Healthcare Informatics and Privacy

Arun Iyengar
IBM T. J. Watson Research Center

Ashish Kundu
IBM T. J. Watson Research Center

George Pallis
University of Cyprus

The digital transformation of healthcare ecosystems on the Internet has been rapid and explosive. While web-based and IoT-driven ecosystems promise a future for universally accessible and more intelligent healthcare, the privacy of patients, doctors, nurses, and healthcare providers is of greater concern today than ever. Regulatory requirements are evolving (such as EU's DPR to GDPR) to address such privacy requirements. This special issue addresses two important topics involved in healthcare informatics and privacy.

The healthcare and life sciences industries have relied on computing and information technology since the 1950s.[1] In the last decade or so, with the advent of mobile computing, cloud platforms as well as machine learning and analytics, the digital transformation of healthcare on the Internet has been rapid and explosive. The Internet has enabled the transfer and communication of healthcare data as well as the delivery of healthcare services and apps connecting patients and healthcare providers. Digital healthcare ecosystems have evolved on top of cloud platforms, mobile computing, and the Internet of Things (IoT), as well as wearable devices distributed across geopolitical and socioeconomic boundaries. While such ecosystems promise a future for universally accessible and more intelligent healthcare, the privacy of patients, doctors, nurses, and healthcare providers is of greater concern today than ever.

Regulatory compliance requirements, such as the US Health Insurance Portability and Accountability Act (HIPAA) and EU General Data Privacy Regulation (GDPR), support the protection of privacy at various levels. Healthcare data breaches not only can cause significant adverse personal and social impacts of patients and their families, but also incur a huge cost: $6.2 Billion USD.[2] Organizations maintain the privacy of the medical status of their C-suite as closely guarded secrets, to protect themselves from adverse reactions from Wall Street and investors.

Protection and management of healthcare privacy has several challenges and open problems; for example, how does the application of AI and machine learning for healthcare analytics affect privacy? How can anonymization of genomic and healthcare data be carried out to preserve utility while protecting privacy? How can we protect healthcare devices from leaking sensitive

Published by the IEEE Computer Society
1089-7801/18/$33.00 USD

healthcare data? How do we develop the next generation of privacy policies and regulatory compliance regimes?

Protected health information (PHI) is often used to refer to any data that can be used to identify a patient. However, HIPAA's current definition of PHI does not include genomic data, which can be used to identify individuals with a great degree of success.[3] Neither does GDPR address such a privacy requirement.[4] Personally identifiable information (PII) also does not have a standards body that identifies genomic data as sensitive personal information.[5] This is an instance where computer science is ahead of industry and regulatory requirements in terms of privacy issues and challenges in HCLS (healthcare and life sciences). Consequently, anonymization of genomic graphs and data is not yet practically used. The anonymization and de-anonymization of PHI/PII is well-addressed in regulatory compliance requirements, both in practice and in the scientific community, but challenges exist regarding the utility of data and level of anonymization (for example, different privacy techniques preserve privacy but may lead to loss of desired utility of the data).

The security of healthcare computing systems involves identifying how to protect healthcare data, processes, and systems from being attacked by different channels. Lack of encryption and associated best practices for encryption schemes and key management have allowed attackers to gain access to millions of data records. Homomorphic encryption and its variants might provide a solution but they are far from being practical at this point. If appropriate cryptographic and security measures are not taken, integrity verification can lead to privacy leakages. When a query result on an electronic medical record comprised of parts of the record is authenticated against a digital signature computed from the complete record, it might leak information about the remaining parts of the record.

Amazon AWS, IBM Watson Health, Microsoft Azure, and so on offer healthcare cloud services across the industry. But the security challenges of multi-tenant environments such as the cloud include computing isolation guarantees, identity and access control management, SIEM analytics, data provenance, and protection from multiple types of insider threats—cloud provider admins, cloud stack managers and solution admins. Use of best practices in security hardening still has a long way to go (e.g., the Equifax data breach and the Apple root login with no password on MacOS). IoT devices used for healthcare provide better functional capabilities but are vulnerable to security attacks. Blockchains are being used for healthcare data management and audits. However, the weakest link is not in defining security capabilities, but in misconfiguration as well as unreliable and negligent software engineering.

