Andrew van der Stock

15/12/2016

Response to Privacy Amendment (Re-identification Offence) Bill 2016

The bill in its current form, whilst tackling an important issue, does not address the root cause of failing to properly protect sensitive personal information as defined by APP 11.2, as it criminalizes disclosure but not inadequate de-identification, nor requires timely resolution of notified breaches as per the risk of the data, thus placing highly sensitive data at risk for far longer than otherwise would be the case.

My credentials

I am writing today in a personal capacity.

I am Chief Technology Officer of a security consultancy firm, and Director of the Open Web Application Security Project (OWASP), which sets the standards and publishes research for application security. I have worked globally, both here in Australia and in the USA for over 20 years, in the health, financial, insurance, telecommunications, legal, logistics, mining, and government sectors over that time.

I researched, developed, and documented many of the common application security controls that all IT security firms use worldwide. I wrote the standards in use for application security, such as the OWASP Developer Guide, OWASP Top 10 (which is used by PCI Data Security Standard to protect all global credit card transactions), the OWASP Application Security Verification standard, which was recently mandated by the UK's National Health Service for verifying the security of clinical software and thus the controls protecting highly sensitive data held within clinical management systems.

I helped set the first SANS GIAC GSSP (Java) certification. I was

the recipient of the AusCERT SC Magazine Individual Award for Information Security Excellence in 2013.

Concerns regarding the bill as drafted

I am broadly supportive of the explanatory objectives of the bill, which is to ensure that re-identification does not lead to unauthorized breaches of sensitive and sensitive data as defined by the Australian Privacy Principles (APPs).

However, the bill's sole objective cannot be met by the legislation as written, as it seeks to criminalize and severely punish all re-identification, regardless of intent, and regardless if any sensitive records were re-identified, transmitted or stored in a way that publicly exposes these records any more than the original de-identification process.

It is my understanding that this bill is due in no small part to the fine work undertaken by University of Melbourne academic researchers (Culnane, Rubinstein, Teague), who reconstructed approximately 10% of two public de-identified data sets. They have submitted an excellent overview of their work to the Committee, and I highly recommend the Committee read it.

Cryptologists, highly qualified mathematicians who can create or analyze cryptographic systems, typically only accept that a new cryptographic algorithm is safe and effective once it has been extensively analyzed by groups other than the group creating the algorithm. No new algorithms should be used prior to extensive peer review and practical attacks. This is how AES encryption standard was selected and verified. We are currently undertaking a similar process for the new Argon2 hashing algorithm, which although theoretically stronger than our current best hashing algorithm, has yet to pass through sufficient cryptanalysis to replace existing hashing algorithms in our standards.

The academics in this instance were doing such work - they set out to see if re-identification was possible, and demonstrated it was possible. This type of research is critical to the continuing privacy and safety of all personal and sensitive private records. Culnane *et al* strengthened security of these records, not weakened it.

In my field, we rely upon skilled people, processes, and algorithms

(such as de-identification routines, cryptography libraries, and hashing algorithms) being in place and effective to protect data. The work of the ASD, the OAIC, OWASP, and others would be made difficult to impossible if we do not have an effective legal framework to work within when testing the security and assumptions of concrete implementations.

In my view, this bill was a rapid but incorrect response to the University of Melbourne's research. The bill can be made far better with a few amendments, whilst protecting good faith research as undertaken by Culnane *et al*.

To that end, I make the following observations:

As these data sets are likely to be unique to Australia, criminalizing only Australian researchers will place an undue burden on Government and its agencies to ensure that de-identification is performed using leading edge statistical, set and information theory, but with the absence of any Australian research into this important topic, they will simply not know.

Weak algorithms and processes, not disclosure, is the root cause and is not addressed by this bill. Security researchers have shown many times that data that is improperly de-identified using a weak algorithm or process is a risk to the privacy and data protection of the original data set.

Researchers need balanced and practical guidelines on how to notify affected data set owners, and how best to protect data they have re-identified to protect both the sensitive personal data, and the researcher who is acting in good faith. Researchers should not need to be affiliated with a university or institution, and should not need to have a contract with an Agency to conduct the research, as this will stifle industry research or notifications from accidental re-identification.

The bill should define a timeline on how long data set owners have to rectify an issue based upon the risk of the data set. For example, a high risk data set such as medical or mental health records should be addressed very rapidly, where a data set of public transport trips might be able to be addressed in a longer time frame. What is not acceptable is the Government being able to

stifle both the publication of the research as well as having an indefinite time frame for resolution.

We have seen many examples of legal threats chilling research to the substantial detriment of potential victims, such as a car manufacturer that sued security researchers who demonstrated weaknesses in a car locking mechanism [1] and delayed a fix for all owners for over two years rather than rapidly addressing the problem and issuing updates to the key locking mechanism or issuing new keys. This caused significant harm to this manufacturer's owners, with criminals being able to duplicate electronic key fobs and steal cars.

The simple act of researchers describing their methods and algorithms would be sufficient for others to repeat the work, so even if the researcher properly protected and redacted the resulting re-identification, other researchers could replicate this work even if they are unaware of the original research or data. This is why research in this area is critical, so such weak de-identification techniques can be retired, such as DES, 3DES, RC2, RC4, RC5, MD5, SHA1, and a whole host of other weak algorithms.

Machine learning may accidentally re-identify insecure data. Advances in machine learning and AI over the last five years have highlighted the ability for off the shelf MLN software to autonomously identify patterns in data they process. Accidental re-identification is highly likely where de-identification is weak. If a business obtains legal advice and finds that discussing this accidental and ever more likely outcome with an agency has some risk of prosecution, the Agency will almost certainly not hear about such automated and intent-less re-identification data breaches.

