Data Management for Future Wireless Networks: Architecture, Privacy Preservation, and Regulation

Xuemin (Sherman) Shen, Cheng Huang, Dongxiao Liu, Liang Xue, Weihua Zhuang, Sheng (Robert) Sun, and Bidi Ying

***Abstract*—Next-generation wireless networks (NGWN) aim to support diversified smart applications which require frequent data exchanges and collaborative data processing among multiple stakeholders. Data management (DM), including data collection, storage, sharing, and computation, plays an essential role in empowering NGWN. However, DM for NGWN faces two significant challenges: (1) stakeholders' data cannot be easily managed across different trust domains under a distributed network architecture; and (2) privacy preservation requirements of personal data become more rigorous under new privacy regulations. To explore possible solutions to address the challenges, we first investigate the state-of-the-art architecture designs for DM and emphasize advantages of a blockchain-based DM architecture. Then, we summarize existing privacy-preserving techniques in terms of advantages and challenges when being applied to DM. In addition, we review recent privacy regulation with their impacts on DM and discuss the existing solutions with privacy regulation compliance based on blockchain. Finally, we identify further research directions for achieving DM with privacy preservation.**

***Index Terms*—Blockchain, next generation wireless networks, privacy regulation, data management**

I. Introduction

With the rapid advancement of wireless communication networks, smart devices can be connected through reliable and seamless wireless connections. It is revolutionizing our daily lives by providing various smart applications, such as smart transportation systems and e-healthcare systems [1]. As the wireless networks continue to evolve, the next-generation wireless networks (NGWN) will further unleash the power of connectivity with a virtualized network architecture and the assistance of artificial intelligence (AI) [2]. To achieve the full potential of NGWN, a key enabler is the wealth of user data. First, massive connected smart devices can lead to the generation of user data at an unprecedented rate. Second, the evolving computing and storage infrastructure at the network edges makes the data collection and pre-processing more convenient. Third, the AI-based data processing poses high requirements on the volume, dimension, and quality of collected user data for accurate training and evaluation of AI models. Therefore, data management (DM) will play a critical role in NGWN, in terms of data collection, data storage, data sharing, and data computation [3].

DM faces significant technical challenges in NGWN [4]. First, multiple data stakeholders from different industrial sectors, such as mobile operators, technology vendors, data centers, and application providers, need to collaboratively manage the lifecycle of user data. For instance, user data can be generated at smart home appliances provided by technology vendors, transmitted through NGWN, stored and processed at data centers, and finally utilized by application providers for marketing analysis and product developments. The complex DM process requires frequent data exchanges and distributed data processing among data stakeholders with dynamic degrees of mutual trust. As a result, a reliable and trustworthy architecture for DM is required. Second, data privacy regulations are taking effect and significantly reshaping the privacy landscapes of NGWN. In particular, the European General Data Protection Regulation (GDPR) [5] defines legal requirements on DM of personal user data from different aspects: 1) it grants users a wide range of legal rights to obtain information and control operations on their personal data; 2) it requires "restricted processing" over personal data, where a set of privacy-preserving techniques can be adopted to enhance user identity privacy and data confidentiality; and 3) it requires privacy compliance for data lifecycle events to enforce obligations of data stakeholders [6]. Any data stakeholder failing to comply with the GDPR requirements on DM may face severe financial and legal consequences. Without proper solutions to DM with privacy preservation under the GDPR, there will be significant data barriers for data stakeholders in NGWN.

This survey article aims at providing a comprehensive understanding of DM in NGWN under privacy regulations. In Section II, we discuss architecture specifications for DM by investigating various existing solutions based on the cloud computing, fog computing, and blockchain. By comparisons, we emphasize the advantages of blockchain-based DM architectures. In Section III, we summarize a wide range of traditional privacy-preserving techniques and discuss the use cases and challenges when applying them to DM. In Section IV, we present state-of-the-art blockchain-based solutions to DM under the GDPR. In Section V, we present research challenges on achieving DM in NGWN under the GDPR and discuss potential solutions. Finally, we conclude this study and discuss further research directions.

This work was supported by research grants from Huawei Technologies Canada and from the Natural Sciences and Engineering Research Council (NSERC) of Canada.

X. Shen, C. Huang, D. Liu (corresponding author), L. Xue, and W. Zhuang are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ontario, Canada.

