Vic Grout, Professor of Computing Futures, Glyndŵr University discusses

IDENTITY **VOYEURISM**

There's more than one type of identity crisis. Conventional identity theft is one thing but what of identity voyeurism? How much of us is in the shop window anyway? Are we in control? What are the risks? And where's it heading?

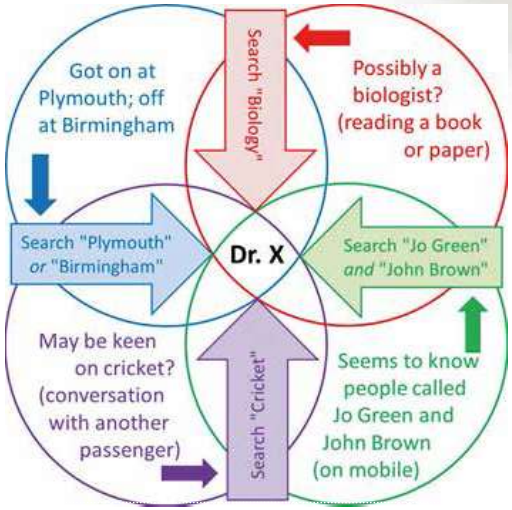

Figure 1: Playing the 'Prof on a Train' game

The next time you're on public transport, try playing the 'Prof on a Train' (PoaT) game. (It doesn't really have to be a train or an academic but it's a good example to work with.) Take a look at the person opposite you. Armed only with your senses, intelligence, intuition and an internet connection, how much can you find out about them?

Well, if they're quietly dozing in the corner, unremarkably dressed, with no distinguishing features whatsoever, you'll probably lose. However, any activity at all or any visible clues might give you a chance. Are they doing, reading, saying or wearing anything? Who's with them? Are they easier to identify? Where did they get on and do you know where they're going? Anything odd? Here's the basic strategy, on which PoaT is based:

They're reading an academic paper on a certain subject (X) and you know where they got on (Y). A quick look at the photos on the 'Department of X' webpage for the 'University of Y' might be enough.

Academics aren't the only people who proclaim their achievements so publicly.

> Data-mining is easy; unique identification is the challenge. (Unusual names help a little.)

Something as simple as a business card in clear sight gives an even easier win. Once, you've named them, Google takes care of the rest. Most information is free but a small amount (county court judgements [1], for example) might be behind a paywall. Data-mining is easy; unique identification is the challenge. (Unusual names help a little.)

It won't always be that straightforward, of course: not everyone has a dedicated webpage...yet and, even then, it won't always succeed. This is where you and the internet have to work together. A cleverly-constructed search based on several key-words might just do the trick. Each extra feature narrows the search. You couldn't process a query so complex but a search engine can. Figure 1 gives the general idea.

Now, there are two fairly obvious things about PoaT:

- It's of dubious ethicality (yes, we were going to say that eventually)
- It's going to get easier to play as time goes by - as more individual information gets put online and search algorithms become more sophisticated.

So how might PoaT be 'played' in a few years' time? More precisely, how could we make sure we win?

Well, as with most such processes, the idea would be to automate it as far as possible. For the determined 'identity voyeur', this would mean:

a) Automatic recognition and identification through hardware [enhancing our senses],
b) Improved searching through semantic software [enhancing our intelligence], and
c) Integrating (a) and (b) to work, not as discrete concepts but, as one – through iteration/feedback [enhancing our intuition].

Possible tools for automatic recognition (a) include: face-recognition; voice-analysis; fingerprints; gait-analysis; body size/shape-analysis; age-estimation; breath-analysis; DNA; biometrics; odour; clothes/uniform/

Images: Thinkstock
doi:10.1093/itnow/bwv067

'Considered separately, some pieces of data from various sources might be innocuous-seeming data or even de-identified datasets, i.e. free from identifiable information. Data analysts may therefore tend to consider them as not covered by existing data protection laws. However, analysts can, intentionally or accidentally, identify new inferences or discover new sets of sensitive information the data subject has not agreed to share. Analysts can do this by correlating supposedly de-identified data sets with publicly or other privately available data sets.'

Hervais Simo Fhom
Big Data: Opportunities and Privacy Challenge

Figure 2: Big data analytics quotation

badges; the car/bike they drive/ride; any identifiable technology on/in their person; location/occupation/known habits; and association with colleagues/friends/family. As far as possible, the ideal would be to build much of this technology into a single device.

Each of these techniques is, at worst, a very real area of research in some stage of development – none are science fiction. Individually, they might be unavailable, inappropriate, unsuccessful or insufficient to produce a 'perfect match'.

However, together, they would combine to form a 'personal fingerprint' of the individual in different settings, in much the same way that identification services, such as Shazam [2], produce acoustic fingerprints [3] for music under different conditions.

The necessary improvements in semantics and data-mining (b) are ongoing. They essentially involve the usual Holy Grail of intelligent searching, answering the question that wasn't asked. 'He's on an early-morning train to London, on the day of a test match, wearing a red and yellow tie: scan MCC comments on Facebook posts/check-ins?'

