spectus™

Data. In sight.

WHITEPAPER

Privacy Enhanced Mobility

The hallmark of Spectus from its inception has been a deep and unwavering belief that the consumer is our partner. Therefore it is not only our ethical responsibility but a fundamental business principle to protect their data and ensure transparent communication. As we continue to be leaders in privacy best practices, compliance with current and future privacy regulations is turn-key for our customers and partners.

Data has become a fundamental tool across a wide range of disciplines. Human mobility data specifically offers valuable insights that can help mar-keters, researchers, investors, and data scientists to understand human be-havior, movement patterns, and global trends. Working with the best privacy practices is of utmost importance as data becomes more valuable and more difficult to come by—without proper measures in place, the widespread use of data can introduce potential privacy risks. Techniques like data aggrega-tion can ameliorate privacy concerns, but these measures also reduce the utility of mobility data for use cases that require greater granularity.

For this reason, Spectus has developed a proprietary and patent-pending data solution called Privacy Enhanced Mobility (PEM). PEM data for custom geospatial analyses and methods development is currently available within the Spectus Data Clean Room.

What Is PEM?

Spectus developed PEM to protect user privacy by mitigating the risk of user re-identification. In order to achieve greater data utility without sacrificing user privacy, PEM improves upon data anonymization techniques to further strengthen our privacy protection practices. With PEM, Spectus aims to increase user privacy preservation while retaining the ability to infer broad user demographics, and to observe mobility behaviors of truly anonymous users across “whitelisted” points of interest—public and commercial venues that do not reveal sensitive information about individual users. Spectus goes beyond simple de-identification of user data by applying a patent-pending privacy enhancement methodology. PEM introduces noise to the inferred home location of users to prevent re-identification by transforming the latitude and longitude values of device location and stops.

How It Works

PEM data consists of geographical coordinates for de-identified smartphone devices that opted-in to data collection under a CCPA and GDPR compliant framework. For each row of data, metadata are collected such as a de-identified user ID, latitude, longitude, timestamp, and accuracy. Whenever a device within the Spectus panel records a data point, producing a row of data, the point is labeled as one of four classifications:

Sensitive POI: a point of interest deemed to be sensitive from a privacy perspective.
Whitelisted POI: a point of interest included in the Spectus database of public and commercial venues.
Personal Area: a point in the neighborhood of the inferred home area for a given device.
Other: points falling in areas not included in the above.

Based on the classification of the point, the following transformations are applied to the data:

Remove: Data points are completely deleted. Such is the case with points falling within sensitive POIs.

Keep: Data points remain unaltered. Such is the case with points within Whitelisted POIs.

Personal Area Centroid: The original latitude and longitude values are transformed to the coordinates of the centroid of the nearest geometry with 600+ households. The resulting point will always remain in the same Census Block Group (in the US) where the original point was located. Such is the case with points falling within the Personal Area Classification.

Other: In the version of PEM under analysis, data points classified as "Other" were maintained. In the most recent version of PEM (see "PEM+) "Other" data points are removed entirely.

Why PEM?

PEM technology allows us to obfuscate data and take privacy protection a step further. In a dataset where the location of an individual is specified hourly, and with a spatial resolution equal to that given by the carrier's antennas, current industry standards allow for just four spatio-temporal points to uniquely identify 95% of the individuals. PEM improves user privacy preservation, with 88% improvement over industry standards.

PEM+

We hold our privacy-protecting technologies to the highest standards and therefore we are constantly looking for ways to improve and innovate. In response to internal and third-party evaluations (see below) of PEM, Spectus implemented steps to improve the already robust privacy protections afforded by its enhancements. A more recent version of PEM, known as “PEM+” filters out low-value data points classified as “other”, which results in the reduction of stay points surrounding the personal areas. Using “PEM+”, evaluations showed that privacy-preservation doubled.

Oxford Artificial Intelligence Study

In the development phase of PEM, our data science team at Spectus performed extensive testing internally. After proving the value and effectiveness, we commissioned an external study to remove inherent bias. We partnered with Oxford Artificial Intelligence Limited (OxAI), a company specializing in Machine Learning and data science techniques based in Oxford, UK, whose team conducted an independent, rigorous, third-party evaluation of PEM’s effectiveness in preserving user privacy.

The Oxford-AI Privacy Attack Challenge

We proposed two problems to solve:

Scenario 1 (find the user): given a home latitude and longitude, find the anonymized device ID living there using upleveled device location. Repeat for 1000 home locations.

Scenario 2 (find the home): given an anonymized device ID, find its home using upleveled device location. Repeat for 1000 anonymized device IDs.

Attack Scenario 1 - “Find the Home”

In the first attack scenario, the OxAI team developed an algorithm to observe points that fall within the pre-identified home census block group for a given device, and then identify high density groupings of points surrounding a 100 meter gap. This approach assumes that the points surrounding the gap are generated when users are entering, leaving, or moving around their inferred home location. The algorithm then assumes that the centroid of the 100 meter gap is the true location of the user.

