More data for the NCP implies more privacy risks

Bart Coppens, Olivier Zendra

In the Next Computing Paradigm (NCP), numerous home sensors will communicate with servers and services. The multitude of data and communication is bound to raise privacy issues and these must be taken care of for the NCP to succeed.

More data for the NCP implies more privacy risks

by Bart Coppens and Olivier Zendra

Key insights

- Even the most innocuous data sources, such as those from IoT-based sensors, can be used to infer personal and private information.
- Integrating different sources of data significantly increases the amount of personal and private information that can be inferred.
- When such data is exposed, this creates an enormous privacy problem for the affected people.
- Systems that are tightly integrated into people's daily lives thus lead to a significant privacy risk.
- The NCP is such a tightly-integrated system which could lead to a significant privacy risk if not properly designed and implemented.

Key recommendations

- Promote research into technologies that enhance people's privacy and reduce the risks and impact of leaks of private data.
- Stand by EU principles of privacy for its citizens, requiring companies to actively adhere to the principles of privacy by design. In particular, functional requirements for systems should include:
    - be designed to not leak data in any form whatsoever, except with the explicit consent of the users; no backdoors should be allowed.
    - give the user the option to decide (with explicit consent) what data is collected, and to understand what that consent implies.

Innocuous Data Sources Can Leak Personal and Private Data

Data stored by systems can come from numerous sources. Once source is users uploading or explicitly sharing their own data (when writing in an online editor, sharing photos, sending messages, etc.). In these cases, users often know that they share information with others and can mentally distinguish between personal and private data. However, this explicitly user-shared data is only the tip of the proverbial iceberg of data. Other sources of data include the (logs of the) user's activity, and the sensors that increasingly pervade our daily lives. Users are typically unaware of these sources and cannot estimate how personal or private data extracted from them can be. In this article, we focus on the privacy of data inferred from actual, physical sensors; however data extracted from digital interactions with a system presents major privacy concerns too.

Different sensors keep being introduced into our lives. Obvious ones include security cameras in our cities and villages, and ANPR cameras that track car license plates. Our mental picture of these cameras seems quite clear: they are on fixed spots, and relatively visible, and we 'know' their purpose is catching criminals. However, the privacy implications might not always be immediately as clear. But these cameras of course do not magically only turn themselves on when a convicted criminal passes by: they are always on.

While such cameras can already lead to some privacy worries, they are in public locations, and we are typically aware of them. Mobile device cameras are more intriguing. Google Glass faced a backlash, with people calling users using Glass in a creepy manner 'glassholes' [19], and venues banning their use because of privacy concerns [20,21,22]. Google Glass was first rebranded to focus on productivity in professional settings, such as technicians and factory workers [6], but was eventually completely discontinued [16].

A more nuanced example of cameras in private places are those found on laptops and smartphones. Many laptops have a front-facing camera for video calls. Unfortunately, this means the laptop camera faces the user when making video calls. This would be fine if the camera just recorded what the user wants, when the user wants it. However, an attacker could illegally access the laptop, enable the camera (which is always there) without consent or awareness of the user, and record whatever private and personal is happening in front of the camera. Because this worries customers, some laptop makers add a privacy screen for laptop cameras so users can physically prevent unwanted recording. Since privacy screens physically obstruct the camera lens, users can clearly see when they're active. More importantly, this cannot be overridden from software. The latter is crucial: if an attacker has gained control of the laptop and secretly enabled the camera, the attacker could also try to override any software-based protection. Note that not all users care about this issue. Some users will trust that their browser's / smartphone's webcam prompts are sufficient protection, others will say they don't think they'll be targeted, or aren't high-profile enough to warrant circumventing protections, and others will say they don't care if they're filmed by an attacker, or that the privacy screen's hassle outweighs its benefits. An additional consideration is that if an attacker takes over a device to control a camera, that attacker probably has access to all files on that device. While people can know upfront what information is (not) stored on a device, they cannot know what will be said or done in front of a camera. It is important to realize that the aforementioned considerations divide people: some find them unimportant, while others find them crucial. These very personal and individual trade-offs have varied outcomes for various persons. This is fine because such users were given the choice, can weigh the advantages and downsides *for their specific situation*, and can make an informed decision.