Once re-issued, researchers should be free from legal sanctions to pursue research in this important field, and publish their results (but not the data) in peer reviewed journals, industry conferences and proceedings. The current bill outlaws disclosure of all re-identification, which is not the outcome required to protect data sets today and into the future, as agencies will not know if a particular process or algorithm is still safe or needs to be changed. Some data cannot be safely de-identified, and research continues into how to identify such data sets.

[1] https://www.wired.com/2016/08/oh-good-new-hack-can-unlock-100-million-volkswagens/

The onus of proof is reversed, which means that prosecution is all but guaranteed even if one or more the various defences are met. This will have a chilling effect on security research, because a researcher who does not have an official contract with an agency, or an industry researcher who is not affiliated with an institution such as myself, or a researcher who absolutely sets out to re-identify data to ensure that the data set is properly protected by the chosen algorithm, will be at risk of going to jail. This has no other effect than preventing benign research and discussion. The likely outcome defeats the underling reasoning for this bill, and also the controls set out in APP 11.44.

The punishment does not match the crime. The penalties are equivalent to the penalties for far more serious crimes such as aggravated property crimes. The punishments should be aligned with the Privacy Act, as it seeks to protect the same exact data de-identified under APP 11. It is incongruous that an agency that fails to perform de-identification correctly and thus exposes the same exact sensitive data would face a penalty far less than this bill contemplates.

The bill's penalties should focus only on prosecuting malicious release of re-identified data, with the stronger end of penalties reserved for not disclosing and working with Agencies to close out the risk. This can only be achieved by ensuring that researchers are free to continue to test the results of re-identification of sensitive data sets and to improve the algorithms in use. There should no crime committed as long as researchers provide timely notification, work with the agency, and provide evidence that they have adhered to necessary data redaction protocols for the re-identified records in their possession.

Taking all of these points into consideration, the downside is if the people, processes or algorithms are actually weak, but academia, industry and researchers such as myself cannot provide independent advice within a sound legal framework, the only people who will know about the weaknesses are those who set out to breach the system but not inform the government, whether within arms reach of Australian law, or from somewhere else on the Internet. After the passage of the bill as written, the legally safest option for researchers, academics and business if they don't disclose re-

identification at all.

I'm sure the likely outcomes of non-disclosure but ongoing data breaches is not the purpose of this bill nor the Government's intention when drafting the bill. Changes to the draft bill are required to achieve the desired outcome. To that end, I make the following recommendations:

Recommended changes

The bill should be amended to include intent as a backbone to the test of prosecution:

- The bill should have an absolute defense limb that permits good faith academic and industry research into this important topic. Security researchers and academics should not face prosecution if there is no evidence of bad intent, but instead protected by the law if they can demonstrate good faith to keep the results of their research from harm, such as notifying the data set owners as soon as they have a result and protecting any re-identified records in their possession.
- The bill should be redrafted to narrowly focus on willful bad faith disclosure of methods, algorithms, and re-identified sensitive personal information prior to an Agency or agencies resolving the issue in a timely fashion. There should no penalty or crime committed for releasing methods, techniques, algorithms or code once the data set is re-issued.
- The bill should include a regulation or additional powers for OAIC to develop unified and public guidelines for the protection and secure destruction of re-identified data held by researchers rather than demand they cease and desist. The clause "cease any other use or disclosure of the re-identified information, and" should be deleted in its current form.
- The bill should be amended to permit the timely publication of re-identification methods, techniques, algorithms, code and papers by academics and industry researchers after Government has had adequate - but not indefinite - time to develop and implement a better de-identification algorithm or process, and re-issue data sets in the new, more secure format.
- The reversal of proof should be removed from the bill, as the

only outcome of reversal of proof is to prevent disclosure of methods, algorithms or re-identified data sets to Agencies, and does nothing to protect the data from unauthorized re-identification by malicious actors outside of Australia.

- The Privacy Act penalties either should be aligned with these new harsh penalties, or this bill aligned with the existing Privacy Act's penalties. This avoids the appearance that this bill sets out to punish researchers who might appear to embarrass the Government or an agency, and yet allows agencies to continue publishing sensitive records without adequate safe guards in place. The current Privacy act penalties are sufficient in my view.
- The bill should contain a single crime covering the unauthorized release of sensitive and personal information via re-identification, misuse, interference or loss, with the same penalty units as the existing Privacy Act. This should always be prohibited, regardless of intent. This should be written in such a way as to not criminalize the product of re-identification by the researcher who is adhering to published data security protocols.

Conclusion

The airline industry did not become the safest mode of transport by criminalizing the disclosure of crashes, but investigating them fully, and removing blame from investigations, so that pilots and others would freely speak about what might have caused the issue.

I ask the Committee to work with industry and academia to come up with the right balance between the need to protect sensitive and personal information, whilst also protecting researchers who act in good faith to improve the security of data sets and algorithms in use, as per APP 11.44, and APP 11 and OAIC[2] resources[3].

[2]https://www.oaic.gov.au/information-policy/information-policy-resources/information-policy-agency-resource-1-de-identification-of-data-and-information

[3]https://www.oaic.gov.au/agencies-and-organisations/business-resources/privacy-business-resource-4-de-identification-of-data-and-information

For more than 70 years, the information security industry and academia have openly researched and improved security in exactly the way the University of Melbourne researchers did. They should not be punished, and no future researcher should be punished for informing the Government and its agencies about the insecurity of processes, configuration or sensitive personal data.

Thank you for considering my submission. I am available for any questions you might have relating to my submission – or on information security, application security, privacy, and data protection more generally – as a subject matter expert of some standing.

Andrew van der Stock
Highton, Victoria