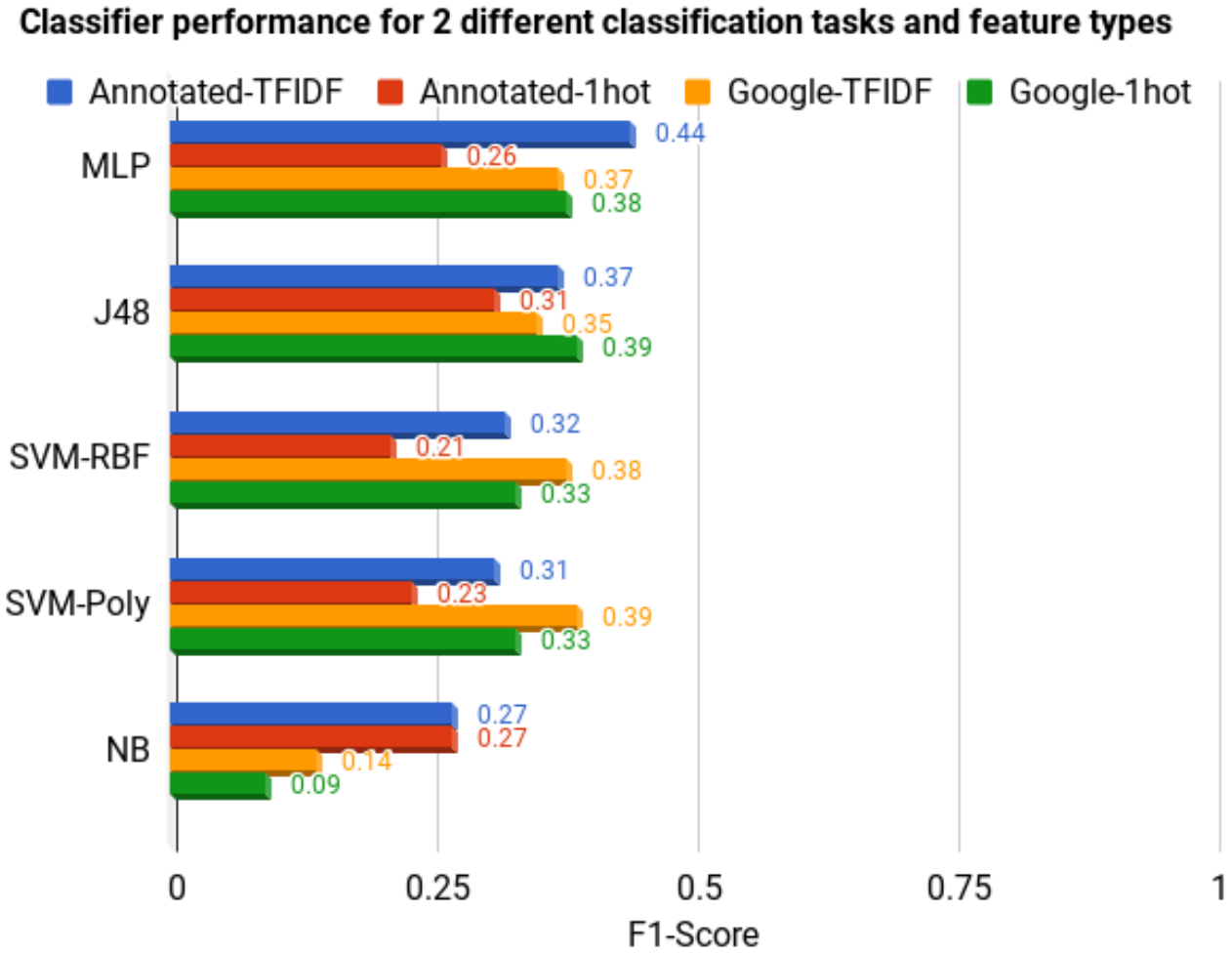
```
open("/var/log/cups/page_log", O_RDWR|O_CREAT|O_APPEND, 0666) = 6
fstat(6, {st_mode=S_IFREG|0640, st_size=0, ...}) = 0
lseek(6, 0, SEEK_END)                   = 0
fcntl(6, F_GETFD)                       = 0
fcntl(6, F_SETFD, FD_CLOEXEC)           = 0
fchown(6, 0, 4)                         = 0
fchmod(6, 0640)                         = 0
open("/etc/papersize", O_RDONLY)        = 7
fstat(7, {st_mode=S_IFREG|0644, st_size=3, ...}) = 0
mmap(NULL, 4096, PROT_READ|PROT_WRITE, MAP_PRIVATE|MAP_ANONYMOUS, -1, 0) = 0x7f063de96000
read(7, "a4\n", 4096)                   = 3
close(7)                                = 0
open("/var/cache/cups/job.cache.N", O_WRONLY) = -1 ENOENT (No such file or directory)
open("/var/cache/cups/job.cache.N", O_WRONLY|O_CREAT|O_EXCL, 0666) = 7
fstat(7, {st_mode=S_IFREG|0644, st_size=0, ...}) = 0
ftruncate(7, 0)                         = 0
fcntl(7, F_GETFD)                       = 0
fcntl(7, F_SETFD, FD_CLOEXEC)           = 0
fchown(7, 0, 7)                         = 0
fchmod(7, 0640)                         = 0
write(7, "# Job cache file for CUPS v1.7.2"..., 64) = 64
```

