Privacy Is Hard and Seven Other Myths

Jaap-Henk Hoepman*

As digital technologies permeate further and further into the fabric of our society, privacy is increasingly threatened. This privacy problem is very well studied and also ever more widely discussed within society at large. Much less public attention is paid to privacy solutions, even though privacy by design and privacy enhancing technologies have been studied in academic circles for decades. A clear explanation for this unfortunate imbalance escapes me: perhaps discussing problems attracts more media attention than describing possible solutions. Perhaps the privacy paradox is real: people may say that a lack of privacy is problematic, but when push comes to shove achieving other goals in life is fundamentally more important. Or perhaps discussing solutions inherently involves a more technical perspective that few people feel comfortable to engage with.

Or perhaps this revering stance towards technology as something holy, that cannot be fully understood anyway, prevents us from even considering the possibility that the privacy paradox can actually be resolved. Let us go through some of the common and persistent myths that surround privacy and let us deconstruct them to show that privacy is not hard, but can actually be achieved through careful design.

Myth #1: We Are Not Collecting Personal Data

It was, and perhaps sometimes still is, common for people to believe that data is only personal if it is immediately obvious who it belongs to. As a result, organisations believed that opaque identifiers like IP addresses were not personal data, and claimed that they were not collecting personal data. In Europe at least, this is not the case. The General Data Protection Regulation (GDPR)1 (similar to earlier data protection laws) defines personal data as:

'any information relating to an identified or identifiable natural person (‘data subject’); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.'

DOI: 10.21552/edpl/2023/2/5

* Jaap-Henk Hoepman, Associated Professor, Radboud University / University of Groningen. Guest Professor, Karlstad University. This editorial is a summary of J.-H. Hoepman, Privacy Is Hard and Seven Other Myths. Achieving Privacy through Careful Design, (MIT Press, 2021). For Correspondence: <jhh@cs.ru.nl>.

1 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016, ‘On the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation),’ Official Journal of the European Union L 119 (May 4, 2016): 1–88; Regulation (EU) 2016/679 (GDPR).
As a result IP addresses and car license plates are personal data: IP addresses are often static, and your ISP keeps a record of which IP address belongs to whom; similarly the owner of the car with a particular license plate is recorded in a central car registry.

Data that allows ‘singling-out’ an individual, i.e. being able to recognise an individual again based on a unique characteristic or numeric label without being able to actually fully identify this person by name, is typically also considered personal data. Cookies and static MAC addresses (the hardware addresses associated with Bluetooth or WiFi enabled devices) are common examples of such personal data.

One strategy to make data less personal therefore is to break, or weaken, the link between a data item and the person it belongs to. Using *pseudonyms*, especially ones that change over time or differ across contexts such that your pseudonym in a health care database is completely unrelated to your pseudonym used in the educational sector, is a common technique to achieve this goal. Be aware that pseudonyms do not render data fully non-personal. When properly constructed they do help to ensure that personal data from different contexts cannot be recombined to a rich personal profile.

**Myth #2: You Have Zero Privacy Anyway — Get Over It**

Data protection is only the most recent lens through which we try to understand and regulate privacy. Technological developments have always had an impact on our privacy. In the late nineteenth century for example cheaper, easier to operate cameras combined with improvements in printing technology allowed newspapers containing pictures to be more widely circulated. Fearing that ‘what is whispered in the closet shall be proclaimed from the house-tops’, Warren and Brandeis formulated the right to be let alone as a response.

The invention of computers and digital networks and their proliferation in all aspects of our lives undeniably has changed the privacy landscape beyond recognition. Archiving information used to be hard. Digital documents on the other hand are easily copied, and it is hard to ensure that all copies are truly deleted. Computers make searching for information a breeze. Networks allow information from different sources to be combined, and to be retrieved anywhere. Social networks and the ‘web 2.0’ allow users to generate content as well. Reflecting on these developments, Scott McNealy, then CEO of Sun Microsystems, famously quipped in 1999 that ‘You have zero privacy anyway — Get over it.’

