“3 is a crowd” : a proposal to reintroduce effective privacy on-line

John Taysom : Harvard University, June 2012

Executive Summary : Privacy on-line is the subject of legislation in EU and expected legislation in US. This legislation seeks to encourage self-regulation while providing punishment for infringers. The major protection offered to prevent privacy invasion before it happens is for users to chose to ‘opt out’ of data collection, usually site-visit by site-visit, which usually also means accepting an inferior service because the technical mechanisms to tailor the user experience rely on being ‘opted in’. Users are faced with the choice of a better service and more information or entertainment now, traded against potential uncertain costs from individual or collective privacy infringement at some uncertain future date. Opting-in on almost always on the basis of arcane self-imposed rules . But technology now permits a ‘privacy protected opt-in’, and that is the focus of this paper. How can a privacy protected opt-in be provided for users (who can still opt out if they wish, but will have less reason to do so) without harming the growing internet eco-system which has become so important for jobs and for wealth creation in many countries.

In the past most work on privacy has focussed on individual harms. These remain real, but the focus in this paper is on the collective harm to society when ever larger parts of the economy are dependent in part on the internet. In economies where healthcare and education, to take two cler examples, are provided by the State, moving them in part on-line will result in the State holding significant information about individuals. That has never been made clear by Governments.

To start with a statement of the problem: in the last decade we have allowed technology to erode our privacy both on-line and off-line. Like the fabled frog in warm water the heat has slowly increased so that we hardly noticed. When we did notice we were told that there had been a shift in values: we would just have to get over it[1]. Meanwhile, a generation has grown up not knowing what is it like to have only an ‘off-line’ world or indeed an off-line mode, except as a punishment. In fact there is good evidence to show that younger users develop their personality in part ‘on-line’ (Turkle; 2011). That alone should make us think again about privacy. As technology changes maybe this will not be off the agenda until we can implement a safe mechanism for opted in users which is flexible to technology change.

---

[1] Facebook’s Mark Zukerberg, Linkedin’s Reid Hoffman, Sun Microsystems’s Scott McNealy, have all said similar things publicly.

3 The focus in US has been on tightening self-regulation following several high profile cases which have resulted in th FTC imposing 20 year audit requirements on Google and others. In EU country level legislation has been enacted in UK and Germany and several other countries following an EU Directive. The emphasis in EU is on what is done with the data, and in US on how the data is collected. The US has long had sepcial rules for healthcare data (HIPPA) and data about children under 14 (COPPA). The balance has changed in favour of more control over data that might infringe individual privacy.

"3 is a crowd" : a proposal to reintroduce effective privacy on-line

Self-regulation does not appear to be working judging by the number of privacy breeches that are being reported[2]. Legislation to protect privacy has been slow in US. In EU although a Privacy Directive has been passed it has been largely ignored at the country level. In US, the FTC is reported to have privacy at the top of its enforcement agenda (WSJ, front-page Friday Nov 21st). But this will be effective only after the fact. And yet in developed and less developed economies Governments are putting more weight onto on-line education and on-line healthcare as a way to reduce spending and increase effective service delivery. There is therefore a pressing need for a new way to tackle this issue. This project proposes a new approach which follows from two simple observations:

> Our similarities make us economically valuable: our differences make us individuals, and over that individuality we have a right to privacy explicitly under the EU Directives and implicitly under the US Constitution. Very few organizations can defend the collecting and storing of personal data (PII)[3] if they can adequately conduct their business without it.

> Private companies do not have long lives as compared to the average life expectancy of 100 years for a new-born in a developed economy. Private companies can be bought and sold. They are probably not the right place to store personal data. A different kind of organization would be preferable. Governments should rarely store personal data. Where it is required, Government use of personal data should be mediated by the Courts.[4] This project therefore proposes an endowed not-for profit structure[5] to ensure robustness and independence of governance from Goverment or private for-profit influence.[6]

---

[2] Apple, Google and Facebook, have all had incidents exposed in 2011. Google has settled and Facebook is negotiating a settlement with FTC. Facebook has responded with a re-vamped privacy and sharing mechanism. Mobile phone tracking and motor vehicle tracking have generated court cases.

[3] Companies like BlueKai state that they do not collect PII and make explicit what data they do collect. http://www.bluekai.com/consumers_privacyguidelines.php

[4] This is not the case now: WSJ Nov 9 2011: "State and federal authorities follow the movements of thousands of Americans each year by secretly monitoring the location of their cellphones, often with little judicial oversight..."

[5] This approach has a venerable history: the Church of England in UK as an endowed non-Governmental, non private sector body was championed by John Stuart Mill[3] in 1834. (Reeves; 2007).

[6] There are many examples; ICANN; the WEF; The Red Cross; Underwriters Laboratories.

“3 is a crowd” : a proposal to reintroduce effective privacy on-line

To borrow from the past, when needed we have developed systems to protect privacy but which still allowed commerce to flourish. Zip codes, as an example, cluster individuals into groups which reveal a lot about the group members collectively, but which don’t allow identification of the individuals within. Just as for zip codes, clusters of similarities (other than simply location) can act as proxies for us in revealing enough information to be actionable for many purposes, but not so much as to allow identification of individuals within the crowd. Yet private companies routinely collect data that can identify individuals. They then aggregate the data to generate economically interesting clusters. In other words, what they wanted was the economic value: but they generated a by-product, personal information which could alone, or in combination, deliberately or inadvertently, invade personal privacy. This does not need to be the case. (Taysom, Cleevely; 2007). Away from the private sector, under the banner of ‘Big Data’, Governments in US and UK have been persuaded to ‘open’ their data to the public to allow the development of new applications. [7] This is data about us but it is not ‘our’ data[8] because the firms that collect data and analyse it claim it as their own property. There are protections in some countries to enable users to ask for and get the data that is held on them, but not universally, not with the requrement tiochange data that is wrong. The argument over who should collect, process and store our data has taken on a strongly partisan political flavor. It re-opens the debate about how big should the State be, and what are the valid non-ideological boundaries of State spending and control. [9]

