CERIAS

The Center for Education and Research in Information Assurance and Security

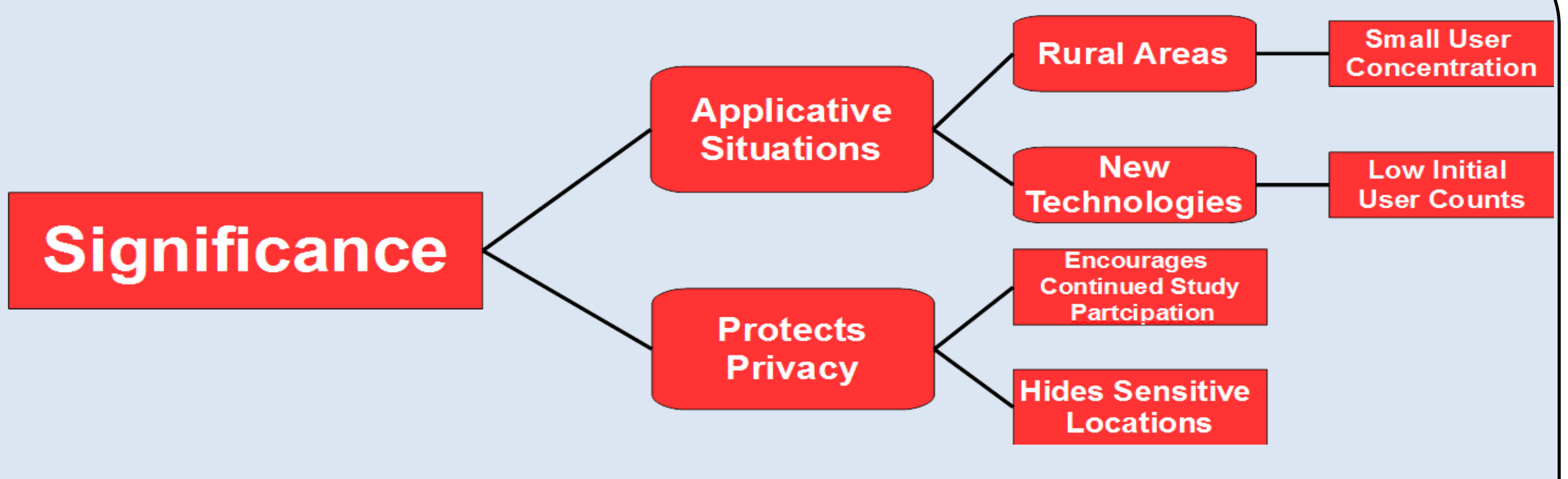
Location Data Anonymization Using Small Data Sets

Daniel Bittner and Isaac Senior

Advisor: Dr. John Springer

Problem Statement

- Privacy invasion and tracking algorithms compromise the privacy of study participants by connecting sensitive locations to driver identities.
  - When locations and identities are linked, conclusions outside the original intention of the data can be drawn.
- We seek to show that small data sets need different data obfuscation techniques in comparison to larger data sets in order to protect privacy and preserve data utility.
- This research tests suppression-based obfuscation algorithms on a small data set against a privacy invasion algorithm.



Significance

Applicative Situations

Protects Privacy

Rural Areas

New Technologies

Encourages Continued Study Partcipation

Hides Sensitive Locations

Small User Concentration

Low Initial User Counts

```
-------------------------------------------------------------------------------------------------------------------
| point id | car id | time | longitude | latitude | odometer | max speed | relative odometer | course | charge | temperature  |
-------------------------------------------------------------------------------------------------------------------
```

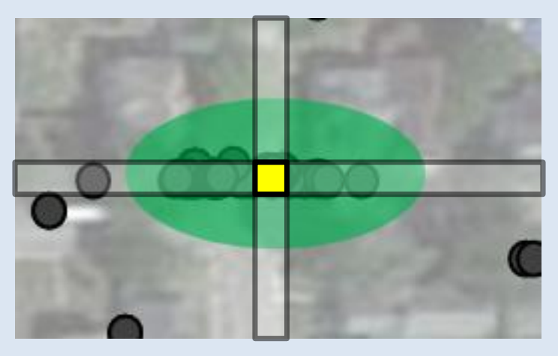
1. Determining Potential Risk to Privacy

- To determine any potential risk, an algorithm was created to find significant locations.
- Locations were found by analyzing the beginning and end points of a driver's trips, similar to the approach used by Iqbal (2010).