Evaluation: Heimdall app analytics

Research Question 1

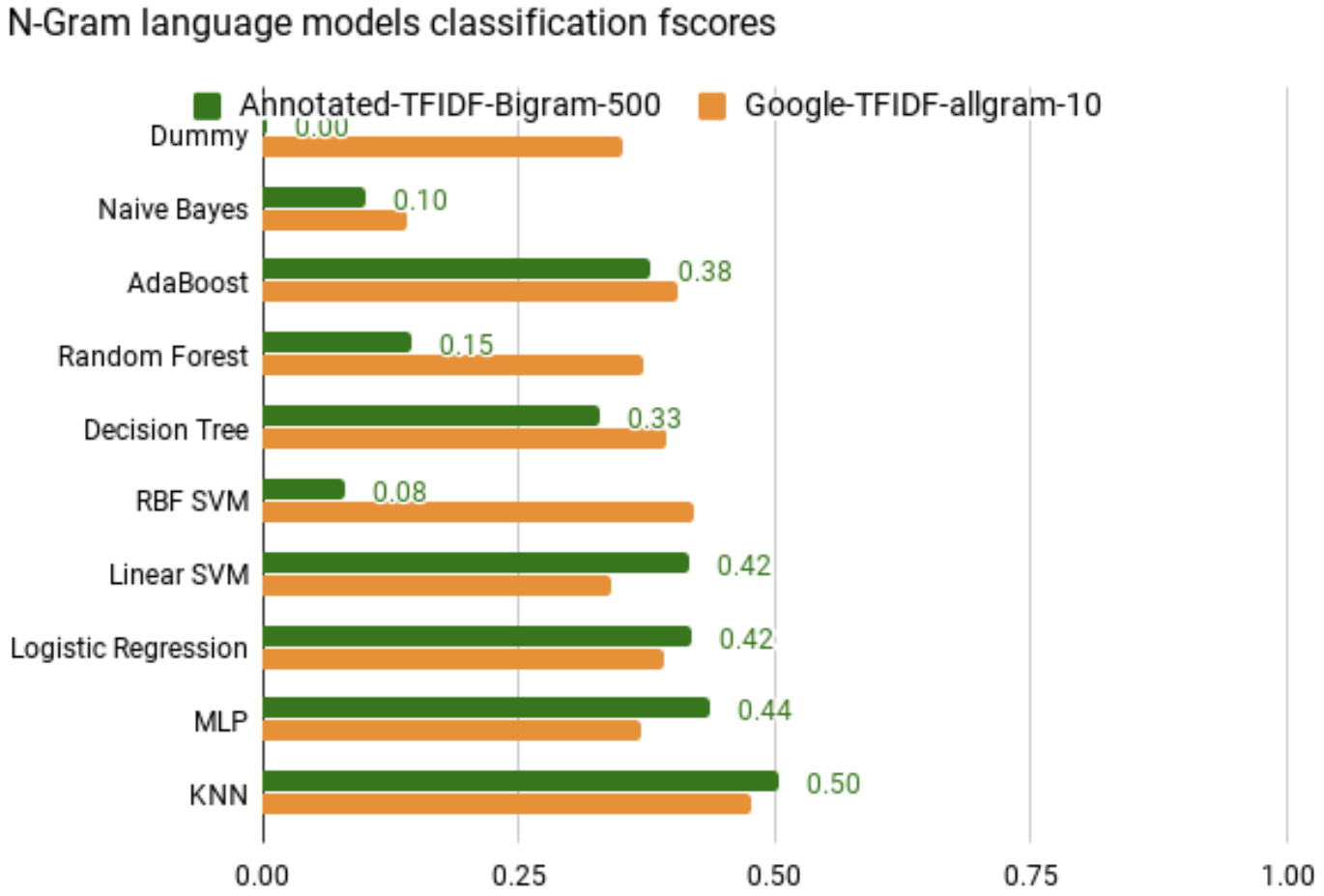
Can system calls be used as features to classify mobile applications into their behavioral classes?

Results: System call analysis round 1



	MLP	J48	SVM-RBF	SVM-Poly	NB
Annotated-TFIDF	0.44	0.37	0.32	0.31	0.27
Annotated-1hot	0.26	0.31	0.21	0.23	0.27
Google-TFIDF	0.37	0.35	0.38	0.39	0.14
Google-1hot	0.38	0.39	0.33	0.33	0.09

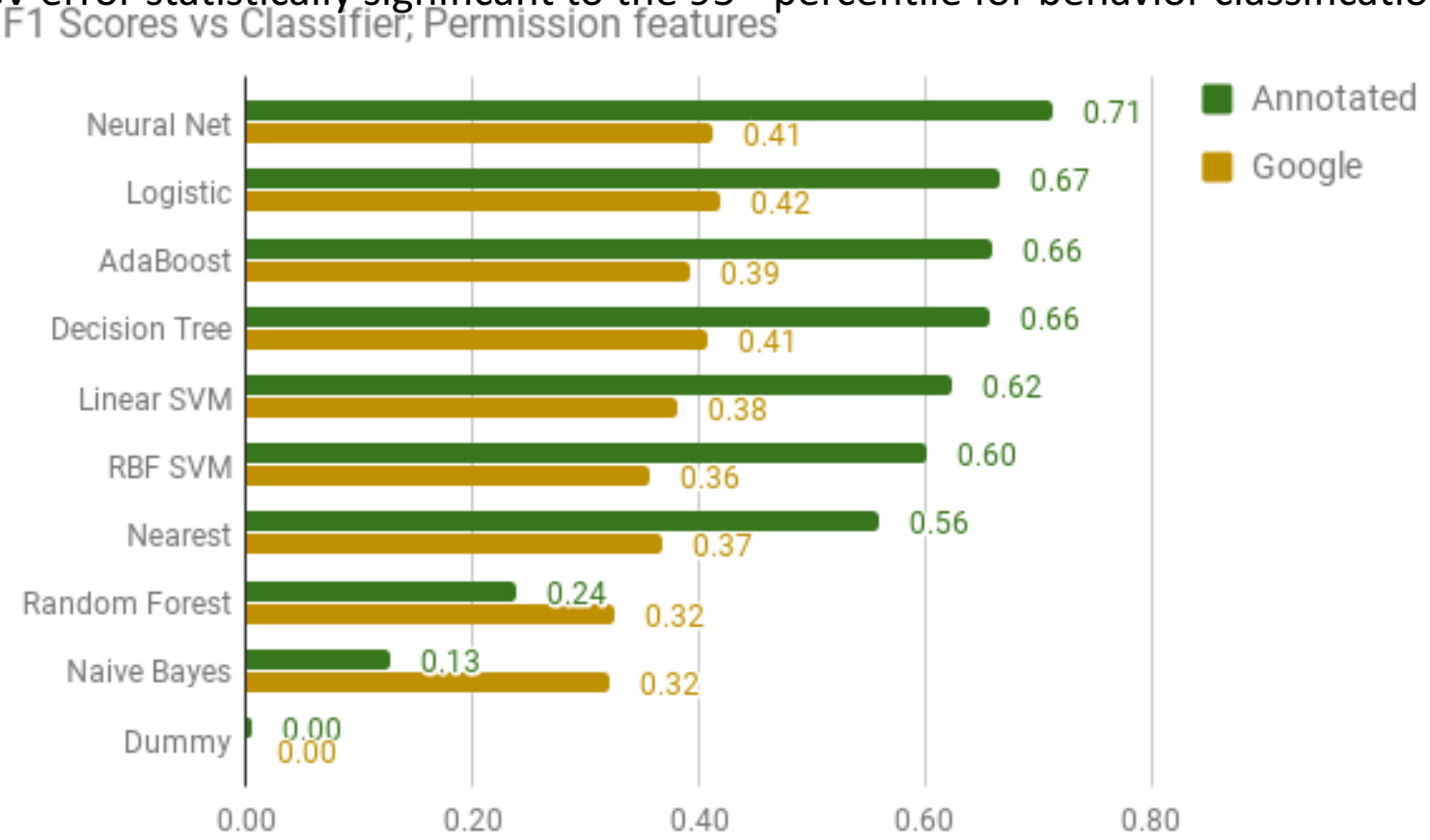
Results: System call analysis round 2



	KNN	MLP	Logistic	Lin-SVM	RBF-SVM	J48	Forest	AdaBoost	NB	Dummy
Annotated TFIDF bigram 500	0.50	0.44	0.42	0.42	0.08	0.33	0.15	0.38	0.10	0.00
Google TFIDF all gram 10	0.48	0.37	0.39	0.34	0.42	0.39	0.37	0.40	0.14	0.35