Meanwhile, storing data has become the organizational default. It is estimated that from 2011 the cost per transaction of finding and deleting a piece of data exceeded the cost of just storing it in perpetuity. It is hard to fully erase a digital record. To take a second example from the past, in any museum of antiquity glazed pottery is abundant. This is not because pottery was especially valuable. Most of it was ephemera of use for daily living and not much more. But fired pottery is hard to destroy. Even when broken it is able to be pieced together again. We need to think of our digital data not as digital footprints, washed away with the next rain, but as more analogous to pottery in prior eras.

---

[7] In November 2011 the UK Government announced its “midata” initiative to extend this beyond Government to 20 commercial organizations.

[8] Liam Maxwell : 2009 CPS paper. I am indebted to Prof. Nigel Shadbolt for highlighting this reference. And in the medical area, Portable Legal Consent is a new initiative that seeks to establish ownership of personal data used in medical studies for the purpose of making it more easily shared.

[9] We have a rule for how big a company should be from R Coase, but there is very little analysis other than ideological posturing about how big should the State be . (Taysom, Harvard 2001 : How big should the State be).

“3 is a crowd” : a proposal to reintroduce effective privacy on-line

There are four components to the project, running in parallel:

> development of a proposed legal structure;

> development of technology prototypes to demonstrate feasibility;

> outreach to multiple stakeholders to solicit support in outline to the direction taken, seeking any objections in principle early in the project;

> outreach to commercial and Government organizations to stimulate demand for the safe data analysis that the proposed platform would enable. This includes the scientific and social science research establishment who are hungry for more data but are often constrained by privacy concerns.

Progress can be gauged in each of these components. The proposed legal structure has received wide ranging support. Expressions of support have come from all stakeholders: business and private users, civil rights organizations, data collectors, data analytic companies. An Advisory Board to represent the key stake-holders has been assembled. The technology approach has received support from various academic institutions in UK and US. Discussions have commenced with three major corporations to be collaborators in pilots to test the technical and proposed commercial details.

Alongside these activities it is necessary to educate policy-makers that a non-private sector, non-Government institution can operate services that simply ‘can’t be evil’. Because a well intentioned slogan of ‘won’t be evil’ really should not be considered sufficient protection for society.

"3 is a crowd" : a proposal to reintroduce effective privacy on-line

Introduction: A majority of users of on-line services, when asked, say that they do care about their privacy.[10] Even when user behavior indicates that they don't care, informed observers believe that they should care. Governments have become concerned that private companies do not always reflect that concern in their products. Even when products do provide privacy settings they are often difficult to use. Research shows privacy controls are widely misunderstood, and even if they were perfectly clear and simple to use, they would generally be specific to one supplier or product. Immediate tangible benefits in the form of richer product features through the use of cookies beacons and 'tags' are offered by service providers if users volunteer or allow collection of unlimited user information. But the costs to set against these immediate benefits are intangible unknown costs at an uncertain future date and this makes informed individual decisions hard.

In USA, legislation is being considered under which users of web services will likely be offered the choice to either 'opt out' of being tracked on-line, or to remain with the typical service default setting and be tracked in as much detail as service providers choose. If passed, and there have been seven draft bill proposed to date, this would be the first US legislation to take a comprehensive approach to privacy with the FTC expected to be given the right of oversight. The choice to opt out fully should always be available, but a full opt out is not a fair choice for many users when simply accepting the default setting delivers a better service.[11]

There has been some pressure in US to start from a position of full 'opt out'. This is considered to be too harmful for the now extensive internet economy, where advertising funds much of the content, to win enough support to become law, the only way such a change is likely to be made. In EU legislation has already been passed in the form of a Directive to be enshrined in national legislation during 2011. Only a few States have complied by October 2011[12]. This legislation is aimed at notifying the user that they are being tracked and allowing a site-specific opt out from tracking. Some EU States have already passed or are committed to passing legislation even

---

[10] Harris Interactive and TRUSTe polls - www.TRUSTe.com

[11] See final FTC Report, March 2012, Report on Protecting Consumer Privacy. I am indebted to Prof. Nolan Bowie for this reference.

[12] Although in Germany a FaceBook user has been able to use new legislation to gain access to the data Facebook stores on him. The extent of the information caused a public outcry. And WSJ Sept 9 2011 reported that .." Thilo Weichert, data protection commissioner for Schleswig-Holstein, a state in northern Germany, declared that Facebook's "Like" button violated German data protection laws, and order that the button be removed from websites in the German state." prompting an agreement to work on a self-regulated code of conduct for social networks.

though there is not universal agreement about how to implement it, of if implemented how to audit compliance with the new law. It is generally understood in UK for example that as long as providers are working towards adopting new privacy standards there will be no reprisals.

