How CSPs Can Manage Metadata Privacy Concerns

Yaakov Stein
Forbes Councils Member
Forbes Technology Council
Aug 10, 2023

Yaakov (J) Stein is CTO at Allot.

Ten years ago, The Guardian and The Washington Post **PUBLISHED ARTICLES** detailing mass surveillance of U.S. domestic telecommunications by the National Security Agency (NSA).

In particular, they raised questions about if and how two types of technology companies had handed over personal data relating to their customers: prominent "over the top" **INTERNET TECHNOLOGY MEGACOMPANIES** (OTTs), such as Google, Facebook and Amazon; and **COMMUNICATIONS SERVICE PROVIDERS** (CSPs), such as Verizon and AT&T.

The information's source was a **HORDE OF INTERNAL NSA DOCUMENTS** leaked by Edward Snowden. The news triggered in-depth debate not only on government surveillance, but on the very nature of privacy in the modern world. In this article, I briefly describe the effect of these revelations on OTTs and CSPs.

The OTTs

The essential OTT business model is based on gleaning as much information as possible from the customer's interactions on its platform. In fact, an OTT **MAY KNOW MORE ABOUT A CUSTOMER** than she knows about herself. This knowledge enables **EXQUISITELY TARGETED ADVERTISING**, and has enabled the OTTs to become the most highly valued companies of all time.

The OTTs quickly **EXPRESSED OUTRAGE** at their data centers' links being hacked, but either denied or remained quiet about their being complicit in giving front-door access to the NSA. **SOME THREATENED LEGAL ACTION** against the NSA in order to restore public confidence.

While the Snowden revelations brought privacy back into the limelight for a short time, **MANY PEOPLE ARE APPARENTLY WILLING TO FORGO PRIVACY** in exchange for services or convenience. While there was large public debate at the time of the leaks, the OTTs **STILL COLLECT** a vast amount of data.

The CSPs

CSPs MOSTLY COLLECT AND STORE METADATA (data about data), such as the start-time of a call or connection, the subscriber's identifier, the amount of data transferred and the time the connection ended. This information is needed for billing purposes. CSPs also collect (but seldom store) more in-depth metadata for traffic management.

While the requirement for billing metadata is easily understood, the aim of traffic management is more subtle, although arguably more crucial. Traffic management is the process of prioritizing particular kinds of user communications under conditions when there are insufficient resources to optimally handle all traffic. To optimally prioritize, for example, the CSP needs to know what the customer is doing in order to provide the necessary bandwidth for video streaming and to ensure low latency for gaming, while delaying non-real-time applications such as email or backups.

Some may think that CSPs oversubscribing resources should occur no more than, say, airline overbooking; but the situations are hardly analogous. For example, you might have a one Gbps internet connection, but the percentage of the time you really use anything approaching that speed is very small. Most of the day your internet connection is dormant, and even streaming four independent UHD movies requires LESS THAN 100 MBPS.

If a CSP multiplies the number of subscribers by 100 Mbps instead of 1 Gbps, the access fiber will still typically be relatively empty. But without traffic management, statistical usage peaks would inevitably lead to noticeable performance degradation. And the need for traffic management has grown dramatically over the past 10 years. At the time of the initial Snowden revelations, the average wired internet speed WAS UNDER 10 MBPS, while Gbps fiber connections are BECOMING MORE COMMON today, a hundredfold increase.

Unlike the OTTs, which are close to monopolies in their market segments, different CSP networks interwork to provide the end-to-end service (which is why the aggregate is called "the internet"). This only works due to protocol standardization by such organizations as the INTERNET ENGINEERING TASK FORCE (IETF) and the INTERNATIONAL TELECOMMUNICATION UNION (ITU).

The professionals at these organizations OBJECTED TO THE IDEA of pervasive monitoring, and immediately started PLUGGING ALL LOOPHOLES discovered by the NSA. This not only made it harder for government surveillance, but makes it harder for CSPs to perform traffic management.

In regards to CSPs managing traffic, the primary mechanism for application identification—deep packet inspection (DPI)—was deemed nearly untouchable in much of the Western world; now, its **EFFICACY IS BEING THREATENED** because of these restrictions. Rather than monitoring intentional protocol artifacts, modern DPI needs to leverage AI to track application and user behavior, inevitably leading to longer application identification times and lower identification precision.

The CSP solution

CSPs are under contract to provide quality service to their subscribers, meaning they need to gather information about these subscribers. Recent technological trends are making it more and more difficult and expensive to gather this information. What can CSPs do about this trend?

The raison d'être of the aforementioned technological trends is shrouding metadata from pervasive long-term collection, but preventing the use of this metadata for traffic management systems is collateral damage. Two solutions are possible.

The simplest one is for CSPs to aggregate metadata at frequent intervals (say every 15 minutes) for monitoring and planning purposes, leaving no subscriber-traceable data that could be demanded by third parties. Unfortunately, there **MAY BE LEGAL REQUIREMENTS** to save any metadata that has been gathered. Even if the CSP wipes potentially private information, invasive third-party taps might store it. These are probably the reasons that the technological mechanisms have been so aggressive.

The second solution would be to open a metadata control channel from service consumer to service provider, allowing consumers to opt out from sharing any data with which they are not comfortable. Those opting out would either accept possible degradation of quality of experience during periods of network congestion or pay for the extra resource utilization needed to blindly guarantee service quality. Combined with technological developments, this approach assures the highest level of privacy for those who desire it, and intermediate levels for those who wish to trade-off privacy for lower cost.

Were such a model to become popular, one could imagine it being applied to OTT services as well.