MEDINFO 2021: One World, One Health – Global Partnership for Digital Innovation
P. Otero et al. (Eds.)

doi:10.3233/SHTI220069

Towards a Toolbox for Privacy-Preserving Computation on Health Data

Suhail Yazijy[a], Reto Schölly[a], Philipp Kellmeyer[a,b]

[a] *University Medical Center Freiburg, Department of Neurosurgery, Neuroethics and AI Ethics Lab*
Breisacher Str. 64, D-79106 Freiburg i. Br. (Germany)
[b] *Saltus! Group "Responsible Artificial Intelligence", Freiburg Institute for Advanced Studies (FRIAS), University of Freiburg*

Abstract

Substantial advances in methods of collecting and aggregating large amounts of biomedical data have been met with insufficient measures of protecting it from unwarranted access and use. Most of the current layers of protection are merely aimed at ensuring compliance with regulations (e.g., the EU's General Data Protection Regulation) but do not represent a vision of privacy-by-design as an efficient and ethical advantage in biomedical research and clinical applications. This not only slows down the pace of such efforts but also leaves the data exposed to a wide spectrum of cyberattacks. This work presents an overview of recent advancements in data and compuation security, along with a discussion of their limitations and potential for deployement in both health care and research settings.

Keywords:

Computer Security, Privacy, Confidentiality

Introduction

Storing health-related big data, be it DNA sequences, EEG signals or electronic health records, in centrally managed servers makes these systems vulnerable to an increasing variety of cyberattacks and other forms of data leakage. Although perfect informational security is unattainable, the attack surface could be reduced.

Encryption alone won't do it. A company with unregulated and unaudited access to encryption keys is probably using encryption for nothing more than marketing. Not to mention that there are weak encryptions and strong encryptions. Without a way to verify what kind of encryption is in place and how it is implemented, it can be safely dismissed as a data security solution.

In late 2020, the private Finnish firm Vastaamo was hit with a 'shocking' hack that affected thousands of psychotherapy records that were later used for ransom [1]. This a de facto attack on informational privacy, in this case highly sensitive data on mental health content, that could have been avoided by using better encryption.

In 2012, an attack vector was demonstrated by using visual stimuli and affordable EEG BCIs (e.g., Emotiv) to extract private information (e.g., banking, addresses, etc.)[12]. In 2018, another research group demonstrated a model for predicting human mental states (e.g., concentration) from EEG data[5].

Even well-protected data will be happily handed over by, or to, actors for whom informational security and privacy is not a priority. Google is notorious for such practices. Project Nightingale with Ascension involved a HIPAA-noncompliant and unconsented transfer of personal medical records of millions of Americans to Google Cloud [15]. DeepMind had a similar deal with the Royal Free NHS Trust, granting the AI giant, once again, unconsented access to sensitive healthcare data of more than a million Brits [8].

The repercussions are thus not only limited to unimaginable data misuse - although that's sufficient on its own - but also extend to exclusivity in use and access.

It is important to stress that the surface of benefits from health-related data is possibly as large as the surface of attacks on it; amongst many others, some studies demonstrated the use of Convolutional Neural Networks for early diagnosis of Alzheimer's Disease [9,13]. It is therefore not the aim here to advocate for less access but rather a more agile and secure one.

The heart of the matter is that if our legal and technical systems are standing in the way of true data protection and potentially impactful research efforts while facilitating the perpetual influence of powerful entities, they must be ripe for a technical alternative.

If we were able to enforce something like the EU General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA), it shouldn't take a miraculous efforts to enforce a reliable, privacy-preserving technology framework for storing and dealing with neuro and medical data, or any kind of sensitive data for that matter.

Methods

We conducted a systematic review of recent relevant literature and press releases on the topic of informational security and privacy.

Results

Our current health data safeguards are mostly convoluted legal hoops that both consumers and scientists have to jump through before getting any meaningful access, either to use a product or service (i.e., "inscrutable EULAs" [10] in the case of consumers) or to use data for research and analysis (i.e., HIPAA and IRB reviews in the case of scientists).

As mentioned, actors with big financial and lobbying power can maneuver through whatever hoops come their way, potentially monopolizing most of the impactful research and slowing it down overall as a result. Whether this is an unavoidable dynamic in nature and free markets (i.e., power law) or one that should be regulated is beyond the scope of this work.

The claim here is that it is possible to have strong technological safeguards that can remove a lot of the legal barriers, making general access to health data even more secure and attainable.

The following are some high-level definitions of such potential safeguards that are necessary to later examine how all the different pieces may fit together.

