Context-Aware Policies for Privacy and Security

Tim Finin
University of Maryland, Baltimore County

UMBC HON 300 / CMSC491, Spring 2012

Joint work with Anupam Joshi, Laura Zavala, Radhika Dharurkar, Pramod Jagtap, Dibyajyoti Ghosh and Amey Sane

Adapted from a presentation from SDMSM 2012 : ICDE Workshop on Secure Data Management on Smart Phones and Mobile Devices

UMBC ebiquity

http://ebiq.org/r/340

Four Billion Little Brothers?

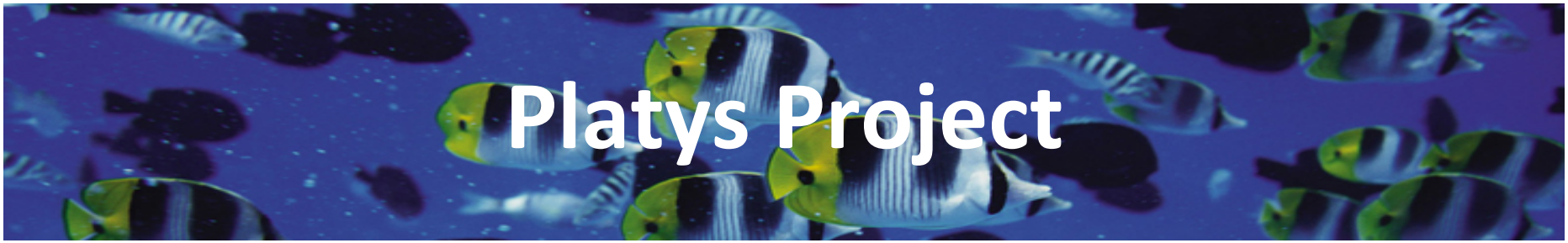
- Katie Shilton, Four Billion Little Brothers?: Privacy, Mobile Phones, and Ubiquitous Data Collection, Communications of the ACM, Vol. 52 No. 11, Pages 48-53 (http://bit.ly/4blitbros)
- Participatory sensing technologies can improve our lives and our communities, but at what cost to our privacy?
- Your smartphone knows more about you than your spouse or your mother (or even yourself?)

The Story

- Smart mobile devices know a great deal about their users including their current context
- Acquiring and reasoning about this knowledge will enable them to provide better services
- Sharing the information with other users, organizations and service providers can also be beneficial
- Context-aware policies can be used to limit information sharing as well as to control the actions and information access of apps

Roadmap

- Introduction
- Semantic context modeling
- Context / situation recognition
- Context-aware policies for privacy and security
- Ongoing and future work
- Conclusion



Platys Project

- Part of an NSF collaborative project with NC state (M. Singh & I. Rhee) and Duke (R. Choudhury)
- Overall theme: enable smartphones to learn and exploit a richer notion of *place*
  - Place is more than GPS coordinates
  - Conceptual places include people, devices, activities, purpose, roles, background knowledge, etc.
  - Use this to provide better services and user experience

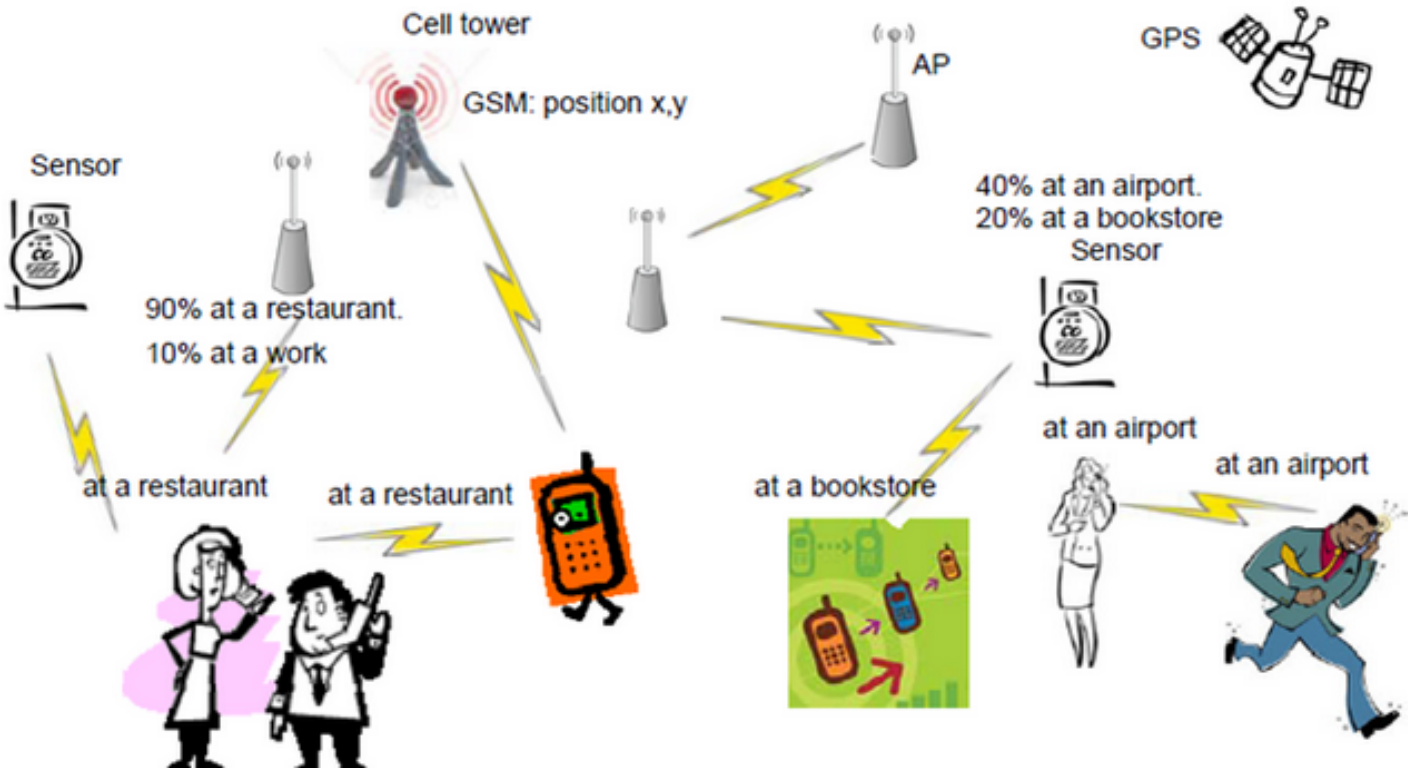
I am …

- at (37.79414, -122.39597)

vs.

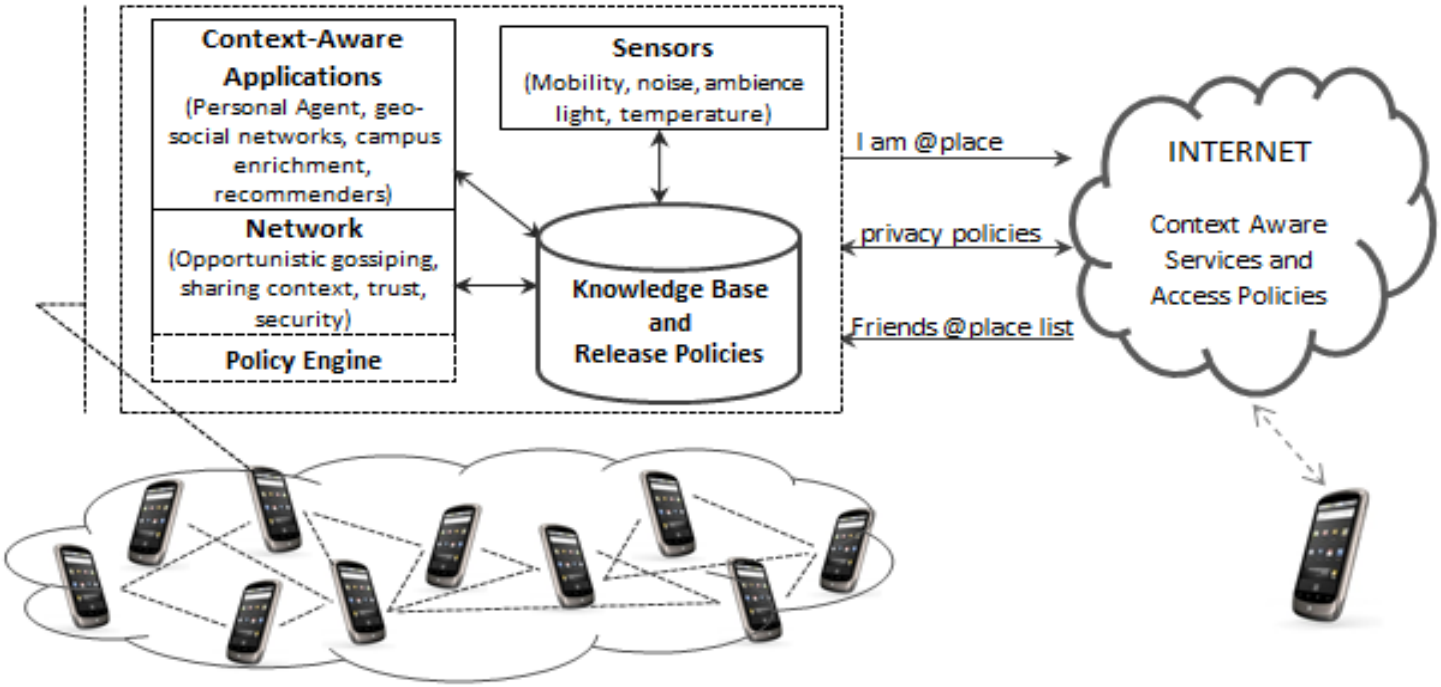
- in a hotel, the Renaissance , RLINGTON VA, …
- participating in a meeting, a workshop, the IEEE SDMSM 2012 , …
- With >10 people including Palani Kodeswaren …
- filling a speaker role, an audience role, …
- remembering I was here Monday from 09:08 to 1:45 …
- seeing many WIFI access point here: …
- …

Sharing place information



- Peer to peer communication
- Opportunistic Gossiping
- User privacy policies control sharing
- Fixed devices acquire, store, share, and summarize