Upon submitting the algorithm’s guesses to Spectus for cross referencing (using unaltered mobility data), the results showed that the algorithm correctly predicted just 5% of inferred home locations within 30 meters, and 20% of homes within 100 meters, with a median error of 259 meters. After manually adjusting these results, which requires a more labor-intensive process, OxAI researchers correctly predicted 14% of homes within 30 meters, and 44% of homes within 100 meters.

By running this attack on Spectus’ improved “PEM+” dataset, privacy-preservation was doubled, with only 2.5% of homes identified within 30m.

Attack Scenario 2 - “Find the User”

Within the second attack scenario, OxAI researchers developed an algorithm to observe all devices that reside within a given census block group, and to predict whether or not they pertain to a specific inferred home location within that census block.

Specifically, the algorithm sets filters to remove devices that have pings closer than 100 meters to the specified inferred home location, as well as those whose nearest pings to the presumed home location would require implausible speed to reach within the timestamps displayed.

Under this scenario, researchers score devices based on the above parameters, and weigh devices based on their scores. They then submit only the highest scoring devices as guesses.

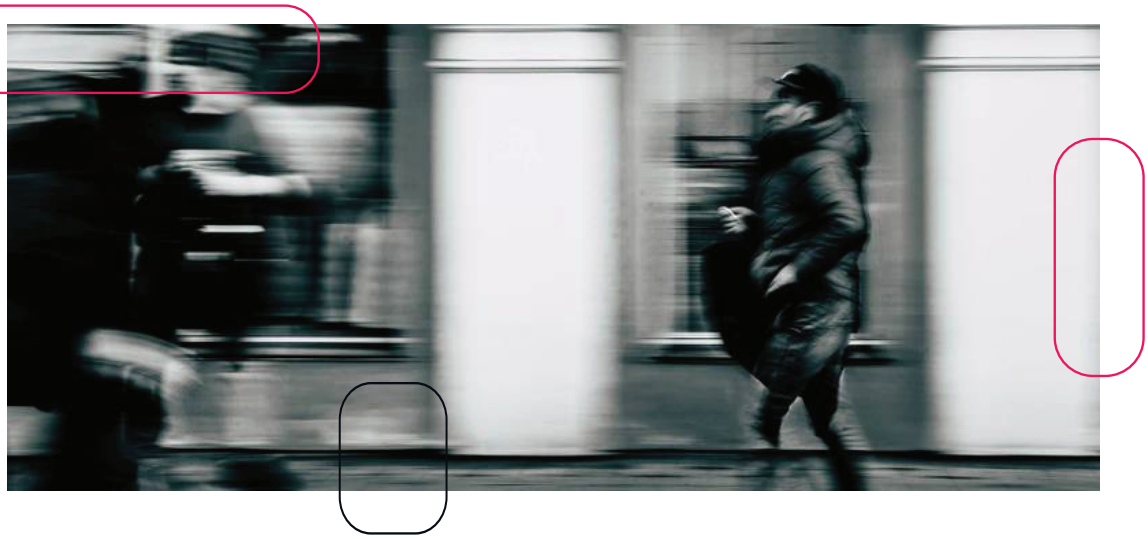
Upon submitting the guesses to Spectus, roughly 74 - 106 out of 848 devices were correct, resulting in an 8.5% - 12.5% accuracy of this approach. By running this attack on Spectus’ improved “PEM+” dataset, privacy-preservation was improved by a factor of 1.75x, with only a 4.9% - 7.1% accuracy of the attack algorithm.

Findings

In OxAI's evaluation of our PEM methodology, researchers demonstrated that Spectus PEM data had a 97.5% success rate in preventing matches between devices and their home locations. This represents a substantial improvement over industry standards of simple de-identification which, according to past studies, may only have a 5% success rate in preventing home-device matches.

When OxAI's algorithm was applied to PEM+ under attack scenario 1 "Find the Home", we observed a 2x improvement in privacy preservation over the original version of PEM. While 5% of inferred home locations were identified within an accuracy of 30m by the algorithm within the PEM dataset, the algorithm was only able to identify 2.5% of inferred home locations within the improved PEM+ dataset. Likewise, when the "Find the User" algorithm was applied to PEM+, we observed a 1.75x improvement in privacy preservation over the original version of PEM, with the re-identification risk reducing from 8.5% to 12% to just 4.9% - 7.1%.

In comparison to unaltered mobility data, in which 95% of users with at least 4 data points can be re-identified, PEM provides a marked improvement in privacy preservation, evinced by the fact that the aforementioned privacy attacks were only successful in associating a device with a precise inferred home location only 5% - 12.5% of the time. These numbers reflect the performance of the algorithm in absence of feedback from Spectus on the accuracy of predictions, which reflects a more realistic attack scenario.



Conclusion

The OxAI study reveals both strengths and areas of continued improvement for PEM. Privacy protection is our priority at Spectus, and it is integral to our approach to data collection. We are committed to providing an excellent user experience and clarity about how we collect location data, what we use it for, and more information about our privacy framework. As we continue to evolve our privacy approach to adhere to each new regulatory advance, such as GDPR and CCPA, we believe it's our responsibility to communicate honestly and openly about our stance and commitment to consumer privacy. We've built our privacy framework around four key principles: consent, transparency, control, and accountability. Learn more in our privacy center or email us with questions at privacy@spectus.ai.