Results: Static permission features

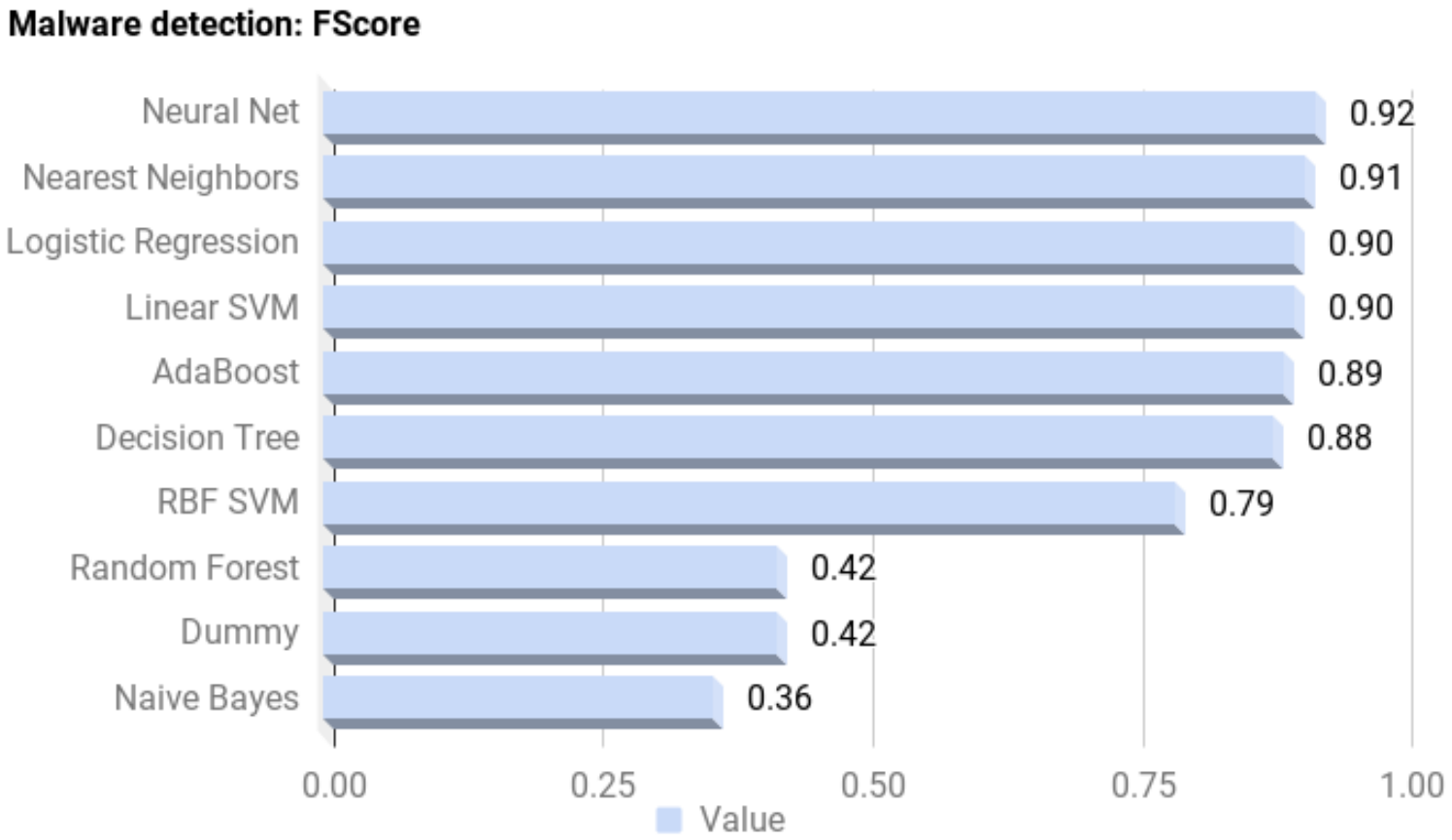
10 fold CV error statistically significant to the 95th percentile for behavior classification



	MLP	Logistic	AdaBoost	J48	Lin-SVM	RBF-SVM	KNN	Forest	NB	Dummy
Annotated	0.71	0.67	0.66	0.66	0.62	0.60	0.56	0.24	0.13	0.00
Google	0.41	0.42	0.39	0.41	0.38	0.36	0.37	0.32	0.32	0.00

The value of *t* is -5.461092. The value of *p* is 1.3E-05. The result is significant at $p \leq 0.01$.

Results: Malware detection

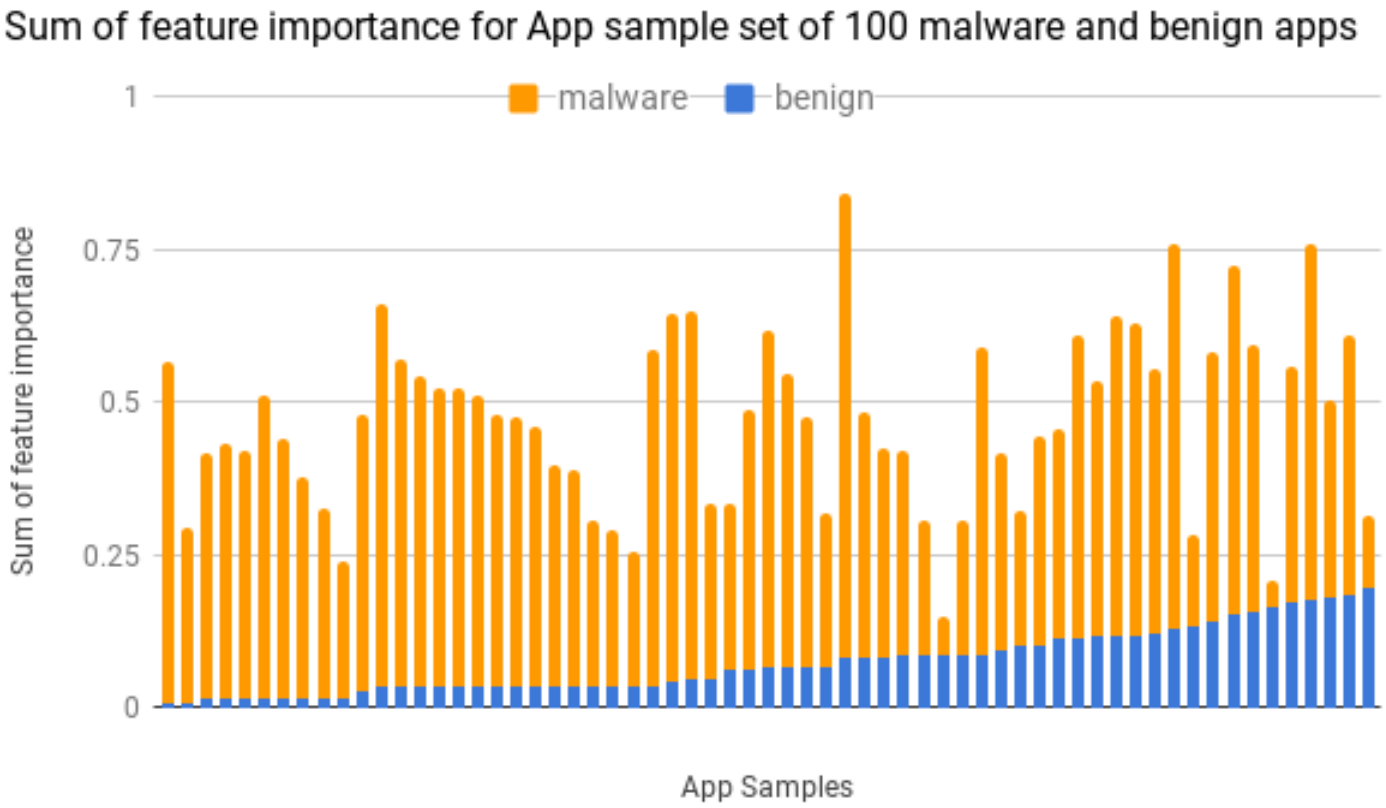


	MLP	KNN	Logistic	Lin-SVM	AdaBoost	J48	RBF-SVM	Forest	Dummy	NB
Malware detection	0.92	0.91	0.90	0.90	0.89	0.88	0.79	0.42	0.42	0.36

The value of *t* is -13.868062. The value of *p* is < 0.00001. The result is significant at $p \leq 0.01$.

