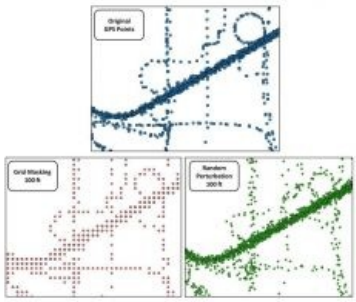
Protecting Location Privacy Brings Opportunities to the Geosector



Since members of the geosector are well versed in potential applications of geoinformation and its fusion with other data sources, they are best equipped to devise and test technology-based solutions for location privacy-related challenges. That is the opinion of Dara Seidl, a young expert on geoprivacy, who won the first pan-American master's thesis contest on geoinformation.

Dara Seidl was selected by a jury of worldwide experts as the winner of the 'Prize for the Outstanding Master's Thesis in Cartography, Geodesy and Geo-Information 2015' for her thesis titled 'Striking the Balance: Privacy and Spatial Pattern Preservation in Masked GPS Data'. The prize is awarded each year by the pan-American Institute of Geography and History (based in Mexico). Seidl wants to help to protect location privacy: the right of individuals to determine how, and the extent to which, their location information is shared with other parties.

Geoprivacy is a hot topic in the face of increasing digital surveillance by governments and as private companies are increasingly exploring the benefits of location-based services. Both private and public organisations encourage people to volunteer as much data as possible in their cars or via their phone to help the organisations to 'improve their service'. Even if they have nothing to hide, many people may prefer to decide for themselves which information they share.

The big data industry sees location as a crucial component to enable its analytical programs to extract meaningful results from the huge heap of data puzzle pieces. The flip side is that location is a strong personal identifier, primarily because of its role of linking disparate sources of data. More and more people are perceiving a risk in the growing capability to combine volunteered and non-volunteered data from various sources into personal profiles. Such profiles, which may be inaccurate, can subsequently be shared without consent, misused by authorities, insurance companies, etc., or even hacked.

The public release of government data with a location component in the interest of open data has a number of important advantages but, even in the case of public records, privacy issues merit attention. For example, in 2013, a newspaper in New York State caused outrage when it released an online map of gun owners from requests listed in public records. Related applications include remote sensing for change detection in the built environment, or mobile mapping images of every dwelling. And most individuals leave digital footprints that are picked up in big data, such as through their social media posts, credit card transactions and travel-card usage.

Digital fingerprint

GPS data from cars, smartphones or other devices create unique personal patterns. A 2013 study by de Montjoye et al, based on an anonymised dataset of mobile phone interactions for 1.5 million people, concluded that the use of just four location points over 15 hours is sufficient to identify 95% of all individuals. In the study, the authors examined the number of location points shared between traces within one hour to determine how many spatiotemporal points are necessary to make the trajectory unique. Consider a GPS dataset that contains waypoints for an individual's daily commute. Even if the GPS dataset contains no other information about the person other than a unique anonymous identifier, specialists can still likely determine that person's home address, place of work and other frequently visited locations, which can in turn be linked to other openly available data. Then, in the auxiliary data, the specialists may discern more sensitive information about the individual, including name, gender, age, marital status, health information, occupation and hobbies. In the health realm, location is termed a 'quasi-identifier' for its ability to establish links between background information and a given dataset. The identifying power of GPS data goes further than database linkage in that a sequence of locations provides clues about individual characteristics, such as whether the person is a parent indicated by visits to a school in the morning and afternoon with other locations in between.

Location remains a personal identifier even if the data is aggregated in a central server and becomes part of 'big data'. If geodata about individuals is not properly encrypted, centralisation in a server creates a high potential privacy risk since it may be linked with other personal attributes. Seidl points out: "Even though a given individual may seem anonymous in a flood of big data, his or her personal location data is unique enough to be identifying, like a digital fingerprint. As we improve and develop new technologies that collect increasingly fine-grained and high-volume geodata, we should not disregard ethical considerations."