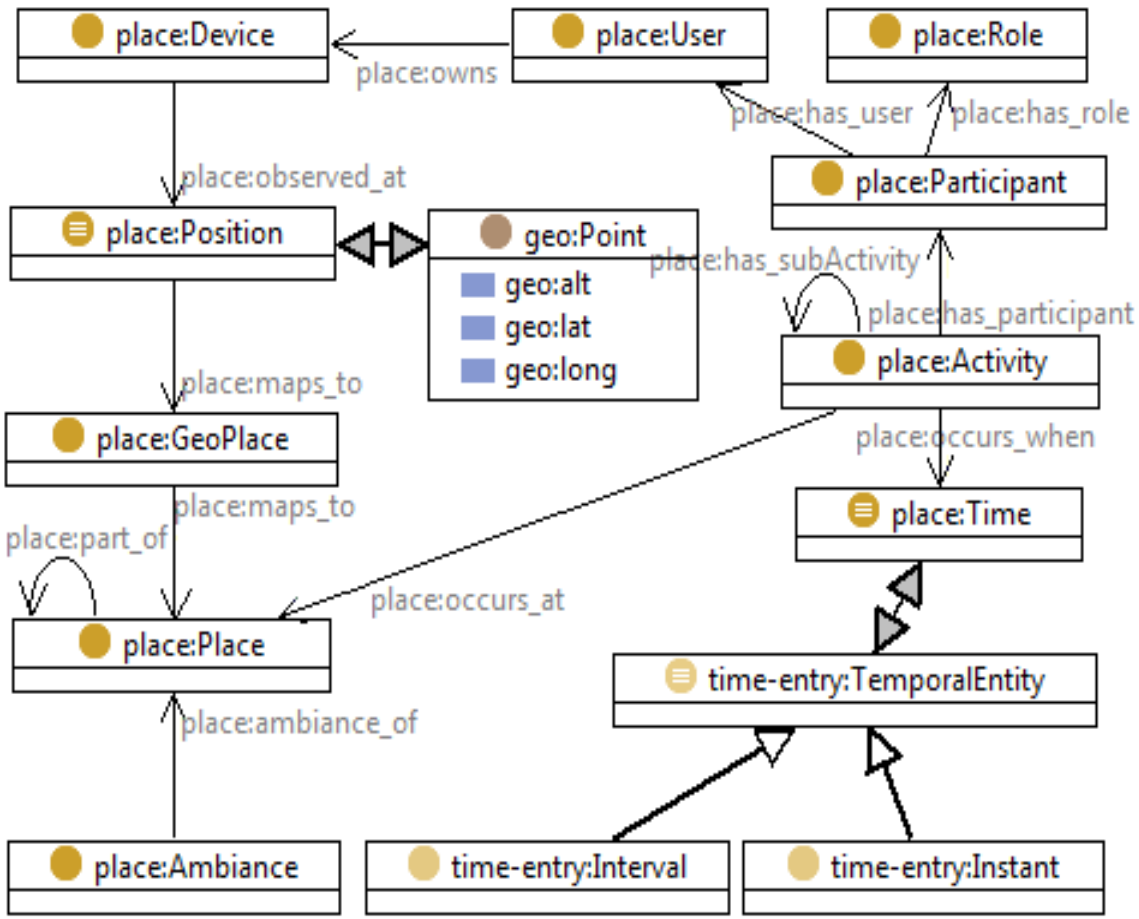
General Interaction Architecture



- Device sensors used for contextual clues
- Context RDF KB on each device
- Context shared with neighboring devices
- Devices interact directly or via Inter-net services
- Privacy policies specify user's information sharing constraints

Our Ontology

- Light-weight, upper level context ontology
- Encoded in OWL
- Centered around the concepts for: users, *conceptual places*, goe-places, activities, roles, space, and time.
- Conceptual places such as *at work* and *at home*
- Activities occur at places and involve users filling particular roles



What's an Ontology?

- Ontologies formally represent knowledge as a set of *concepts* and possible *properties* and *relationships* within a *domain*
- They provide a *vocabulary* to *describe* and *reason* about things in the domain
- Example: an ontology for a social networking environment
- Ontologies are often expressed in a specialized formal language but can also be simple object-oriented models

What's the Semantic Web?

- A set languages and protocols allowing data and knowledge to be published on the Web
  - The Web of documents (text and pictures) has made people smarter
  - Adding or embedding data supported by ontologies will make machines smarter
- Currently based on W3C standards like RDF (data), OWL (ontologies) and SPARQL (queries)
  - Microdata is a dialect being used by major search systems for a small set of ontologies in schema.org

What kinds of Reasoning?

- RDF and OWL are logic-oriented knowledge representation languages that support reasoning
- This makes them more powerful than, say, XML or relational databases

```
@prefix rdfs: <http://www.....>.
@prefix : <genesis.n3>.
:parent a rdf:Property;
    rdfs:domain :Person;
    rdfs:range :Person.
:mother
    rdfs:subPropertyOf :parent;
    rdfs:domain :Woman;
    rdfs:range :Person.
:eve :mother :cain.
```

```
:Person a rdfs:Class.
:Woman a rdfs:Class.
:Woman rds:subClassOf :Person.
:mother a rdfs:Property.
eve a Person;
      a Woman;
      :parent :cain.
:cain a Person.
```

The Device's Context KB

- RDF KB on device which conforms to the ontology
- Imports FOAF and GeoNames onto-logies
- Uses Geonames Linked Data to assert further spatial knowledge in the KB
- RDF for tools, standards, data, interoperabilty…

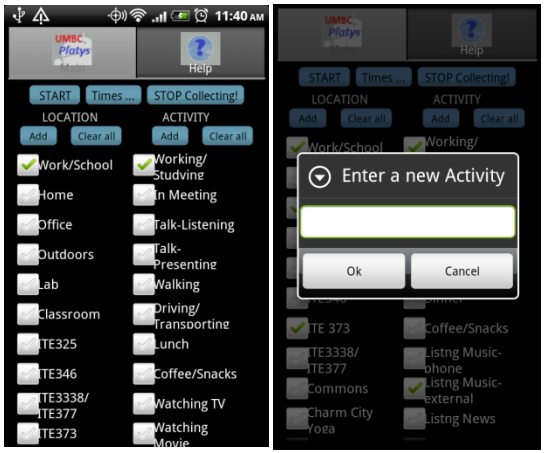
```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix place: < http://ebiquity.umbc.edu/ontologies/platys#>.
@prefix geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>.
@prefix kb: <http://example.org/kbdevice/>.
@prefix gn: < http://www.geonames.org/ontology#>.
---
kb:droid1 place:observed_at kb:anon01f
kb: anon01f rdf:type place:Position
kb: anon01f geo:lat 39.253525
kb: anon01f geo:long -76.710706
```

```
<gn:Feature rdf:about="http://sws.geonames.org/4372143/">
<gn:name>UMBC</gn:name>
  <wgs84_pos:lat>39.25543</wgs84_pos:lat>
  <wgs84_pos:long>-76.71168</wgs84_pos:long>
  <wgs84_pos:alt>61</wgs84_pos:alt>
  <gn:parentFeature rdf:resource="http://sws.geonames.org/4347790/"/> ←Baltimore County
  <gn:parentCountry rdf:resource="http://sws.geonames.org/6252001/"/> ← United States
  <gn:parentADM1 rdf:resource="http://sws.geonames.org/4361885/"/> ← Maryland
  <gn:parentADM2 rdf:resource="http://sws.geonames.org/4347790/"/> ← Baltimore County
</gn:Feature>
```

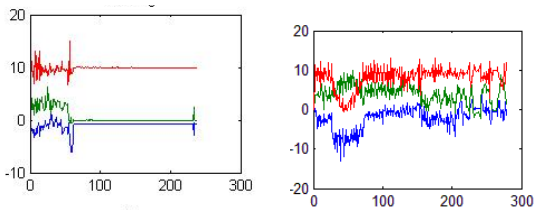
Context / situation recognition

- Focus on individual activity and place recognition
- Using smartphones as sensors we use probabilistic models for context recognition
  - noise, ambience light, accelerometer, Wifi, Bluetooth, call stats, phone settings, user calendar
- Data collection program used to collect training data to learn to recognize context
  - Five users, one month, logging TRUE activity and place attached to phone readings (noise, light, wifi, etc.)
  - Naive Bayes, decision tree, SVM, and bagging+decisiontrees

Context / situation recognition




```
23:31:00,MONDAY,False,False,False,False,True,Home,Working/Studying,False,0,0,-14.5,18.6576,-22.61,-44.8202,-2.58523,
00:02:00,TUESDAY,False,False,False,False,True,Home,Working/Studying,False,184.438,-0.568493,-12.7397,17.9576,-23.615
00:33:00,TUESDAY,False,False,False,False,True,Home,Working/Studying,False,198.958,0.09375,-11.5208,18.5931,-22.8312,
22:05:00,THURSDAY,False,False,False,False,True,Home,Sleeping,False,19,-2,-3.16667,-7.38406,24.7824,-44.0506,-0.57071
22:35:00,THURSDAY,False,False,False,False,True,Home,Sleeping,False,0,0,-4,-6.36577,23.7348,-44.5852,-0.753992,0.3503
23:06:00,THURSDAY,False,False,False,False,True,Home,Sleeping,False,21,0,-3.6,-6.37239,23.7091,-44.6825,-0.753755,0.3
23:37:00,THURSDAY,False,False,False,False,True,Home,Sleeping,False,23.9396,-1.51007,-4.48993,-7.2726,22.1923,-43.649
00:02:00,FRIDAY,False,False,False,False,True,Home,Sleeping,False,0,-1.54545,-4.25175,-7.27059,22.3573,-43.9119,-0.80
00:36:00,FRIDAY,False,False,False,False,True,Home,Sleeping,False,96.7545,-5.37273,-27.6818,6.20096,22.3808,-41.5671,
```



Feature Vector

Time, Noise level in db (avg, min, max), accel 3 axis (avg, min, max, magnitude, wifis, …

Decision Trees
Naïve Bayes
SVM

Train Classifiers

```
Afternoon = False
| Evening = False
| | Place = Outdoors: Walking
| | Place = Elsewhere: Sleeping
| Evening = True: Walking
Afternoon = True
| Wifi Id8 = True: In Meeting
| Wifi Id8 = False
| | Place = Home: Lunch
| | Place = Restaurant: Lunch
| | Place = Movie Theater: Watching
                          Movie
| | Place = Work/School: Working
| | Place = ITE346: Lunch
| | Place = Outdoors: Walking
| | Place = ITE3338/ITE377: Lunch
```