However, given the success of these privacy screens, this example should be considered more broadly. While users are aware and actively decide whether or not their laptop's camera can record them, they are typically less aware of their *smartphone* cameras. These however may partially face the same scenes that the laptop's webcam was deemed too risky or private to capture. People rarely even *think* about their phones' cameras, let alone privacy screens for them. Even when people mask their laptop's camera, their laptop and smartphone contain microphones able to record private talks. As with cameras, some individuals trade off that this is not a real threat, while others do.

Microphones may indeed constitute a privacy risk. Academics were (and still are) reluctant to provide recordings of online classes due to privacy concerns [24] and the risk of excerpts being

taken out of context [23], even though such classes are semi-public. In intimate conversations, privacy becomes more obvious: who wants an automated assistant listening in on their MD or partner in bed? While some might not care, others would. This is similar to Orwell's "1984" novel, where telescreens gather video and audio [25], except that now it is primarily corporations doing so (although some governments still strive to get that information).

Once pointed out, most people can reason about the privacy implications of cameras and microphones spying on private scenes and conversations. But not all sensors and types of data will elicit the same reaction, because even though the privacy implications are high, it can be harder to intuitively grasp attackers' imaginative possibilities. For example, not only may microphones be used to listen to conversations, but an eavesdropper can recognise the sounds a keyboard makes when typing, which can reveal confidential messages or passwords [26,27]. A smartphone could thus even leak information typed on a non-internet-connected computer. A microphone isn't even required for such an attack: an accelerometer in a smartphone on the same table as a laptop can capture what's being typed on the laptop through table vibrations [17].

One last example: smart building CO_2 sensors. These can monitor building air quality and control ventilation. However, these sensors can detect a room's occupant [18].

Again, these sensors being present does not mean data *will* be utilised against user privacy expectations. But there is a non-zero *risk* that an attacker will surreptitiously enable sensors when they should not be, or that sensors' data will be exploited to deduce and disclose personal information. How much risk is tolerated varies by user.

Integrating Different Data Sources Increases Privacy Issues

More sensors and data sources increase privacy danger. The main reason is that all of these sources have a privacy risk, and having more combines them. However, that is not the end of the story.

Multiple data sources that *individually* do not reveal privacy-sensitive information may *leak* it when *combined*. Consider GPS position. As an individual data source, it already allows attackers to determine that people's homes are where their GPS signals remain most nights, their workplaces are where they stay during the day, etc. GPS combined with *public map data* reveals vast amounts of private information. A GPS location staying near churches, gay clubs, abortion clinics, etc., can reveal very sensitive information about religion, sexual preferences, health, etc. People may be wary of sharing such precise location data with huge firms like e.g. Google, but they often forget that when using Google Maps or Waze they transmit the same information to Google.

One example is de-anonymizing (potentially innocent) media-reported crime suspects. This reporting can include initials, age, profession, location, etc. While each of these bits of information is *individually* shared harmlessly with many people, *combined* they deanonymize reports, identifying the report's person. This information can be further integrated with semi-public sources like white pages, data breaches, and leaks [3].

Fitness monitoring social networks like Strava are another example. Some people want to share their cycling and running successes, but they don't want others to know where they live or work. Sharing their entire track would however reveal that. This caused social media platforms to implement endpoint privacy zones that hide track ends. However, these websites reported the track length and middle part of the track. This information can be used with maps to confine the start and end positions of the tracks, revealing critical information [4]. Similarly, soldiers' exercise paths have been used to locate military bases abroad, raising concerns about location data shared with fitness trackers like Strava [2]. Using runners' itineraries, Figure 1 even shows a military base's internal map, in Helmand Province, Afghanistan [2].

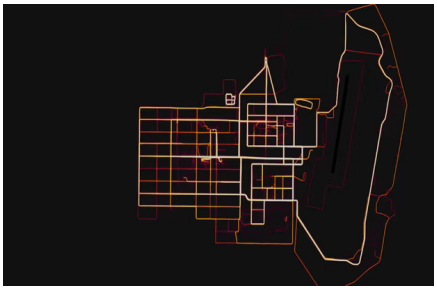

Figure 1. A military base in Helmand Province, Afghanistan with routes taken by joggers highlighted by Strava. Photograph: Strava Heatmap. Caption & image from The Guardian [2].

This also ties in with the difficulty of anonymization of data sets. If one collects precise GPS information, but 'anonymizes' it by removing the name, it is really not anonymized in any practical sense. The more information can be combined, the more options to deanonymize, the more options to reveal privacy-sensitive information [28].