IN THIS ISSUE

This special issue addresses two important topics involved in healthcare informatics and privacy. The protection of patients' privacy while linking their medical records is essential for better healthcare.

Patient health data is often distributed across different systems and regulatory domains. However, precision and personalized healthcare requires access to medical records in a holistic way. It is essential to enable the linkage of disconnected health records. In the first article, "Perfectly Secure and Efficient Two-party Electronic Health Record Linkage," Feng Chen, Xiaoqian Jiang, Shuang Wang, Lisa M. Schilling, Daniella Meeker, Toan Ong, Michael E. Matheny, Jason N. Doctor, Lucila Ohno-Machado, Jaideep Vaidya propose a technique for healthcare record linkage using garbled circuits, which is privacy-preserving while also offering better utility of data.

Analytics of healthcare data from IoT devices and the cloud have the potential to offer real-time and more accurate and efficient healthcare solutions. In "Towards Practical Privacy-Preserving Analytics for IoT and Cloud Based Healthcare Systems," Sagar Sharma, Keke Chen, and Amit Sheth discuss several research challenges that could allow the realization of the full potential of IoT devices in personalized medicine and real-time healthcare.

We hope that readers will find these articles interesting and informative. We would like to express our gratitude to all of the authors and reviewers for their contributions to this special issue.

REFERENCES

1. J.P. Sewell and L.Q. Thede, "Computer Development and Health Care Information Systems 1950 to Present," *Informatics and Nursing: Opportunities and Challenges*, LWW, 2015; http://dlthede.net/informatics/chap01introni/healthcare_computers.html.
2. *Sixth Annual Benchmark Study on Privacy & Security of Healthcare Data*, Ponemon Institute, 2016; https://www.ponemon.org/library/sixth-annual-benchmark-study-on-privacy-security-of-healthcare-data-1.
3. M. Gymrek et al., "Identifying personal genomes by surname inference," *Science*, vol. 339, no. 6117, 2013, pp. 321–324.
4. G. Chassan, "The Impact of the EU General Data Protection Regulation on Scientific Research," *ecancermedicalscience*, vol. 11, no. 709, 2017; doi.org/https://doi.org/10.3332/ecancer.2017.709.
5. P.M. Schwartz and D.J. Solove, "The PII Problem: Privacy and a New Concept of Personally Identifiable Information," *New York University Law Review*, vol. 86, no. 6, 2011, p. 1814.

ABOUT THE AUTHORS

Arun Iyengar does research and development into distributed computing, Web and cloud performance, and artificial intelligence at IBM's T.J. Watson Research Center in Yorktown Heights, NY. His techniques for caching, load balancing, and serving dynamic content are widely used for Web and distributed applications. Iyengar received a PhD in computer science from the Massachusetts Institute of Technology. He is a fellow of the IEEE. Contact him at aruni@us.ibm.com.

Ashish Kundu is a master inventor and research scientist in security research at the IBM T. J. Watson Research Center, Yorktown Heights, NY. His research interests include security, privacy, compliance and AI ethics. Kundu's long-term research vision is: "How to weave security, privacy, compliance, and ethics requirements with the functionality." Kundu is currently associate editor for *IEEE Transactions on Dependable and Secure Computing*. His work has resulted in several publications, more than 100 patents filed, and 55 or more patents granted so far. Kundu is an ACM Distinguished Member, an ACM Distinguished Speaker and an IEEE Senior Member. Contact him at akundu@us.ibm.com.

George Pallis is an assistant professor of computer science at the University of Cyprus, Nicosia. His research interests include distributed systems, cloud computing and big data analytics. Pallis has a PhD in computer science from the Aristotle University of Thessaloniki. Contact him at gpallis@cs.ucy.ac.cy.