

Caché: Caching Location-Enhanced Content to Improve User Privacy

Carnegie Mellon University

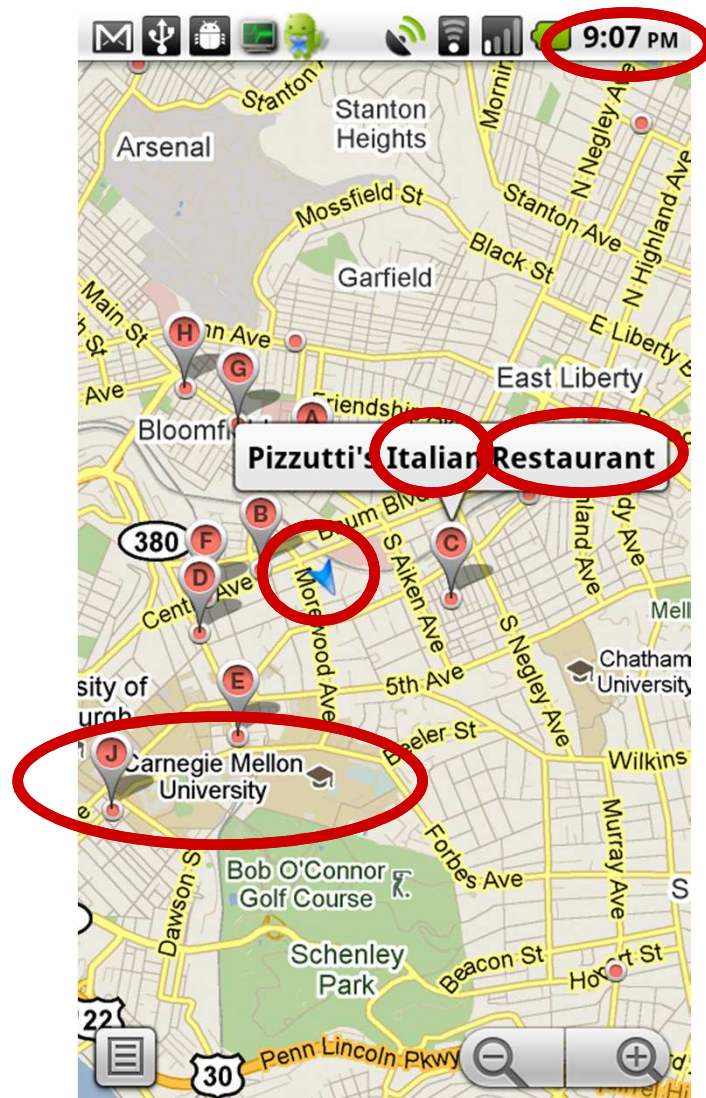
Shahriyar Amini, Janne Lindqvist, Jason Hong
Jialiu Lin, Eran Toch, Norman Sadeh

Widespread Adoption of Location-Enabled Devices

- 2009: 150M GPS-equipped phones shipped
- 2014: 770M GPS-equipped phones expected to ship (~5x increase!)
- Future: Every mobile device will be location-enabled (GPS or WiFi)



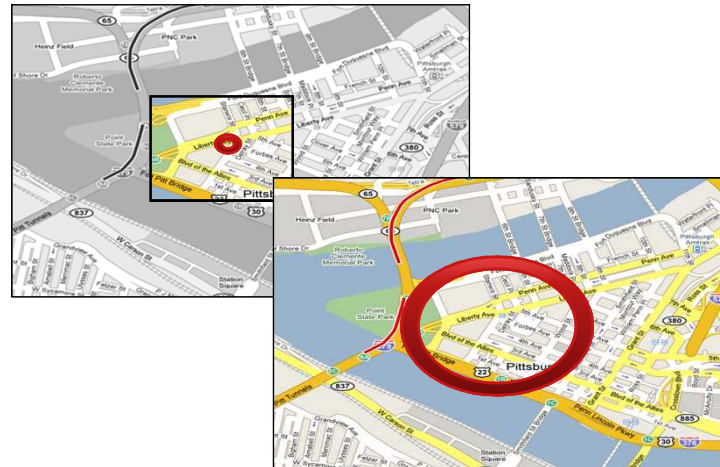
Apps Reveal Private Information



- App reveals:
 - Time of Use
 - User Interest
 - Current Location
- Over time:
 - Mobility Behavior
 - Significant Locations
 - Socioeconomic Status?