---

2 This is where the GDPR differs from earlier legislation, offering stronger protection than before. Recital 26 of the GDPR specifically mentions singling out individuals, as opposed to the corresponding recital 26 of the previous Data Protection Directive; see Directive 95/46/EC of the European Parliament and of the Council of October 24, 1995, ‘On the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data,’ Official Journal of the European Communities L 281 (November 23, 1995): 31–50. But the Article 29 Working Party did consider such identifiers to fall under the scope of the definition of personal data under the old directive even back in 2007; see Article 29 Working Party, ‘Opinion 4/2007 on the Concept of Personal Data,’ June 20, 2007. See also Purtova, ‘The Law of Everything’; and F. J. Zuiderven Borgesius, ‘Singling Out People without Knowing Their Names—Behavioural Targeting, Pseudonymous Data, and the New Data Protection Regulation,’ Computer Law & Security Review 32, no. 2 (2016): 256–271.

3 S. D. Warren and L. D. Brandeis, ‘The Right to Privacy: The Implicit Made Explicit,’ Harvard Law Review 4, no. 5 (December 15, 1890): 193–220.

4 P. Srenger, ‘Sun on Privacy: “Get Over It,”’ Wired, January 26, 1999.
Such a distortion of how technology develops is common among Silicon Valley tech entrepreneurs, in an attempt to hide their own role in the direction of those technological developments. But technology never develops in a vacuum, as a force of nature: it is constructed by people in response to perceived needs and according to their own norms and values. It embeds these beliefs in how it functions, what it affords us to do, what it prevents us from doing, and what it does regardless of our own intents and wishes. At the moment, information technology is used to invade our privacy. But this is not necessarily so: technology can also be used to protect privacy, through a process called privacy by design. No need to ‘get over it’.

Privacy by design is an engineering approach popularised by Ann Cavoukian in the 1990s. The essential idea is that privacy should be considered as a design requirement from the very beginning and then throughout the life-cycle of a system. This is necessary for two reasons. Early design decisions have a strong privacy impact that cannot easily be changed later on in the design process. Moreover, by considering privacy together with all other requirements from the outset, designers will be forced to think of how to meet all other requirements in a privacy-friendly way.

The power of privacy by design thinking is best illustrated by an example. Suppose you want to leave your coat at a cloakroom when visiting the theatre. You’d probably be unpleasantly surprised if the assistant would ask for your name and record that while taking your coat, and checking your name when you retrieve it after the show. If a cloakroom would work this way, it would be quite privacy invasive, compiling a record of who visited the theatre, and when. (It would also be much less efficient.) Luckily cloakrooms have applied privacy by design (although probably for efficiency reasons, and not because they care about privacy): when taking your coat, the assistant hands you a numbered token in return. With that token, you can reclaim your coat after the show. This systems records no personal information whatsoever.

This is of course a trivial example, but it does show the kind of tilted perspective needed to properly design systems in a privacy friendly way. Many technologies and approaches exist to support such privacy friendly designs, as will be illustrated by busting a few more privacy myths.

**Myth #3: I’ve Got Nothing to Hide**

Nothing is more pervasive than the ‘If you’ve got nothing to hide, then what do you have to fear?’ myth. This myth has been dispelled many times, so we won’t go into

---

5 A. Cavoukian, Privacy by Design: The 7 Foundational Principles, report, revised version (Ontario: Information and Privacy Commissioner of Ontario, January 2011).
6 Taken from R. J. Cronk, Strategic Privacy by Design (Portsmouth, NH: IAPP, 2018).
7 I would like to express my deepest gratitude to the lovely people of TILT (The Tilburg Institute for Law, Technology, and Society), and especially Bert-Jaap Koops, for hosting me and allowing me to write the book this editorial is an excerpt of.
8 See e.g. D. J. Solove, ‘I’ve Got Nothing to Hide’ and Other Misunderstandings of Privacy,’ San Diego Law Review, no. 44 (2007): 745.
the counterarguments in detail here. Suffice it to say that hiding is natural, and is part of being human. In fact, it is impossible, unavoidable even, to hide our thoughts and feelings. Moreover, the myth assumes that the question what should or should not be hidden is easily answered: either ‘you did something you shouldn’t have done in the first place’, 9 or there is really nothing to worry about. In practice, this is not clear at all. What is wrong or not, depends a lot on context and interpretation: is something you did illegal, or morally wrong? In which jurisdiction, or which cultural context? Is being gay wrong? Smoking pot?