But it's important to realise that an automatic system for PoaT, doesn't have to wait for a future when we have (a) and (b) in place. What we have now is good enough to start with.

An 'Auto-PoaT' or 'Shazam for People' [4] service or application will simply improve its accuracy as the hardware and software technologies in (a) and (b) develop.

A Shazam for People (SfP) app doesn't yet have the simple, indexed online look-up for individuals that conventional Shazam has for music.

Now, (c) is the interesting one because it's both:

- The point at which (a) and (b) are integrated into a complete, usable system, and
- The only part of the process where legislation may be effective if we want to stop this happening.

Because the process that's required here is:

a) The hardware at your disposal captures what distinguishing information it can and combines it with intelligent (probably local) software to produce a personal fingerprint
b) This personal fingerprint is matched against an online database to identify the individual, then
c) As much information as is available can be returned to the 'user' of the 'system'.

C) is trivial. A) is developing. The component that's missing, at the moment, is the database in b) – or possibly, its primary key, because the internet may already be the database.

A Shazam for People (SfP) app doesn't yet have the simple, indexed online look-up for individuals that conventional Shazam has for music. Indeed, this is where legislation may have some chance of preventing it. But is such a perfectly coherent organisational structure really necessary? And could it be effectively prevented?

Collecting information in piecemeal form and making it available (mostly) isn't illegal: the internet has been doing it for years; and searching for it obviously isn't (on the whole). So can the reduction of an individual person to a single, global identifier – with its associated data repository, necessary for SfP, be either achieved...or stopped?

Well, an email address, just as one example, is already a unique global identifier – used by many existing systems as ID. (No, not everyone currently has anything like an email address but project this whole discussion into the future.)

Also, for some of us, information networks are springing up everywhere: in some cases, by our own hand (LinkedIn, Facebook, etc.); in others, independently (Google Scholar, for example). It's not difficult to imagine these, and newer – more general, systems both expanding to cover larger numbers of people over the next few years and widening their spheres of interest as internet of things technology makes more information available.

It's debatable whether these individual sources will ever get merged into a single 'information super-page' for everyone but the SfP software probably doesn't need this – or an index to it: it will know where to look anyway.

Unfortunately, this harvesting and storing of information centrally, or centralised indexing to multiple sources, may have been the only points at which data-protection legislation could have made any of this process illegal. Enforceably illegal, that is. (Figure 2)

But, without some form of protection, what might the future hold?

Internet-based 'big-data' type analysis [5] is already dredging up material many had hoped was buried in the past [6]. What could happen in real-time when we combine accurate individual recognition, huge personal data-banks, fully integrated searches, opportunistic capitalism and anytime/anywhere delivery of results? (Figure 3)

One thing's for sure. Now may not be a good time to be hiding new skeletons in the cupboard; the old ones may be hard enough to keep in there!

Consider a future scenario, not too many years distant.

You meet someone in the street, they introduce themselves as 'John Green'. As they're speaking, the headset or smartwatch you're wearing (maybe even an implant) scans their features and analyses their voice and chemistry. It matches this against a global database and reports back to you:

'No, this isn't John Green. This is Paul White. He's 45 years old and lives in Sheffield; married with two children (he's not the father of one, although he thinks he is). He was arrested in 1996 for shoplifting and declared bankrupt in 2005. He works as a landscape gardener but his attendance record isn't very good. He smokes and has a chronic lung condition, which is making it difficult for him to get insurance. He voted Liberal Democrat at the last two elections. His favourite type of porn is...'

Figure 3: The ultimate 'identity voyeur'?

Vic Grout CEng CSci FBCS CITP is Professor of Computing Futures at Glyndŵr University, Wrexham, serving on the UK Committee of the Council of Professors and Heads of Computing as Chair of the Council of Heads of Computing in Wales. He has worked in senior positions in academia and industry for many years and has published over 300 research papers, patents and books on various applications of internet technologies - with a particular focus on the social and ethical dimensions of increasing connectivity.

www.bcs.org/security

References

[1] Trust Online
www.trustonline.org.uk/
[Accessed: 20/06/2015]

[2] Shazam
www.shazam.com/
[Accessed: 20/06/2015]

[3] Wikipedia, 'Acoustic Fingerprint',
https://en.wikipedia.org/wiki/Acoustic_fingerprint
[Accessed: 20/06/2015]

[4] Vic Grout, 'Shazam for People', 'Turing's Radiator',
http://vicgrout.net/2014/08/20/shazam-for-people/
[Accessed: 20/06/2015]

[5] Hervais Simo Fhom, 'Big Data: Opportunities and Privacy Challenge', in: Richter (ed.), Privatheit, Öffentlichkeit und demokratische Willensbildung in Zeiten von Big Data, Nomos, Baden-Baden 2015, pp13-44.

[6] The Moscow Times, 'Accusations of Plagiarism Become a Political Weapon',
www.themoscowtimes.com/news/article/accusations-of-plagiarism-become-a-political-weapon/476800.html
[Accessed: 20/06/2015]