CUSPE Cambridge University Science and Policy Exchange

CJSP Cambridge Journal of Science & Policy

Innovation vs. Privacy: tech's false dichotomy

Communication | Editorial | Invited contribution | Perspective | Report | Review

Adam Ó Conghaile
Department of Computer Science and Technology
University of Cambridge

Abstract

When it comes to data, privacy and innovation often appear to pull in opposite directions, as seen in the differing policy approaches taken in the US and EU [1]. Fully homomorphic encryption offers to resolve this conflict by making third-party data analysis possible without revealing user data. Policymakers need to be ready for this paradigm shift. This article describes this new technology and its potential policy implications, particularly in the context of recent EU and US data protection policy.

Science ⇒ Policy

When it comes to data, privacy and innovation often appear to pull in opposite directions. Full homomorphic encryption offers to resolve this conflict by making third-party data analysis possible without revealing user data. Policymakers need to be ready for this paradigm shift.

Keywords Privacy · Data Protection · Machine Learning · Encryption

Policy Background

With the rise of modern telecommunications networks over the past 50 years, consumers have entrusted more and more private data to companies, turning encryption from a mathematical niche into an everyday business necessity. Initially, policy focused on encouraging companies to encrypt their data so that customers' personal information would be stored in a coded form which kept it private. The US introduced a Data Encryption Standard (DES) as early as 1972. However, encryption traditionally limits the usefulness of data for applications, such as training machine learning models. These data-driven technologies have become increasingly important and economically valuable in the early 21st century, introducing a potential tension between privacy and economic value. The EU and US, with the exception of California, have taken different policy approaches to this problem. The EU's General Data Protection Regulation (GDPR) champions privacy while restricting data sharing, whereas the US has opted for a more open approach. This divergence has caused tension particularly around transatlantic data sharing. In the Schrems II case [2] the EU found that existing data sharing protocols with the US were incompatible with GDPR.

Fully homomorphic encryption (FHE) is a technology with the potential to completely re-write this narrative. Proved possible by Craig Gentry in 2009, FHE allows data to be analysed while still securely encrypted, meaning that companies would have the ability share their customers' data with third parties for analysis without ever giving them access to sensitive information. Technical

advances will make Gentry's theoretical discovery an increasingly viable technology in the coming decade. Policymakers need to be prepared for the challenges and opportunities this presents.

Scientific Background

From a technical point of view, the field of homomorphic encryption (HE) is best understood in light of two other important concepts from data security: encryption systems and data anonymisation. As is shown in this section, HE combines these two concepts.

An encryption system in its purest form is a pair of algorithms: one algorithm ('enc') encodes data into ciphertext, a format which is unintelligible to anyone without access to the second algorithm ('dec') which decodes ciphertext back to data. As depicted in Figures 1 and 2, it helps to conceptualise 'enc' as a lock that can be put on data and 'dec' as the corresponding key. This is the basis of many practical applications of encryption, including private communication and secure storage of data work. The security of such a system can be defined as how difficult it is to work out the "key" (or some part of the key) by studying the "lock". In general, computationally hard mathematical problems are used as the basis for encryption systems. For example, the widely-used Rivest–Shamir–Adleman (RSA) cryptosystem [3] is based on the difficulty of factorising numbers into prime factors. Breaking into, or exploiting, such encryption systems involves coming up with algorithms to solve these difficult problems and find keys - potentially even using new hardware such as quantum computers.

Unlike encryption, data anonymisation is currently used to protect privacy in scenarios where data needs to be shared or analysed en masse. Anonymisation means removing all personally identifiable information from data. This means not only removing names, phone numbers and addresses but also information which could be combined with other data to deduce the identity of the subject, such as location data, internet history or phone records. As depicted in Figure 4, anonymisation is often done before sharing data with a third party who may then aggregate it with other anonymised data before performing analysis – this is in contrast with the process of analysing non-anonymised data, as depicted in Figure 3. In the scenario of Figure 4, it should be noted that the anonymisation will prevent the third party from performing some kinds of analysis which may have been possible on the full data set. Also note that the results of the analysis are known to the third-party.