What is important in the context of privacy by design is that there are many technologies out there that can help hiding and protecting personal data. Symmetric and public key encryption, and their application to secure communications (virtual private networks, secure browsing, and end-to-end encryption) and to securely store data (in the cloud and on your local devices), are the prime example. 10 In the context of databases, techniques like statistical disclosure control 11 and differential privacy 12 exist to make them more privacy friendly.

**Myth #4: It’s Merely Metadata**

The techniques described in the previous section are effective to protect the *content* of the messages we send or the *data* that we store, but are insufficient to protect the *metadata* necessary to establish the connection between the sender and the recipient (e.g., an address, or a phone number) or to store the data (the file name, for example). Unlike its metadata, the data itself is often considered private and offered stronger protection: the secrecy of correspondence is enshrined in the constitution of many countries. But shouldn’t metadata be given similar protections? Is it really ‘merely’ metadata?

Perhaps it should even be considered to be more sensitive than the actual data it pertains to. First because it is collected *surreptitiously*, without people being aware (and being able to control it). Second because it is more *structured* and therefore more easily analysed. And third because metadata really is *behavioural data* and as such is telling everything about our behaviour: where we are, what websites we visit, who we are communicating with and when. And as such reveals a lot more about our deepest desires and needs than we explicitly divulge when talking our writing.

Given the ‘leaky’ design of our digital devices and networks, limiting the collection of metadata is somewhat challenging. For devices like smartphones and smart watches,
it is important to review their default settings and severely limit the permissions we give to the apps we use: block cookies and trackers, do not share your location, block access to your contacts, etc. Manufacturers and service providers should make these settings the default.

To properly prevent metadata collection on networks involves a much more fundamental overhaul of how the Internet works: it should for example incorporate mixing techniques first tried out by Tor (The Onion Router)\textsuperscript{13} and now popularised through services like Apple Relay. This would make sender anonymity the default for the whole internet. With that, many web services suddenly become much less privacy invasive, and new privacy friendly services become much easier to deploy.

**Myth #5: We Always Need to Know Who You Are**

Many of the online services we use require us to create an account and provide some basic information like a name, address and phone number. This even happens for simple, one-shot, services like booking a table at a restaurant or when buying a concert ticket. Earlier we explained how a cloakroom could be made more privacy friendly. Similarly, buying concert tickets online should be possible without creating an account by delivering the ticket directly as a pdf in the browser or storing it in our smartphone wallet.

This unfortunate state of affairs can perhaps be explained by looking at how traditionally access to resources (mostly computers) and later services in general was controlled by first needing to prove you are the owner of an account (using some form of authentication to log in, for example). After this, the authorisations linked to that account determine whether you are granted access or not.\textsuperscript{14}

A particular technology aimed at making access control decisions much more privacy friendly are attribute based credentials. The attributes describe a certain property of an individual, like their age, their qualifications (education, skills), their sex, their nationality, their name, their address, etc. The essence of the idea is that in order to prove a certain attribute, and to authorise access, it is no longer necessary to first prove your identity. Attributes are securely stored in attribute based credentials. Credentials are issued by an issuer, that vouches for the fact that the attributes it contains belongs to the holder.\textsuperscript{15}

Users obtain credentials with attributes that are relevant to them from the necessary issuers once, beforehand. In that sense they function much like traditional paper based

\textsuperscript{13} R. Dingledine, N. Mathewson, and P. F. Syperson, ‘Tor: The Second-Generation Onion Router,’ in 13th USENIX Security Symposium, ed. M. Blaze (USENIX Association, 2004), 303–320. See also <https://www.torproject.org> accessed 1 July 2023.