Feature importance

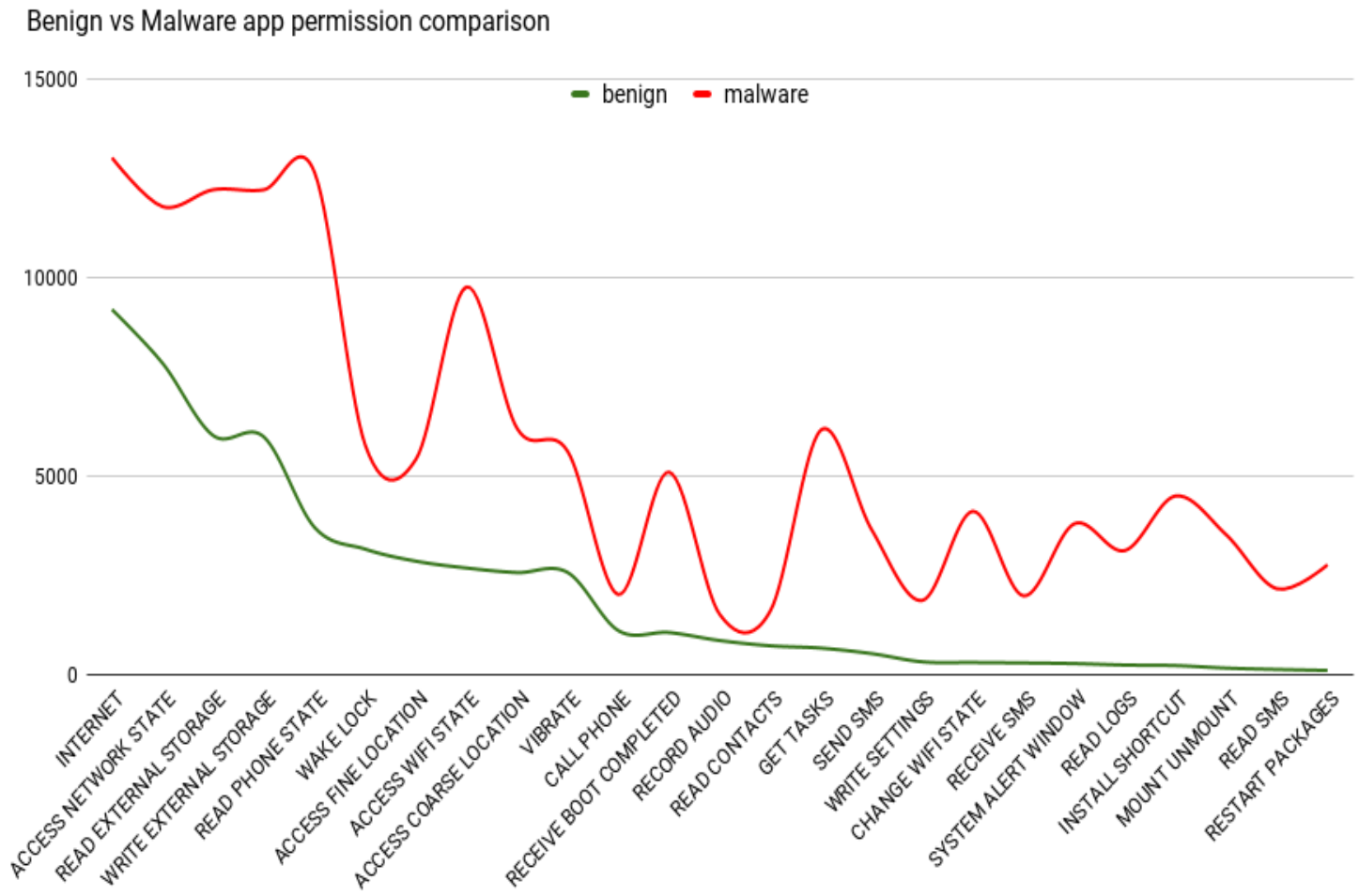
- App risk: Sum of computed feature importance from malware detection task
- Sample set of 100 apps out of 15k malware samples from the Android genome project, Virus share, Drebin dataset
- High sum for malwares
- Can be used directly for classification



Top malware characteristics

Features that critical for malware detection include

- access running apps
- access wifi state
- change wifi state
- see what's on screen
- mount filesystems
- relaunch applications



Evaluation: Mithril Mobile Middleware

Research Question 2

Given an initial policy P and user goal policy P', can violation metric be used to determine the completion of the capture process?

Use Violation Metric or Policy Precision (range [0,1])