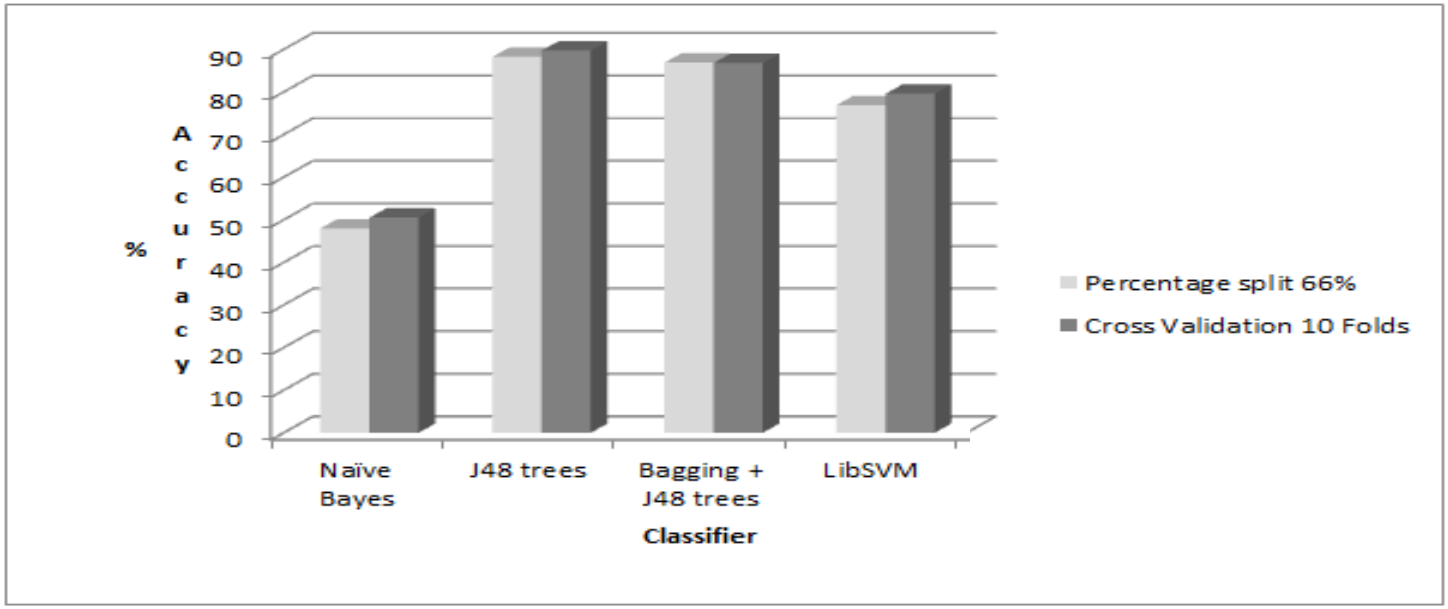
```
loc2 = '(-inf-39.17259]': Watching
Movie
loc2 = '(39.17259-39.18528]': Sleeping
loc2 = '(39.18528-39.19797]': Lunch
loc2 = '(39.24873-39.26142]': Walking
```

Evaluation Experiments

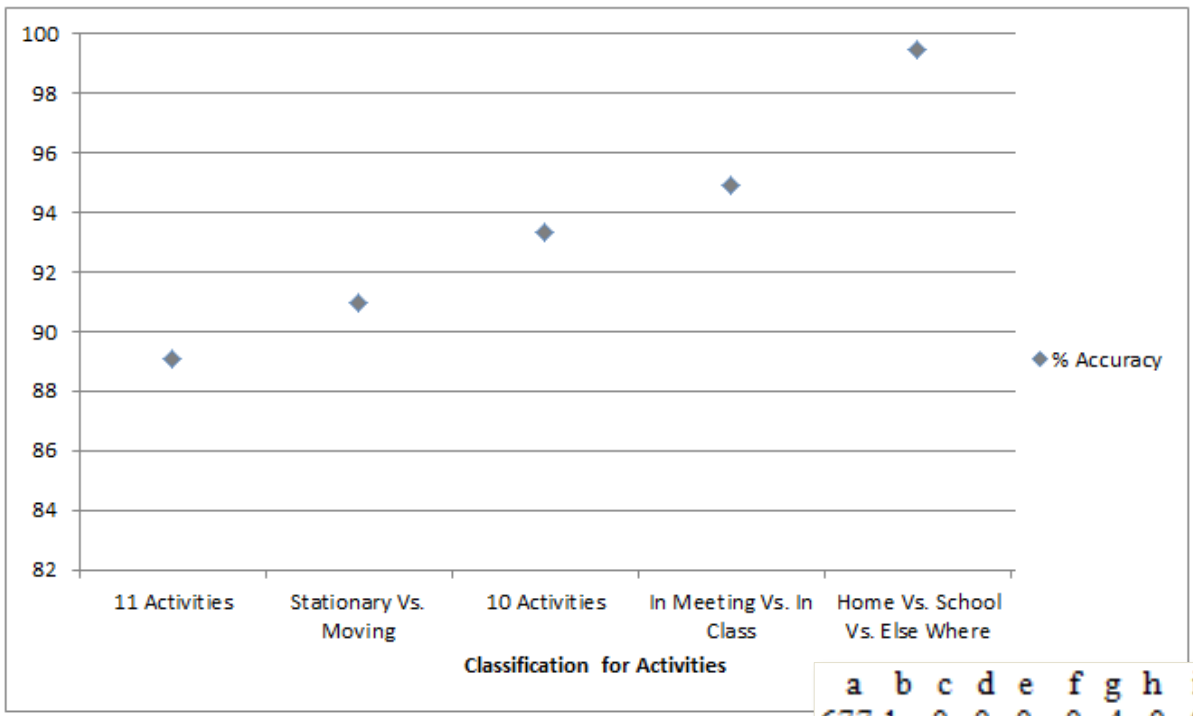
- Varying granularity level on activities
  - Motion, Stationary
  - Work, Home, Outdoors, Other
  - In meeting, in class, watching TV, reading, sleeping, etc.
- Two different schemes
  - Individual: training and testing on one person's data
  - Across users: training with one person's data and testing it with other's

Results – Comparing classifiers

- Accuracy higher for decision tree classifiers
  - Improved with bagging
- SVMs slightly below decision trees
- Weak performance of Naive Bayes



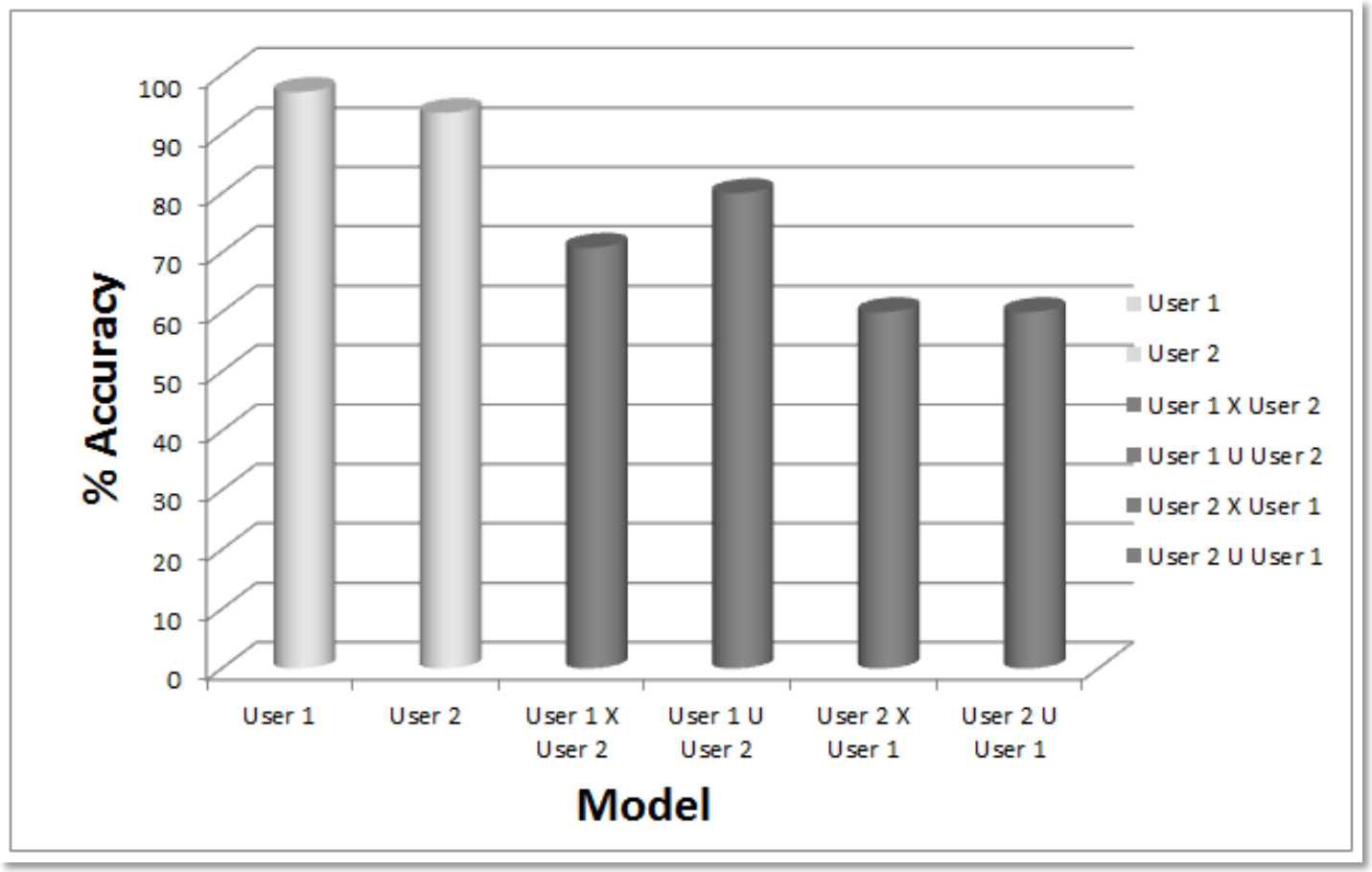
Results – Generalizing activities



a	b	c	d	e	f	g	h	i	j	k	<-- classified as
677	1	0	0	0	0	4	0	0	0	2	a = [Sleeping]
0	186	0	0	20	0	3	0	5	0	0	b = [Walking]
0	0	27	0	0	0	0	0	0	0	0	c = [In Meeting]
0	2	0	65	0	4	0	0	0	0	0	d = [Playing]
0	**37**	**0**	**0**	**37**	**0**	**0**	**0**	**4**	**0**	**0**	**e = [Driving/Transporting]**
0	0	0	2	0	146	1	0	0	2	0	f = [Class-Listening]
8	0	0	0	0	2	52	2	0	0	8	g = [Lunch]
9	**0**	**0**	**0**	**0**	**0**	**8**	**11**	**0**	**0**	**0**	**h = [Cooking]**
0	**11**	**0**	**0**	**6**	**0**	**0**	**0**	**13**	**0**	**0**	**i = [Shopping]**
0	**2**	**0**	**0**	**0**	**5**	**0**	**0**	**0**	**7**	**0**	**j = [Talk-Listening]**
5	0	0	0	0	0	1	0	0	0	34	k = [Watching Movie]