\textsuperscript{14} G. Alpár, J.-H. Hoepman, and J. Stilje, ‘The Identity Crisis: Security, Privacy and Usability Issues in Identity Management,’ Journal of Information System Security 9, no. 1 (2013): 23–33.

\textsuperscript{15} D. Chaum, ‘Security without Identification: Transaction Systems to Make Big Brother Obsolete,’ Communications of the ACM 28, no. 10 (1985): 1030–1044.
credentials like diplomas or drivers licenses. After that subset of the attributes in a credential can selectively disclosed to a particular service provider in an attempt to access the service based on the value of these attributes. For example, a person wishing to access age restricted content from the national broadcasting corporation could use attribute based credentials to prove that they are over eighteen and English, without revealing their name at all. It is important to mention that the actual privacy protection offered by attribute based credentials is much stronger than implied here. In fact an attribute based credential is unlinkable, meaning that subsequent uses of the same credential cannot be distinguished from uses of different credentials that happen to contain the same values for the disclosed attributes. Practical implementations of attribute based credentials are now available.16

Myth #6: Your Data is Safe With Us

A recent trend for many big platforms like Google, Facebook and Apple is to frame privacy as security: ‘trust us, your data is safe with us’. And at some level this makes sense: individuals and smaller organisations may not have the tools or resources to properly protect their own personal data. Also, for such companies our personal data (and the profiles and insights they derive from that) form their most priced asset. But at a more fundamental level this view is hugely problematic, as privacy does not mean that Google or Facebook keep our data private. Privacy means that we ourselves are in control. Adding insult to injury is the fact that companies like Google and Facebook are actively subverting our abilities and efforts to do so.

In fact, decentralised or fully distributed, peer-to-peer, designs offer a strong form of privacy protection almost as a matter of principle: instead of processing the data on a central server, the data is processed locally on the device of the user him or herself. Many common services are amenable to such a design. This is the case for fitness apps for example, that track your daily fitness performance (as long as you do not want to share your runs with others, and even that could be implemented in a fully peer-to-peer fashion), or menstruation apps that track your menstrual cycle. Even location based services, for example map-like apps that show you restaurants or other points of interest in your area, can be implemented in a much more privacy friendly fashion as follows: your phone only shares your coarse location with the service provider (e.g. your location rounded to an area of a few square kilometres in size), and refines the results returned by the service provider based on your exact location and interests.

A word of warning is in order here, however: decentralisation is not a silver bullet. All approaches that propose to store personal data locally throw all responsibility for their privacy back at the feet of the users. This assumes that people always are willing and able to make the right decisions. And this is not always the case. There is also a more fundamental limit to the protection offered by processing data locally. By moving in-

16 See e.g. <https://privacybydesign.foundation> accessed 1 July 2023.
telligence to our own devices through approaches like federated learning, this line of defence breaks. Although there is no central place where information about our personal details resides, the system as whole (i.e., the system comprised of all our own devices as well) still knows about us, and can predict, judge, and nudge us.

**Myth #7: Privacy and Security Are a Zero-Sum Game**

Security and privacy are often seen as opposite, irreconcilable goals; as a zero-sum game. Because the stakes involved are high, the debate is often heated and polarised. This is an unfortunate situation, both for our privacy and our security, and for society at large:

'It is highly unlikely that either extreme — total surveillance or total privacy — is good for our society.'

One way out of this war of trenches is to consider approaches that satisfy both security as well as privacy requirements at the same time. Contrary to commonly belief this is certainly possible.

Consider the problem of digital cash. Traditional cash, and especially coins, are untraceable: looking in someone’s wallet will not allow you to tell where they got their money from, nor can the supermarket tell at the end of the day who its customers were and what each of them bought. Digital payment schemes on the other hand are typically traceable because they are account based: credit card numbers, for example, allow a payment to be traced to the card holder. This traceability in essence ensures that any money cannot be spent twice: once it leaves the account it is no longer under control of the user.