$$VM = \frac{TV}{FV + TV}$$

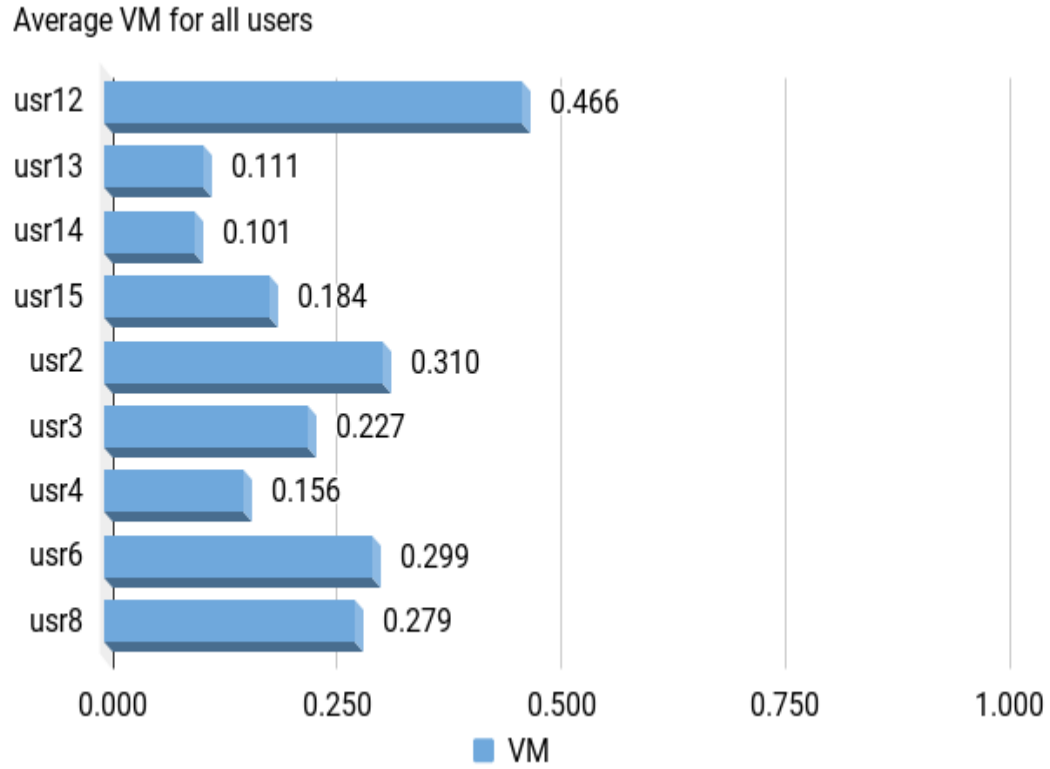
User study: Stats

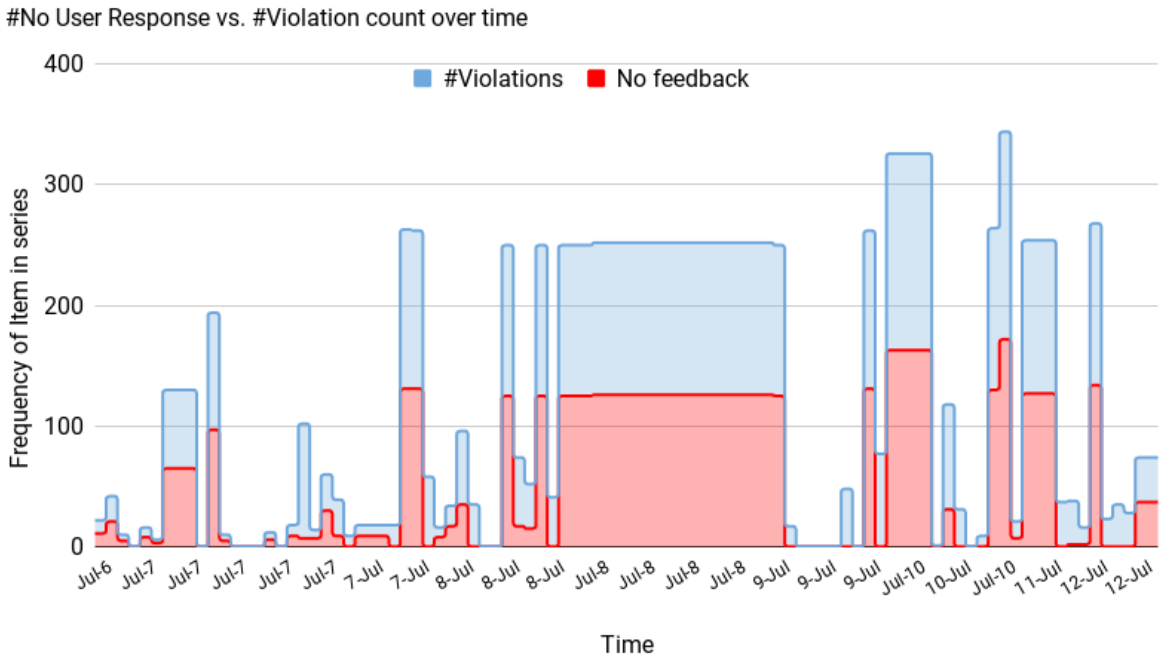
- Number of participants: 24 (graduate students from the department)
- Max number of policies Round 1: 3200
- Max number of policies Round 2: 800
- Average number of apps per user: 48
- Total stats for two round of user study

	#Users	#Violations	#TV	#FV
Round 1	14	778	228	550
Round 2	10	347	300	47

Key observation: Round 2 has more true violations and less false violations

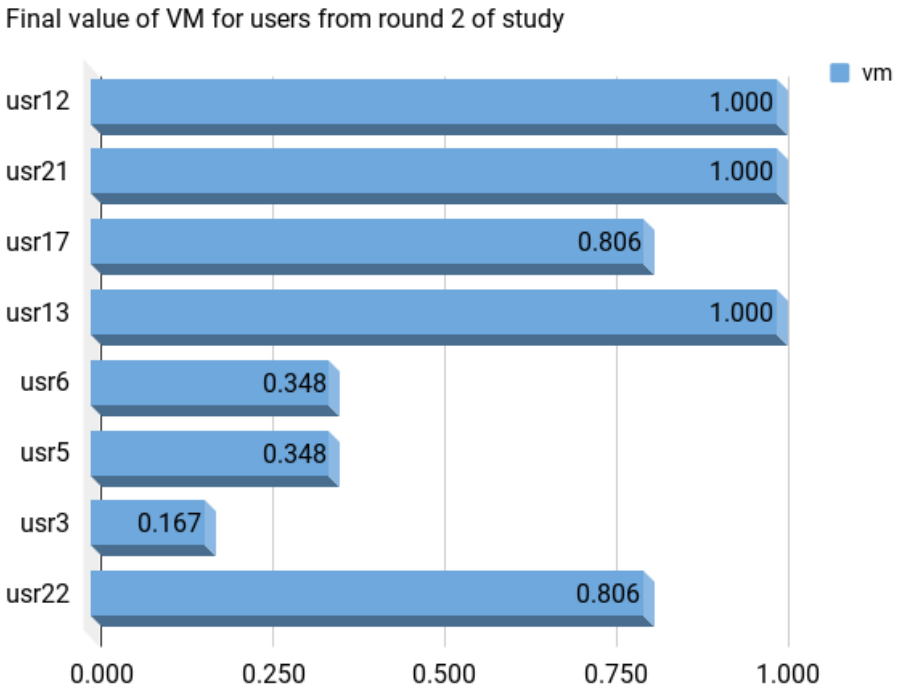
User study round 1

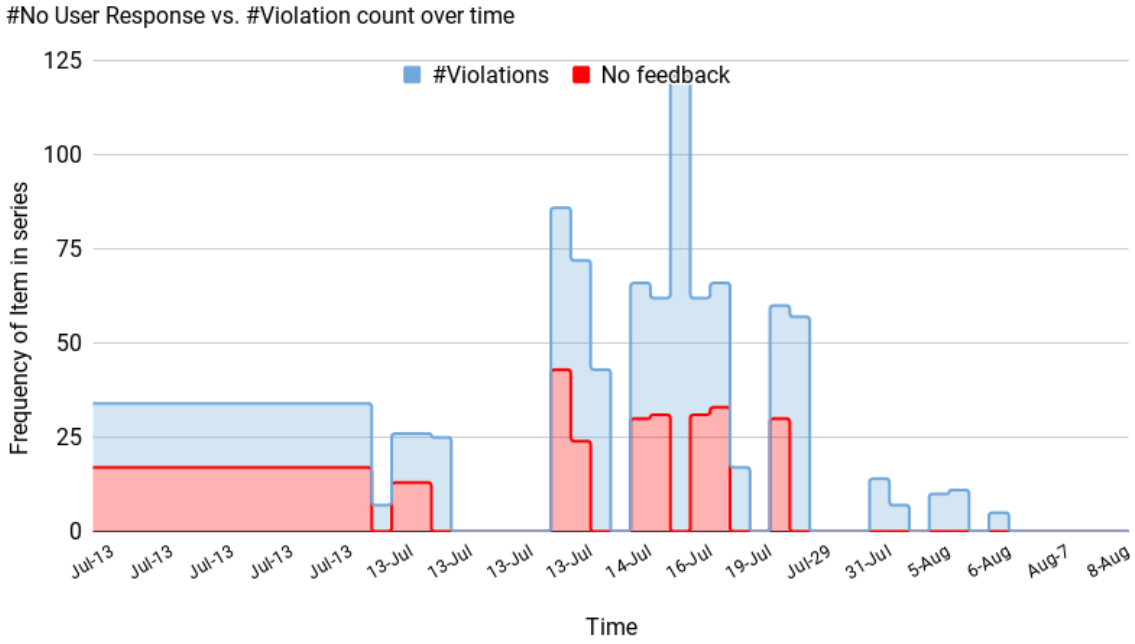




- Strategy used: Default deny policy
- Key lesson: Default deny causes issues with legitimate app activity

