DATA CAN BE USED FOR MACRO-LEVEL PLANNING

Using smartphone location data without infringing privacy

A new series in TODAY's Science section, in collaboration with the National University of Singapore School of Computing, starts today. This series explores a selection of computer science research projects carried out here and how they can improve people's lives.

COMMENTARY BY
STEPHANE BRESSAN

I have a smart phone; you have a smartphone; Alice has a smart phone. If our smartphones could talk they would tell where we are and where we have been. For better or for worse, they talk — and they have a lot of information to share.

With the emergence of smart cities — which is of particular relevance to a city-state like Singapore, which has positioned itself as a Smart Nation and a national test-bed for innovation — the question of how to open up such public and private data repositories without compromising the privacy of individuals is paramount.

The massive amounts of location data recorded by our mobile devices are a continuous and abundant source of data that can fuel the development of a smart city. Location data can assist city planning, make transportation systems intelligent and feed a multitude

of smart location-based services, offering utility and convenience to everyone.

Yet, this incessant chirping of our whereabouts, if in the wrong hands, may breach our privacy. Growing privacy concerns worldwide have made it difficult to fully unlock the potential of location data. Data managers are reluctant to make location data available for the development of innovative services by the entrepreneurial community.

To tackle this challenge, my team at the National University of Singapore (NUS) School of Computing, in collaboration with the SAP Innovation Cen-

Assoc Prof Bressan (right) and his team have developed an algorithm that 'cuts' longer trajectories into shorter segments and time windows. PHOTO: NUS

tre, has developed a method that allows the massive amounts of data stored in smartphones to be used for macro-level planning purposes without compromising the privacy of individuals.

The research shows how modifying data from smartphones can better protect the privacy of individuals.

My team, which also comprises Dr Song Yi from NUS and Dr Daniel Dahlmeier from the SAP Innovation Centre, first studied a large-scale mobility dataset with time-stamped location data from more than half-a-million users.

To determine the possibility of

• Stephane Bressan is an Associate Professor in the Computer Science department at the NUS School of Computing.
• The research was done at the SAP Innovation Centre located at the National Research Foundation's CREATE campus.

identifying individuals — even if the data contains no personal information such as names or NRIC numbers — my team examined the uniqueness of the location trajectories. For example, it is possible to find out where Alice has been by comparing her known home and work location with the recorded locations in the data set.

The results show that human mobility patterns are highly unique, confirming findings of previous studies. This means that the possibility of identifying individuals is high, but only if such data is anonymised by removing personal information.

To reduce this risk, my team developed an algorithm that "cuts" longer trajectories into shorter segments and shorter time windows. The modified data still represents the movements of crowds but makes it harder to trace one individual's movement.

The results of the study allow the value of smart devices to be unlocked without compromising the privacy of individuals. This algorithm was completed last year and governments, telecommunication companies and other data owners can apply this algorithm when publishing smart device data and other personal data. Entrepreneurs, small and medium enterprises may then tap the data and deploy tailored services.

As governments around the world look for ways to engage their citizens better, my team continues to work on developing privacy-preserving solutions to connect and correlate different data sources as one of the central enablers for smart cities.

Next week:
• How natural language processing technology is used to enhance computer-aided learning for the English language