This proposal offers a middle way. “Privacy-protected tracking” can provide an immediate benefit to both the user and the advertiser. The key concept is that what makes us economically valuable is our similarities, whereas our differences make us individuals. Privacy protected tracking is accomplished by generating dynamic ‘clusters’ of similar individuals (similar with respect to any observed or volunteered characteristic) that are rather like real-time virtual Zip codes, but not limited to the user location. The unexpected benefit to service providers and users is that in many cases there is potential for more fine-grained targeting without invading privacy. Tracking does happen, but in a manner that protects privacy by design, rather than as now where large numbers of users can provide a kind of accidental privacy, at least until service users go mobile. Adding location data can make re-identification more likely. It is proposed that for several reasons neither Governments nor the private sector firms could credibly offer this middle ground service. Instead, this proposal calls for a financially independent endowed not-for profit entity to be the vehicle for storing the mapping of user attributes to ‘real’ ID. This could be an extension of the role of an existing not for profit organization, or it could be one that has this purpose only. The likelihood is that a new dedicated organization would be required.

Other approaches were considered. Some have advocated full reciprocal transparency (Brin; 1999). Some see the solution being driven by a rebalancing of the rights of the customer and the vendor (Doc Searles, Berkman Centre at Harvard University)[13]. Others (Pentland; Clippinger, both at MIT) have proposed a ‘market solution’ - recompenses for reductions in privacy. Some see this as being levied before the fact in what has been characterized as an ‘insurance’ approach.

There is a view that the problem of privacy can easily be solved by interposing a ‘market mechanism’. Services can pay to have (or rent) your PII. But this is likely to result in a market failure : the seller does not know how to price the data nor who is an honest agent. The amounts may be very small and there is no obvious way to aggregate value and distribute it. However others have analyzed the fundamental problem with privacy to be the inherent market failure that

---

[13] “In its description of ProjectVRM[2], the Berkman Center says "The primary theory behind ProjectVRM is that many market problems (including the widespread belief that customer lock-in is a 'best practice') can only be solved from the customer side: by making the customer a fully empowered actor in the marketplace, rather than one whose power in many cases is dependent on exclusive relationships with vendors, by coerced agreement provided entirely by those vendors." wikipedia

occurs when there is significant information asymmetry: the benefit now cannot be weighed against the future cost. (NYU: Helen Nissenbaum).

The better solution is to have a privacy protected default, which this paper proposes. This does not take away the opportunity to add market mechanisms but it does provide a broader protection even for those whose activity may have little economic value.

Project rationale: Privacy underpins autonomy which is itself the basis for freedom. Yet privacy is a nuanced concept intertwined with trust hierarchies and notions or what is a public space and what a private space. Many younger people in the developed world when faced with the option to cede some privacy in order to access more functional, more ‘social’, internet services, act as if they feel the notion of privacy is no longer relevant. Yet in a Harris Poll in July 2001 repeating a similar poll in 2008, 94% said they were concerned about privacy.[14] This seems to stem in part from a belief that “this is just the way things are’ and there is no alternative. To attempt to provide some level of protection, in EU, services must by law give users an ‘opt out’ from being tracked on-line. But this has yet to be reflected in country-level legislation in most EU countries. There is draft legislation in process in US. In US, laws have addressed specific problems, COPPA[15] for children under 13, and HIPPA[16] for medical records, but these fall far short of a general comprehensive solution. They are often subverted. Parents for example collude with their children to subvert the COPPA rules. But although the option to opt out completely should always exist, it is not a practical option for most users who may calculate that the immediate benefit of richer services is always worth any potential indeterminate harm. And wanting to opt out is not always respected within the self-regulatory regime now in force.[17]

Corporations and Governments seek to analyze ‘Big Data’, which is to say data about ‘us’, to help with everything from better advertising to better disease control. But for commercial applications,

---

[14] Harris Interactive and TRUSTe polls - www.TRUSTe.com

[15] Childs On-line Privacy Protection Act. Websites that are collecting information from children under the age of thirteen are required to comply with Federal Trade Commission (FTC) Children's Online Privacy Protection Act (COPPA). www.coppa.com

[16] Health Insurance Portability and Accountability Act: From 2003 the HIPPA Privacy Rule requires ‘covered entities’ to : notify individuals of uses of their personal health information (PHI): must disclose PHI to the individual within 30 days upon request: may disclose PHI to facilitate treatment, payment, or health care operations, or if the covered entity has obtained authorization from the individual. However, when a covered entity discloses any PHI, it must make a reasonable effort to disclose only the minimum necessary information required to achieve its purpose.

[17] Jonathan Meyer at Stanford Law School Centre for Internet and Society studied 61 internet service providers and publishers and found that many continue to track users even if they have chosen ‘do not track’

and for certain medical record applications, and maybe even for genomic data, it is our similarities that are valuable, not our individual characteristics. Legislatures seem confused and the commercial lobby to reduce privacy is heard loudly whereas the counter-voice is easily labeled Luddite and dismissed. How should civil society act when both the Government, in the guise of more national security, as well as commercial organizations, in the guise of newly social products capable of ‘sharing’ information automatically, both deliver the message which is heard as : “Privacy is for old people: and they will soon be dead” (Class contribution from a Harvard undergraduate CompSci student 2011, class on Privacy). Does it even matter if with no privacy left we have achieved total security?

This project is a call to action for those who believe that there is a compromise: better privacy and better sharing of information. I argue that the solution requires both a new legal structure to house personal data, and a new technical solution extending the current internet protocols to allow ‘privacy by design’ rather than by accident. It may require a change in the status of personal information in law as has been proposed at MIT by Prof. Sandy Pentland.[18] It certainly requires a contextual approach to privacy recognizing that ‘on-line’ is not a new country.[19]

Theory of Change: To permanently change the level of privacy built into on-line services, rather than to police and prosecute breaches after the fact, a diverse groups of stakeholders needs to be engaged: users, in business, Government, academia, and the general public; content publishers; advertisers; data collection and analytic companies acting on behalf of advertisers; the Government (itself a major publisher of data). The internet eco-system needs to embrace ‘privacy by design’ solutions in the products they offer. A dual approach is needed: to persuade existing actors that there is a way to reconcile privacy with targeting, and to develop some new services which embody this concept based on a new platform.

