Ancile: Privacy-Aware Programming for Microscale Data

Jason Waterman
Vassar College

Eugene Bagdasaryan
Cornell Tech

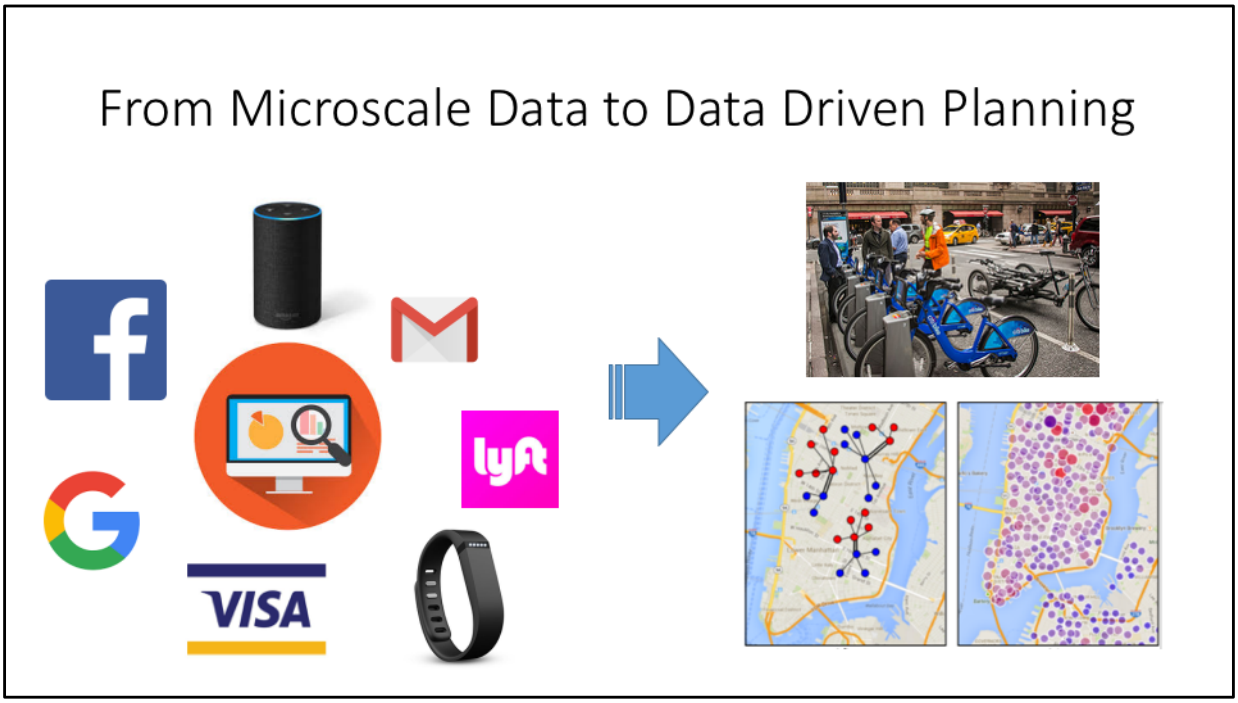
Matthew Griffith
Cornell Tech

Griffin Berlstein
Vassar College

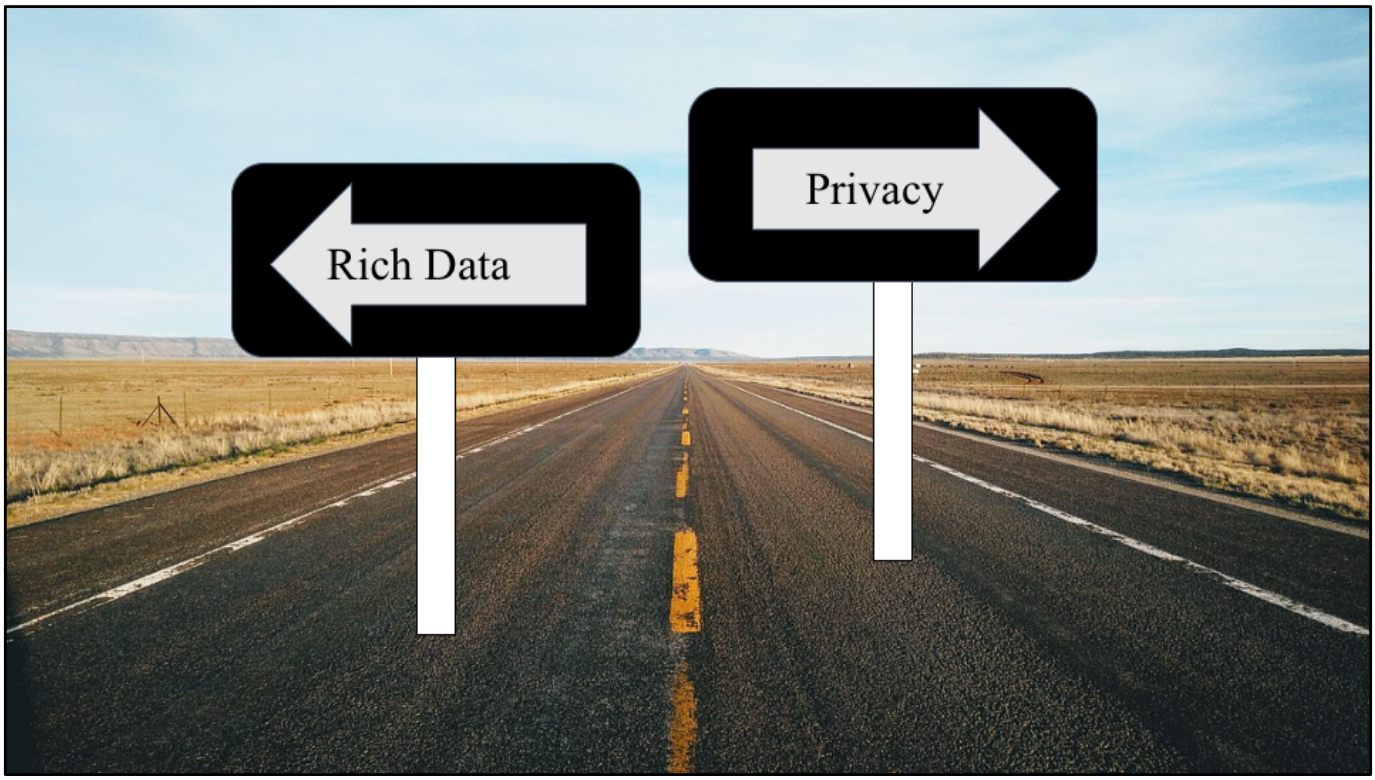
Nate Foster
Cornell University

Eleanor Birrell
Pomona College