- Recordings containing longitude, latitude, and time are selected for each driver.

- Terminal points are identified and grouped into clusters.
- Significantly large clusters are isolated.
- All other points are disregarded.

- Confidence intervals are found for the longitude and latitude of each cluster to pinpoint the most probable coordinates of the location.

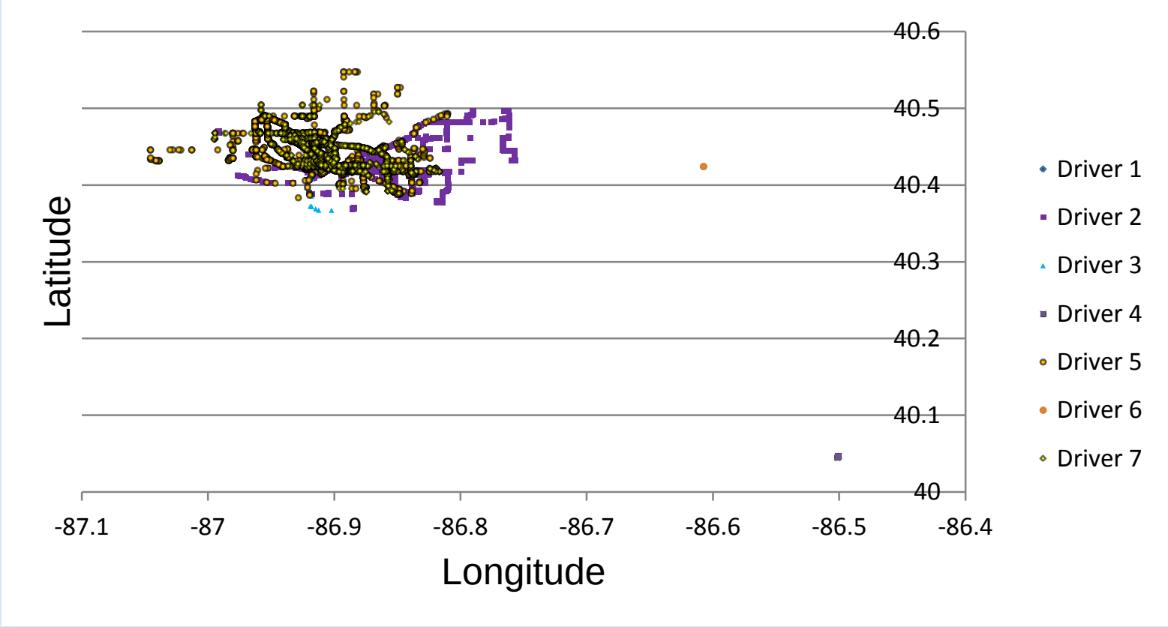