To start to make the change, support from vocal academics and the institutions in which they work is needed. Some academics have been in effect co-opted into the camp of the entrepreneurs who claim that society has somehow changed and there is less emphasis on privacy among the connected young. This is to mis-characterize the debate, as Helen Nissenbaum at NYU has

---

[18] Pentland advocates a new class of information assets where the presumption is that the individual has ownership rights over their personal information, volunteered or observed. MIT lecture Oct 2011

[19] Helen Nissenbaum : “A contextual approach to Privacy online” Fall 2011 Journal of the American Academy of Arts and Sciences.

argued.[20] Willingness to share with peers is not the same as willingness to share with all observers. Polls confirm that all ages are concerned about their privacy.

Next, the support of civil rights groups is essential. Any resolution to the tension between targeted information and privacy must pass their scrutiny.

Members of the legislature and their advisers need to be aware that technology solutions exist to enable privacy protection other than via the crude and ineffective ‘opt out’; and that will require working prototypes.

Commercial organizations, potentially using pressure from both public opinion and lawmakers, must be reassured that their profitability will not be reduced by implementing ‘privacy by design’ : in fact it is likely that better targeting and more privacy can co-exist if design objectives for products change.

Technical alternatives to ‘Do Not Track’ must be shown in pilot deployments to work and perform without impacting user experience before they can be widely adopted.

And finally for there to be wide adoption it will be necessary to show commercial viability for these technical alternatives via one or more commercial product.

The new regime proposed in this project is in effect a platform on top of which a range of new ‘privacy by design’ products can be built.

Funding strategy : Funding for the project will come from three sources: private donations from individuals and from institutions to fund a permanent endowment; from Government grants to develop the technology; and from commercial organizations to develop pilots and to deploy the technical solution proposed here. It is proposed that some services provided by the NFP, for example to host observed and volunteered information including Pii on behalf of users themselves, or for data collectors, will be fee earning. This is anticipated to enable the organization to be self reliant and not have cause to fund-raise annually nor to rely on Government subsidy once it is fully operational. Examples include: ICAAN; Underwriters Laboratories; Mozilla Foundation; The Scott Trust which operates The Guardian Newspaper in UK.

Stakeholders: Stakeholders, many of which have already been contacted as part of the project, include: internet users of all types; Civil rights groups like EPIC, CDT, ACLU, EFF, and others; publishers; e-commerce services; ISPs; trade bodies for service providers like The Interactive Advertising Bureau and BITS, who represent the financial industry; data collection companies; data analytic companies; and existing projects that seek to open Government and commercial data

[20] Helen Nissenbaum interviewed by Jonathan Zitran - MIT 2011 Future of Entertainment.

to a wider audience. An example of such a multi-stakeholder model is ICAAN the organization which manages internet domains (http://www.icann.org/en/structure/

Implementation plan: “It will not be possible to realize the full potential of the next generation of the Internet and Cloud Computing without developing better ways of establishing digital identity and protecting privacy.” (A Cavoukian, September 2008.)

There is an urgency to this proposal. But there is a catch. Technology providers won’t spend money providing privacy protected solutions if regulation favors services with little regard to privacy. And law-makers won’t legislate unless they feel there are credible technology alternatives. There has been very little development of the view that there is an alternative in which where service operators simply can’t be evil.

Policy-makers formulate policy in the light of advice on technology as it exists and is anticipated. Policy is rarely affected until adoption at scale has taken place, which means policy is likely to lag technological change. Until a credible alternative to existing technology is offered how can policy-makers act to ease the path of a better solution? This project has so far shown that there is a reason to cherish and nurture privacy and that there was wide support for the approach that has been called ‘Privacy by Design’. By contacting civil rights groups and major commercial actors to gauge their views prior to announcing a major new initiative , whilst at the same time socializing the proposed technical approach with experts in technology and law, this project set out to build a foundation for the introduction of a new technical solution to on-line privacy (and potentially some off-line privacy) and the foundation for the adoption of a new legal structure that would redress the balance between civil society on the one hand, and Government and commerce on the other.

Who will use the new services provided by a NFP? : Customers for this proposed new service will come from several part of the eco-system. And perhaps surprisingly there is widespread support for the idea. Willing supporters include publishers worried that they might be holding PII data and be liable for its mis-use; and advertising networks wanting to target narrow sub-sets of users more tightly than they can now do whist maintaining privacy within well defined threat levels.

There is a greater awareness now that digital data is much easier to create than it is to destroy. Such that data collected becomes to the process of advertising targeting like toxic waste is to the nuclear energy industry. This proposal promises to deal with the toxic waste. The user will always

be able to access his own data, in line with long-standing OECD guidelines[21]. And in addition properly accredited legal or regulatory bodies can require access to the individual data that is tracked and stored but never revealed downstream to commercial users because it is only ever made available in safe ‘clusters’.

Users will also be customers for a range of products and services as now but where the publisher now can give them the chance to simply control the extent to which data about them is revealed on-line. Advertising planning companies wanting detailed data that has been de-identified within known levels of re-identifiability adequate for the purpose of protecting privacy will likely be customers. There are applications in medical health records and genetic databases[22]

Providers of on-line education and healthcare in developing countries might be particularly sensitive to the need to provide privacy for those who are among the poorest. Free services should not mean privacy invasion.

