AGENDA: REPORTING ON ETHICAL CONSIDERATIONS – RADHIKA IYENGAR-EMENS

Privacy Considerations

Widespread and prolific Internet connectivity as well as the use of Apps means that more Americans are continually generating a high volume of personal data (Congressional Research Office R45631). Today, private companies can and do leverage this data to offer personalized capabilities to users. In fact, highly lucrative business models depend entirely on the value of consumer data, but the inherent trade-off is that consumer data is mined and sold. While consumers get the convenience of online services, their data and online behavior is harvested, analyzed, mined and sold often without consent and certainly without disclosure. With these practices, there has also been a large gray area around consent. We are at a crossroads of privacy, security, law, and data sovereignty. There are ethical and privacy considerations of how personal "data is collected, protected and used." (US Government Accountability Office (GAO) 19-52, "Internet Privacy").

Privacy has become a global concern, reverberating around breaches of data and privacy that seem to dominate the news. Egregious examples of privacy breaches include the Cambridge Analytica scandal: the world woke up to the devastating exposure and misuse of personal data without consent. "The Facebook–Cambridge Analytica data scandal was a major political scandal in early 2018 when it was revealed that Cambridge Analytica had harvested the personal data of millions of people's Facebook profiles without their consent and used it for political advertising purposes. It has been described as a watershed moment in the public understanding of personal data and precipitated … calls for tighter regulation of tech companies' use of personal data." (Wikipedia)

In the aftermath of Cambridge Analytica, consumers and policymakers grappled with the understanding that private companies could provide third party access to data without adequate disclosure and proper user consent. Europe was already ahead of the game in adopting General Data Protection Regulation (GDPR) in 2012, but it became enforceable in May 2018. GDPR guidelines address data protection and privacy for all forms of personal data in the European Union (EU) and the European Economic Areas (EEA). Among other principles, GDPR brings the "right to be forgotten" into the mainstream lexicon – an individual can download their personal data from a company and then request the company to delete their personal data from the company's servers. Furthermore, GDPR offers data protection and privacy to transfers of personal data outside the EU and EEA.

In the US, privacy has been historically shaped around constitutional protections for government intrusion on personal privacy. However, there is no uniformity in the law at the federal level as the laws only protect certain categories of data or in certain industries, and "no single federal law comprehensively regulates the collection and use of personal data." (Congressional Research Office R45631). For example, there is privacy and data protection for health data through HIPAA (Health Insurance Portability and Accountability Act). Certainly, the Cambridge Analytica scandal and other similar large-scale privacy breaches have underscored the importance of data collection, data protection and explicit consent for the intended use of personal data.

California is leading the way with the California Consumer Privacy Act (CCPA), offering California residents privacy protection by requiring entities to declare intent to collect personal data and the intended use of that data as well as requiring users to opt-in consent for the collection and use of their personal data. California enacted the CCPA in large part to overcome the gaps in federal law, but there are questions that must be considered:

- Will this influence other states to follow suit?
- On a broader scope, should privacy be mandated at the federal level, and is this overarching approach appropriate?
- On the other hand, if states each have their own laws, are there implications from the lack of cohesion among the states?

Personal data also comes into different data categories, such as health or financial data. In fact, there is a collection of agencies and regulations that deal with only certain types of data, for example HIPAA or FCRA. Each type of data can come with specific compliance requirements.

With the compliance frameworks around personal data, we have to consider the consequences of non-compliance or transgression. What are the current implications of security breaches for entities that store personal data and who oversees security vulnerabilities? What are the obligations at a federal or state level towards breaches and what are the consequences of non-compliance? Currently, there is a required disclosure that a breach has occurred, but the recourse for the transgression is limited. For example, GDPR fines non-compliance in certain cases such as for data breaches with fines of 20 million Euros or 4% of annual revenues, whichever is greater. CCPA does mirror GDPR in the need for disclosure, but the fines are capped. Are these fines enough to deter transgressions? What non-compliance consequences can deter repeated transgressions without becoming overreaching and burdensome?

Solving Privacy Issues with Blockchain

Blockchain technology can effectively mitigate pervasive privacy issues. There are different types of blockchains, both permissionless (sometimes called public blockchains) and permissioned (sometimes called private blockchains).

In permissionless blockchains, the goal is full transparency. The nodes are pseudonymous (known by an alias), which means that anyone with sufficient hashing or computing power can be a node. All the data on the blockchain is transparent which means any node has visibility to the records on the blockchain. Privacy can be ensured cryptographically by mechanisms such as zero knowledge proof (ZKP) which authenticate identity without revealing the actual identity. Permissionless blockchains are generally the preferred blockchains for consumer-facing applications. To preserve data privacy, the big rule of thumb is to never store personal data on the blockchain, rather only a hash of the data.