2. Performing Data Obfuscation

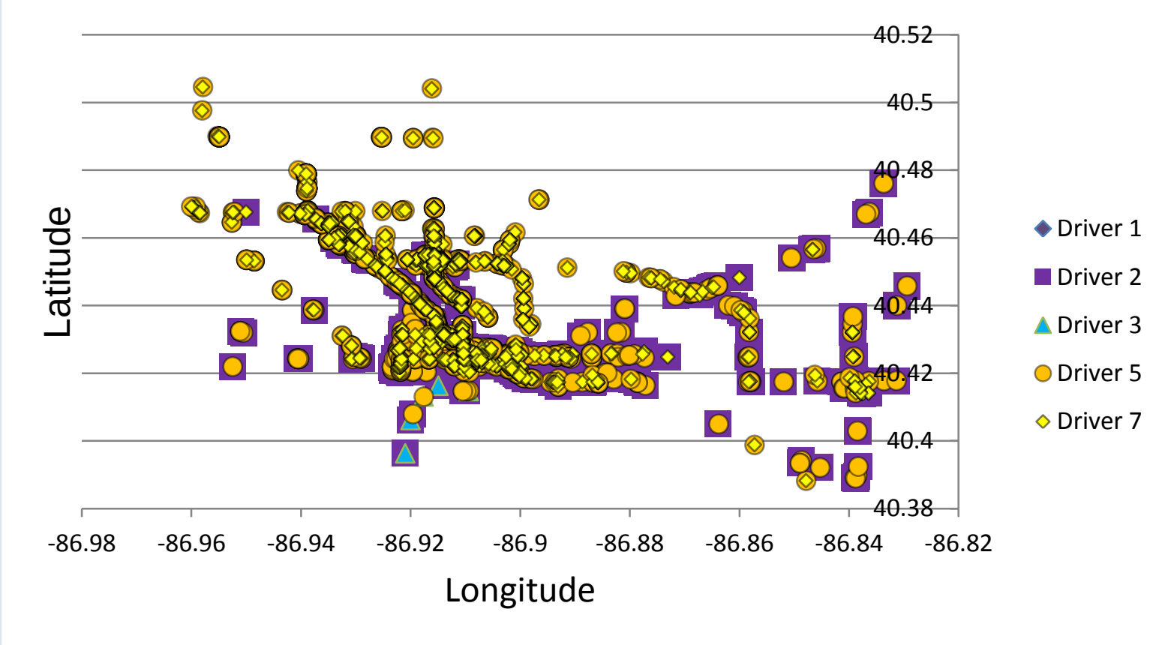
- *k*-anonymity states that identifiers cannot be unique to some group smaller than size *k*, according to Samarati and Sweeney (1998).

- An obfuscation algorithm was created to apply the suppression technique to the electric vehicle data in order to achieve *k*-anonymity where *k*=2 for each point.