R. Sun and B. Ying are with Huawei Technologies Canada, Ottawa, Ontario, Canada.

II. Data Management Architecture: from Centralization to Decentralization

In this section, the existing architectures for DM are discussed in terms of centralized and decentralized architectures.

A. Centralized Architecture: Fog to Cloud

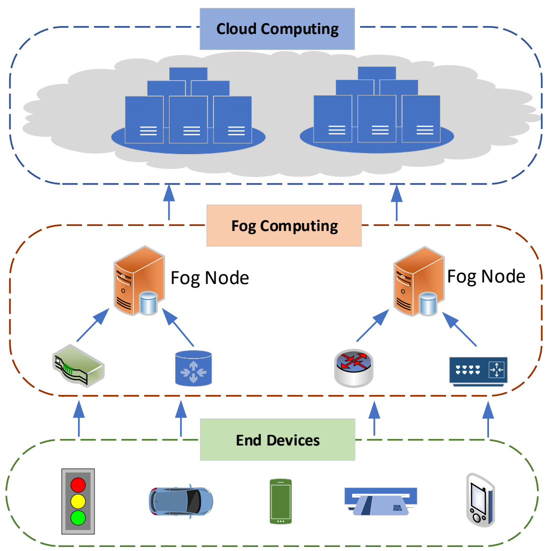

Fig. 1: Fog-to-cloud Data Management Architecture

One of the most popular DM architectures is based on the cloud computing [7], where applications are deployed on virtual machines in cloud servers to store and process their data in centralized databases. This architecture has brought many advantages, such as cost-effective storage and efficient data analytics, since cloud servers have massive computational and communication resources. But the cloud-based architecture may not be suitable for the complicated DM in NGWN, since a centralized solution may suffer from various attacks, such as the single point of failure and remote hijacking attacks, which can cause unexpected data leakage. At the same time, the concept of fog computing [8] is introduced and integrated with a cloud-based architecture to meet the DM requirements of location awareness, low latency, and real-time data processing. A fog-to-cloud DM architecture [8] is illustrated in Figure 1. In the architecture, data are collected from distributed sources including vehicles, sensors and computers, to be processed and temporarily stored in the fog layer. After being pre-processed, the data are uploaded to a cloud platform for data sharing and analysis. Compared with cloud-based architectures, this fog-to-cloud DM architecture is more hierarchical and provides more flexibility. At the same time, data privacy policies specified by the data owners can be enforced in fog computing to achieve fine-grained data access control. However, as fog nodes are usually restricted in storage and processing capabilities and the provided functionality can be highly vendor-dependent, they may pose limitations on DM architecture design.

The centralized architecture is widely considered in many research works for DM under a common assumption that the cloud and fog nodes are honest in data storage and processing. Nevertheless, in reality, they may be subject to potential security breaches and may misuse user data for self-interests without user awareness. Under this circumstance, not only the sensitive user data may be exposed, but also the functionalities of the deployed services are affected. Hence, to improve the security and robustness of DM, decentralized architectures are proposed in recent years.

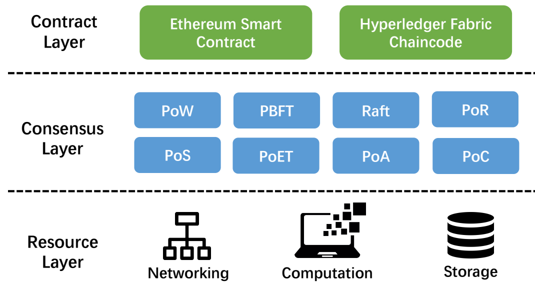

Fig. 2: Blockchain Layers and Components

B. Decentralized Architecture: Blockchain

In comparing with the centralized DM architectures, a decentralized architecture can mitigate the reliance of a single trusted entity and is a preferred approach to DM. In particular, blockchain is a promising distributed architecture that mainly consists of three layers: resource layer, consensus layer, and contract layer, as shown in Figure 2. The resource layer defines three basic components in the blockchain. Networking components can facilitate Peer-to-peer (P2P) communication among distributed nodes. Computation components allow the distributed nodes to perform necessary computation operations, such as data hash and signature. Storage components are vital for storing transaction data of the blockchain. The consensus layer includes different consensus protocols, such as proof of work (PoW) and proof of stake (PoS) to provide different security and scalability guarantees. The layer relies on the networking components for distributed nodes to communicate and maintain a consistent view of the blockchain. Many cryptographic computations can also be implemented by consensus protocols with the computation components. Upon the consensus layer, smart contracts and chaincodes [9] are deployed to support various functions and applications for blockchain users.