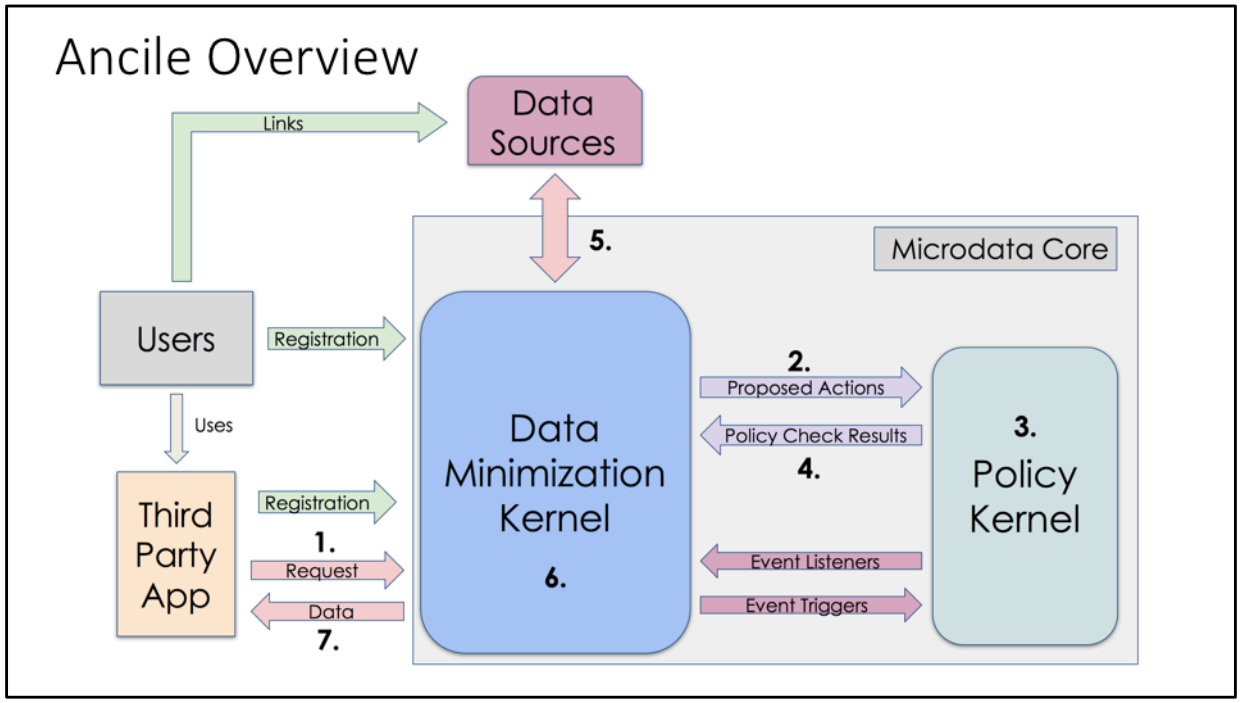
Deborah Estrin
Cornell Tech



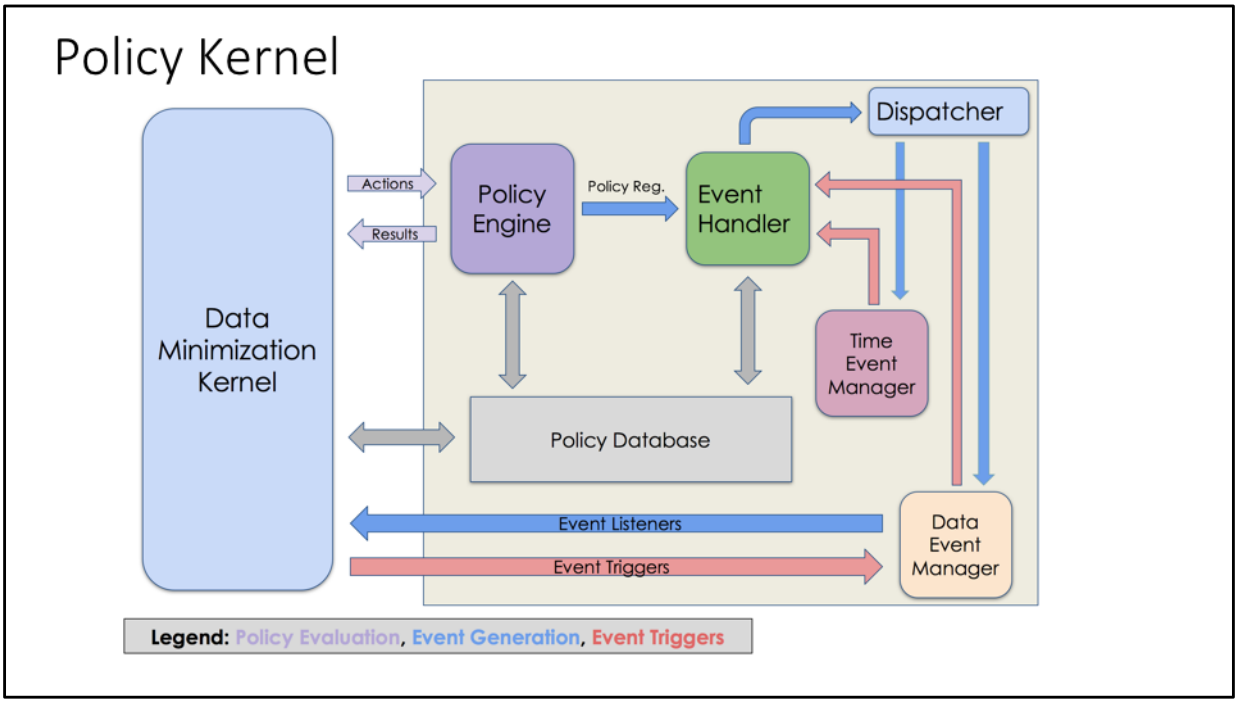
We use the term Microscale data to refer to all the personal and sensitive data generated as we live our lives in the digital age. This data can be used to address an array of space management and sharing needs, but can also raise serious privacy concerns. One example of microscale data being used for data driven planning is the Bike Angel's program, which uses real-time tracking of city-share bikes to build models that reward users who move bikes from more congested areas to less congested ones as shown in the pictures above.



