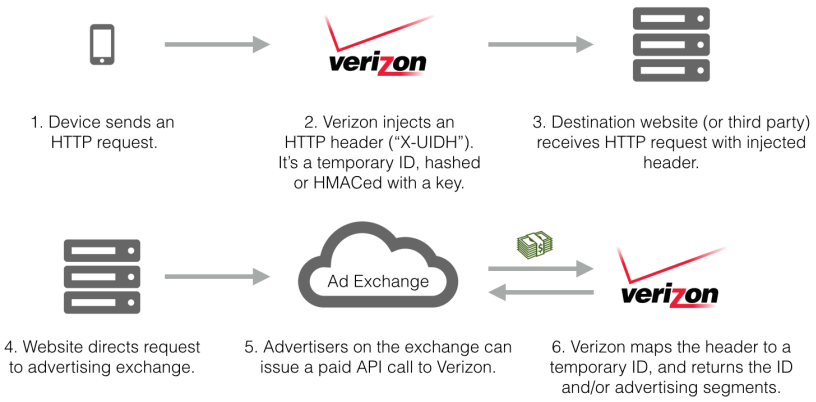
Handout 7 (Privacy)

The first motor car was invented around 1886. For ten years, until 1896, the law in the UK and elsewhere required a person to walk in front of any moving car waving a red flag. Cars were such a novelty that most people did not know what to make of them. The person with the red flag was intended to warn the public, for example horse owners, about the impending novelty—a car. In my humble opinion, we are at the same stage of development with privacy. Nobody really knows what it is about or what it is good for. All seems very hazy. The result is that the world of "privacy" looks a little bit like the old Wild West. For example, UCAS, a charity set up to help students apply to universities, has a commercial unit that happily sells your email addresses to anybody who forks out enough money in order to bombard you with spam. Yes, you can opt out very often, but in case of UCAS any opt-out will limit also legit emails you might actually be interested in.[1]

Verizon, an ISP who provides you with connectivity, has found a "nice" side-business too: When you have enabled all privacy guards in your browser, the few you have at your disposal, Verizon happily adds a kind of cookie to your HTTP-requests.[2] As shown in the picture below, this cookie will be sent to every web-site you visit. The web-sites then can forward the cookie to advertisers who in turn pay Verizon to tell them everything they want to know about the person who just made this request, that is you.



How disgusting? Even worse, Verizon is not known for being the cheapest ISP on the planet (completely the contrary), and also not known for providing the

[1] The main objectionable point, in my opinion, is that the *charity* everybody has to use for HE applications has actually very honourable goals (e.g. assist applicants in gaining access to universities), but in their small print (or better under the link "About us") reveals they set up their organisation so that they can also shamelessly sell email addresses the "harvest". Everything is of course very legal…moral?…well that is in the eye of the beholder. See:
http://www.ucas.com/about-us/inside-ucas/advertising-opportunities or http://www.theguardian.com/uk-news/2014/mar/12/ucas-sells-marketing-access-student-data-advertisers

[2] http://webpolicy.org/2014/10/24/how-verizons-advertising-header-works/

fastest possible speeds, but rather for being among the few ISPs in the US with a quasi-monopolistic "market distribution". Well, we could go on and on…and that has not even started us yet with all the naughty things NSA & Friends are up to.

Why does privacy matter? Nobody, I think, has a conclusive answer to this question. Maybe the following four notions clarify the picture somewhat:

- **Secrecy** is the mechanism used to limit the number of principals with access to information (e.g., cryptography or access controls). For example I better keep my password secret, otherwise people from the wrong side of the law might impersonate me.
- **Confidentiality** is the obligation to protect the secrets of other people or organisations (secrecy for the benefit of an organisation). For example as a staff member at King's I have access to data, even private data, I am allowed to use in my work but not allowed to disclose to anyone else.
- **Anonymity** is the ability to leave no evidence of an activity (e.g., sharing a secret). This is not equal with privacy—anonymity is required in many circumstances, for example for whistle-blowers, voting, exam marking and so on.
- **Privacy** is the ability or right to protect your personal secrets (secrecy for the benefit of an individual). For example, in a job interview, I might not like to disclose that I am pregnant, if I were a woman, or that I am a father. Similarly, I might not like to disclose my location data, because thieves might break into my house if they know I am away at work. Privacy is essentially everything which 'shouldn't be anybodies business'.

While this might provide us with some rough definitions, the problem with privacy is that it is an extremely fine line what should stay private and what should not. For example, since I am working in academia, I am very happy to be essentially a digital exhibitionist: I am happy to disclose all 'trivia' related to my work on my personal web-page. This is a kind of bragging that is normal in academia (at least in the CS field). I am even happy that Google maintains a profile about all of my academic papers and their citations.

On the other hand I would be very peeved if anybody had a too close look on my private live—it shouldn't be anybodies business. The reason is that knowledge about my private life usually is used against me. As mentioned above, public location data might mean I get robbed. If supermarkets build a profile of my shopping habits, they will use it to *their* advantage—surely not to *my* advantage. Also whatever might be collected about my life will always be an incomplete, or even misleading, picture—I am sure my creditworthiness score was temporarily(?) destroyed by not having a regular income in this country (before coming to King's I worked in Munich). To correct such incomplete or flawed data there is, since recently, a law that allows you to check what information is held about you for determining your creditworthiness. But this concerns only a very small part of the data that is held about me/you.

This is an endless field. I let you ponder about the two statements that are often float about in discussions about privacy:

- *"You have zero privacy anyway. Get over it."*

  Scott Mcnealy (CEO of Sun)

- *"If you have nothing to hide, you have nothing to fear."*

There are some technical problems that are easier to discuss and that often have privacy implications. The problem I want to focus on is how to safely disclose datasets. What can go wrong with this can be illustrated with three examples:

- In 2006 a then young company called Netflix offered a 1 Mio $ prize to anybody who could improve their movie rating algorithm. For this they disclosed a dataset containing 10% of all Netflix users (appr. 500K). They removed names, but included numerical ratings as well as times of ratings. Though some information was perturbed (i.e., slightly modified).

  Two researchers took that data and compared it with public data available from the International Movie Database (IMDb). They found that 98 % of the entries could be re-identified: either by their ratings or by the dates the ratings were uploaded.

- In the 1990, medical databases were routinely made publicised for research purposes. This was done in anonymised form with names removed, but birth dates, gender, ZIP-code were retained.