Homomorphic encryption incorporates features of data anonymisation into a new encryption system. As shown in Figure 5, a homomorphic encryption system creates a ciphertext which is still unintelligible to anyone without access to the decoding algorithm but which retains enough structure to be analysed in encrypted form. This means that if a third-party can offer superior data analysis but the data owner doesn't want them to have access to the original data, the owner can allow the third-party to perform the analysis on an encrypted copy of the data and return an encrypted answer, which only the owner can then decode.

Homomorphic encryption as a concept has existed since the 1980s but it was previously thought that only certain limited kinds of analysis of the encrypted data were possible. So-called "fully" homomorphic encryption (FHE), where any reasonable computation can be performed on encrypted data by the third party, was something of a holy grail in the field for decades. This grail was found in 2009 when Craig Gentry published the first scheme for FHE [4].

From Theory to Practice

Gentry's scheme was a major achievement in cryptographic research but in its original form the computational overhead for doing analysis on homomorphically encrypted data was too large for practical applications. Basic operations on encrypted data took between 30 seconds and an hour [5]. However, there have been further breakthroughs in the past 10 years which have narrowed this performance gap and fully homomorphic encryption has transformed from an unlikely theoretical breakthrough to a functional early-stage technology. There are now industrial implementations of the technology, such as Microsoft's Simple Encrypted Arithmetic Library (SEAL) [6]. Additionally, several start-ups have emerged offering

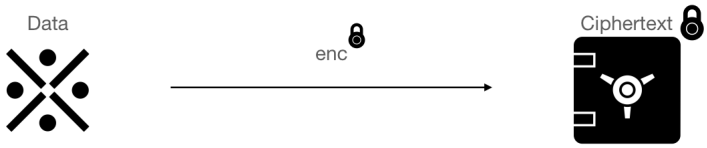

Figure 1: Data encryption.

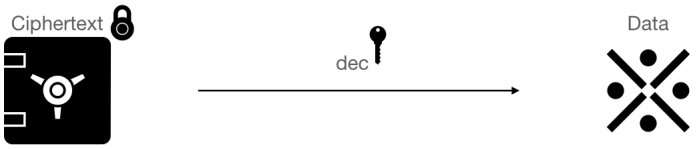

Figure 2: Data decoding.

Figure 3: Data analysis without data anonymisation.

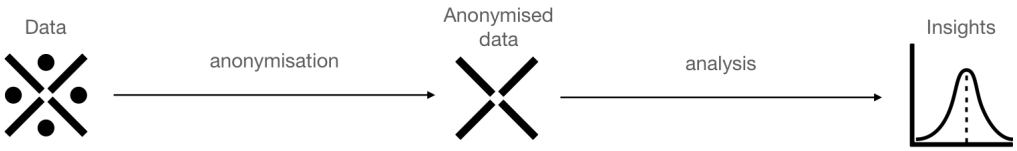

Figure 4: Data analysis with data anonymisation.

Figure 5: Data analysis using FHE.

FHE services in the financial sector [7]. The offerings of such companies look set to grow in technical sophistication and scale in the coming decade.

From Practice to Policy

The advent of FHE and its increased adoption in industry raises important policy challenges in the short and medium term.

In the medium term, homomorphic encryption could transform transatlantic data diplomacy. As applications of FHE mature and the range of data processing services that can be offered homomorphically increases, FHE will help to ease current tensions between the privacy-focused EU and the innovation-focused US by offering an alternative channel to move data between European data collectors and US data analysts without compromising privacy. Policymakers should look to promote this new technology in establishing a secure transatlantic trade in data and analysis services. There are also opportunities in the medium term for regions in Europe where the focus on privacy has limited innovation in machine learning to use FHE as a focus for a new indigenous secure tech sector which can compete with US-based tech monopolies.