How will success be measured? The “3 is a crowd” project is intended to provide the basis on which to establish a new Not For Profit which will house the trusted ‘vault’ for personal data.. Ultimately success can only be claimed when private data stays private by design and there are many projects running in parallel that in combination will deliver this over time. But progress on the route to get there requires an institution to be the ‘home’ of personally identifiable data (Pii). The establishment of this organization will be a major milestone. A legal framework has been identified for this NFP vehicle and a potential Chair of the Governance Board and a potential CEO identified. This organization will also propose reference technical designs for a proposed technical solution, and potentially develop products which illustrate consumer demand for better privacy. These will be based on a patent application which has already been filed.

On the way to the bigger goal small victories have been had. A first tangible output from the project has been a simple commercial product[23] to enforce ‘Do Not Track’ (www.monitr.com). The premise of the project is that ‘do not track’ is insufficient but tracking ‘do not track’ will need to be part of the solution. A UK Government grant application is being filed to start to test some technology solutions; and discussions have started with potential commercial partners willing to take small scale proof of concept demonstrators to the next level on the way to full commercial adoption.

---

[21] This begs the question of how to identify the user. Many companies are addressing this issue, for example OneiD. http://www.oneid.com/

[22] It has been suggested in discussion with MDs that In these cases therapeutic value also may arise from the similarities with other like individuals rather than the differences.

[23] Developed by www.Kontexto.com

“3 is a crowd” : a proposal to reintroduce effective privacy on-line

Progress has been made in 2011 in gathering support in advance of formal lobbying from influential actors in USA and EU in commerce, in Universities, within Government, and within civil rights groups and supra-national groups like WEF.

“3 is a crowd” : a proposal to reintroduce effective privacy on-line

Some background on Privacy. Privacy in EU is a legal right. In US it is an inferred right whereas, for example, the right to free speech is explicit in the Constitution. In US, laws have addressed specific problems, COPPA[24] for children under 13, and HIPPA[25] for medical records, but these fall far short of a general comprehensive solution.

Privacy was always a mark of status in the past. Slang of all sorts evolved as a way to converse privately in ‘public’. Cockney rhyming slang for example is a form of simple message encryption to allow private speech in front of the ruling class[26]. From the home to the work-place, closed doors were the right of the powerful. In a modern setting ‘hoodies’ empower youths to be anonymous in an age of CCTV. An 1881 legal case which turned on the right to give birth unobserved appears to be the first reference to a claimed ‘right’ to privacy in US. (Clapham; 2007).

On-line privacy is quite a recent concern for many observers of the evolution of the Web. In his seminal work ‘The Wealth of Networks’ published in 2006, and as comprehensive as its title and its reference to Adam Smith would suggest, Prof. Yochai Benkler makes no reference in the index to ‘privacy’.

And yet contrary to the view frequently prevalent in the press up to late 2010, it seems that individuals do care about privacy. Parents care; employees care; and employers care. The Military cares and Civil Rights organisations like ACLU on behalf of civil society care. Zuckerberg was proved wrong; the kids do care as Gallop polls in February 2011 confirm. 75% of those polled were concerned about privacy on Facebook and Google. The WSJ cares: it ran a 6 part series on Privacy in early 2011. The Economist and Time magazine have recently run extended analyses of the erosion of privacy and its impact. Even those who have been relaxed about on-line privacy become more animated about privacy on the ‘mobile web’ where location becomes an attribute that is discoverable. Zuckerberg for FaceBook and Reid Hoffman for Linkedin have both said

---

[24] Childs On-line Privacy Protection Act. Websites that are collecting information from children under the age of thirteen are required to comply with Federal Trade Commission (FTC) Children's Online Privacy Protection Act (COPPA). www.coppa.com

[25] Health Insurance Portability and Accountability Act: From 2003 the HIPPA Privacy Rule requires ‘covered entities’ to : notify individuals of uses of their personal health information (PHI): must disclose PHI to the individual within 30 days upon request: may disclose PHI to facilitate treatment, payment, or health care operations, or if the covered entity has obtained authorization from the individual. However, when a covered entity discloses any PHI, it must make a reasonable effort to disclose only the minimum necessary information required to achieve its purpose.

[26] In Cockney, nouns are replaced by pairs of words one of which rhymes with the respective noun: but only the non-rhyming half of the pair of words is used. You have to know the pairs to decrypt the message: So ‘I lost his Lordship’s daisies‘ expands to ‘I lost the bosses daisy roots’ ie: his boots.....

publicly that they believe privacy is dead, an out-moded concept. This of course favours their own business interests directly.

But in fact the concern for privacy is widespread, not just in advertising-driven on-line commerce but also in the data collection required to provide portable medical records and the handling of genetic information derived from DNA analysis.

“Patients are not currently being adequately informed about possible secondary uses of their medical data for medical research; are not asked to give clear, specific, free, and informed consent; are not offered unambiguous and effective opt-outs; and are misled about the degree of anonymisation of their data and the likelihood of re-identification.” (British Medical Journal, Feb 2011: Oxford Internet Institute : Prof. Ian Brown

In part this concern results from the delayed impact of giving up privacy in the form of data collected. The costs are not well understood, and are hard for users to put an economic value on. The choice to ‘opt out’ may be there but for most of the potentially vulnerable it is not really much of a choice - rather like being offered the chance to ‘opt out’ of having a credit history. In situations like this Governments often feel the need to legislate. And this is the case in the EU and likely to be the case in US. Washington cares: most recently Senators Kerry and McCain have circulated a draft Bill. The US Courts meanwhile have begun to enforce the right of users to be protected from providing more information than is required to enable a transaction to occur. In a related ruling, a California Court ruled in March 2011 that a vendor concluding a credit card transaction is not entitled to ask the buyer for their Zip code. That is the underlying principle of this proposal, to provide commercial users with only sufficient information to enable services to derive commercial (or medical) value from attributes, and never with enough detail to identify an individual among a ‘crowd’ of similar users. Some web services companies say that they do this already and that a kind of ‘accidental’ anonymity is provided because large numbers are involved. This is fine as far

"3 is a crowd" : a proposal to reintroduce effective privacy on-line

as it goes but over a year ago the "Privacy Commissioners"[27] called for 'Privacy by design' to be proposal calls for this to be codified and standardized, to be designed in rather than accidental, and to be auditable.