The blockchain architecture has many desired features including distribution, transparency, immutability, traceability, programmability, and automated verification. These features bring clear advantages for DM in NGWN:

- *Distribution and transparency*: When applying the blockchain architecture to DM, the conventional single point of trust can be avoided. More data stakeholders in NGWN are able to participate in data collection, storage, and processing through different consensus protocols in the consensus layer. In this way, DM becomes transparent across these stakeholders, and decentralized trust can be built among them. Honest stakeholders can be viewed

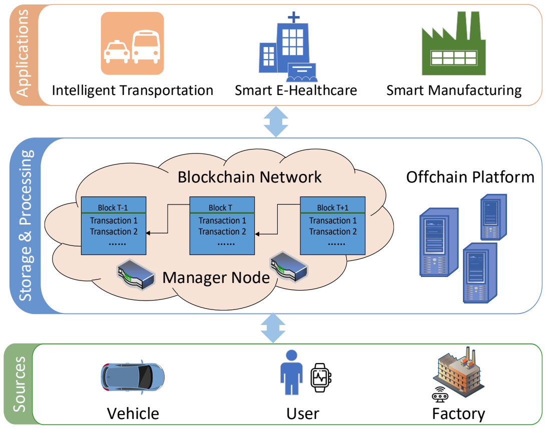

Fig. 3: An Illustration of A Blockchain-based Data Management Architecture

as honest custodians to maintain the validity of DM, provided that the deployed consensus protocols are secure under reasonable assumptions.

- *Immutability and traceability*: By utilizing the cryptographic techniques, such as Merkle hash tree, blockchain storage is immutable. It guarantees that lifecycle events, including reading, writing, modification, and authorization, on the blockchain cannot be maliciously modified. Moreover, with digital signatures on lifecycle events, data operations can be traced to their sources. Data provenance and digital forensics can also be adopted to detect malicious behavior and privacy breaches.
- *Programmability and automation*: The blockchain has two additional features: programmability and automation from the contract layer. Different DM procedures can be coded into executable programs and can run in a decentralized manner on the blockchain. If DM procedures involve many data stakeholders, only one data stakeholder is required to invoke the deployed program when conditions or terms are met.

In the following, we briefly summarize existing literature on the blockchain-based DM. A blockchain-based DM architecture is illustrated in Figure 3, which consists of three components: data sources, data storage and processing, and applications. Data sources include a large number of end devices, including smart vehicles and sensors, with constrained storage and computing capabilities. Normally, they cannot afford the computational overhead of running consensus protocols and the increasing data storage overhead of maintaining the blockchain ledger. As a result, end devices are not directly connected to the blockchain network. Instead, they first connect to manager nodes [10] which are responsible for getting updates of transactions and sending/retrieving necessary data to/from end devices. To increase the throughput and reduce the transaction processing delay, the manager nodes can be further classified according to their roles in different functions, such as proposing, validation and confirmation [11]. Note that, the blockchain network can be partitioned into multiple network channels (ledgers) based on intended transactions and functions, such as user enrollment and data analysis [12]. Each channel is maintained by designated stakeholders (consortium) who can enforce fine-grained polices through deploying properly designed smart contracts. That is, the consortium blockchain can strike a balance between network robustness and efficiency for DM [13].

To mitigate storage bottleneck of the blockchain, an off-chain data storage system is often integrated with the blockchain, where large-scale data are stored off blockchain, and only pivotal data are recorded on blockchain to verify correct executions of off-chain operations [14]. The off-chain storage system does not participate in consensus protocols in the blockchain, and can be viewed as an optional system to assist the blockchain in terms of data storage. In reality, the off-chain storage systems may be deployed by different entities, such as clouds and fog nodes. Although an off-chain storage enables users to revoke their consents on data usage and delete data, the on-chain data modification and deletion are still challenging [15] due to the immutability property of the blockchain.