Before Obfuscation

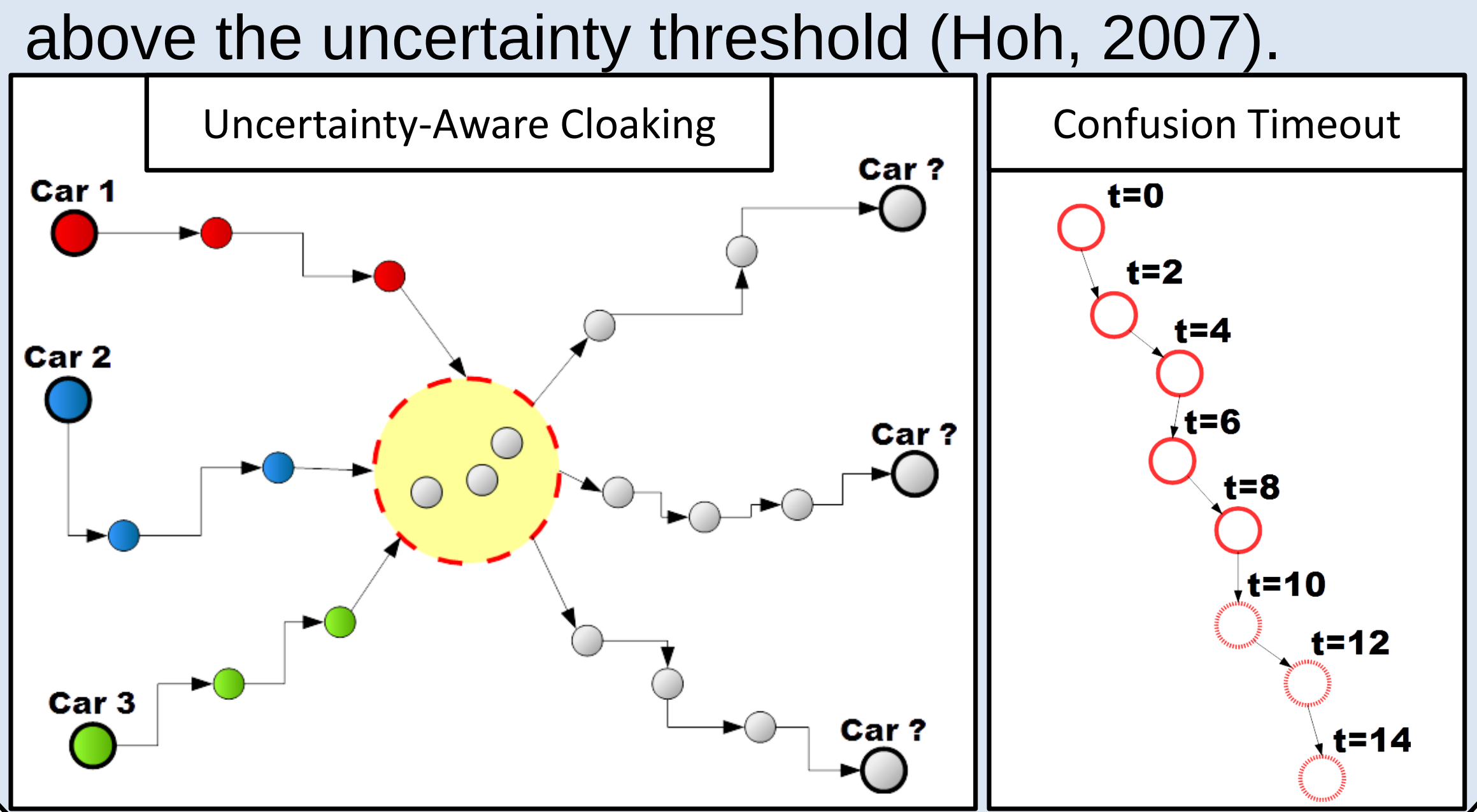


After Obfuscation



Uncertainty Aware Path Cloaking Algorithm

- Points are suppressed when the time since the car's last confusion point is above the confusion timeout threshold. The last confusion point for cars is updated at the start of trips and when a point is above the uncertainty threshold (Hoh, 2007).