Our Approach

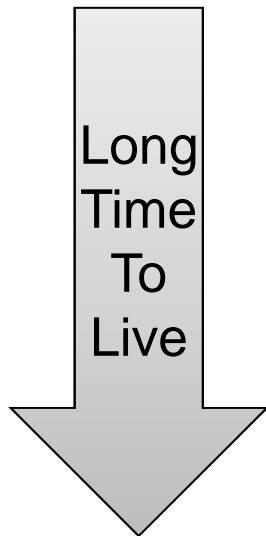
- Pre-Fetch content for large regions
- Store content on mobile device
- Determine location using GPS/trusted source
- Respond to queries using stored content
- Periodically update content



Pre-Fetching Insight

- Some location-based content are still useful even when old (time to live)

Update Rate	Data Type
Real-Time (STTL)	Traffic flow, parking spots e.g. Loopt, PeopleFinder, Reno, Bustle
Daily	weather forecasts , social events, coupons e.g. Dede
Weekly	movie/theatre schedules , advertisements, crime rates e.g. Yelp!, GeoNotes, Placelts, PlaceMail
Monthly	restaurant guides, bus schedules , geocaches e.g. Wikipedia (geo-tagged pages)
Yearly	maps , points of interests, tour guides, store locators e.g. Google Maps, Starbucks, Wal-Mart



Feasibility of Pre-Fetching

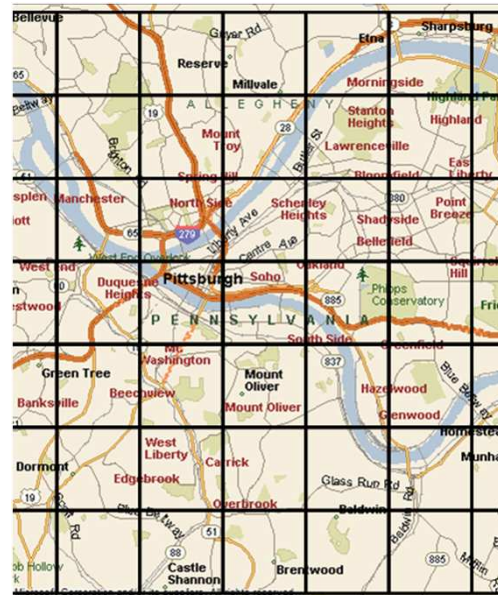
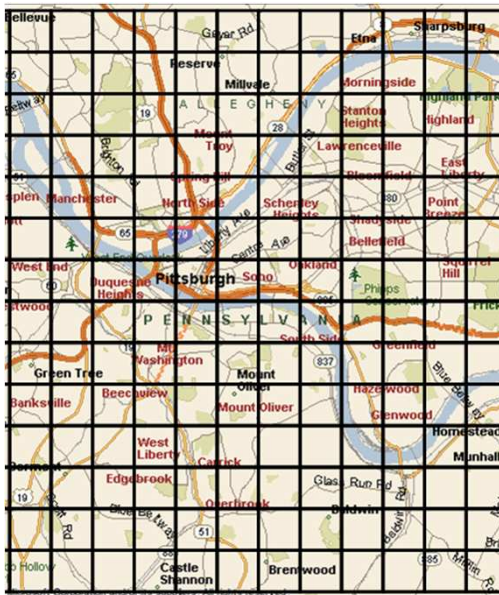
- Content doesn't change too often
 - Average daily amount of change over a 5 month period

Data Type	Added %	Removed %	Modified %
Weather	25.00	25.00	67.26
Events	5.28	5.35	11.75
Yelp POI	0.15	0.06	0.04
MSN POI	6.69	6.80	1.43
Bus Schedule	0.00	0.00	0.15
Map Tiles	0.00	0.00	0.00

