Study shows how easy it is to determine someone's identity with cell phone data

March 25 2013, by Lisa Zyga

An artistic representation of being "unique in the crowd," representing the new finding that human mobility traces are highly unique, and mobility data can be used to reconstruct individuals' movements in space and time. The original photo is on the right. Credit: Yves-Alexandre de Montjoye, et al. Original image "Grand Central Station" by theotter

(Phys.org) —While most people know that using a cell phone means that the phone's location is being recorded, a new study has revealed just how little information is required to determine an individual's personal identity. By analyzing 15 months of cell phone mobility data from 1.5 million people, researchers have found that only four spatio-temporal points (an individual's approximate whereabouts at the approximate time when they're using their cell phone) are all that's needed to uniquely identify 95% of the individuals. The study has implications for modifying privacy law in order to keep pace with technological advances.

The researchers, Yves-Alexandre de Montjoye at MIT in Cambridge, Massachusetts, and Université Catholique de Louvain in Belgium, and his coauthors, have published their paper in Nature's *Scientific Reports* on how cell phone data places fundamental constraints on the privacy of an individual's mobility traces.

The reason that four locations are sufficient to identify most people is simply that human mobility is unique. Just as how everyone has unique fingerprints, everyone has unique daily travels. In the case of fingerprints, Edmond Locard showed in 1930 that only 12 points are needed to uniquely identify a fingerprint. Likewise, the researchers' data shows that just four spatio-temporal points are needed to uniquely identify the mobility trace of an individual. In other words, it's not likely that someone else will be in the same locations as you are at four different times of day. In fact, the researchers found that knowing just two randomly chosen points can uniquely identify more than 50% of the individuals.

As the researchers noted, these four points that allow for the identification of individuals could come from information that is publicly available, such as the individual's home address, workplace address, or Twitter posts.

One might expect that the data would provide more anonymity by decreasing the resolution of the data, which is done by increasing the time range from one hour to several hours and increasing the spatial range from a few square meters to several hundred square meters. As an analogy, decreasing the resolution of a photograph causes people in the photograph to appear blurry and unidentifiable.

But this is not what happens when decreasing the resolution of mobility data. Surprisingly, the researchers found that decreased resolution does not make the data that much more anonymous; a few more pieces of

information are all that is needed to identify individuals.
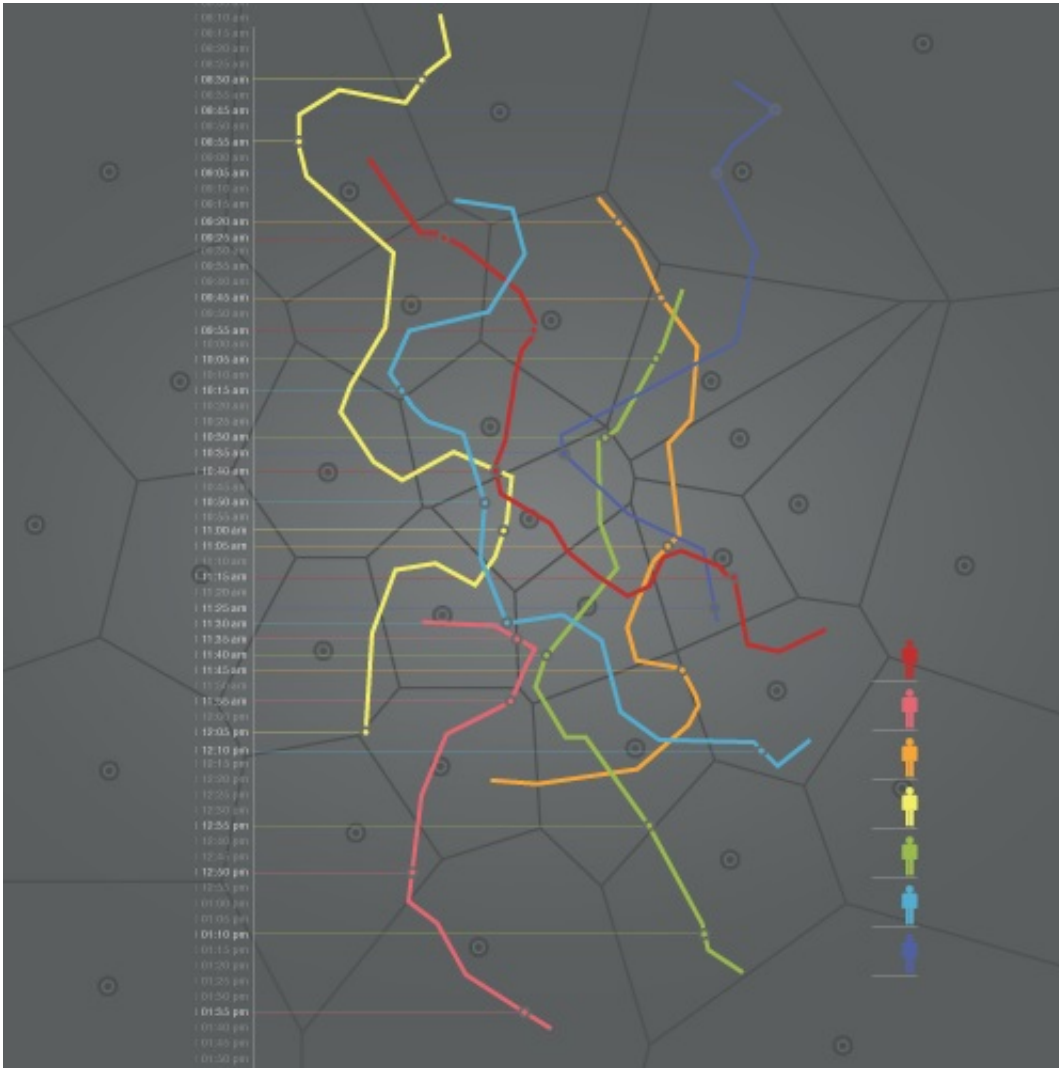
The unique traces of mobile phone users, along with the information that can be inferred from them, highlight the importance of understanding the privacy bounds of human mobility. Credit: Yves-Alexandre de Montjoye, et al.

As the focal point of their paper, the researchers used these results to develop a mathematical formula that tells the probability of uniquely identifying an individual based on the data's temporal and spatial

resolution. Essentially, they found a formula for estimating privacy.

These results highlight the potential risk to individual privacy and anonymity from mobility data combined with publicly available information. The researchers hope that the findings will help inform the design of future policies and technologies.

"Our results help us understand what is possible and what is not possible," de Montjoye told *Phys.org*. "We all have a lot to gain from this data being used. Our formula allows us to estimate privacy, so now the question is how do we use it to balance things out and make it a fair deal for everybody?"

In the future, de Montjoye plans to continue to investigate the consequences of what large amounts of personal data mean for an individual's privacy.

"Another area we want to pursue is to apply this same methodology to other sources of human-generated data," he said.

More information: Yves-Alexandre de Montjoye, César A. Hidalgo, Michel Verleysen, and Vincent D. Blondel. "Unique in the Crowd: The privacy bounds of human mobility." *Scientific Reports*. DOI: 10.1038/srep01376