In summary, blockchain is a promising architecture for DM in NGWN, to achieve a wide range of security functionalities, including identity authentication, access control, provenance tracking, and record logging [16]. However, due to the transparency of blockchain, it is not straightforward to achieve blockchain-based privacy-preserving DM. Note that, many important privacy protection properties, such as anonymity and confidentiality, are not considered in the three layers of the blockchain. The privacy issues include how to protect anonymous nodes in the blockchain network, how to hide users' identities among transactions, and how to protect data confidentiality of on-chain data, etc. To bridge this gap, we next discuss the state-of-the-art privacy-preserving techniques and DM solutions with privacy regulation compliance.

III. Privacy-preserving Techniques for Data Management

This section focuses on four major stages of DM: data collection, data storage, data sharing, and data computation. We summarize existing privacy-preserving techniques throughout these stages, as shown in Table I. These privacy-preserving techniques include cryptographic techniques and non-cryptographic techniques, which can be adopted by users, network operators, and service providers to protect data privacy from not only external adversaries but also internal attackers.

A. Data Collection

Data collection can have two major components: data uploading at user sides and data transmission in NGWN.

Data uploading: Anonymization is a widely recognized privacy-preserving method, which enables users to anonymize their identity and data during data uploading. Two main techniques can be applied to achieve anonymization: pseudonymization [17] and group/ring signature [18]. Pseudonymization-based anonymization methods, though more efficient in terms of computational cost,

TABLE I: Summary of Privacy-preserving Techniques for Data Management

Stages	Functions	Techniques
Data Collection	Data Uploading [17]–[22]	Pseudonymization; Group/Ring Signature; Local Differential Privacy; AI-assisted Taint Analysis.
	Data Transmission [23], [24]	Public/Symmetric Key Encryption; Mix-networks; Onion Routing.
Data Storage	Data Retrieval [25], [37]	Private Information Retrieval; Oblivious RAM; Searchable Encryption.
	Data Auditing [26], [38]	Homomorphic Authenticator.
Data Sharing	Data Publishing [27]–[29]	Anonymity Set; Differential Privacy.
	Data Access Control [30], [31]	Symmetric Key Encryption; Proxy Re-encryption; Attributed-based Encryption.
Data Computation	Outsourced Data Computation [32]	Homomorphic Encryption.
	Collaborative Data Computation [33]–[35]	Federated Learning; Trusted Execution Environment; Secure Multi-party Computation

achieve weaker anonymity than group/ring-signature-based anonymization methods. An improper pseudonym change strategy has a great chance to cause serious privacy leakages. A straightforward way is to use a fresh one-time pseudonym per uploading to overcome the drawback at the huge cost of data storage and communication at user sides, since they need to update their pseudonyms frequently. In contrast, group/ring-signature-based anonymization methods [36] provide more robust privacy protection with lower storage and communication overhead at the cost of computation efficiency. Specifically, heavy cryptographic operations are performed at the user side, and each revocation of an existing anonymous identity credential belonging to one user can affect other users. As there can be a large number of end devices with limited capabilities in DM, it is necessary to have an anonymization solution that strikes a balance between privacy protection and efficiency. Different from anonymization, local data obfuscation is a lightweight non-cryptographic privacy-preserving approach for users to mask their data before uploading. A typical technique for local data obfuscation is local differential privacy [20], which does not rely on any trusted party and allows users to encode and perturb their data using Laplacian or Gaussian noise before data submission. Many variants of the local differential privacy, such as geo-indistinguishability [21], have also been proposed with different focuses. Additionally, taint analysis [22] is another non-cryptographic method that can assist users in detecting privacy leakages at a system level. By tracking the information flow, sensitive user inputs can be identified based on static and dynamic analysis. AI-based models have been applied for the taint analysis to automatically discover the information flow and predict potential privacy leakage. A challenge of this technique is to configure particular classifiers for identification accuracy.

Data transmission: Encryption is a general cryptographic approach to protect data content privacy during data transmission. Generally, public-key encryption techniques are utilized by two parties for negotiating a short-term symmetric key, who can then use symmetric key encryption techniques with the negotiated key for private communications. For data transmission in DM, it is a challenging task to effectively manage multiple keys that belong to different devices, owners, and groups [23]. A centralized solution based on a traditional key server may suffer from many vulnerabilities, such as the single point of failure. Therefore, it is desirable to have a hierarchical and decentralized key management mechanism. From another perspective, privacy-preserving communication techniques such as mix-networks and onion routing, offer routing path privacy for users [24]. By shuffling or hiding the routing paths, adversaries cannot distinguish networking packet sources and destinations, as long as one of the routing nodes is not compromised. However, real-world implementations in NGWN may face many difficulties since the anonymous routing techniques can introduce more computation and communication overheads.