Privacy is a culturally determined concept and it changes over time. There is a complex relationship between privacy, freedom, and security. Villages afford little privacy from comings and goings whereas in a city neighbors may be strangers. It is not however the case that culturally determined changes in privacy always trend towards less. US law focuses more on the protection of individual privacy from the Government, where UK and EU law more generally is more focused on privacy as between private individuals and corporations. This leads to a different approach to solutions. There is a wide divergence both within countries on how different elements of identity are protected and between countries for similar data types. Images in US are treated quite differently than text for example. Off-line this has had little consequence as long as you knew what the rules were in your jurisdiction. On-line the position is more complex as the notion of 'where' is blurred. Off-line the mass adoption of CCTV cameras in public places[28] is changing the off-line privacy assumptions in parallel. In a small UK town near Cambridge, Royston, the local government now records all number plates of vehicles entering and leaving Royston. They know if you have been: if you have left: and if you are still there. Facial recognition is now very sophisticated, as witnessed by the success in prosecuting the London 2011 rioters from CCTV footage. New technology from companies like SceneTap, 'facial detection', is already in use in 50 bars in Chicago. (NYT, Sunday Nov 13, 2011). But at least these CCTV cameras are not installed outside their homes, let alone inside them. On-line usage is trackable in your den, your study, or your kids bedrooms.

On the whole, off-line, to twist the well known saying, know-one knows you have a dog: unless they see you out walking with one. On-line things have developed quite differently. On-line your browsing and buying habits would label you as a dog-owner quite quickly. In fact on-line service

---

[27] The Privacy Commissioners : source www.privacylaws.com
The resolution proposed by Ontario's Privacy Commissioner, Dr Ann Cavoukia was passed in November 2010 which`;

i) Recognises Privacy by Design as an essential component of fundamental privacy protection;
ii) Encourages the adoption of Privacy by Design to establish privacy as an organization's default mode of operation;
iii)Invites Data Protection and Privacy Commissioners/Authorities to promote Privacy by Design in their jurisdictions.

28

providers will know that you may be thinking of getting a dog before you tell your family. Off-line, no-one knows when you board the bus to town whether you are going to shop, to catch a plane, or to the ball-game: on-line this can be inferred with considerable accuracy. For the generation who grew up with ‘connectedness’ as part of their daily routine, as Sherry Turkle has written based on her own primary research, the ‘web’ is not a memory prosthetic and an aide to research, as it has been for most people over thirty, so much as it is an extension of personal psyche. Prof. Jonathan Zittrain in response to a question at the 2011 Harvard Berkman Law School iLaw seminar notes that this changes the context in which law should be developed to ensure ‘..the highest protection under the Constitution’. Which cannot be said to be the status quo in October 2011.

Rapid growth from 1995 in the use by businesses, Governments, and individuals of services based on the internet has created the conditions for an unintentional acceleration of the erosion of privacy not supported by public opinion[29]. The potential for internet use to erode privacy came about because of the design of the internet protocols on which internet services run which allow for traceability of users unless they take specific actions to prevent this. The acceleration in the potential for loss of privacy has been compounded by the commercial discovery that individual actions can not only be traced on-line and used to deliver targeted advertising, but also can be shared with other users automatically to create the so called ‘social’ internet products of the last five years. Until 2011 there had been little vocal outcry from users adopting popular ‘social’ products. This appears to be in part because of the phenomenon of the entrepreneur as cultural hero, which can vest in companies and their leaders a key role in setting new norms for privacy. Multimillionaire Scott McNealy (CEO Sun Microsystems) whose company was arguably one of the key commercial drivers in the adoption of the internet and its underlying protocol TCP/IP declared : “Privacy is dead, get over it!”, but this seems more as a statement of fact arising from the implications of the technology than a manifesto. Billionaire Mark Zuckerberg (founder of Facebook) and Reid Hoffman (founder of Linkedin) echo this view, but in their cases as a desirable trade-off for a more social world made social by the adoption and use of the products of their for-profit companies. The young appear to see the erosion of privacy as inevitable even if it is not desirable. Its just the way the world is. This would explain their concerns when polled but their actions online which seem to contradict.

[29] Several public polls in 2010 and 2011 establish this point. Gallop results were reported in WSJ and the Economist

“3 is a crowd” : a proposal to reintroduce effective privacy on-line

The proposal to use a Not For Profit. Privacy and security are in obvious tension, and this exposes party political leanings. But privacy is an issue which has become further politicized as Governments seek to harness Big Data. For example in UK, Liam Maxwell in a report for the Centre for Policy Studies in 2009 wrote:

“ A clear choice is emerging for the future of government IT: – Either to continue with the Transformational Government agenda. This relies on the State holding, in the words of the Treasury’s adviser, a “deep truth about the citizen, based on their behavior, experiences, beliefs, needs and rights”, with huge centralized databases directing public services to the point of need (as judged by the State). – Or to abandon expensive and failing centralized IT projects and yield control of personal information to individual citizens. This is the approach that has been increasingly effective in the private sector.”