The Risks of Systems that Tightly Integrate with People's Lives

What does this mean for systems that tightly integrate into people's lives? We discussed smartphones, computers, fitness trackers, etc., and their sensors before. The privacy risks of all the data such technologies can collect are obvious. Perhaps surprisingly, cars too contain multiple sensors that are tightly integrated into our daily lives. Systems are being developed to measure driver eye activity, tension, and well-being [9]. The Mozilla Foundation examined car manufacturers' privacy practices [1,7]. The outcomes were appalling. Modern cars come with numerous sensors inside and outside and can connect to your phone. They can monitor your music, access your contact list, even record and intercept text communications [8]. This leads to privacy policies relating to music preferences, employment, and sometimes 'sexual activity', 'sex life', 'psychological trends', 'intelligence,' 'genetic characteristics', and many other characteristics that many people would find creepy to think their car can collect or infer [1,7,10]. These policies also explain how car companies can not only collect such data, but also sell them [1,7,10].

It should be clear that systems which tightly integrate in our daily lives can collect a dizzying and overwhelming amount of private data. Once collected, what can be done with the data? Of course they can be sold. Even when they are not sold, but are merely collected for 'internal' or 'business' purposes, these sensitive data persist as a privacy risk. Who has access to such data is unclear even in such cases. A few years ago, for Siri's voice assistant quality control, Apple supplied contractors portions of recorded conversations, even those where Siri misidentified its activation instruction [13]. These fragments included recordings of people having sex, or discussing confidential medical information [13]. Another, worse example is Tesla personnel internally exchanging and joking about embarrassing or intimate customer recordings collected by Tesla cars [11]. Even when all employees and contractors behave, a breach could expose this data, since automotive manufacturers are also involved in data leaks and breaches [12,14,15].

Even if not all such information would be explicitly kept or inferred for European customers due to regulations, it's important to realize that *the sensors are still there*. The sensors will still be there, recording data. Even if not everything is *sent* or *stored* by default, it's still possible to *enable* recording or storing of data. The data, once recorded, still allow for inferences about privacy-sensitive topics to be made, if not now, then perhaps later if it is stored. The sensors, even if they *should* be disabled and not recording, are not necessarily so. For example, consider the Pegasus spyware, which amongst its many spying features allows a remote attacker to surreptitiously activate cameras and microphones on an infected smartphone [5]. This spyware was used by authoritative governments, allowing them to spy not only on their own citizens, but also on foreign citizens [5]. Importantly, if governments are able to do so, criminals can try to make use of the same mechanisms to spy on people. This should be a dire warning to *not* require (and even *forbid*) such backdoors, as they also introduce a significant additional risk to the privacy of end users because that back door could also enable criminals to listen in on them directly.

Some of the above privacy risks may surprise many people. Often individuals are ignorant of the sensors and the data they record or infer, alone or in combination. They are often unaware that talking about sensitive information near a seemingly-inactive phone might still be sending fragments of that conversation to contractors or that their car might take nude pictures of them and send them to the car manufacturer who can then make fun of them. They are often unaware that their car's GPS (built-in or smartphone app) can deduce their sexual preference, pregnancy, health, etc.

Complex systems with many sensors always require a balance: the more you want a system integrated into people's life to measure, the more sensors it needs, and the greater the privacy risk. Not everyone has this issue, everyone's risk assessment trade-off is different. Users should be able to determine their own trade-offs and *accept risk* if relevant. But to do so, ***the user must be aware of the risk***. Too often, privacy policies hide these risks in tiny corners with opaque legal wording

and/or imprecise language, obscuring the actual impact. Users can authorise or decline data collection only if they understand it. Furthermore, their ***consent must be a meaningful choice***, not simply 'either use our product with all sensor data being collected, or don't use it at all'. Expecting customers to either have their car collect mental state data or not drive a (modern) car is a false choice. Users must have a ***meaningful choice to disable unnecessary sensors and collected and inferred data without losing the entire system access***. In addition, systems should ***never disclose or leak data without the explicit, informed and meaningful consent of the user.***

Finally, it is worth considering the design space of possible approaches to fulfil such goals. Webcam privacy screens are an interesting design choice, though not always applicable. It plainly reveals whether explicit or accidental video recording is allowed, in a way that is not corruptible or subvertible in software. This visual cue makes the webcam difficult to overlook. The laptop's non-webcam functions can still be used when the webcam is disabled. Thus, a system in which the sensor whose functionality a user does not need can be *physically* removed (either temporarily or permanently, e.g. by physically disabling or removing it) can give the user a clear *choice* that is not subvertible by software, hence *reducing the attack surface* in a meaningful way.