A natural idea to avoid account based schemes is to mimic traditional cash using digital equivalents. Such digital coins need to satisfy three properties: they need to be unforgeable, untraceable, and double spending needs to be prevented: unlike physical coins, digital coins are trivial to duplicate. A natural way to make coins unforgeable is to let the issuer sign them using a digital signature. Such signatures also make the coin unique, so the bank can check for double spending. Unfortunately this also means coins are traceable. This apparent paradox was resolved by David Chaum through his invention of blind signatures. Such signatures allow a bank to sign a coin, without knowing (and later recognising) which coin it actually signed. This way, issued coins cannot be linked to coins that are spent later, thus guaranteeing untraceability.

---

17 M. Smith and M. Green, ‘A Discussion of Surveillance Backdoors: Effectiveness, Collateral Damage and Ethics,’ February 5, 2016.

18 J.-H. Hoepman, ‘Revocable Privacy,’ ENISA Quarterly Review 5, no. 2 (June 2009): 16–17; W. Luks, M. Everts, and J.-H. Hoepman, ‘Revocable Privacy: Principles, Use Cases, and Technologies,’ in Annual Privacy Forum (APF 2015) (Berlin: Springer, 2016), 124–143.

19 D. Chaum, ‘Blind Signatures for Untraceable Payments,’ in Advances in Cryptology—CRYPTO, ed. D. Chaum, R. L. Rivest, and A. T. Sherman (New York: Plenum Press, 1982), 199–203. D. Chaum, A. Fiat, and M. Naor, ‘Untraceable Electronic Cash,’ in Advances in Cryptology—CRYPTO ’88, 8th Annual International Cryptology Conference, ed. S. Goldwasser (Berlin: Springer, 1988), 319–327.
Myth #8: Privacy is Hard

A common myth is that ‘privacy is hard’. Indeed, designing totally ‘private’ systems is next to impossible even under ideal circumstances. (The same is true for designing 100% secure systems by the way.) But let perfect not be the enemy of good. A little bit of effort and consideration can actually prevent a lot of privacy harm. In fact, just as technology can be used to invade our privacy, it can also be used to protect our privacy by applying privacy by design. Existing privacy-friendly technologies and privacy by design approaches can be used to create privacy friendly alternatives to the systems we commonly use today. The previous sections have offered many examples of this.

Significantly improving the privacy of the apps and services we use should be our first priority. This may require some effort, and may squash some extractive business models, but it is not at all as hard as many make us believe. An important tool to support privacy by design in the very early stages of such system development, during the ideation and definition phases, are privacy design strategies.20 These help the designers to translate fuzzy privacy norms into concrete engineering decisions.

But this is only the first step, that focuses on the short term. The next step is much more fundamental, but one that is sorely needed to guarantee proper protection of privacy and other human rights and societal values in the long run. This step requires us to dig deeper down into the technology stack and look beyond the products and services we use, and reconsider the designs for the underlying computers and networks, both at the hardware and the operating system levels. These designs are half a century old by now and never fundamentally changed, while the world in which they are used has changed beyond recognition. We are stretching the boundaries of their use beyond the breaking point—not only in terms of privacy, by the way, but also in terms of security and reliability. It’s time to start redoing the plumbing, instead of applying Band-Aids to temporarily stop some leakage while we frantically mop the floor against all odds.

20 J.-H. Hoepman, ‘Privacy Design Strategies,’ in IFIP TC11 International Information Security Conference 2014, ed. N. Cuppens-Boulahia et al. (Berlin: Springer, 2014); M. Colesky, J.-H. Hoepman, and C. Hilken, ‘A Critical Analysis of Privacy Design Strategies,’ in 2016 International Workshop on Privacy Engineering—IWPE’16 (IEEE Computer Society, 2016), 33–40; J.-H. Hoepman, Privacy Design Strategies: The Little Blue Book (May 2018).