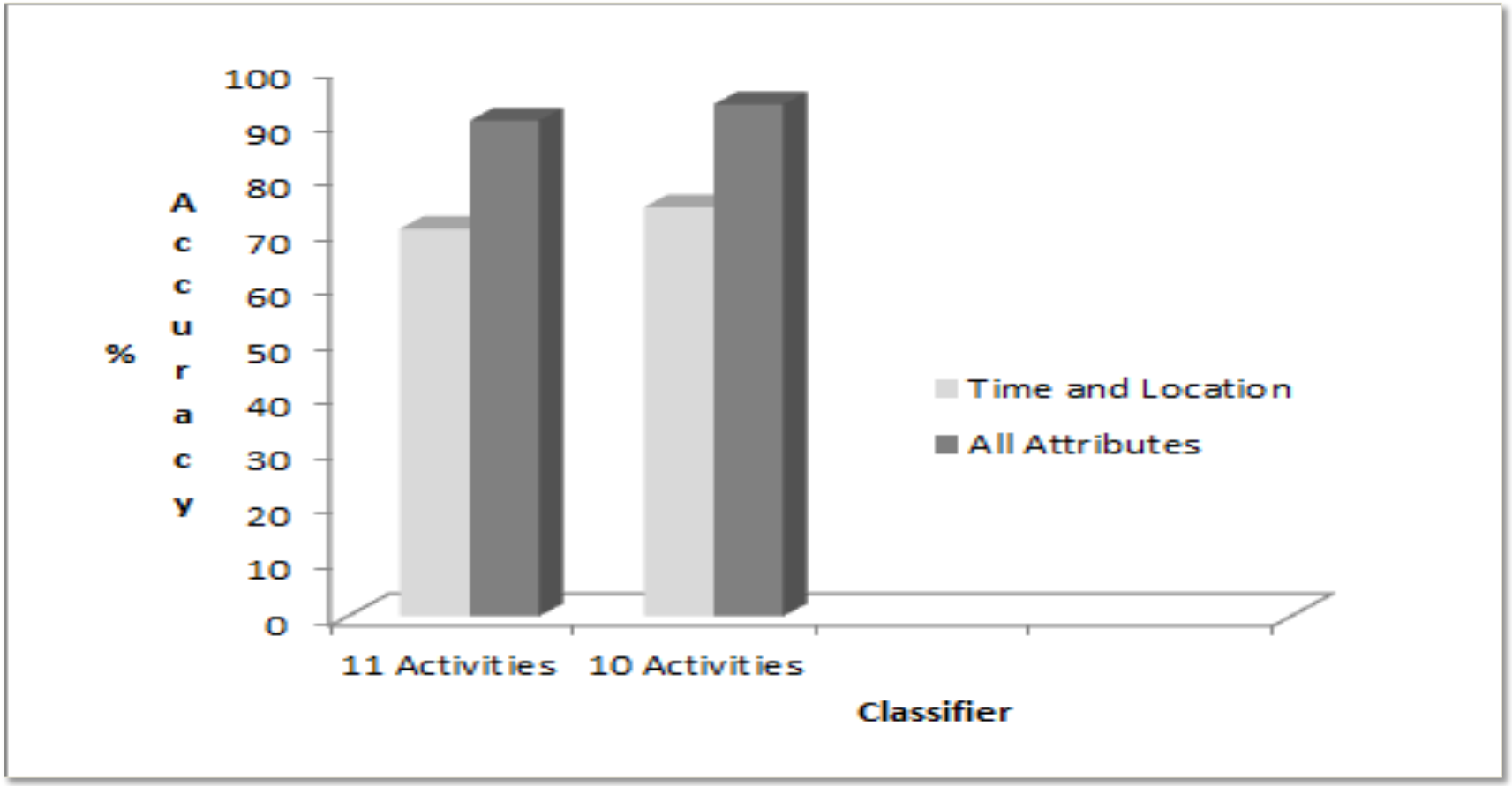
- Some states hard to distinguish (e.g., walking, shopping)
- Fewer states => greater accuracy

Results – Testing across users



Accuracy drops when using a general model or one trained on a different user. Integrating data from many users should lead to a reasonable initial model that can then be adapted with some individual's training data.

Results – Time and Location



Time and location attributes important for predicting activity, but adding more significantly improves accuracy

Ex:Decision tree output model

```
Place = Home: Sleeping
Place = ITE346: In Meeting
Place = Outdoors
| G1 = False
| | Morning = True: Walking
| | Morning = False: Driving/Transporting
| G1 = True: Walking
Place = Home
| Evening = False: Sleeping
| Evening = True
| | noise = '(-inf-28.19588]': Cooking
| | noise = '(28.19588-32.71862]': Cooking
| | noise = '(32.71862-inf)': Watching Movie
Place = Restaurant: Lunch
Place = Movie Theater: Watching Movie
Place = Elsewhere: Walking
Place = ITE325: Talk-Listening
Place = ITE3338/ITE377: In Meeting
Place = Groceries store: Shopping
```

```
Wifi Id8 = True: In Meeting
Wifi Id8 = False
| Afternoon = False
| | Evening = False: Sleeping
| | Evening = True: Walking
| Afternoon = True
| | Place = Work/School: Working
| | Place = ITE346: Lunch
| | Place = Outdoors: Walking
| | Place = Home: Lunch
| | Place = ITE3338/ITE377: Lunch
```

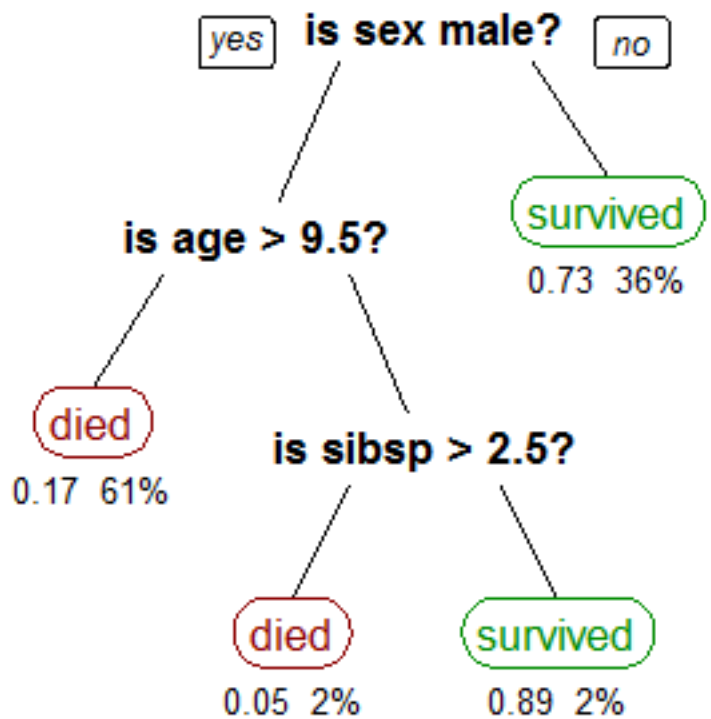
```
Afternoon = False
| Evening = False
| | Place = Outdoors: Walking
| | Place = Elsewhere: Sleeping
| Evening = True: Walking
Afternoon = True
| Wifi Id8 = True: In Meeting
| Wifi Id8 = False
| | Place = Home: Lunch
| | Place = Restaurant: Lunch
| | Place = Movie Theater: Watching Movie
| | Place = Work/School: Working
| | Place = ITE346: Lunch
| | Place = Outdoors: Walking
| | Place = ITE3338/ITE377: Lunch
```

```
loc2 = '(-inf-39.17259]': Watching Movie
loc2 = '(39.17259-39.18528]': Sleeping
loc2 = '(39.18528-39.19797]': Lunch
loc2 = '(39.24873-39.26142]': Walking
```

The decision tree model was among the best and is easy to inspect and to apply to predict a person's current activity

What Are Decision Trees?

- Decision tree learning is popular in datamining
- Given labeled examples, learn an "optimal" decision tree to predict the outcome of a new, unlabeled case
- A case is represented as a vector of feature values (e.g.: sex, age, sibsp)
- Decision tree classifiers can have two or more categories as outcomes (labels)



A tree showing survival of passengers on the Titanic. (*sibsp*: # of spouses/siblings aboard) Figures under leaves show survival probability survival and % of observations in the leaf. (Wikpedia)