To achieve such goals, research into technologies that improve privacy and prevent data leaks is crucial. This should cover both the *design* of systems, as well as their *implementation.*

The NCP is precisely such a system integrating tightly into people's lives, comprising many sensors, and combining many different data sources. From a privacy perspective, it is crucial that its design includes the above ***requirements*** for informing the user clearly about the potential risks, allowing the user to enable or disable certain sensors and information sources from being integrated into the system, and only sharing information with *any* party with the user's *explicit, informed and meaningful consent*, so that each user can make their own trade-off. Again, the NCP's design and implementation must both meet privacy criteria. Its functionalities must gracefully degrade depending on user privacy choices.

Conclusion

Even innocuous data sources like IoT sensors can reveal personal information. Integrating data sources greatly increases the amount of personal and private information inferred. When such data is exposed, affected users face huge privacy issues. Integrating systems tightly into people's daily life increases privacy risks. Due to its intimate integration, the NCP could present a considerable privacy risk if not properly designed and implemented. Thus, the NCP must be designed and implemented to let users consent — or deny — to data collection and data sharing knowingly. In addition, the NCP must make that choice a meaningful one, by having only the specific features, in the narrowest sense, that actually need the denied information degraded by that refusal.

References

[1] It's Official: Cars Are the Worst Product Category We Have Ever Reviewed for Privacy. Jen Caltrider, Misha Rykov and Zoë MacDonald. Privacy Not Included, Mozilla. 2023-09-06 (accessed 2023-11-23) https://foundation.mozilla.org/en/privacynotincluded/articles/its-official-cars-are-the-worst-product-category-we-have-ever-reviewed-for-privacy/

[2] Fitness tracking app Strava gives away location of secret US army bases. Alex Hern. The Guardian. 2018-01-28 (accessed 2023-11-27). https://www.theguardian.com/world/2018/jan/28/fitness-tracking-app-gives-away-location-of-secret-us-army-bases

[3] De Boeck, Kevin, et al. "Poster: The impact of public data during de-anonymization: a case study." 7th IEEE European Symposium on Security and Privacy, Date: 2022/06/06-2022/06/10, Location: Genoa. 2022.

[4] Dhondt, Karel, et al. "A Run a Day Won't Keep the Hacker Away: Inference Attacks on Endpoint Privacy Zones in Fitness Tracking Social Networks." Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security. 2022.

[5] Stephanie Kirchgaessner, Paul Lewis, David Pegg,Sam Cutler,Nina Lakhani and Michael Safi. "Revealed: leak uncovers global abuse of cyber-surveillance weapon". The Guardian, 2021-07-21 (accessed 2023-11-23). https://www.theguardian.com/world/2021/jul/18/revealed-leak-uncovers-global-abuse-of-cyber-surveillance-weapon-nso-group-pegasus

[6] Google Glass 2.0 Is a Startling Second Act. Steven Levy. Wired. 2017-07-18 (accessed 2023-11-23) https://www.wired.com/story/google-glass-2-is-here/

[7] What Data Does My Car Collect About Me and Where Does It Go? Jen Caltrider, Misha Rykov and Zoë MacDonald. Privacy Not Included, Mozilla. 2023-09-06 (accessed 2023-11-23) https://foundation.mozilla.org/en/privacynotincluded/articles/what-data-does-my-car-collect-about-me-and-where-does-it-go/

[8] Court rules automakers can record and intercept owner text messages. Suzanne Smalley. The Record. 2023-11-08 (accessed

2023-11-23) https://therecord.media/class-action-lawsuit-cars-text-messages-privacy

[9] Harman's driver-monitoring system can measure your heart rate. Rebecca Bellan. TechCrunch. 2023-01-04, accessed 2023-11-23 https://techcrunch.com/2023/01/04/harmans-driver-monitoring-system-can-measure-your-heart-rate/

[10] Nissan. Privacy Not Included, Mozilla. 2023-08-18 (accessed 2023-11-23) https://foundation.mozilla.org/en/privacynotincluded/nissan/

[11] Tesla workers shared sensitive images recorded by customer cars. Steve Stecklow, Waylon Cunningham and Hyunjoo Jin. Reuters. 2023-04-06 (accessed 2023-11-23) https://www.reuters.com/technology/tesla-workers-shared-sensitive-images-recorded-by-customer-cars-2023-04-06/