Data Anonymization

This is the process of simply removing personally identifiable information from a dataset so that any sensitive data cannot be linked to any specific data subject.

Differential Privacy (DP)

Differentially private algorithms are used for protecting individual private information when aggregating sensitive data by adding some noise to the dataset. Balancing how much noise should be added is a non-trivial task that is specific to the data type or the use case.

Federated Learning (FL)

The key principle behind federated learning is sending the computational model (e.g., neural network) to train itself wherever the data lives, instead of sending the data to where the model is.

Zero Knowledge Proofs (ZKP)

Zero Knowledge Proofs are essentially probabilistic proofs of possessing some information without revealing the information itself. A famous example of this is proving that the salary of a loan applicant is within a certain range without revealing the exact salary.

Homomorphic Encryption (HE)

HE allows for performing computations on encrypted data, without having to decrypt it. The computed results, once decrypted, are identical to those performed on the same data in an unencrypted form.

Secure Multiparty Computation (MPC)

Secure Multiparty Computations are done using cryptographic protocols that allow multiple parties to jointly compute the output of a function over their distributed inputs, without the need for a trusted third-party and without revealing information about what the function inputs are.

All the technologies presented so far, given their imperfections, seem to stack up perfectly in a Swiss cheese model [22] that allows privacy-preserving analysis on sensitive data.

Indeed, such a hybrid model has been previously put forward, for example, by Truex et al. [17]. It is still however a leaky one overall. Let's start with data anonymization.

Anonymized data is vulnerable to re-identification or linkage attacks with the help of other datasets. This was famously demonstrated on the anonymized Netflix Prize dataset that was re-identified using public data from IMDb [14].

While Federated Learning protects the raw data from being exposed altogether, it actually exposes the model's details and parameters. The problem here is two-fold. First, if the model is proprietary, its architecture should not be exposed. Second, such a setting would allow for model inversion attacks, where the model could be reverse engineered to leak the private data it was trained on. Zhu et al. demonstrated such an attack on publicly shared model gradients [21]. They suggested a few techniques to mitigate the attack (e.g., Differential Privacy with Gaussian and Laplacian noise). However, they all came at the cost of sacrificing the model's accuracy.

FL is also prone to data poisoning and backdoor attacks. Tolpegin et al. demonstrated how malicious participants can send partial model updates derived from mislabeled data, causing significant negative impact on the performance and accuracy of the global model [16]. Bagdasaryan et al. demonstrated how any FL participant can introduce a hidden backdoor to poison and manipulate the joint global model [2].

Lastly, similar to MPC and HE, FL generally requires a large computation and communication overhead [19].

The problems identified above could be mitigated with a combination of the following techniques.

Split Learning (SplitNN)

This approach, pioneered by Gupta and Vepakomma et al. [7,18], is a method that allows each FL data owner (e.g. a hospital) to train the model or neural network up to a certain layer (i.e., "cut layer"). The analysts can then pick up the outputs of this training to train the rest of the network on their side. After that, they backpropagate gradients until the cut layer, at which time the gradients are sent to the hospitals where the rest of the backpropagation, and thus the training process, is completed.

Split Learning has proven to be effective in reducing computation and communication costs as well as protecting the model details (i.e., architecture and weights), without sacrificing the model accuracy. However, it still requires a relatively large communication bandwidth when the training is done within a smaller network of data owners [19]. Something that could be addressed with advanced methods for neural network compression [11].

It is not clear however if SplitNN could still overcome some of the general FL attacks, namely model inversion, poisoning and backdoors.

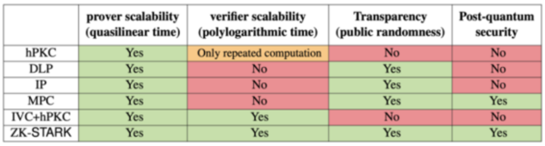
Federated Learning of Cohorts (FLoC)

Although this is a new technique currently pushed for by Google as a replacement for web advertisement cookies, it would be interesting to bring it to health and brain data. Especially since most of its downsides seem to apply only in a web browser context [4].