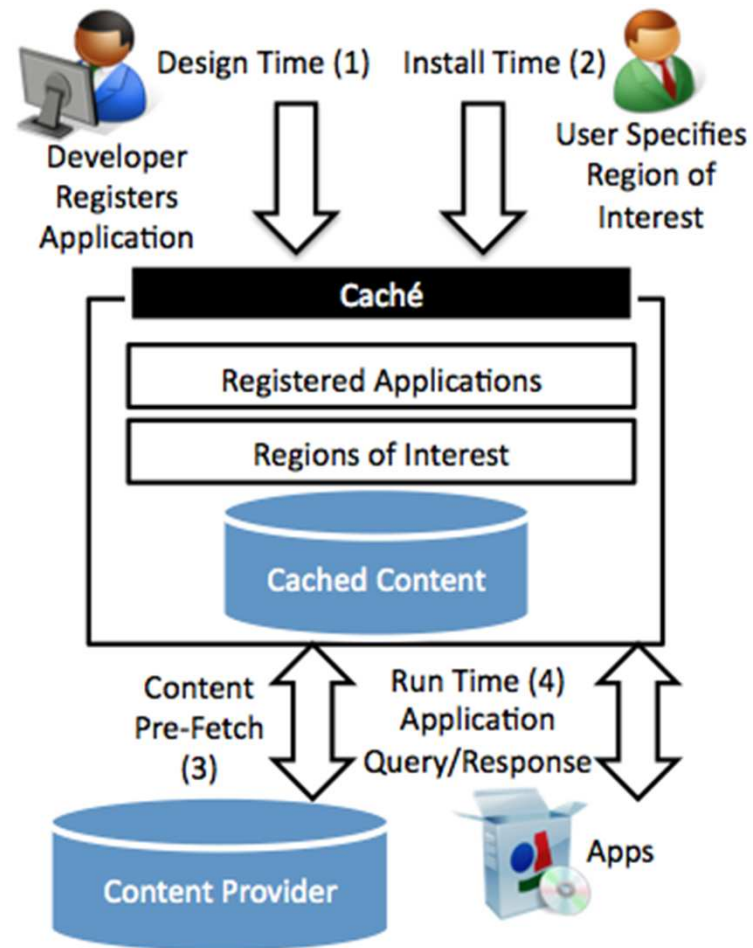
- Requires <20 MB for Pittsburgh, ~100 MB for NYC

Pre-fetching Content

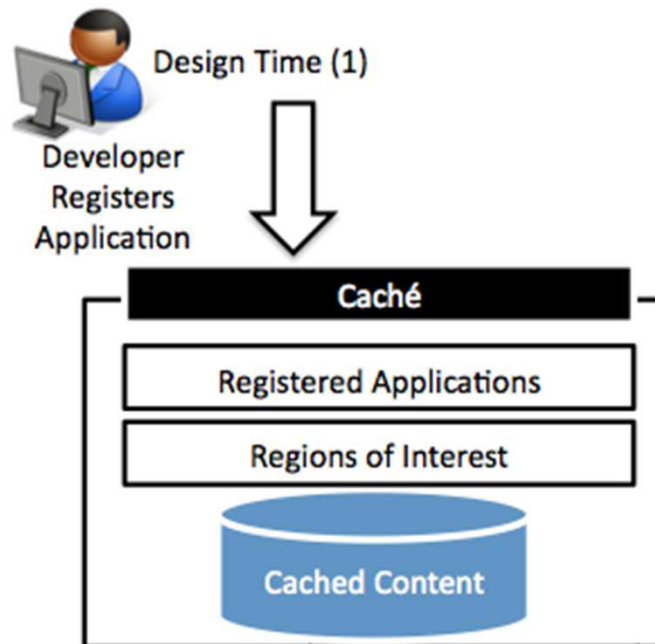
- Geo-coordinates are continuous
- Content cannot be pre-fetched for every point
- Use a grid to discretize space