It is hard to develop tailored State delivered healthcare delivery and tailored education, to take just two examples, without developing detailed databases on individuals. Should this really be maintained within Government departments? A new approach is needed.

It is proposed that the ‘quasi proxy’ routing which is the essence of the technical approach proposed should not offered by a commercial entity, Rather it is proposed that the key element of the service which protects id or personally identifiable information (PII information) is housed in an unusual but fundamentally very robust legal vehicle: an endowed charity or not-for profit. Capital will be raised to form an endowment which will pay the costs of building and maintaining a database of on-line users and their preferences behind military-grade security. This entity will license to service providers real-time assembled sub-sets of the data collected in such a way that users with a reasonable level of confidence appropriate for the transaction undertaken and the domain cannot be individually identified. The size of these ‘crowds’ will be actively managed by the system in real-time to ensure the individual is always in a crowd big enough to remain anonymous. Profits from this license income will be applied to top the endowment up for inflation to ensure that its purchasing power remains the same and to fund ongoing research to ensure that the system remains at the cutting edge of the best technological solutions available.

Government as the possible ‘owner’ of the database that enables this service is clearly a non-starter, especially in US. In UK its has been estimated that central government spends £300 per person per year to manage personal data, Google, MSN and online banks spend between £10 and

£60 per person per year.[30] Two quotes from influential UK reports[31] are probably sufficient to make the case against public sector ownership:

“The public are neither served nor protected by the increasingly complex and intrusive holdings of personal information invading every aspect of our lives.” (The Joseph Rowntree Reform Trust, “Database State”, 2009)

and:

“The people who are handling the amounts of data , because they are in contact with them every day are utterly blasé about the risks associated with the data... and have no understanding about the impact of disclosure or leaking... has on the lives of individuals... That is something that has to change.” (Professor Angela Sasse, House of Lords Surveillance Committee)

Private companies would seem to be a good vehicle because they are nimble, can take the required risks, and can be long-lived. But on reflection the fate of companies, even those which seem to be essential parts of the landscape like Amazon, Google and Facebook, all of whom have within the last month announced changes to their privacy polices, is short-lived. There is concern today that Yahoo may be acquired by a Chinese company. The concern is not the usual one of losing know-how to the Chinese (or whoever the current bete-noir happens to be) but of losing control over information that Yahoo has about US citizens.

Private companies are run for the benefit of their shareholders: the power of the NFP as the provider for this service could not have been better highlighted. Private companies do not last for generations. Data from medical records and even more so DNA data has implications across generations. By way of contrast with private for-profit companies, Underwriters Laboratory (UL), has been operating since 1894 and has developed a worldwide business in certifying product safety. UL is a not for profit trust with a for profit subsidiary providing test services.[32]

---

[30] Liam Maxwell: 2009 CPS

[31] ibid

[32] Other possbile models include ICAAN, Wikipaedia, Mozilla

“3 is a crowd” : a proposal to reintroduce effective privacy on-line

Some background to ‘on-line’. To understand how the potential for wholesale loss of privacy on-line has happened it is necessary to look at what on-line has come to mean. When we say ‘on-line’ we mean usually using a network which is based on the TCP/IP protocol. TCP/IP has as its underlying structure a series of physical nodes which switch data addressed from one node to another. The origin and destination are both explicit and traceable. Quite a good analogy is that TCP/IP as a network for sending messages is that it is more analogous to a postcard than a letter inside an envelope. Not only do you know the sender and the recipient, with only a little effort you can read the message.

An approximation of fully anonymous use is possible using many commercial services by re-routing data via a proxy server, an intermediate ip address. When this is done the benefits of tracking, presenting users with relevant information and remembering what they did last time a site was visited are lost. Some anonymity is quite ‘strong’. But there remain issues even using proxy servers. The TOR protocol, arguably the most sophisticated developed to date, provides anonymity of the activity undertaken, but the nodes are still discoverable.

The overlay of the World Wide Web protocols (HTTP, HTTPS, HTML etc) onto TCP/IP has inherited the properties of the underlying technology. Web 2.0, a short-hand both for more user-friendly features and also for more sophisticated distributed approach to the provision of applications over TCP/IP, adds a measure of ‘memory’ by way of cookies and ‘tags’, snippets of data and sometimes bits of code, being left on the device of a user or the page of a provider of information by an upstream web service when the service is visited by a browser. This is true whether the service is delivered to a desk-top device or a mobile device: and it is true whether the service delivered is a recreational game or is an on-line health advice or an on-line education service. Cookies can be ‘read’ and ‘dropped’ by any upstream device and they are linked to the device by way of the IP address on which they reside.

So how can privacy be re-introduced without offering a full opt-out as the only alternative, and without harming the ability of companies to offer web 2.0 services at a profit funded by advertising? It is proposed that an intermediary is interposed between the browser and the user which ‘clusters’ like individuals into real-time assembled ‘crowds’ within which no individual can be identified. These clusters then act rather like a proxy server routing messages and content based not on a single ip address or user but on a cluster of ip addresses which generate an ephemeral ‘pseudo ip’ address representing more than one user where the number of users in the ‘crowd’ is

determined by the information being passed and the use to which it is being put. [This approach is the subject of a 2007 UK patent application which has been extended to EU and US.]

Project Advisors. During the Spring and Summer of 2011, as part of a project at Harvard University, a summary of this note has been circulated and discussions have been started with existing service providers and potential customers of this new interposed service layer, and discussions have been had with advocacy groups, advisors, researchers, and law-makers, to enlist support for the idea as a valid middle ground which allows users to get the benefits of targeting and analysis without compromising personal privacy.