B. Data Storage

We focus on two major research issues on data storage: data retrieval and data auditing.

Data retrieval: To conceal the access pattern of personal data, privacy-preserving techniques such as private information retrieval and oblivious RAM have been proposed [25]. These techniques utilize oblivious transfer and shuffle techniques, such that users can retrieve an item from a database without revealing the item. When applying the above techniques to data storage in DM, the main goal is to reduce the computational complexity and communication rounds among data stakeholders. Moreover, privacy-preserving techniques, such as searchable encryption (SE) [37], have been proposed to protect the search content and search patterns. By combining these techniques, the stakeholders can outsource and store data into a remote database, and retrieve data in a privacy-preserving manner. That is, the data content is retrieved without leaking search keywords or indexes to the remote database. For various data types in DM, it is required to have a versatile data indexing mechanism with adaptable privacy protections.

Data auditing: The most common method for achieving privacy-preserving data auditing is provable data possession (PDP) with homomorphic cryptographic authenticators [26]. Based on this technique, a third-party auditor can help users verify the integrity of their outsourced data in the remote database by checking homomorphic signatures generated from the data without knowing the data content. Even if the auditor is not always trustworthy, the property of public verification can still be guaranteed [38]. Nevertheless, when data auditing requirements become complicated in DM, it is more difficult to achieve desirable privacy guarantees for user data. A practical and lightweight data auditing scheme for DM should achieve privacy preservation, support dynamic data updating, enable batch auditing, and achieve auditing for multiple replicas.

C. Data Sharing

Data sharing mainly involves two procedures: data publishing and data access control.

Data publishing: Data masking is a lightweight non-cryptographic method that can achieve data publishing with privacy preservation to some extent. Different principles have been proposed to achieve data masking, including k-anonymity, l-diversity, and t-closeness. The basic idea of these techniques is to conceal personal identifiable information (PII) and generate a large equivalence class to reduce the risks of privacy exposing [27]. Unfortunately, there exist some attacks to de-anonymize a unique user in the dataset. Differential privacy can be introduced to resolve the issue in the data publishing to generate high-utility synthetic data, which is suitable for high-dimension data without sacrificing the utility cost. [28], [29]. The research challenges when applying these techniques to DM is how to build a data generative model that fully complies with the privacy requirements, since different datasets may have different privacy and utility requirements,

Data access control (DAC): File-based DAC and policy-based DAC are two kinds of DAC mechanisms. To achieve file-based DAC, privacy-preserving techniques, such as symmetric key encryption and proxy re-encryption, can be applied [30]. Based on symmetric key encryption, each file is encrypted with a unique symmetric key. If a user is authorized to access the file, the user will obtain a shared symmetric key from the data owner. Since this technique is key-based, when the number of files increases, the key distribution cost also increases. To reduce the cost, proxy re-encryption has been proposed, which enables a user to encrypt files using a public key before uploading them to a remote data-sharing center. A proxy re-encryption key can be generated based on a user's private key and a receiver's public key to transform the ciphertext, such that the receiver can decrypt it. Policy-based DAC provides a more flexible data sharing approach as compared with file-based DAC. An example of policy-based DAC is attribute-based encryption (ABE) that allows a user to encrypt their data following specific policies. By doing so, only receivers who satisfy the policies can decrypt the ciphertext to recover the shared data [31]. However, the revocation of attribute-based credentials becomes a challenging task as there exist a large number of participants in NGWN.

D. Data Computation

Data computation involves two main areas: outsourced data computation and collaborative data computation. The former focuses on offloading heavy computational tasks to powerful servers without leaking task requester's data inputs. The latter deals with cooperative computation tasks with multiple participants without exposing each participant's data input.

Outsourced data computation: Homomorphic encryption is a traditional cryptographic technique for achieving privacy preservation in outsourced data computation [32]. The data are encrypted before being outsourced to untrusted servers which perform computations directly on the ciphertext. This approach has two widely accepted designs: partially homomorphic encryption (PHE) and fully homomorphic encryption (FHE). The PHE supports specified computations and is more efficient than the FHE, while the FHE can support arbitrary computations at the cost of a time-consuming bootstrapping.