Caché Architecture



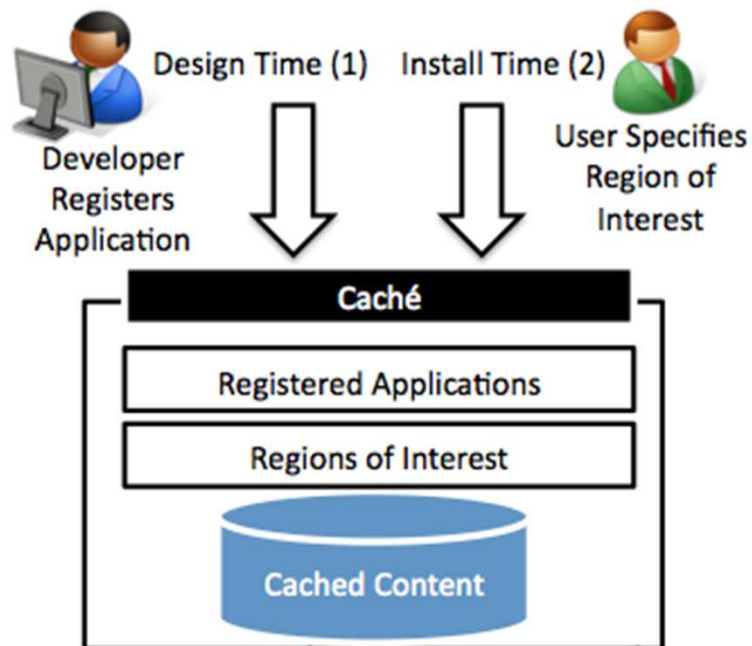
Application Design



- REST-based content
- Developer defines:
 - Size of cells
 - Content update rate
 - Query string

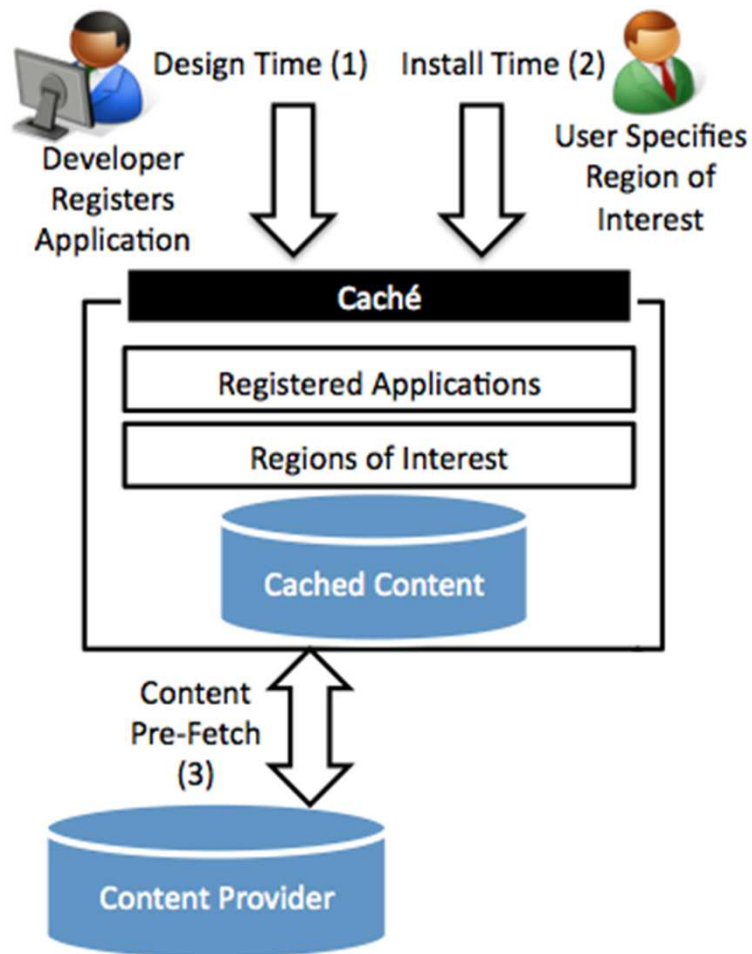
http://api.yelp.com/v2/search?term=food&ll=#SLL_LAT#,#SLL_LON#