[12] More than 2 million Toyota users face risk of vehicle data leak in Japan. Daniel Leussink and Kantaro Komiya. Reuters. 2023-05-12 (accessed 2023-11-23) https://www.reuters.com/business/autos-transportation/toyota-flags-possible-leak-more-than-2-mln-users-vehicle-data-japan-2023-05-12/

[13] Apple contractors 'regularly hear confidential details' on Siri recordings. Alex Hern. The Guardian. 2019-07-26 (accessed 2023-11-23) https://www.theguardian.com/technology/2019/jul/26/apple-contractors-regularly-hear-confidential-details-on-siri-recordings

[14] Volkswagen, Audi disclose data breach impacting over 3.3 million customers, interested buyers. Charlie Osborne. ZDNet. 2021-06-14 (accessed 2023-11-23) https://www.zdnet.com/article/volkswagen-audi-disclose-data-breach-impacting-over-3-3-million-customers-interested-buyers/

[15] 1.6 million hit in possible Mercedes-Benz data breach — what you need to know. Paul Wagenseil. Tom's guide. 2022-10-20, (accessed 2023-11-23) https://www.tomsguide.com/news/mercedes-benz-data-breach

[16] RIP (again): Google Glass will no longer be sold. Samuel Axon. Ars Technica. 2023-03-16 (accessed 2023-11-23) https://arstechnica.com/gadgets/2023/03/google-glass-is-about-to-be-discontinued-again/

[17] Marquardt, Philip, et al. "(sp)iPhone: Decoding vibrations from nearby keyboards using mobile phone accelerometers." Proceedings of the 18th ACM conference on Computer and communications security. 2011.

[18] da Silva, Marlon P., et al. "Impact of using a privacy model on smart buildings data for CO2 prediction." arXiv preprint arXiv:2306.00766 (2023).

[19] Google Glass advice: how to avoid being a glasshole. Samuel Gibbs. The Guardian. 2014-02-19 (accessed 2023-11-27) https://www.theguardian.com/technology/2014/feb/19/google-glass-advice-smartglasses-glasshole

[20] Theater chain bans Google Glass. David Kravets. Ars Technica. 2014-06-10 (accessed 2023-11-27) https://arstechnica.com/tech-policy/2014/06/theater-chain-bans-google-glass/

[21] Unbearable wearable: Bar bans Google Glass, boots 'rude' user. Devin Coldewey. NBC News. 2013-11-26 (accessed 2023-11-27) https://www.nbcnews.com/technolog/unbearable-wearable-bar-bans-google-glass-boots-rude-user-2d11660837

[22] San Francisco bar bans Google Glass for fear of secret recordings. Gale Holland. Los Angeles Times. 2014-03-04 (accessed 2023-11-27) https://www.latimes.com/local/lanow/la-me-ln-bar-bans-google-glass-wearers-20140304-story.html

[23] Is This Thing On? Jeffrey Aaron Snyder and Amna Khalid. Inside Higher Ed. 2020-10-13 (accessed 2023-11-27) https://www.insidehighered.com/views/2020/10/14/thorny-issues-surrounding-classroom-recordings-reasonable-accommodations-and

[24] Should Professors Still Record Lectures? Maybe. Maybe Not. Susan D'Agostino. Inside Higher Ed. 2022-09-06 (accessed 2023-11-27) https://www.insidehighered.com/news/2022/09/07/should-professors-still-record-lectures-maybe-maybe-not

[25] 1984. George Orwell. Published 1949-06-08 by Secker & Warburg.

[26] Asonov, Dmitri, and Rakesh Agrawal. "Keyboard acoustic emanations." IEEE Symposium on Security and Privacy, 2004. Proceedings. 2004. IEEE, 2004.

[27] Zhuang, Li, Feng Zhou, and J. Doug Tygar. "Keyboard acoustic emanations revisited." ACM Transactions on Information and System Security (TISSEC) 13.1 (2009): 1-26.

[28] Data anonymization in Big Data scenarios: an open challenge to become GDPR compliant. Sara El Kortbi Martínez. Gradiant. 2021-11-11 (accessed 2023-11-29) https://www.gradiant.org/en/blog/infinitech-data-anonymization-big-data-gdpr/

Bart Coppens is a part-time assistant professor and a post-doctoral researcher in the Electronics department of Ghent University, Ghent, Belgium.

Olivier Zendra is a Tenured Computer Science Researcher at Inria, Rennes, France.