In permissioned chains, the network nodes are all trusted nodes which means every node has to be vetted and approved. The network nodes agree to share certain data, and that data is visible to the entire network. However, permissioned blockchains also permit network nodes to keep certain information private and confidential across ecosystems involving competitors. Confidential data, such as pricing data or trade secrets, remains private, making permissioned chains the preferred blockchains for enterprises. Governance is also something enterprises care deeply about. Accountability is of great importance, and as the nodes are known, it is easier to investigate appropriately when things go wrong. Malicious actors can be identified, and where appropriate, ousted from the network.

Privacy and Data Sovereignty

No discussion of privacy can occur without a discussion of data sovereignty (self-sovereign identity or SSI and decentralized identifiers or DID). Proponents of data sovereignty declare that individuals not only need access to their records or data, but that individuals should have the control and the right to selectively disclose the data with entities they are transacting with. “SSI is the concept that individuals and entities should own and control their identity and data, independent of a central authority. DIDs are global unique identifiers and provide a [decentralized] foundation for individual identity. DIDs make identity portable, private and persistent.” The key considerations of data sovereignty are addressed in the digital identity section.

In the context of state legislation, both privacy and security are important, but the two go hand-in-hand with data sovereignty. Secure access mechanisms like private cryptographic key management processes are the key to ownership of online digital identities. Specifically, private and corresponding public cryptographic keys enable authentication and a digital identity reference. The key considerations of these processes are discussed further in the digital identity section.

The Right to be Forgotten: Concerns Around Blockchain Regarding Immutability

Some blockchain critics voice concern around blockchain’s immutability, arguing that immutability defeats the right to be forgotten. How does an individual who requests to be forgotten accomplish that with an immutable blockchain? The solution lies in the fact that personal data should not be stored on the blockchain, rather only a hash of the record. Furthermore, the personal data should be stored offline in a decentralized manner, otherwise a centralized data store will be vulnerable to cyberthreats and attacks. Any intent to be forgotten means that the offline personal data must be deleted **and** the key to the hash must also be deleted. The hash that remains is itself immutable, but it is inconsequential as it cannot be decrypted and so does not contain any personal data.

Sustainability Considerations of Blockchains

Sustainability is another key consideration for blockchain, and many blockchain critics point to sustainability being an Achilles heel. Sustainability concerns actually do not apply to all blockchains. Permissionless blockchains are different from permissioned blockchains. Major permissionless chains like Bitcoin and Ethereum are based on Proof of Work (PoW) consensus so rely on massive hashing or compute power to establish trust. The high energy consumption for PoW compute power makes these major permissionless blockchains less attractive from an environmental sustainability perspective. Other consensus models for permissionless chains are being touted as being more sustainable, such as Proof of Stake (PoS) consensus models, but PoS is currently experimental and is not yet mature enough for deployment.

By contrast, the major permissioned chains such as Hyperledger Fabric and R3 Corda do not rely on PoW, rather they rely on other consensus models such as Byzantine Fault Tolerance (BFT) or Crash Fault Tolerance (CFT) consensus. These consensus mechanisms are more sustainable as there is no reliance on massive hashing power to establish trust. Trust is established by vetted network nodes with KYC/AML. As a result, the BFT/CFT consensus protocols are quite lightweight and more performant. Enterprises go through many lengths to avoid uncertainty, so enterprise permissioned systems are designed to base trust on extensive verification and authentication protocols and processes. Therefore, the concerns that many voice about sustainability do not apply to the major permissioned chains.

Conclusion

Privacy is an important issue not just for California but also for the US. There is a need for GDPR-like comprehensive privacy law for the US. CCPA is a great step forward in the right direction, but there is an opportunity to clarify the consequences for non-compliance to deter transgressions. With data being generated at increasing volumes and at an accelerated pace, it is important to develop and uphold strong measures to protect our data, the use of our personal data as well as access to data by third parties. In this last point, we will need to strengthen our stance to hold private companies accountable for transgressions.

Sustainability is an increasingly important issue with any technology initiatives going forward. With the heightened attention on climate change and the climate crisis, we should be mindful of minimizing negative impact on the environment and to manage the state’s resources carefully. Therefore, in selecting blockchain systems, sustainability will be a very important consideration.

Recommendation

As this section discusses some of the considerations for blockchain technology, the recommendation is for members of the CA State Legislature to incorporate these considerations when evaluating any blockchain implementation, ensuring that important privacy and sustainability measures are built-into the implementation roadmap.

Bibliography

1. Congressional Research Service RS45631
   https://fas.org/sgp/crs/misc/R45631.pdf
2. United States Government Accountability Office
   https://www.gao.gov/assets/700/696437.pdf