Collaborative data computation: We introduce three main techniques as follows: federated learning, trusted execution environment (TEE), and secure multi-party computation (SMC). In federated learning, participants can contribute their training parameters rather than their data with personally identifiable information (PII) for co-operative data computation, which is especially suitable for collaborative machine learning on resource-limited mobile devices [33]. TEE implements algorithms with sensitive data input in a secure enclave, such that external attackers cannot access them to break the data privacy [34]. It requires to have a trusted entity to attest the code integrity before loading the code to the enclave. SMC is built on the oblivious transfer and garble circuit [35], where several parties can compute arbitrary functions without exposing individual input. However, SMC may incur a high communication overhead. For resource-restrained devices in NGWN, it is desired to utilize the SMC for collaborative data computations while ensuring low communication and computation overheads.

Existing privacy-preserving techniques have a wide range of functionalities for DM. At the same time, there are emerging challenges to utilize the techniques for DM in NGWN, as discussed in Section V.

IV. Data Management under Privacy Regulation

In this section, we discuss the recent privacy regulation, i.e., GDPR, its impacts on DM in NGWN, and existing literature on the blockchain-based DM under the GDPR.

A. Privacy Regulation: General Requirements and Impacts

GDPR defines legal requirements for data processing related to any "identifiable natural person" [5]. GDPR grants individual users two main categories of rights over their personal data: (1) right to be informed: users can require information about the purpose of data processing, the period of data storage, and the existence of data exchange; and (2) right to control: at any point of the data lifecycle, users can access their data, restrict/object any data processing, and require to delete their data. Therefore, explicit terms about user data management must be specified and user consent must be obtained before any data operation is performed.

Since the GDPR took effect in 2018, many application providers have been supplying consent notices before collecting user data. For DM in NGWN, as data exchanges among different data stakeholders are strictly regulated, traditional business models relying on collaboratively discovering the knowledge from user data are affected.

B. GDPR-compliant Data Management

According to Article 26 of the GDPR, DM in NGWN involving multiple controllers (data stakeholders) should be conducted in a collaborative and transparent manner. As data stakeholders in NGWN usually come from different trust domains, the blockchain can offer a promising solution for building DM platform under the GDPR [3], [39]. First, the blockchain can help data stakeholders decide data usage agreements to provide users with trustworthy access of data processing information. Second, smart contract techniques enable data stakeholders to securely update the data usage status to grant users applicable control of their data. Third, the storage immutability of the blockchain ensures reliable logs of critical data lifecycle events, to monitor privacy breaches and purse joint accountability on misbehaving data stakeholders.

Privacy-preserving techniques can be applied in blockchain-based DM. First, anonymous identity management techniques can enhance user identity privacy in the data uploading phase, while traditional data encryption techniques can help ensure data confidentiality during transmission. Second, secure data retrieving and auditing techniques can achieve efficient on-chain data search and integrity check. Third, various access control techniques can be adopted for consent-based data sharing on the blockchain. Finally, secure data computation techniques can help data stakeholders to collaboratively discover data knowledge without exposing user data privacy.

A summary of state-of-the-art works on blockchain-based GDPR-compliant solutions is given in Table II, in terms of system model, trust assumption, GDPR compliance, and implementation. The early research efforts mainly focus on adopting the blockchain as an add-on component to regulate existing cloud-based DM platforms [40], [41]. In a conceptual framework [40], users can manage the storage and trading of their data on application providers, e.g., cloud servers, with consent-based access control via smart contract techniques. Subsequently, the roles of application providers can be further divided into a service provider (SP) for data collection and a resource server (RS) for data storage [41]. Integrated with data encryption and identity authentication techniques, fine-grained data access control can be achieved. In addition to consent-based data access control, a blockchain platform can be utilized to record data lifecycle events performed by the cloud [6]. By designing a mechanism to translate the GDPR terms to smart contracts, logged data lifecycle events on the cloud can be automatically checked for the GDPR compliance. It is also promising to utilize the blockchain and construct DM platform [42] for a wide range of data sharing applications in the smart city, in terms of health data, smart car data, smart meter data, surveillance data, and financial transactions. Data sharing domains can be modeled as 'organizations' in the Hyperledger Fabric, where business processes are developed as 'chaincode' and a certificate authority (CA) is implemented for identity management.