However, the main short term issue for policymakers is classifying homomorphically encrypted data relative to existing data privacy regulations. Under current legislation such as the EU's GDPR and the California Consumer Protection Act (CCPA), sharing personal data with third parties is strictly controlled, limiting the kinds of analysis companies can commission on the data they collect. Policymakers will need to decide to what extent homomorphically encrypted data should be treated as personal data with respect to these regulations. On the one hand, as this data cannot be used to identify individuals, allowing sharing may enable valuable aggregation and analyses of data without compromising consumer privacy, particularly in fields where current regulation greatly restricts data sharing, e.g. in healthcare. On the other hand, as companies gain more freedom to share data in homomorphically encrypted form, consumer's encrypted data could be analysed by unknown third parties which are subject to less scrutiny than the original platform responsible for the data collection. For third-party companies offering FHE data analysis, there is a risk that their products could be used to provide analysis on unethically sourced data as they have no window into the content of the data. This may lead to complicated issues of liability for damage caused by any such analysis or machine learning performed on unethical data.

Uncertainty ahead

As with all emerging technologies, the roadmap of challenges laid out above is far from complete and far from certain. Further regulatory divergence between the US and EU, for example, could push homomorphic encryption into the spotlight sooner as an alternative method of trading data and analytics across the Atlantic. Stagnating processor speeds [8], on the other hand, could mean that FHE remains too slow for many applications. Either way governments need to keep their eyes on this area.

Acknowledgements

The Author acknowledges the helpful comments and input of both anonymous reviewers and the feedback of the judges of CUSPE's Horizon 2030 competition where an earlier version of this work won joint second prize.

References

[1] R. F. Fefer and K. Archick, "EU data protection rules and U.S. implications," https://crsreports.congress.gov/product/pdf/IF/IF10896, July 2020.

[2] H. A. Mildebrath, "The CJEU judgment in the Schrems ii case," https://www.europarl.europa.eu/

thinktank/en/document.html?reference=EPRS_ATA(2020)652073, September 2020.

[3] R. L. Rivest, A. Shamir, and L. Adleman, "A method for obtaining digital signatures and public-key cryptosystems," *Commun. ACM*, vol. 21, no. 2, p. 120–126, Feb. 1978. [Online]. Available: https://doi.org/10.1145/359340.359342

[4] C. Gentry, "Fully homomorphic encryption using ideal lattices," in *Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing*, ser. STOC '09. New York, NY, USA: Association for Computing Machinery, 2009, p. 169–178. [Online]. Available: https://doi.org/10.1145/1536414.1536440

[5] C. Gentry and S. Halevi, "Implementing gentry's fully-homomorphic encryption scheme," in *Advances in Cryptology – EUROCRYPT 2011*, K. G. Paterson, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 129–148.

[6] "Microsoft SEAL (release 3.6)," https://github.com/Microsoft/SEAL, Nov. 2020, microsoft Research, Redmond, WA.

[7] N. Maxwell, "Innovation and discussion paper: Case studies of the use of privacy preserving analysis to tackle financial crime," https://www.future-fis.com/uploads/3/7/9/4/3794525/ffis_innovation_and_discussion_paper_-_case_studies_of_the_use_of_privacy_preserving_analysis_-_v.1.3.pdf, Jan 2021.

[8] H. N. Khan, D. A. Hounshell, and E. R. H. Fuchs, "Science and research policy at the end of Moore's law," *Nature Electronics*, vol. 1, no. 1, pp. 14–21, Jan. 2018. [Online]. Available: https://doi.org/10.1038/s41928-017-0005-9

About the Author

Adam Ó Conghaile is a 3rd year PhD student in Theoretical Computer Science at the University of Cambridge's Computer Lab, funded by the EPSRC. He recently led the Open Climate Policy Database project with the Global Student Policy Alliance. His first paper "Game Comonads & Generalised Quantifiers" was published earlier this year at CSL 2021. He can be contacted at `ac891@cam.ac.uk`.

Conflict of interest The Author declares no conflict of interest.