Scientists show telephone metadata can reveal surprisingly sensitive personal information

May 17 2016, by Bjorn Carey

Credit: AI-generated image (disclaimer)

Most people might not give telephone metadata – the numbers you dial, the length of your calls – a second thought. Some government officials probably view it as similarly trivial, which is why this information can be obtained without a warrant.

But a new analysis by Stanford computer scientists shows that it is possible to identify a person's private information – such as health details – from metadata alone. Additionally, following metadata "hops" from one person's communications can involve thousands of other people.

The researchers set out to fill knowledge gaps within the National Security Agency's current phone metadata program, which has drawn conflicting assertions about its privacy impacts. The law currently treats call content and metadata separately and makes it easier for government agencies to obtain metadata, in part because it assumes that it shouldn't be possible to infer specific sensitive details about people based on metadata alone.

The findings, reported today in the *Proceedings of the National Academy of Sciences*, provide the first empirical data on the privacy properties of telephone metadata. Preliminary versions of the work, previously made available online, have already played a role in federal surveillance policy and have been cited in litigation filings and letters to legislators in both the United States and abroad. The final work could be used to help make more informed policy decisions about government surveillance and consumer data privacy.

The computer scientists built a smartphone application that retrieved the previous call and text message metadata – the numbers, times and lengths of communications – from more than 800 volunteers' smartphone logs. In total, participants provided records of more than 250,000 calls and 1.2 million texts. The researchers then used a combination of inexpensive automated and manual processes to illustrate both the extent of the reach – how many people would be involved in a scan of a single person – and the level of sensitive information that can be gleaned about each user.

From a small selection of the users, the Stanford researchers were able to infer, for instance, that a person who placed several calls to a cardiologist, a local drugstore and a cardiac arrhythmia monitoring device hotline likely suffers from cardiac arrhythmia. Another study participant likely owns an AR semiautomatic rifle, based on frequent calls to a local firearms dealer that prominently advertises AR semiautomatic rifles and to the customer support hotline of a major firearm manufacturer that produces these rifles.

One of the government's justifications for allowing law enforcement and national security agencies to access metadata without warrants is the underlying belief that it's not sensitive information. This work shows that assumption is not true.

"I was somewhat surprised by how successfully we inferred sensitive details about individuals," said study co-author Patrick Mutchler, a graduate student at Stanford. "It feels intuitive that the businesses you call say something about yourself. But when you look at how effectively we were able to identify that a person likely had a medical condition, which we consider intensely private, that was interesting."

They also found that a large number of people could get caught up in a single surveillance sweep. When the National Security Agency examines metadata associated with a suspect's phone, it is allowed to examine a "two-hop" net around the suspect. Suspect A calls person B is one hop; person B calls person C is the second hop. Analysts can then comb the metadata of anyone within two hops of the suspect.

By extrapolating participant data, the researchers estimated that the NSA's current authorities could allow for surveilling roughly 25,000 individuals – and possibly more – starting from just one "seed" phone user.

Although the results are not surprising, the researchers said that the raw, empirical data provide a better-informed starting point for future conversations between privacy interest groups and policymakers.

For instance, the authors point to the recent shift to reduce the metadata retrieval window from five years to 18 months. By drawing accurate and sensitive inferences about participants from roughly six months-worth of calls and texts, the study suggests that metadata are more revealing than previously thought.

Similarly, the government's two-hop call sweep was previously three hops; that reduction was implemented to reduce the number of people caught in a sweep. Shortening the time window could reduce that number further, Mutchler said.

"If we're going to pick a sweet spot as society, where we want the privacy vs. security tradeoff to lie, it's important to understand the implications of the polices that we have," Mutchler said. "In this paper, we have empirical data, which I think will help people make informed decisions."

More information: Jonathan Mayer et al. Evaluating the privacy properties of telephone metadata, *Proceedings of the National Academy of Sciences* (2016). DOI: 10.1073/pnas.1508081113

Provided by Stanford University