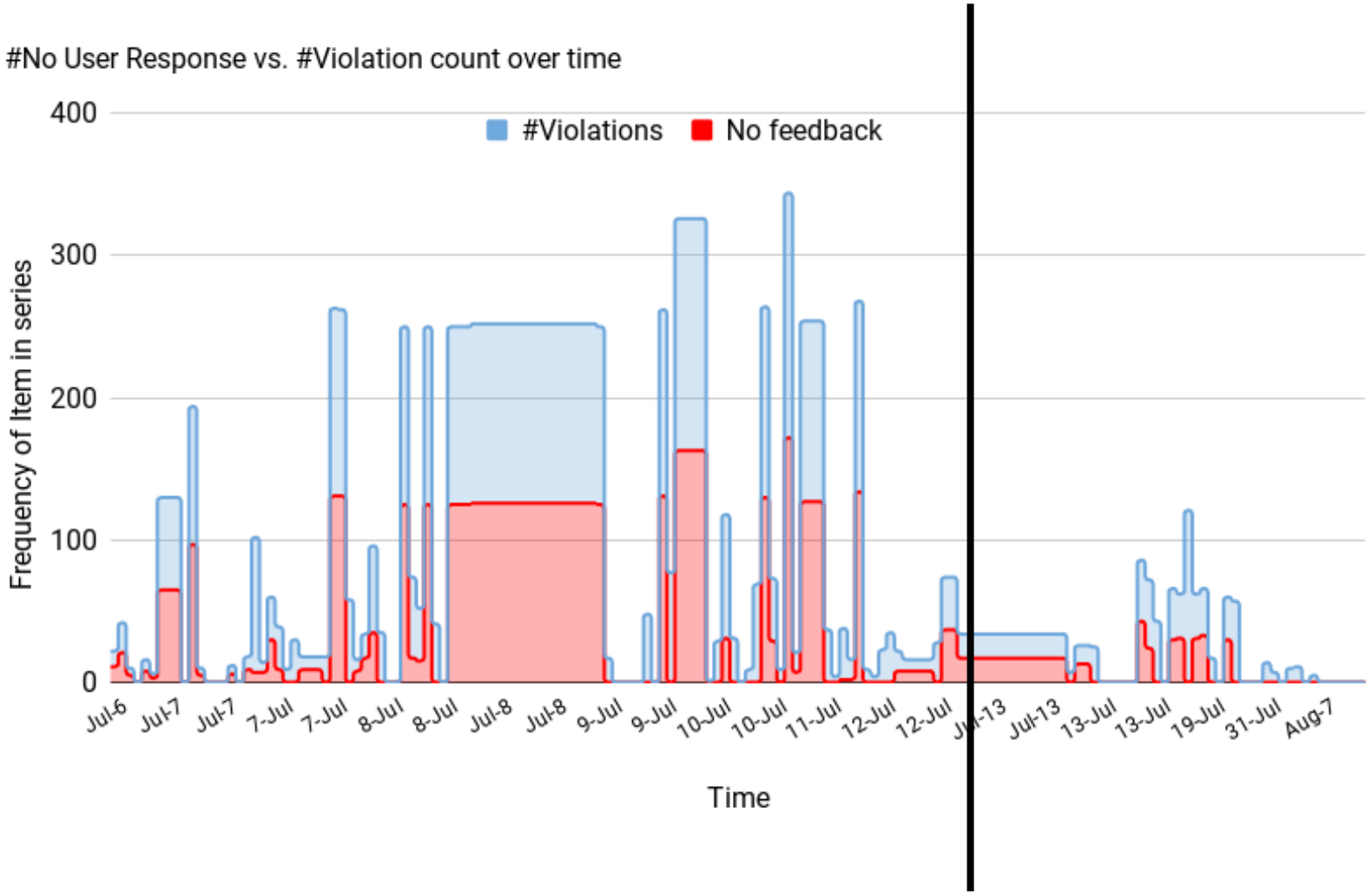
User study round 2



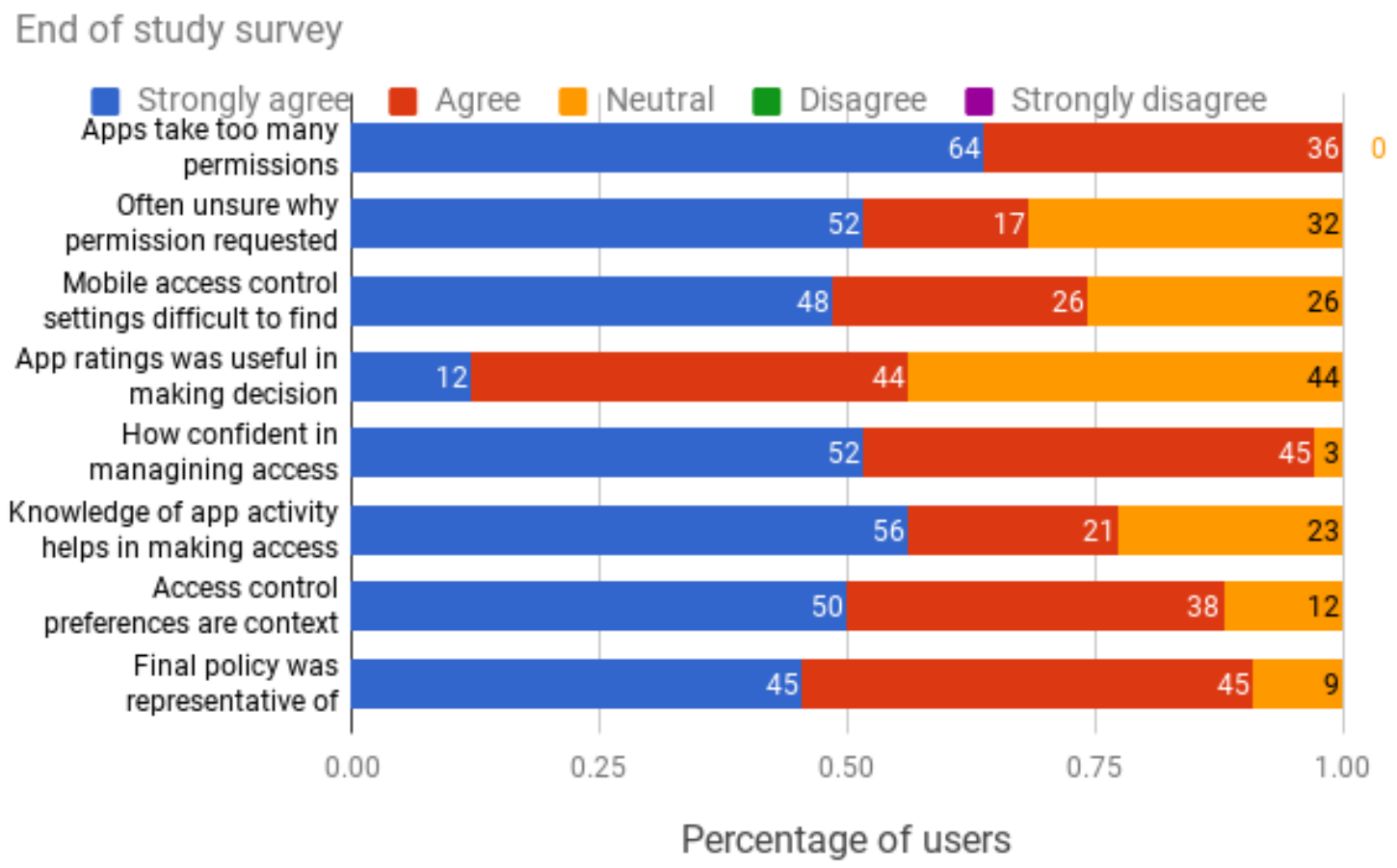


- Strategy used: Crowdsourced policy
- Key lesson: Crowdsourced policy creates less violations thus less user interaction required but still results in fairly high values of Policy precision or Violation metric

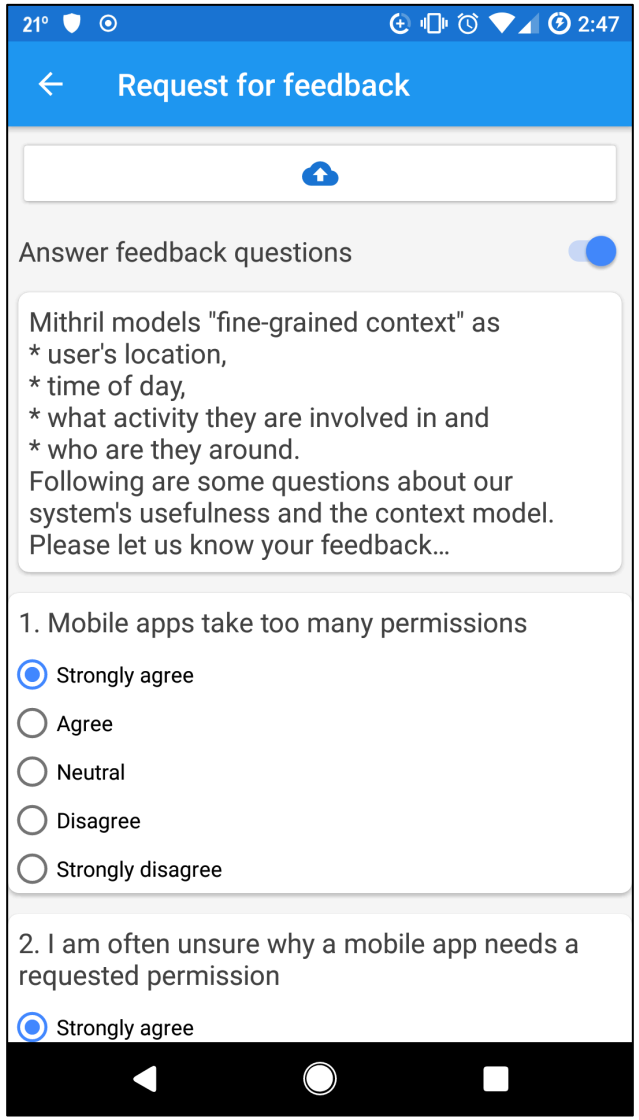
Reduction in required user interaction



End of study survey