The following have agreed to lend their support as advisers to the project.

Prof. Jim Waldo, CTO Harvard University, Sun Distinguished Engineer, investor of Jini. McKay Professor of the Practice of Computer Science .

Prof. Dame Wendy Hall, Professor of Computer Science at the University of Southampton, UK, and Dean of the Faculty of Physical and Applied Sciences. President of the Association for Computing Machinery (ACM) in July 2008, and the first person from outside North America to hold this position. Recipient in 2011 of the Oxford Institute ‘lifetime achievement award’

Dr. David Cleevely FRS. Founding Chair, Cambridge University Technology Policy unit. Adviser to the UK Government on spectrum and security.

Prof. Alex `Sandy’ Pentland director MIT’s Human Dynamics Laboratory and the MIT Media Lab Entrepreneurship Program, and advises the World Economic Forum

Prof. John Clippinger, MIT was Co-Director of The Law Lab at Harvard University Previously, Dr. Clippinger directed Social Physics project at the Berkman Center that supported the development of an open source, interoperability identity framework called Higgins to give people control over their personal information. He is the author of A Crowd of One: The Future of Individual Identity, 2007) and consults with Equifax and other companies, foundations, and government agencies on technology, policy and business strategy.

Anna Burger, 2011 Harvard Fellow: Anna was Secretary-Treasurer of the Services Employees International Union (SEIU), the nation's largest and fastest growing union, and the first chair of America's newest labor federation, Change to Win. Burger serves on President Obama's Economic Recovery Advisory Board.

General Gale Pollock, 2011 Harvard Fellow. Gale was the Commander of the US Army Medical Command and the Acting Surgeon General (the first woman, non-physician to have this role in any

of the military services) and the 22nd Chief of the Army Nurse Corps. Gale is also a Fellow in The American College of Healthcare Executives (FACHE) and the American Academy of Nursing (FAAN).

XXX A member of the Boston Health Service (name available on request), an MD with a previous career as a ComSci graduate.

XXX A member of the MIT faculty and previously head of security for a major IT corporation. (name available on request)

Prof. Latanya Sweeney, Harvard and Carnegie-Mellon. She is also appointed to the Privacy and Security Seat of the Federal HIT Policy Committee. Her work involves creating technologies and related policies with provable guarantees of privacy protection while allowing society to collect and share person-specific information for many worthy purposes. She also founded the C-M Data Privacy Lab.

Dr. Judith Murciano. Harvard Law School. Judy has served as Legislative Director and Acting Executive Director of the New Jersey ACLU, chaired the New Jersey Bar's Juvenile Justice Committee, and clerked for a criminal court judge in the Bronx. She has written human rights articles for the International Herald Tribune and Radio Free Europe while working for Amnesty International in Paris. Judith was a recipient of the 2008 Dean's Award for Excellence at Harvard Law School

Prof Kathrine Heller, MIT : NSF Postdoctoral Fellow in the Computational Cognitive Science group at MIT. previously an EPSRC Postdoctoral Fellow at the University of Cambridge.

Al Zollar: 2011 Harvard Fellow. formerly IBM General Manager, Tivoli Software; member of the board of The Chubb Corporation.

David Weinstein, 2011 Harvard Fellow: most recently Chief of Administration at Fidelity Investments, where he also led its national government relations efforts. Previously, he practiced corporate, securities and tax law. David serves on the Board of Trustees and is now Chairman of Boston College Law School.

Tom Santel. 2011 Harvard Fellow. Tom was President and CEO of Anheuser-Busch International, Inc. up to its acquisition.

Terrence Straub, 2011 Harvard Fellow. Terry was most recently the Senior Vice President of Public Policy and Governmental Relations of U.S. Steel. Prior to joining U.S. Steel, he served at the White House as Special Assistant for congressional affairs under President Carter and has been involved in various state and national political campaigns. He also serves as Co-Chairman of the Washington DC Metropolitan Police Federation.

“3 is a crowd” : a proposal to reintroduce effective privacy on-line

“3 is a crowd” : a proposal to reintroduce effective privacy on-line

Partial bibliography:

NRC 2007 Engaging Privacy and Information Technology in a digital age:

Andrew Clapham: “Human Rights : a very short introduction.” 2007

“ The Constitution of the United States of America” and amendments, 1791

Richard Reeeves: “John Stuart Mill : Victorian firebrand: 2007

Sherry Turkle : “Alone Together” : 2011

NRC “ Putting people on the map: protecting confidentiality with linked social-spatial data” 2006

Nigel Shadbolt and Kieron O’Hara : “The Spy in the Coffee Machine: The end of privacy as we know it” 2008

Philip Sheldrake : “The Business of Influence” : 2011

David Brin : “The transparent society” : 1998

John Taysom and David Cleevely : 2007 **METHOD OF ANONYMISING AN INTERACTION BETWEEN DEVICES** WIPO Patent Application WO/2009/068917

Helen Nissenbaum : “A contextual approach to Privacy online” Fall 2011 Journal of the American Academy of Arts and Sciences.

Liam Maxwell ‘Its Our’s” ‘ Why we, not the Government, must own our data” Centre for Policy Studies 2009

A Cavoukian, Privacy in the clouds – A White Paper on Privacy and Digital

Identity: Implications for the Internet, September 2008.

2011 ‘iLaw’ conference : Berkman

2011 Harvard Law School symposium ; The Democratization of Entertainment.

2011 MIT conference ‘The Future of Entertainment”

The Economist

The Financial Times

Wall Street Journal “What they know” series - from July 2010

The Guardian

The New York Times