The core concept of FLoC is to introduce longitudinal privacy, where users are no longer tracked individually but rather in large groups (i.e., cohorts) instead. Cohort membership changes over time with changes in browsing behavior [20].

zk-STARK

This is a scalable, transparent, and post-quantum secure Zero Knowledge system introduced by Ben-Sasson et al. in 2018 [3]. It overcomes a lot of the inefficiencies of previous Zero Knowledge proofs (e.g., zk-SNARK) and could be, at least theoretically, a superior alternative to other Zero Knowledge (ZK) systems.



	prover scalability (quasilinear time)	verifier scalability (polylogarithmic time)	Transparency (public randomness)	Post-quantum security
hPKC	Yes	Only repeated computation	No	No
DLP	Yes	No	Yes	No
IP	Yes	No	Yes	No
MPC	Yes	No	Yes	Yes
IVC+hPKC	Yes	Yes	No	No
ZK-STARK	Yes	Yes	Yes	Yes

Figure 1 – Theoretical comparison of universal realized ZK systems [3]

hPKC = Homomorphic public-key cryptography
DLP = Discretelogarithmproblem
IP = Interactive Proofs
MPC = Secure multi-party computation
IVC = Incrementally Verifiable Computation

This is another system that hasn't been explored in a medical or health-realted data context yet. It could be particularly powerful in overcoming the aforementioned FL attacks (e.g., inversion, poisoning, etc.) as it allows for secure, verifiable and tamper-proof computation.

Discussion

A privacy-preserving Swiss army knife, albeit so far theoretical, would consist of the following stack of technologies and characteristics:

- Open Source - vetted by the community
- zk-STARKs
- Federated Split Learning (if feasible)
- Differential Privacy (if feasible)

While this could be applied to any kind of data, it could bring tremendous privacy and accessibility benefits to biomedical or health data. With this set of techniques, most of the GDPR and HIPAA requirements pertaining to personally identifiable information (PII) and sensitive data can be easily and ethically met.

Conclusions

The privacy-preserving technologies presented here are still immature in many respects and will witness continuous improvements over time. This is a beam of hope for potential solutions to dealing with advances in data extraction and analysis.

It will be important however to investigate whether the tradeoff between privacy and utility can be completely avoided. That is, whether we will always have to substantially sacrifice privacy in order to keep data accessible enough for research and analysis, or we will be able to have strong "privacy-by-default"[10] without standing in the way of health research.

It is important to highlight that even with the presented technology stack, encryption keys are still prone to loss or leakage. This could be potentially mitigated via a Multisignature (multisig) scheme.

Another puzzle that remains unresolved at this point is how do we protect and consent data about other people who are cross referenced within one person's health data (e.g., in psychotherapy records).

Acknowledgements

The research for this paper was performed in the project "Data Access and Data Use in Medical Institutional and Consumer Health Settings. An Ethical, Legal, and Social Analysis [DaDuHealth]" funded by the German Ministry for Education and Research (BMBF).

References

[1] AFP in Helsinki, "Shocking" hack of psychotherapy records in Finland affects thousands, *The Guardian*. (2020). https://www.theguardian.com/world/2020/oct/26/tens-of-thousands-psychotherapy-records-hacked-in-finland.

[2] E. Bagdasaryan, A. Veit, Y. Hua, D. Estrin, and V. Shmatikov, How To Backdoor Federated Learning, *ArXiv:1807.00459 [Cs]*. (2019). http://arxiv.org/abs/1807.00459 (accessed April 2, 2021).

[3] E. Ben-Sasson, I. Bentov, Y. Horesh, and M. Riabzev, Scalable, transparent, and post-quantum secure computational integrity, *IACR Cryptol. EPrint Arch.* **2018** (2018) 46.

[4] B. Cyphers, Google's FLoC Is a Terrible Idea, *Electronic Frontier Foundation (EFF)*. (2021). https://www.eff.org/de/deeplinks/2021/03/googles-floc-terrible-idea.

[5] D.R. Edla, K. Mangalorekar, G. Dhavalikar, and S. Dodia, Classification of EEG data for human mental state analysis using Random Forest Classifier, *Procedia Computer Science*. **132** (2018) 1523–1532. doi:10.1016/j.procs.2018.05.116.

[6] Futuro 360, Unanimously: [Chilean] Senate approves regulation of NeuroRights, *Columbia NeuroRights Initiative*. (2020). https://nri.ntc.columbia.edu/news/unanimously-chilean-senate-approves-regulation-neurorights.

[7] O. Gupta, and R. Raskar, Distributed learning of deep neural network over multiple agents, *ArXiv:1810.06060 [Cs, Stat]*. (2018). http://arxiv.org/abs/1810.06060 (accessed April 2, 2021).

[8] H. Hodson, Revealed: Google AI has access to huge haul of NHS patient data, *New Scientist*. (2016). https://www.newscientist.com/article/2086454-revealed-google-ai-has-access-to-huge-haul-of-nhs-patient-data/.