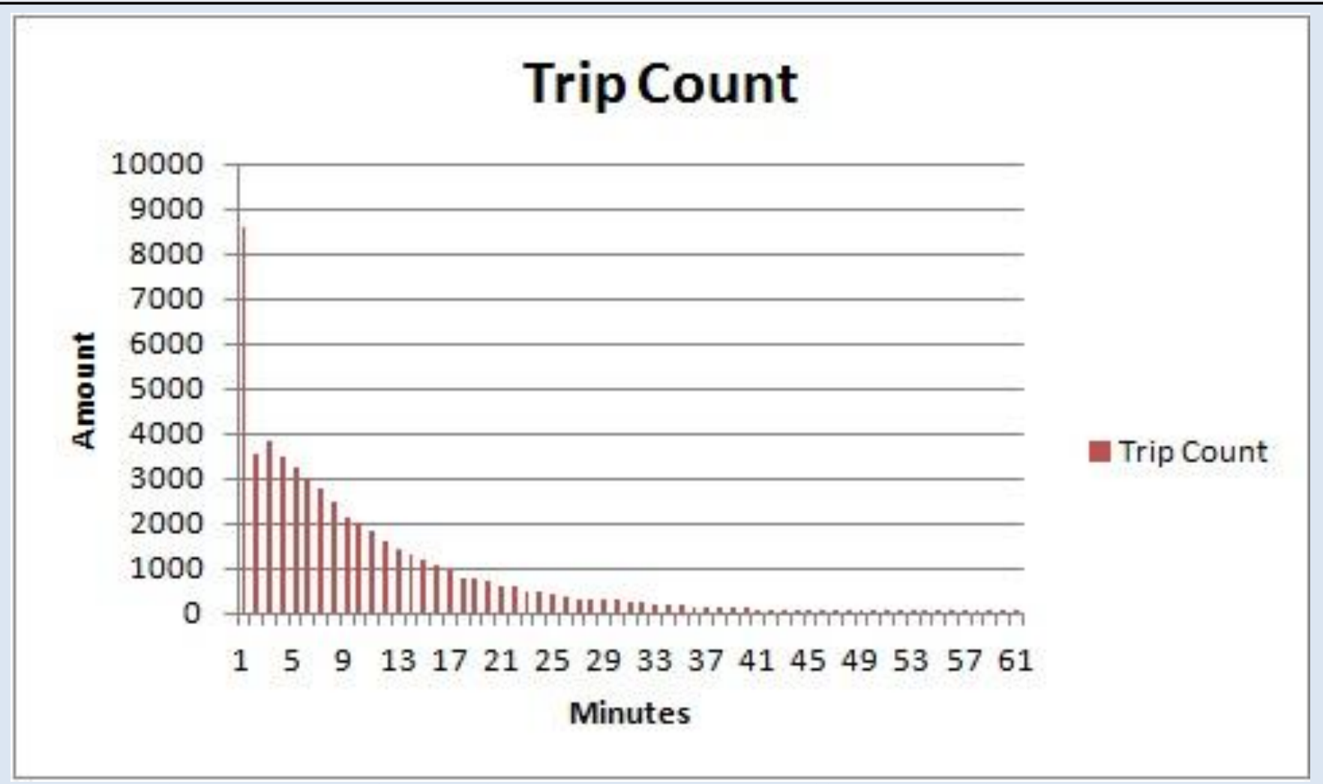
3. Analysis

Analysis of Conventional Obfuscation Techniques

Interval length	Deletion %			Success %			False Positive %			Adjusted Success %		
<30m	0.0	60.8	84.6	89.9	70.9	57.7	14.1	17.2	27.5	83.5	63.3	51.3
<10m	0.0	60.8	84.6	88.6	70.9	57.7	12.4	15.3	25.7	82.3	63.3	51.3
<5m	0.0	60.8	84.6	70.9	49.4	38.5	6.1	7.4	12.3	70.9	48.1	37.2
<2m	0.0	60.8	84.6	20.3	12.7	6.4	0.0	0.0	0.0	20.3	12.7	6.4