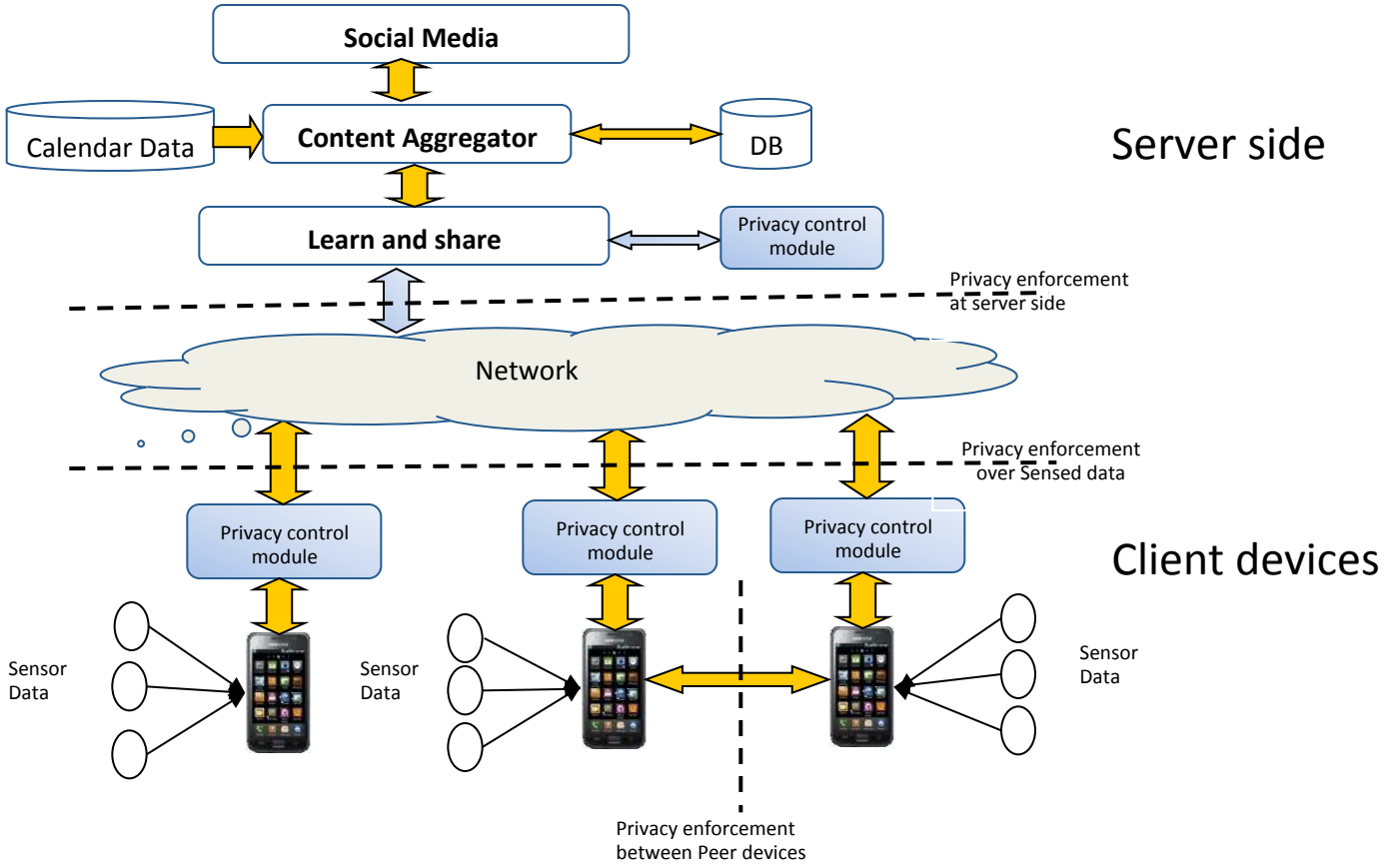
Context-aware Privacy Policies

- We use declarative policies that can access the user's profile and context model for privacy and security
- Privacy: One use is to control what user-sensitive information we share with whom and in what context
- Privacy and Security: We use the same policy infrastructure to control the actions that an app can take (e.g., turn on camera, access SD card)

What's a Declarative Policy?

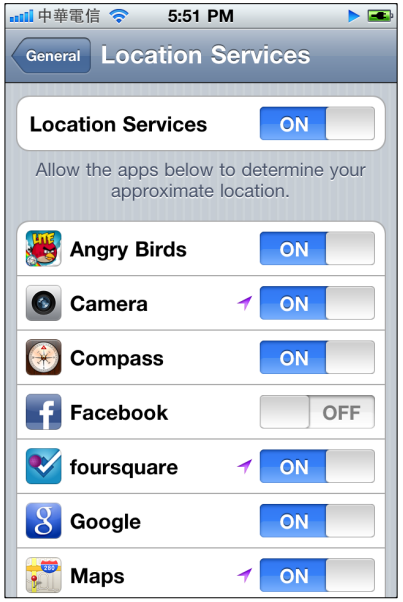
- In computing contexts, a policy is a set of rules or constraints governing what to do in a situation
- Procedural policies are often written as code (e.g., if X do Y else do Z) and trigger actions
- Declarative policies are often written as logical constraints on a (requested) action and decide whether it is permitted, prohibited or required
- Access control mechanisms in an OS or DB are examples of declarative policies

System Architecture



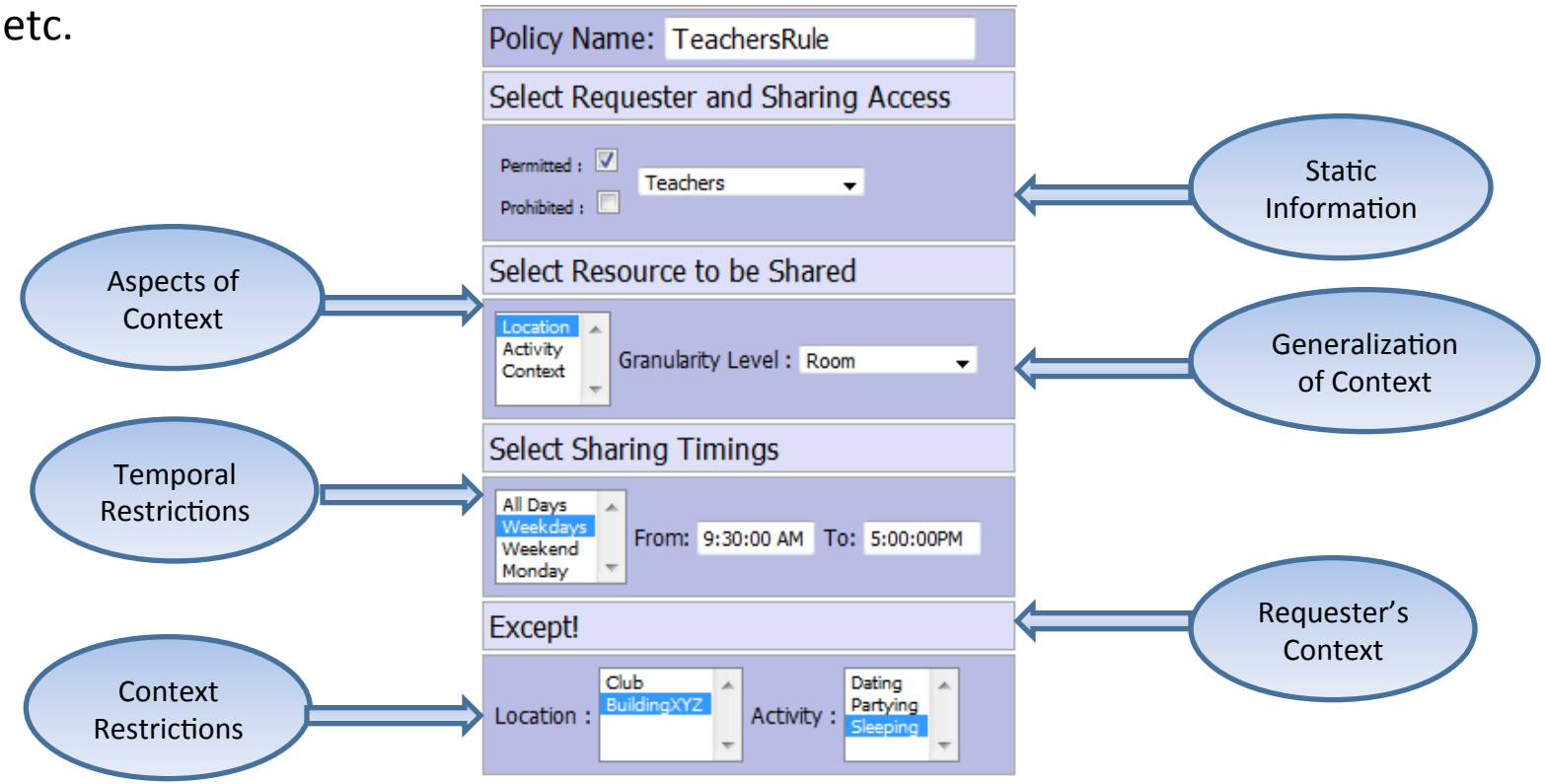
Android's Limited Privacy Controls

- Privacy controls in existing location sharing applications are limited
  - *Friends Only* and *Invisible* restrictions are common
  - Not context-dependent but static and pre-determined
- Controls for sharing other data are largely non-existent



Context-aware Policies for Sharing

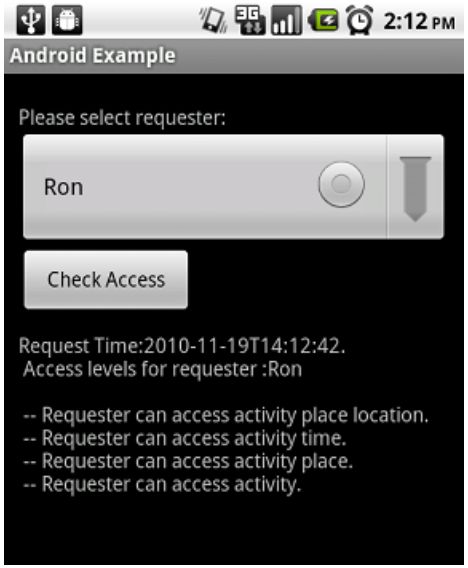
- Need for high-level, flexible, expressive, declarative policies
  - Temporal restriction, freshness, granularity, access model (optimistic/pessimistic)
  - Context dependent release of information
  - Obfuscation of shared information
  - etc.



Privacy Policies

Requests come from other devices asking to share contextual information

- A specified protocol
- SPARQL queries



Jena prototype on Android

Rules using context model and KB on device

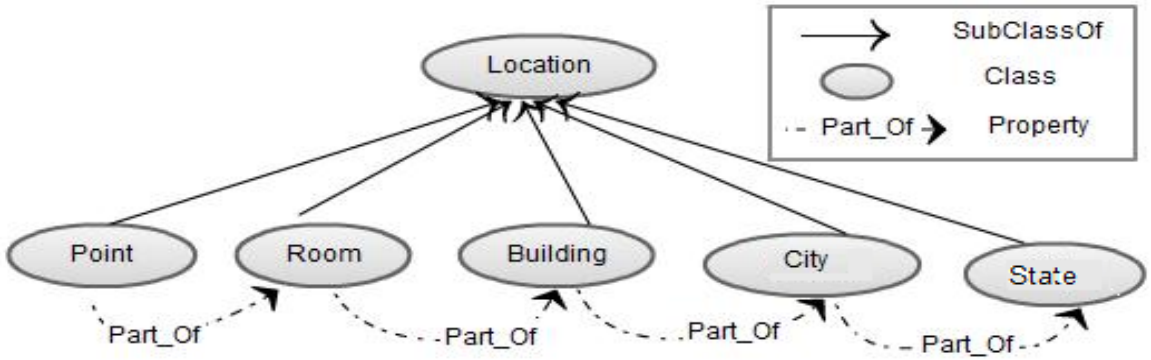
```
@prefix kb: <http://semantic-context.org/device#>.
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix foaf: <http://xmlns.com/foaf/0.1/> .

[AllowFamilyRule:
  (?requester kb:contextAccess kb:userPermitted)
    <-
    (?requester rdf:type kb:requester)
    (?groupFamily foaf:member ?requester)
    (?groupFamily foaf:name "Family")
]
```

Location Generalization

- *Share my location with teachers on weekdays from 9am-5pm*
  - User's exact location in terms of GPS co-ordinates is shared
  - The user may prohibit sharing GPS co-ordinates but permit sharing city-level location
- *Share my **building-wide** location with teachers on weekdays from 9am-5pm*