[9] A. Karwath, M. Hubrich, S. Kramer, and The Alzheimer's Disease Neuroimaging Initiative, Convolutional Neural Networks for the Identification of Regions of Interest in PET Scans: A Study of Representation Learning for Diagnosing Alzheimer's Disease, in: A. ten Teije, C. Popow, J.H. Holmes, and L. Sacchi (Eds.), Artificial Intelligence in Medicine, Springer International Publishing, Cham, 2017: pp. 316–321. doi:10.1007/978-3-319-59758-4_36.

[10] P. Kellmeyer, Big Brain Data: On the Responsible Use of Brain Data from Clinical and Consumer-Directed Neurotechnological Devices, *Neuroethics*. (2018). doi:10.1007/s12152-018-9371-x.

[11] Y. Lin, S. Han, H. Mao, Y. Wang, and W.J. Dally, Deep Gradient Compression: Reducing the Communication Bandwidth for Distributed Training, *ArXiv:1712.01887 [Cs, Stat]*. (2020). http://arxiv.org/abs/1712.01887 (accessed April 2, 2021).

[12] I. Martinovic, D. Davies, M. Frank, D. Perito, T. Ros, and D. Song, On the Feasibility of Side-Channel Attacks with Brain-Computer Interfaces, in: 21st USENIX Security Symposium (USENIX Security 12), USENIX

Association, Bellevue, WA, 2012: pp. 143–158. https://www.usenix.org/conference/usenixsecurity12/technical-sessions/presentation/martinovic.

[13] E. Moradi, A. Pepe, C. Gaser, H. Huttunen, and J. Tohka, Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects, *NeuroImage*. **104** (2015) 398–412. doi:10.1016/j.neuroimage.2014.10.002.

[14] A. Narayanan, and V. Shmatikov, Robust De-anonymization of Large Sparse Datasets, in: 2008 IEEE Symposium on Security and Privacy (Sp 2008), IEEE, Oakland, CA, USA, 2008: pp. 111–125. doi:10.1109/SP.2008.33.

[15] E. Pilkington, Google's secret cache of medical data includes names and full details of millions – whistleblower, *The Guardian*. (2020). https://www.theguardian.com/technology/2019/nov/12/google-medical-data-project-nightingale-secret-transfer-us-health-information.

[16] V. Tolpegin, S. Truex, M.E. Gursoy, and L. Liu, Data Poisoning Attacks Against Federated Learning Systems, in: L. Chen, N. Li, K. Liang, and S. Schneider (Eds.), Computer Security – ESORICS 2020, Springer International Publishing, Cham, 2020: pp. 480–501. doi:10.1007/978-3-030-58951-6_24.

[17] S. Truex, N. Baracaldo, A. Anwar, T. Steinke, H. Ludwig, R. Zhang, and Y. Zhou, A Hybrid Approach to Privacy-Preserving Federated Learning, in: Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security - AISec'19, ACM Press, London, United Kingdom, 2019: pp. 1–11. doi:10.1145/3338501.3357370.

[18] P. Vepakomma, O. Gupta, T. Swedish, and R. Raskar, Split learning for health: Distributed deep learning without sharing raw patient data, *ArXiv:1812.00564 [Cs, Stat]*. (2018). http://arxiv.org/abs/1812.00564 (accessed April 2, 2021).

[19] P. Vepakomma, T. Swedish, R. Raskar, O. Gupta, and A. Dubey, No Peek: A Survey of private distributed deep learning, *ArXiv:1812.03288 [Cs, Stat]*. (2018). http://arxiv.org/abs/1812.03288 (accessed March 31, 2021).

[20] Y. Xiao, and J. Karlin, Federated Learning of Cohorts (FLoC), *Federated Learning of Cohorts W3C Specification*. (n.d.). https://wicg.github.io/floc/.

[21] L. Zhu, Z. Liu, and S. Han, Deep Leakage from Gradients, *ArXiv:1906.08935 [Cs, Stat]*. (2019). http://arxiv.org/abs/1906.08935 (accessed April 1, 2021).

[22] Swiss cheese model, *Wikipedia*. (n.d.). https://en.wikipedia.org/wiki/Swiss_cheese_model.

Address for correspondence

Dr. med. Philipp Kellmeyer
Department of Neurosurgery
University Medical Center Freiburg
c/o FRIAS
Alberstr. 19, 79104 Freiburg
Germany
Phone:+49-(0)761-203-97446
Email: philipp.kellmeyer@uniklinik-freiburg.de