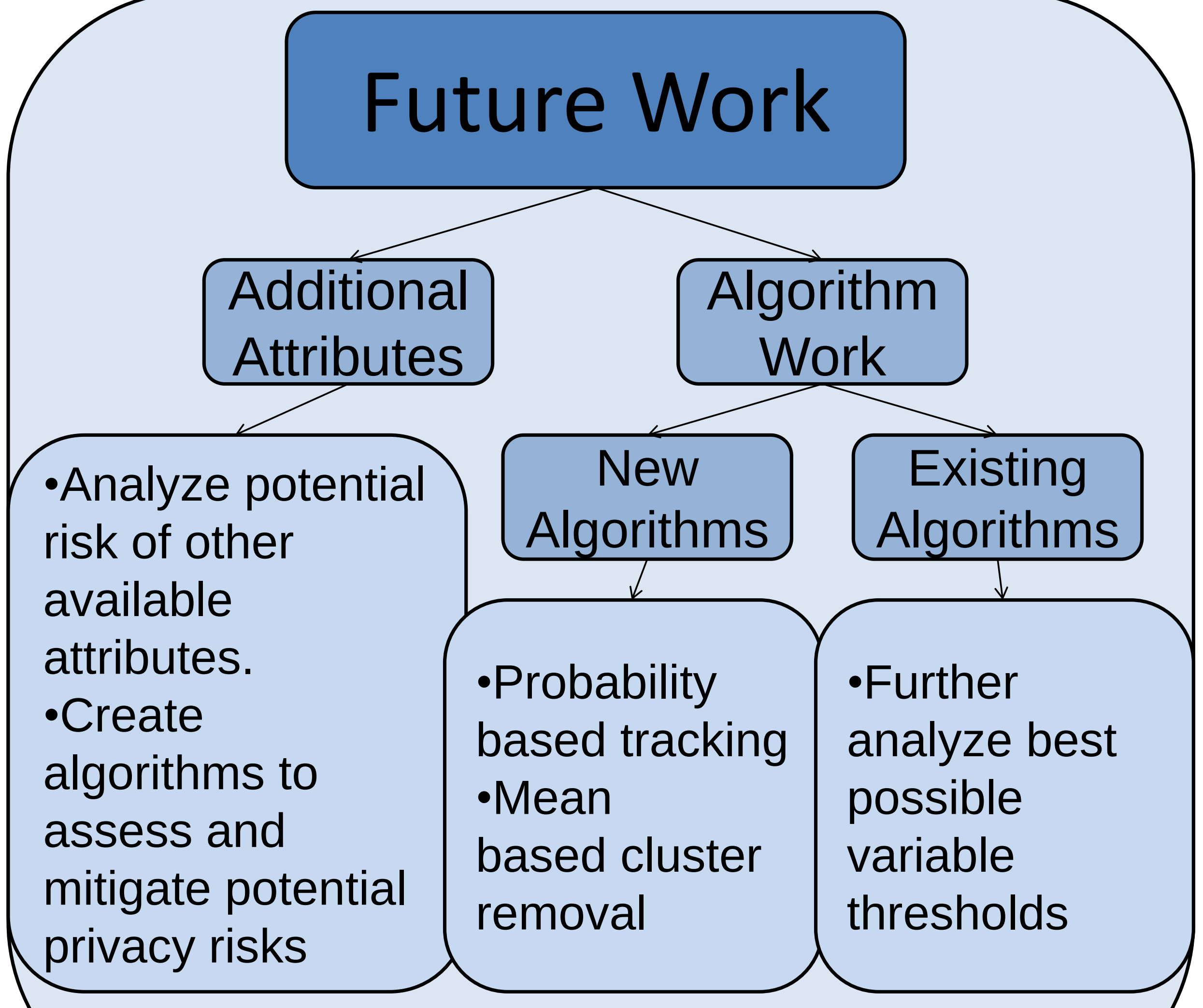
Legend: Unobfuscated k=2 k=5

Analysis: The privacy provided by conventional obfuscation techniques is not proportional to the significant amount of data that is lost.



Analysis: Trip length is exponentially distributed

Future Work



Additional Attributes

Algorithm Work

New Algorithms

Existing Algorithms

- Analyze potential risk of other available attributes.
- Create algorithms to assess and mitigate potential privacy risks

- Probability based tracking
- Mean based cluster removal

- Further analyze best possible variable thresholds

Hoh, B., Gruteser, M., Xiong, H., & Alrabady, A. (2007). Preserving privacy in gps traces via uncertainty-aware path cloaking. *Proceedings of the 14th ACM Conference on Computer and Communications Security - CCS '07*, 161–171.doi:10.1145/1315245.1315266

Iqbal, M. U., & Lim, S. (2010). Privacy Implications of Automated GPS Tracking and Profiling. *Technology and Society Magazine*, *29*(2), 39–46. doi:10.1109/MTS.2010.937031

Samarati, P., & Sweeney, L. (1998). Protecting Privacy when Disclosing Information : k -Anonymity and Its Enforcement through Generalization and Suppression 1 Introduction. *Proceedings of the IEEE Symposium on Research in Security and Privacy*, 1–19.

CERIAS®

NSF

This material is based upon work supported by the National Science Foundation under grant #1062970. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

PURDUE UNIVERSITY