V. Research Challenges & Potential Solutions

As the blockchain can serve as a promising architecture for DM in NGWN, there are many interesting ideas in existing literature for blockchain-based DM with privacy-preserving designs. However, many research challenges exist to be addressed for blockchain-based DM under the GDPR.

Versatile blockchain architecture: Trust degrees among stakeholders and regulation requirements for different use cases in NGWN may change dynamically. A versatile blockchain architecture that can be tailored for different use cases is required, based on flexible consensus protocols, cross-chain operations, distributed membership management, and data provenance. Current consensus protocols cannot easily balance security guarantees with blockchain scalability, considering the large number of resource-limited devices in NGWN. Therefore, a new switching and scaling mechanism should be established to improve consensus protocols in an adaptable way. Moreover, cross-chain operations are necessary for DM since many industrial applications require functions like data sharing through transactions. Cryptographic techniques and non-cryptographic techniques, can be adopted to provide interoperability between different blockchains. In addition, distributed membership management and data provenance through cryptographic accumulator and TEE should be considered to improve the security of blockchain, especially for the consortium blockchain.

One case - one technique: Privacy protection is required for every stage of DM in NGWN. Many solutions have been proposed using different techniques with various properties. When applying them to specific use cases in NGWN, it is challenging to integrate them into a unified platform under the same trust assumptions. Thus, a hybrid solution for privacy-preserving DM is essential, where a trust model should be clearly defined according to various entity's roles and attributes. Also, suitable privacy-preserving techniques should be chosen and deployed as modules to address particular privacy requirements, such that configurable privacy protection can be achieved to resist different kinds of attacks.

On/off chain computation model: On-chain computation and storage resources are expensive since every blockchain node needs to store whole blockchain storage and update blockchain state when new transactions are added. This can result in huge on-chain overhead, especially when a consensus protocol with a high security level is implemented. Hence, it is critical to design an on/off chain computation model, where data operations are performed off chain and verified on chain efficiently [43]. Besides using hash functions for data integrity checking, it is desired to have more functional on/off-chain models that support complex computations.

On/off chain privacy model: The on-chain data access is open to the public in a permissionless blockchain, or is restricted to specific nodes in a permissioned blockchain. As a result, for each data use case in NGWN, it is essential to have an on/off chain privacy model that determines what specific data operations should be revealed to the public or individuals. For example, in a data-sharing case, data providers and consumers should have access to the shared data in an off-chain manner. The blockchain only knows that if the data sharing follows the GDPR requirements [41]. Moreover, the trust levels of blockchain participants and data sensitivity can change dynamically. Therefore, it is essential to tailor the

TABLE II: Summary of Blockchain-based DM under GDPR

	System Model	Trust Assumption	GDPR Compliance	Implementation
[40]	Cloud	N/A	Consent validation	N/A
[41]	Cloud	Honest-but-curious RS & malicious SP	Right of access/restricted processing Right to be informed/forgotten	Hyperledger Fabric
[42]	Smart city	Trusted membership service	Consent-based data processing Right to be informed/forgotten	Hyperledger Fabric
[6]	Cloud	Trusted log generator	Compliance monitoring	Ethereum

designs of privacy-preserving techniques for DM in NGWN with delegatable and fine-grained operation verifications, time-embedded cryptography primitives, and updatable and verifiable secret sharing.

Trusted blockchain input. It is usually required to have a trusted component in the blockchain-based DM, such as a trusted off-chain storage manager, to correctly upload data lifecycle events to the blockchain storage. For DM in NGWN, weaker trust assumptions are more practical that data stakeholders may not always honestly interact with the blockchain. In this case, verifiable computation techniques, such as TEE and succinct non-interactive argument (SNARG) [14], can be utilized to ensure trusted blockchain input. In the meantime, data provenance based on blockchain storage can analyze causal relationships between data lifecycle events to detect dishonest blockchain input.

VI. Conclusion and Future Directions

In this article, we have investigated DM for NGWN. From the perspectives of architecture requirements, privacy-preserving techniques, and privacy regulation compliance, we have conducted a comprehensive survey on the existing DM solutions and highlighted the research challenges.