Application Installation



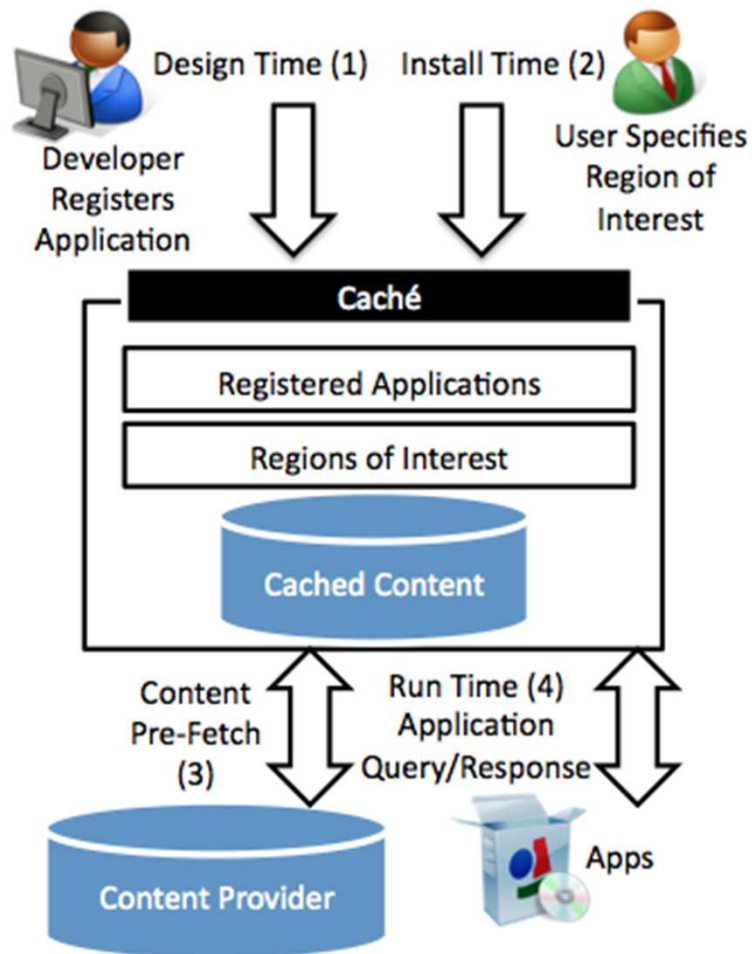
- Regions of interest:
 - 15213
 - Pittsburgh, PA
- Pre-fetch radius
 - 1 km

Content Download



- Pre-fetch only when:
 - Plugged in
 - Connected to WiFi
- Pre-fetch every cell
- Update content at defined update rate

Content Retrieval



- Assume fresh content
- Retrieve content from a single cell
- Content miss results in a live request to LBS

High Content Hit Rate!

- How often will queries be cached?
 - Locaccino: Top 20 people, 460k traces
 - Place naming: 26 people, 118k traces

Radius (miles)	Locaccino	Place Naming
5	86%	79%
10	87%	84%
15	87%	86%

Caché Android Service

- Android background service for apps
 - Apps modified to make requests to service

Application	SLOC		
	Original	Added	Removed
mixare	4692	18 (0.4%)	4 (0.1%)
Panoramio	1268	18 (1.4%)	7 (0.55%)
Restaurant Request	411	12 (2.9%)	8 (1.9%)

- User specifies home and work locations
- Service only pre-fetches when device is plugged in and connected to WiFi

Limitations

- Doesn't work for
 - Rapidly changing content (STTL)
 - Apps with client/server interaction (Facebook)
 - Apps with server computation (Navigation)
- Burden falls on the developer
 - Developer has to effectively sweep content
- New regions have to be specified before use

Related Work

- Content Pre-fetching
 - Coda
- Anonymity
 - k -anonymity, Spatial and Temporal Cloaking, Privad
- Obfuscation
 - SybilQuery
- Route prediction and caching
 - CacheCloak

Conclusion

The most private and energy efficient request is the one you don't make.

- Tradeoffs: Privacy vs. Utility vs. Cost
- Current solutions present challenges
- Comprehensive privacy solution would rely on several approaches
- Consider development and deployment