5 point Likert scale survey
Average rating received out of 5 – 4.3



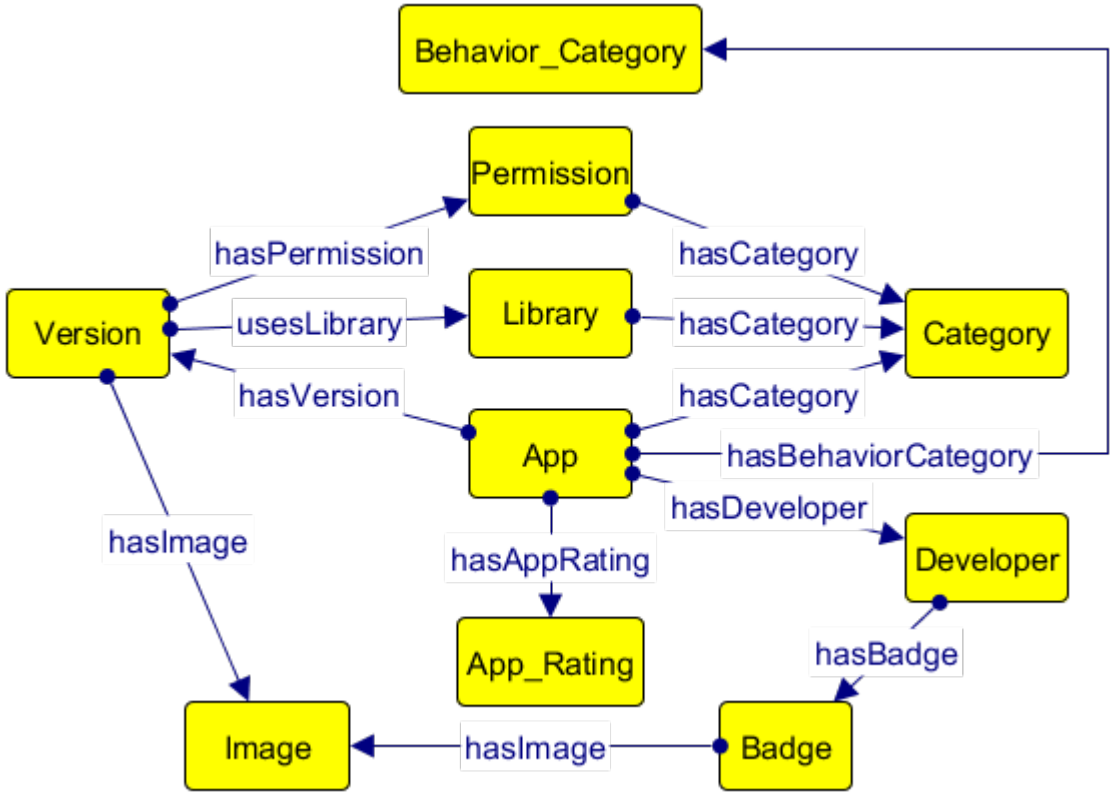
User feedback

User A: “…There is a difficulty understanding the time of old notification. Permissions requested in prior day can no longer be configured since the exact time unclear. What does allowed ignore and running mean? All in all its a good starting effort and the setting requires more explanation and further ease to configure.”