Location Generalization

- Hierarchical model of location to support location generalization

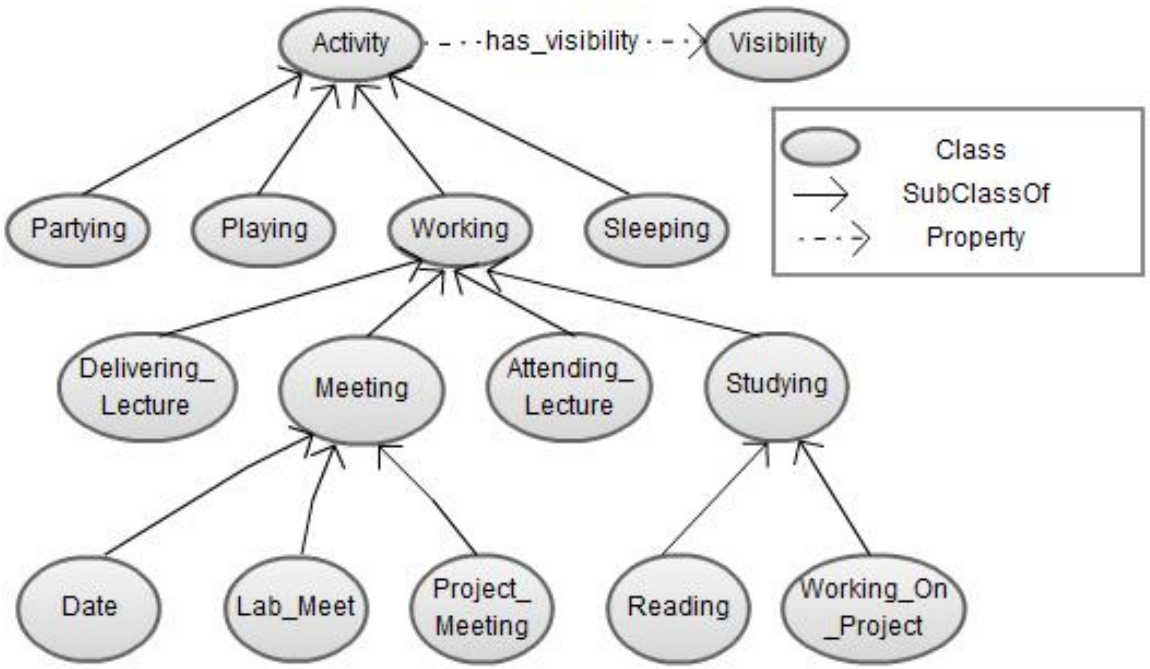


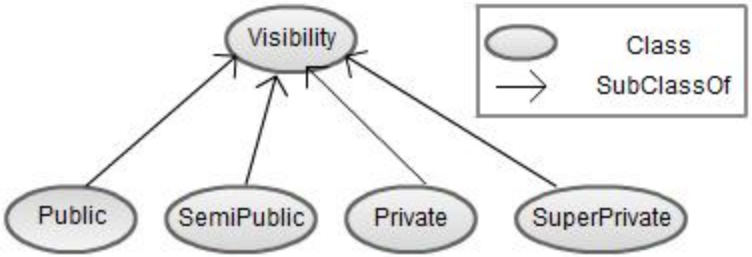
- The transitive *part_Of* property creates the location hierarchy
- GeoNames spatial containment knowledge from the linked data cloud is also used when populating the KB

Activity Generalization

- **Share my activity with friends on weekends**
  - User's current activity shared with friends on weekends
  - Share more generalized activity rather that precise
  - *confidential project meeting => Working, Date => Meeting*
- User clearly needs to obfuscate certain pieces of activity information to protect her context info
- **Share my public activity with friends on weekends**
  - *Public* is a visibility option

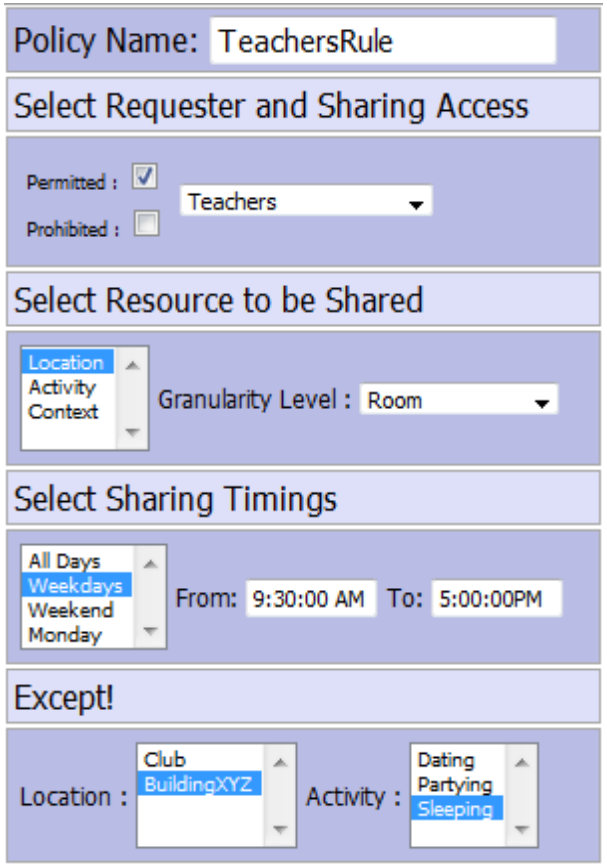
Activity Generalization





Policy Editor

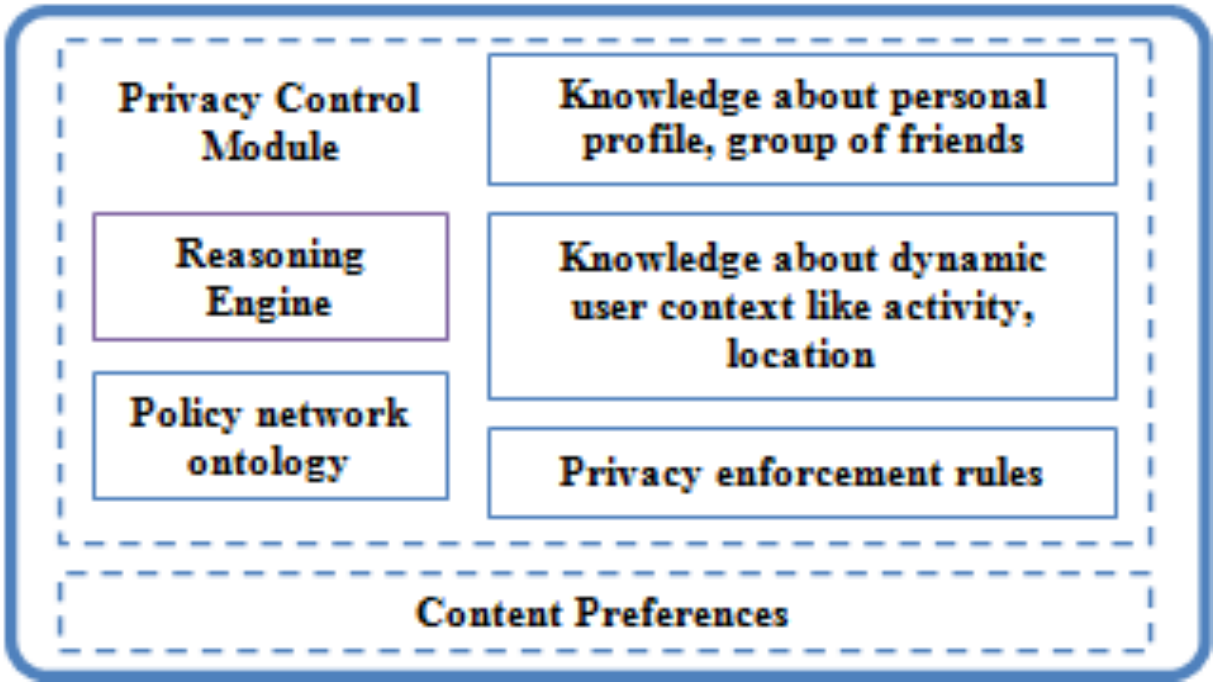
- A simple policy editor allows a user to define, view and edit her policies
- The policies are stored as rules encoded in RDF
- Their conditions & conclusions are RDF triples that specify constraints on the requestor, the context, and the information to be shared



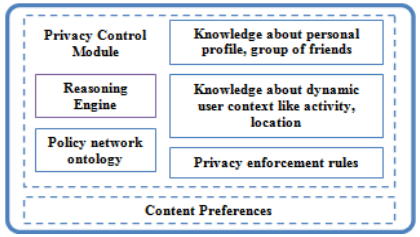
Much work remains to be done here!

Privacy Control Module

- Deals with resources to be protected, the owner of a resource and the requester who wants to access it
- Aims to protect user privacy in a context-aware system by enforcing user privacy policies

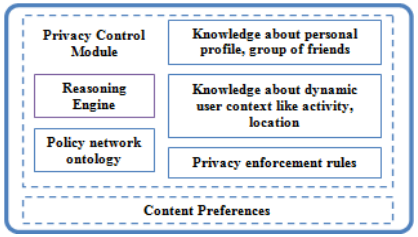


Privacy Preferences

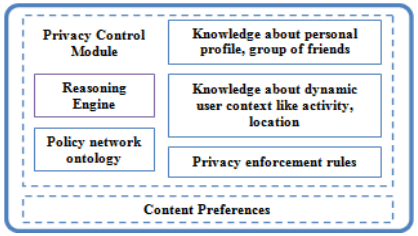


- Access control rules that describes how the user wants to share her information, with whom, and under what conditions
  - Information can be profile information, context
  - Different groups of requesters
  - Condition can be user's or requester's context
- Represented as N3 rules
- User-defined and System-defined privacy policies