It is clear that this microscale data is useful for a wide range of applications, but there is a tension between an application's desire for rich data and a person's willingness to trust that application with their data. Contextual Integrity is a framework allows users to express their privacy wishes, but support for privacy policies is needed at the application level. It can't be done by applications on top of currently existing infrastructure, it requires a system architecture and framework that supports policies. Otherwise, you end up with the lower levels of the system retaining far too much detail beyond what apps and users need. Our system, Ancile, is a framework that provides this support.



Ancile consists of a Microdata Core that sits between third-party applications and a user's microdata. (1) Applications must explicitly declare any microdata requested from a user as well as a pipeline transformations performed in the Data Minimization Kernel on this data. This allows for filtering and minimization of a user's data before it is sent to the application. (2) The request is checked by the Policy Kernel (3) to see if has been authorized by the users privacy policy for this data. If the request is allowed (4), data is fetched from the data source (5) and the requested pipeline operations are performed (6) and the result is sent to the application (7). This model forces applications to declare specifically what data is used and gives a trusted location to perform any data minimization to reduce the amount of data needed by an application.



Here's a walkthrough of the basic parts of the Policy kernel:

Policy Engine: Authorizes requests from the DMK based on the current policy context.
Event Handler: Modifies policies in reaction to receiving contextual events, generates new events as needed.
Time Event Manager: Tracks internal time-based events (e.g., don't share my location after 3pm). At 3pm send an event to change the policy context for location.
Dispatcher: Routes event generation signals to the appropriate event manager.
Data Event Manager: Tracks external data-based events and communicates with the Data Minimization Kernel to set up data collection for such events.

Prototype Applications

Application	Example Policy	Data Shared with App
Roaming office Hours	If I am on campus and it is is between 2 PM and 3 PM, share my location with my students.	Location only in the proper context
Spontaneous Study Group	If I am in the library at the same time as at least two members of my group, notify the entire study group.	Location only in the proper context
Selective Study Participation	Allow my anonymized data to be included in the study only if at least 100 other people are participating.	Anonymized user data or nothing

We are currently focusing on location-aware applications in slack, but Ancile is a general purpose framework and is suitable any app that can make http requests. The first two applications have been implemented, we are currently working on the third one.

Future Work

- Implementation of more complex use cases
  - Deployment outside of our lab
- Providing tools for visualizing and creating more complex policies
- Moving from single instance to a federated version of Ancile
  - Scale to larger deployments

We are working on getting others to use our slack apps, hope soon to have developers using our framework to develop their own privacy-aware applications.