Obfuscation

The anonymisation techniques Dara Seidl applied in her thesis are not computationally expensive, even for large volumes of data. The datasets she used contained close to one million points, and the masking procedures themselves took very little time to run. More complex procedures may require a longer processing time to evaluate levels of privacy and spatial distribution. For example, some anonymisation techniques are coded as optimisation problems, maximising the preservation of spatial pattern while maintaining some threshold privacy value. In Seidl's view, these conceptualisations tend to require more time to complete, but should not be cost-prohibitive. And, of course, the decision to use obfuscation techniques should always be dependent on the particular considerations of the dataset and the estimated privacy risks.

How does obfuscation work? Obfuscation, also known as 'masking', introduces slight inaccuracy to geographic data with the goal of protecting privacy. Each point in a point set is displaced, typically within some distance threshold, so that the associated individual cannot be re-identified based on location, but the spatial pattern of the dataset is preserved at all times. A number of methods exist to achieve this balance, some of which weight the distance of displacement by some characteristic of the underlying population, such as population density. In Seidl's winning thesis, the techniques applied are random perturbation, which displaces each point by a random distance in a random direction within a distance threshold, and grid masking, which snaps each point to the centroid of grid cells of a specified size overlaid with the data.

New growth areas

"Privacy considerations should not be seen as restrictive to the growth of the geomatics sector. Instead, the incorporation of privacy-preserving techniques can be a new area of growth," states Dara Seidl. "Those in geomatics are best equipped to understand how personal identities can be compromised in geodata. One argument posed by those who study location privacy is that it is better that those in the geosector focus now on developing effective technical privacy solutions before more stringent or over-protective rules on data collection are imposed through government regulations. I believe that governments and the geosector must work together to familiarise each other with how privacy protection strategies may work in tandem to prevent identity disclosure without stifling the benefits of new technology developments."

Remote sensing and mobile mapping images represent vast and precise sources of geodata that can produce virtual copies of the world. These industries have already incorporated some important privacy protections, such as by blurring faces and licence plates or omitting data around government or military sites. However, privacy strategies should ideally be in place at the data collection stage. Data collection necessitates storage and, if encryption is inadequate, repositories of data may be vulnerable to disclosure – potentially through hacking activity – before the data even reaches the analysis stage. "Location privacy is not just relevant to clients of collected data, but to anyone who plays a role in data collection," is Seidl's strong opinion.

Geo-ICT companies are looking for new markets and they could certainly offer obfuscation techniques as a selling point. The realms of healthcare, location-based services, banking and transportation have no doubt experienced tensions between data utilisation and privacy concerns, and may even have faced legal action with regard to privacy. It is commonplace in the United States for the geosector to offer privacy protection strategies in project proposals. The most common privacy solutions for geodata are to aggregate or omit data in sparsely populated regions, to establish data enclaves or secure remote access, or to use software agents to perform calculations and provide answers. These solutions are strongly privacy-protective, but remain somewhat at odds with data democratisation and the open access movement, as the public must typically apply for access. Dara Seidl is certain: "This is where obfuscation could become a selling point, by allowing the release of masked geodata without administrative barriers. We are seeing increasing interest in obfuscation techniques particularly in health and transportation research, as well as in citizen science projects."

Pan-American Cartography, Geodesy and GIScience Thesis Contest

An annual prize is awarded by the distinguished Pan American Institute of Geography and History (PAIGH) for the best master's thesis in the specific fields of cartography, geodesy and geographic information science in general, including aspects such as data capture, manipulation, presentation and dissemination. Besides the thesis covered in this article, two further distinctions were awarded in 2015: to Ms Sol Pérez form Mexico (thesis titled '*Atlas de la nueva geografía de la minería en México y los conflictos asociados a ella'*) and to Mr Bruno Lara from Argentina (thesis titled '*Fragmentaticíon de pastizales en el centro de la provincia de Buenos Aires mediante imágenes landsat'*). For 2016, a prize will be awarded at both MSc and PhD level. The goal of the prize is to promote and recognise high-quality academic and scientific work by students on official postgraduate programmes led by organisations or universities located in PAIGH's member states, as well as citizens of such member states who have graduated elsewhere. More information here.

https://www.gim-international.com/content/article/protecting-location-privacy-brings-opportunities-to-the-geosector