Privacy Preferences

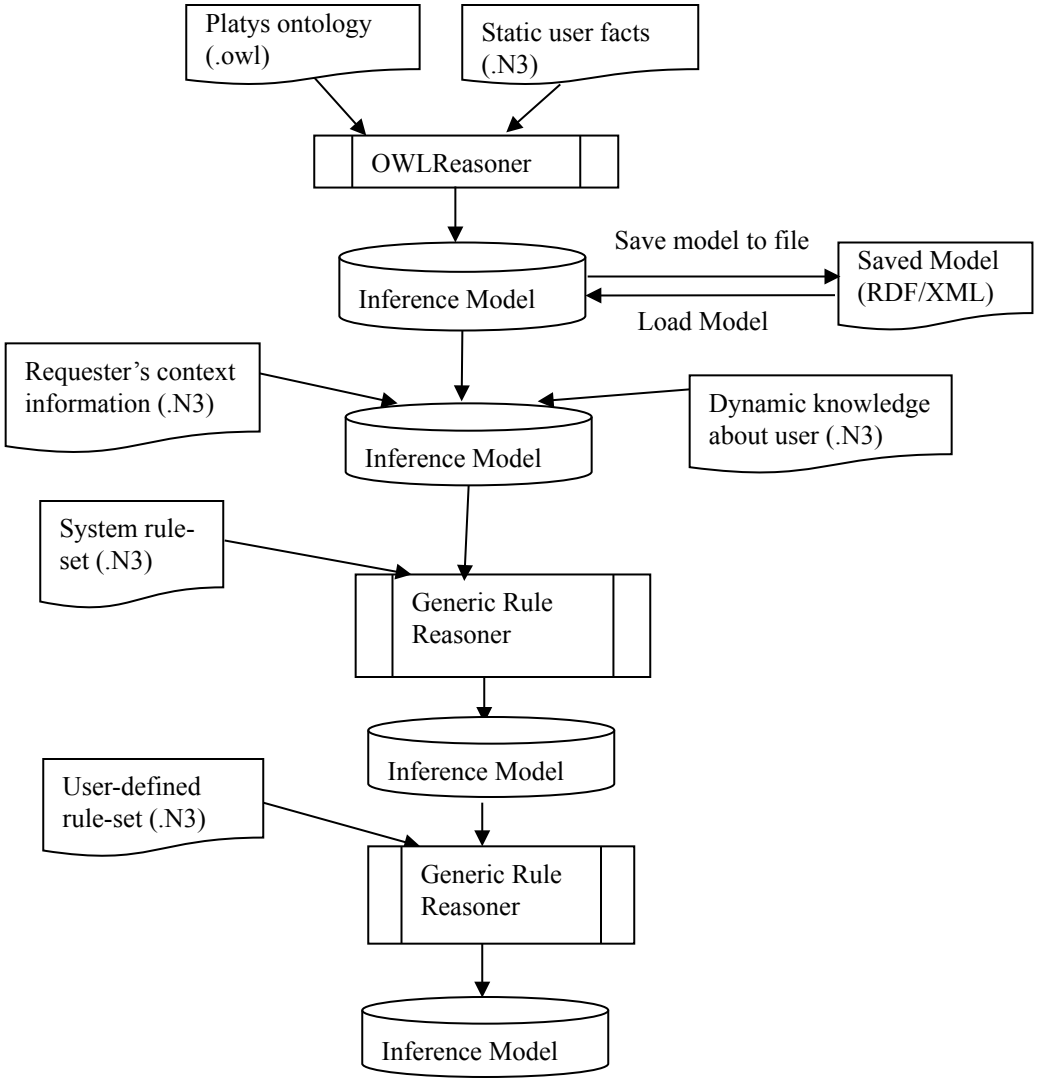


- User-defined policies : specified by the user to protect her information
  - *Share my context with family members all the time*
- System-defined policies
  - Can be needed for military domains or organizations
  - Multi-level secure systems where the system-level policies must override the user-level policies
  - *Do not share the user's context if she is inside a military building BuildingXYZ*

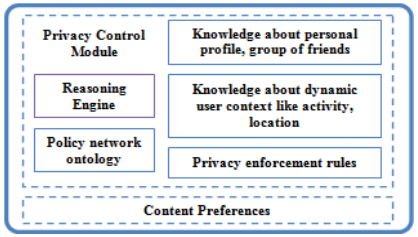


Reasoning Service

- Handles the requester queries and performs reasoning for access control decisions
- Uses the Jena Semantic Web framework
  - Uses both RDFS and OWL reasoner
  - These reasoners are used to infer additional facts from the existing knowledge base coupled with ontology and rules



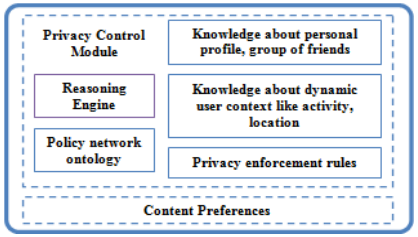
Ex: Context Sharing Policies



- Policy to share context information based on user's profile and group information
  - *Share detailed contextual information with family members all the time*

- Policy to share context information based on the user's context
  - *Share my activity with friends all the time except when I am attending a lecture*

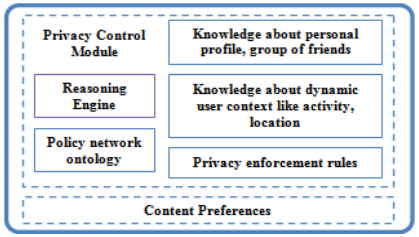
Ex: Context Sharing Policies



- Policy for sharing information based on temporal restriction
  - *Do not share my sleeping activity with teachers on weekdays from 9am-9pm*

- Policy for information sharing based on requester's context
  - *Share my context with anyone attending same class as me*

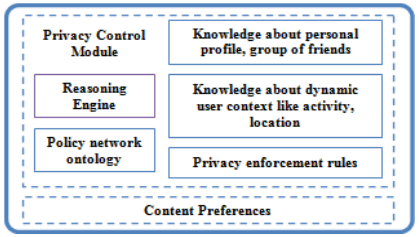
Ex: Context Sharing Policies



- Policies using generalization for sharing
  - *Share my activity with friends if it has public visibility*
  - *Share my public activity with friends*
  - *Share my city-wide location with everyone*

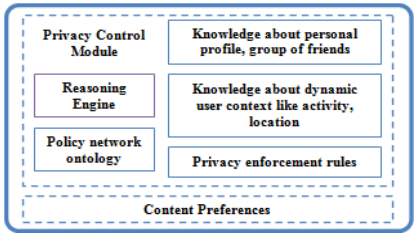
- System-level policies
  - *Do not share user's context if she is inside BuildingXYZ*

Ex: Sensor Data Access Policies



- Let users decide how their sensor information is released
- Sample Privacy policy
  - *share GPS co-ordinates on weekdays from 9am-5pm only if he is in office*
  - *Do not allow access to recorded audio but allow access to accelerometer and WiFi AP ids on weekdays*

Ex: Sensor Data Access Policies



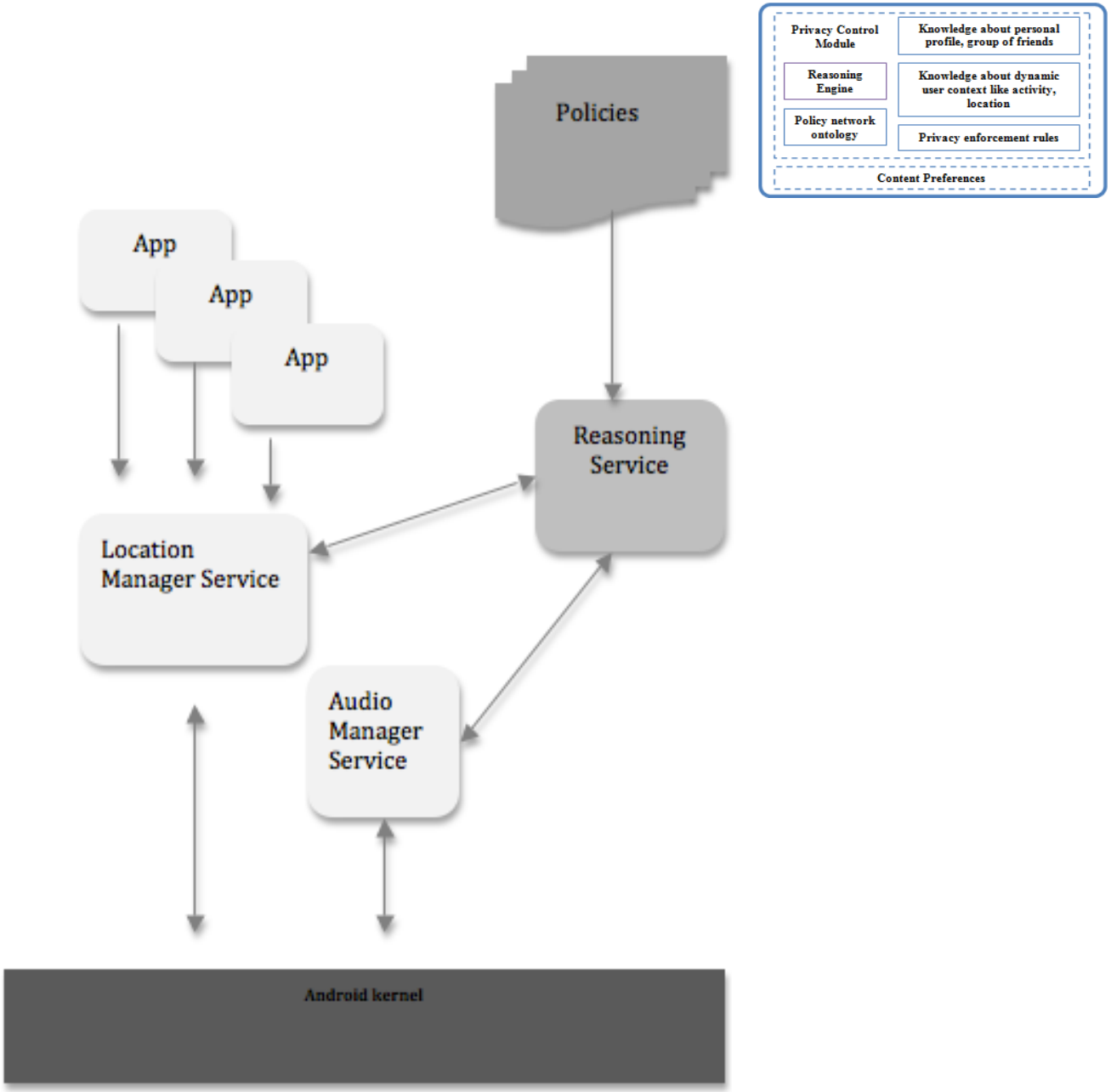
- **Location Manager control**

  Share true or false GPS coordinates, depending on the requesting app and true location

- **Audio Manager control**

  Depending on GPS location audio manager set's the phone to vibration/silent mode.