User B: “... not sure why wikipedia needs my contacts, call log and calendar details. I was happy that mithiril allows context modeling. …, I would not mind if waze asks for contact information and call log access when I am driving. ... Having context is helpful. I also like that number of feedbacks asked by mithiril reduced overtime. … would prefer to add context … customize button … in same window, …”

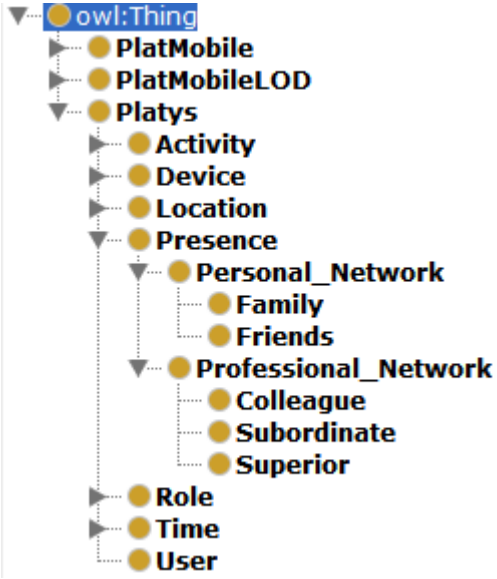
Extending Mobipedia Ontology

https://mobipedia.science



Building a mobile applications knowledge base for the linked data cloud; MoDeST'15

Presence context: Physical Web



- Nearby Messages API
- Beacons store organizational context:
  - sitsIn object property defines where people are located in the organization

Semantic knowledge and privacy in the physical web, PrivOn'16

Key contributions and conclusions

- Mithril: Semi-automated access control approach that
  - Application monitoring using mobile middleware leads to better policies
    - “violation metric” to determine completion of policy capture
    - Reduced user interaction required – Good initial policy required; curated from application analytics
    - Users find Mithril a usable access control system
    - Custom ROM built for executing context-dependent policies
  - “application behavior” for creating curated policies
    - Dynamic features: System calls produce fair classification precision and recall
    - Additional features over system calls are required
    - Static features: Permissions improve classification precision recall for behavior classification
    - Application risks can be computed using features important to malwares
  - …captured better access control policies that are fine-grained and context-dependent.

Future research goals

- Combine static, dynamic and network features to detect application behavior
- Incorporate user feedback from survey
- Use computed risk to completely block “too unsafe” apps
- Perform study of things that were allowed in policy and blocked in OS

Thesis statement – Revisited

A semi-automated approach that combines *mobile application analysis* with *violation monitoring techniques* can *reduce* the amount of *user interaction required* in *capturing better access control policies* that are fine-grained and context-dependent.

Publications

- Prajit Kumar Das, Anupam Joshi, and Tim Finin. “Capturing policies for fine-grained access control on mobile devices”. In *2nd IEEE International Conference on Collaboration and Internet Computing (CIC 2016)*, Pittsburgh, PA, USA, November 1-3, 2016, November 2016.
- Prajit Kumar Das, Anupam Joshi, and Tim Finin. “App behavioral analysis using system calls”. In *2017 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS): MobiSec 2017: Security, Privacy, and Digital Forensics of Mobile Systems and Networks (INFOCOM17 WKSHPS MobiSec 2017)*, Atlanta, USA, May 2017.
- Prajit Kumar Das, Dibyajyoti Ghosh, Pramod Jagtap, Anupam Joshi, and Tim Finin. “Preserving User Privacy and Security in Context-Aware Mobile Platforms”, chapter 8, pages 166--193. *IGI Global*, October 2016.
- Prajit Kumar Das, Abhay Kashyap, Gurpreet Singh, Cynthia Matuszek, Tim Finin, and Anupam Joshi. “Semantic knowledge and privacy in the physical web”. In Proceedings of the *4th Workshop on Society, Privacy and the Semantic Web - Policy and Technology (PrivOn 2016) co-located with the 15th International Semantic Web Conference (ISWC 2016)*, Kobe, Japan, October 17, 2016., volume ISWC 2016, 2016.
- Primal Pappachan, Roberto Yus, Prajit Kumar Das, Sharad Mehrotra, Tim Finin, and Anupam Joshi. “Mobipedia: Mobile applications linked data”. In Proceedings of the *ISWC 2015 Posters & Demonstrations Track co-located with the 14th International Semantic Web Conference (ISWC-2015)*, Bethlehem, PA, USA, October 11, 2015., volume 1486, pages 2--5. CEUR Workshop Proceedings (CEUR-WS.org), 2015.
- Prajit Kumar Das, Anupam Joshi, and Tim Finin. “Energy efficient sensing for managing context and privacy on smartphones”. In Proceedings of the *1st Workshop on Society, Privacy and the Semantic Web - Policy and Technology (PrivOn 2013) co-located with the 12th International Semantic Web Conference (ISWC 2013)*, Sydney, Australia, October 22, 2013., volume 1121. CEUR Workshop Proceedings (CEUR-WS.org), 2013.

- Prajit Kumar Das, Sandeep Narayanan, Nitin Kumar Sharma, Anupam Joshi, Karuna P. Joshi, and Tim Finin. Context-sensitive policy based security in internet of things. In 2016 IEEE International Conference on Smart Computing (SMARTCOMP 2016), St Louis, MO, USA, May 18-20, 2016, pages 1--6, 2016.
- Sudip Mittal, Prajit Kumar Das, Varish Mulwad, Anupam Joshi, and Tim Finin. Cybertwitter : Using twitter to generate alerts for cybersecurity threats and vulnerabilities. In International Symposium on Foundations of Open Source Intelligence and Security Informatics (FOSINT-SI 2016) co-located with the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2016), San Francisco, CA, USA, August 19-20, 2016, 2016.
- Primal Pappachan, Roberto Yus, Prajit Kumar Das, Tim Finin, Eduardo Mena, and Anupam Joshi. A semantic context-aware privacy model for faceblock. In Proceedings of the 2nd Workshop on Society, Privacy and the Semantic Web - Policy and Technology (PrivOn 2014) co-located with the 13th International Semantic Web Conference (ISWC 2014), Trento, Italy, October 20, 2014., volume 1316, pages 64--72. CEUR Workshop Proceedings (CEUR-WS.org), 2014.
- Roberto Yus, Primal Pappachan, Prajit Kumar Das, Tim Finin, Anupam Joshi, and Eduardo Mena. Semantics for privacy and shared context. In Proceedings of the 2nd Workshop on Society, Privacy and the Semantic Web - Policy and Technology (PrivOn 2014) co-located with the 13th International Semantic Web Conference (ISWC 2014), Trento, Italy, October 20, 2014., volume 1316. CEUR Workshop Proceedings (CEUR-WS.org), 2014.
- Roberto Yus, Primal Pappachan, Prajit Kumar Das, Eduardo Mena, Anupam Joshi, and Tim Finin. Demo: Faceblock: privacy-aware pictures for google glass. In The 12th Annual International Conference on Mobile Systems, Applications, and Services, MobiSys'14, Bretton Woods, NH, USA, June 16-19, 2014, page 366, 2014.
- Prajit Kumar Das, Dibyajyoti Ghosh, Anupam Joshi, and Tim Finin. Acm hotmobile 2013 poster: an energy efficient semantic context model for managing privacy on smartphones. Mobile Computing and Communications Review, 17(3):34--35, 2013.
- Prajit Kumar Das, Anupam Joshi, and Tim Finin. Energy efficient sensing for managing context and privacy on smartphones. In Proceedings of the 1st Workshop on Society, Privacy and the Semantic Web - Policy and Technology (PrivOn 2013) co-located with the 12th International Semantic Web Conference (ISWC 2013), Sydney, Australia, October 22, 2013., volume 1121. CEUR Workshop Proceedings (CEUR-WS.org), 2013.

Questions