Controlling Android Services

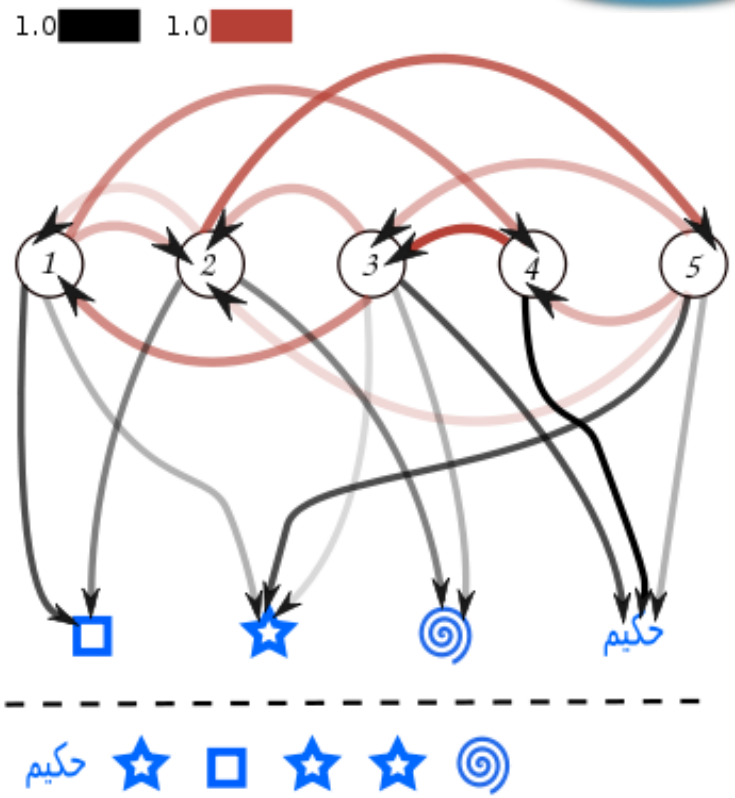


Ongoing Work

- Collecting more data for context recognition
- Including software events (e.g., read email)
- Running the model on the device
- Use HMMs for state recognition
- Incorporating common sense knowledge as priors
  - A person has only one home
  - A person has only one workplace
  - A meal is usually not repeated during the day
  - Sleeping usually occurs at night
  - Students study frequently

HMM: Hidden Markov Model?

- A Markov model is a state transition model; transitions have probabilities; states have probabilities of emitting an observable
- In a HMM we see observables, but don't know the current state
- Viterbi algorithm predicts most likely sequence of states given observations
- Good for finding most likely 'explanation' for an observation sequence



Possible state sequences are:

5 3 2 5 3 2
4 3 2 5 3 2
3 1 2 5 3 2

HMM data for a user

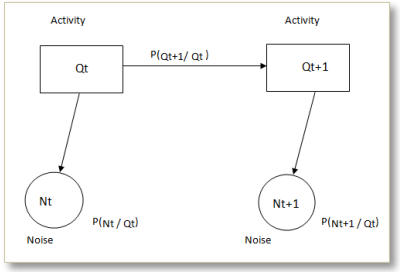
87%	1%	0%	7%	0%	0%	0%	0%	0%	0%	1%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%
1%	91%	0%	4%	0%	0%	0%	0%	0%	0%	3%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
17%	0%	67%	17%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
10%	5%	0%	64%	11%	2%	1%	0%	1%	0%	1%	1%	0%	0%	1%	0%	0%	0%	1%	0%	0%
3%	0%	0%	41%	49%	1%	0%	0%	3%	1%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%
15%	1%	0%	3%	0%	81%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%
25%	0%	0%	0%	0%	0%	50%	0%	0%	0%	0%	0%	0%	0%	0%	0%	25%	0%	0%	0%	0%
50%	0%	0%	0%	0%	0%	0%	50%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
4%	0%	0%	16%	4%	0%	0%	0%	72%	0%	0%	0%	0%	4%	0%	0%	0%	0%	0%	0%	0%
7%	0%	0%	13%	0%	0%	0%	0%	0%	73%	0%	0%	0%	0%	7%	0%	0%	0%	0%	0%	0%
0%	1%	0%	0%	0%	5%	0%	0%	0%	0%	94%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
13%	0%	0%	5%	0%	0%	0%	0%	0%	0%	0%	82%	0%	0%	0%	0%	0%	0%	0%	0%	0%
2%	0%	0%	3%	0%	0%	0%	0%	0%	0%	6%	0%	89%	0%	0%	0%	0%	0%	0%	0%	0%
5%	0%	0%	2%	0%	0%	0%	0%	0%	0%	0%	0%	2%	90%	0%	0%	0%	0%	0%	0%	0%
15%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	5%	80%	0%	0%	0%	0%	0%	0%
0%	0%	0%	17%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	83%	0%	0%	0%	0%	0%
0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	0%	0%	0%
25%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	75%	0%	0%	0%
0%	3%	0%	8%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	89%	0%	0%
14%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	86%	0%
0%	0%	0%	0%	0%	0%	0%	0%	0%	33%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	67%

State transition probability data

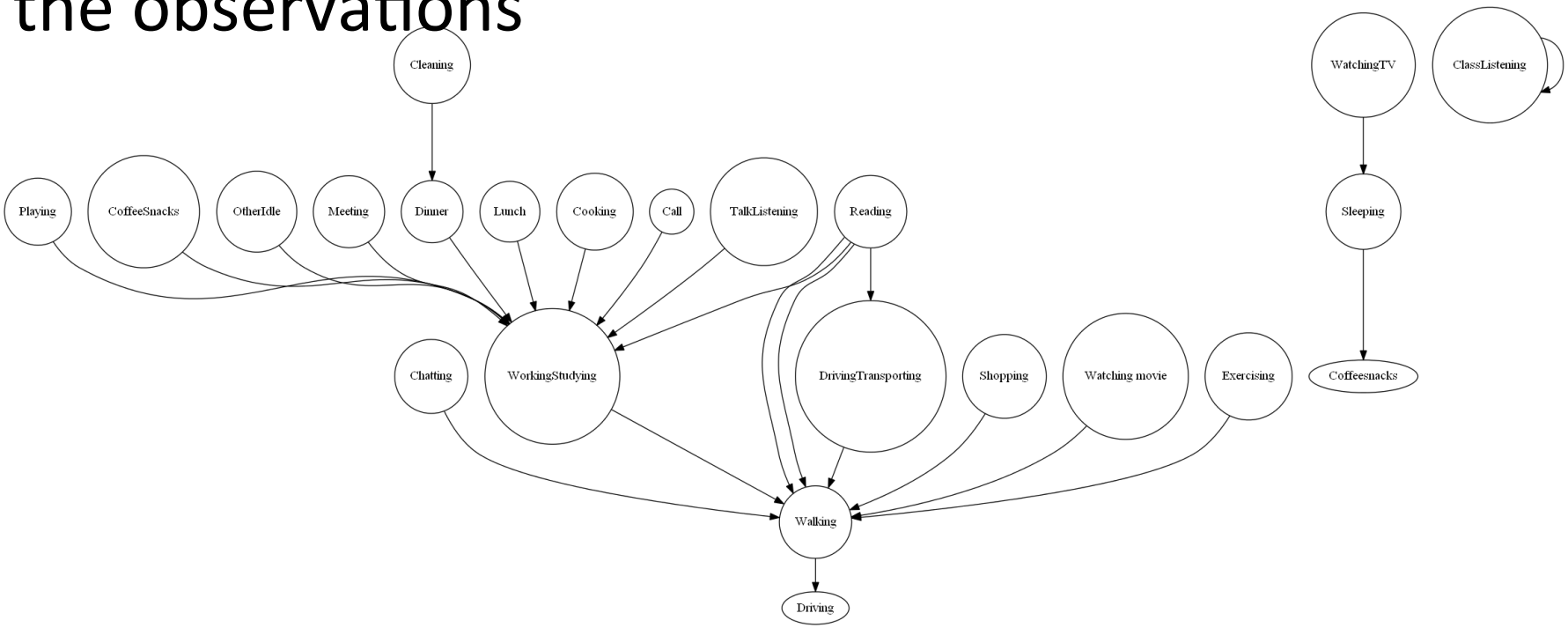
	Noise level				
Working	0.610	0.355	0.021	0.007	0.007
Chatting	0.125	0.851	0.014	0.007	0.003
Playing	0.000	0.833	0.167	0.000	0.000
Walking	0.099	0.792	0.085	0.011	0.014
Driving	0.000	0.956	0.015	0.029	0.000
Coffee	0.570	0.397	0.020	0.000	0.013
Other	0.250	0.750	0.000	0.000	0.000
Meeting	1.000	0.000	0.000	0.000	0.000
Shopping	0.000	0.920	0.000	0.080	0.000
Dinner	0.333	0.667	0.000	0.000	0.000
Sleeping	0.858	0.134	0.008	0.000	0.000
Lunch	0.579	0.395	0.026	0.000	0.000
Watching TV	0.092	0.862	0.031	0.015	0.000
Cooking	0.119	0.786	0.071	0.024	0.000
On Call	0.450	0.400	0.150	0.000	0.000
Movie	0.000	1.000	0.000	0.000	0.000
Class	0.000	1.000	0.000	0.000	0.000
Conversation	0.250	0.750	0.000	0.000	0.000
Exercising	0.000	1.000	0.000	0.000	0.000
Reading	0.000	0.714	0.286	0.000	0.000
Cleaning	1.000	0.000	0.000	0.000	0.000

Emission probabilities

HMM representation



Activities are the states and sensor readings such as noise and accelerometer data are the observations



A partial state transition graph with the most likely non-loop transition for each state

Future Work: Context Sharing

Policy language for policy declaration and enforcement integration

- Richer policies at the triple level
- Protect the inferences that can be drawn from the information that is shared
- A mix of rich pattern matching such as SPARQL and rules, with First Order semantics

Future Work: Service Management

- Location manager

  Depending on time/location/day of week report obfuscated location, e.g., randomized coordinates at city level / state level / country level as per the rule specified by the device user in Jena

- Camera access

  Enabling/disabling camera device by rendering camera resource unavailable to the android apps

- SD card access

  visible / invisible to the requester app depending on device context

Depending on the kindness of strangers

- People are cooperative and ask one another for information
  - Stanger on the street: *Does this bus go to the aquarium?*
  - Random classmate: *When is HW6 due?*
- Devices can use ad hoc networks (e.g., Bluetooth) to query nearby devices for desired information
- Each device has an info. sharing policy for what triples can be used to answer the query based on context and requester's information
- ➔ Mobile Ad Hoc Knowledge Network

Conclusion

- We established our baseline system for simple activity recognition in a university environment
- Our description logic representation enables
  - Inferences and rules
  - An expressive query language (SPARQL)
  - More expressive policy languages for information sharing and privacy
  - A natural way to give less general responses to queries
- The same model can be used to secure access to Android services (e.g., camera)

http://ebiquity.umbc.edu