For future research, more efforts should be directed to the designs, implementations and evaluations of blockchain-based DM with three essential requirements: First, the blockchain-based DM should have a flexible architecture to satisfy various security and scalability requirements in NGWN. Second, the blockchain-based DM should support modular designs from privacy-preserving techniques to adapt to privacy regulation requirements under different use cases. Third, on/off-chain privacy and computation models for DM should be developed to strike a balance between privacy protection levels and processing efficiency.

References

[1] X. Shen, J. Gao, W. Wu, K. Lyu, M. Li, W. Zhuang, X. Li, and J. Rao, “Ai-assisted network-slicing based next-generation wireless networks,” *IEEE Open Journal of Vehicular Technology*, vol. 1, pp. 45–66, 2020.

[2] Y. Dai, D. Xu, S. Maharjan, Z. Chen, Q. He, and Y. Zhang, “Blockchain and deep reinforcement learning empowered intelligent 5g beyond,” *IEEE Network*, vol. 33, no. 3, pp. 10–17, 2019.

[3] G. P. Freund, P. B. Fagundes, and D. D. J. de Macedo, “An analysis of blockchain and gdpr under the data lifecycle perspective,” *Mobile Networks and Applications*, pp. 1–11, 2020.

[4] R. Li and H. Asaeda, “A blockchain-based data life cycle protection framework for information-centric networks,” *IEEE Communications Magazine*, vol. 57, no. 6, pp. 20–25, 2019.

[5] General Data Protection Regulation (GDPR). https://gdpr-info.eu. Accessed November 2020.

[6] M. Barati and O. Rana, “Tracking gdpr compliance in cloud-based service delivery,” *IEEE Transactions on Services Computing*, Early Access, pp. 2075–2092, 2020.

[7] M. Zaharia *et al.*, “Apache spark: a unified engine for big data processing,” *Communications of the ACM*, vol. 59, no. 11, pp. 56–65, 2016.

[8] S. Zeuch *et al.*, “The nebulastream platform: Data and application management for the internet of things,” *arXiv preprint arXiv:1910.07867*, 2019.

[9] E. Androulaki, A. Barger, V. Bortnikov, C. Cachin, K. Christidis, A. De Caro, D. Enyeart, C. Ferris, G. Laventman, Y. Manevich *et al.*, “Hyperledger fabric: a distributed operating system for permissioned blockchains,” in *Proc. of EuroSys*, 2018, pp. 1–15.

[10] O. Novo, “Blockchain meets iot: An architecture for scalable access management in iot,” *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 1184–1195, 2018.

[11] V. K. Bagaria, S. Kannan, D. Tse, G. C. Fanti, and P. Viswanath, “Prism: Deconstructing the blockchain to approach physical limits,” in *Proc. of ACM CCS*, 2019, pp. 585–602.

[12] O. Choudhury, N. Fairoza, I. Sylla, and A. Das, “A blockchain framework for managing and monitoring data in multi-site clinical trials,” *arXiv preprint arXiv:1902.03975*, 2019.

[13] Z. Ma, X. Wang, J. D. Kumar, K. Haneef, H. Gao, and Z. Wang, “A blockchain-based trusted data management scheme in edge computing,” *IEEE Transactions on Industrial Informatics*, vol. 16, no. 3, pp. 2013–2021, 2019.

[14] M. Campanelli, D. Fiore, and A. Querol, “Legosnark: Modular design and composition of succinct zero-knowledge proofs,” in *Proc. of ACM CCS*, pp. 2075–2092, 2019.

[15] D. Deuber, B. Magri, and S. A. K. Thyagarajan, “Redactable blockchain in the permissionless setting,” in *Proc. of IEEE S&P*, 2019, pp. 124–138.

[16] D. Liu, J. Ni, C. Huang, X. Lin, and X. Shen, “Secure and efficient distributed network provenance for iot: A blockchain-based approach,” *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 7564–7574, 2020.

[17] R. Lu, X. Lin, T. H. Luan, X. Liang, and X. Shen, “Pseudonym changing at social spots: An effective strategy for location privacy in vanets,” *IEEE Transactions on Vehicular Technology*, vol. 61, no. 1, pp. 86–96, 2011.

[18] C. Huang, R. Lu, X. Lin, and X. Shen, “Secure automated valet parking: A privacy-preserving reservation scheme for autonomous vehicles,” *IEEE Transactions on Vehicular Technology*, vol. 67, no. 11, pp. 11 169–11 180, 2018.

[19] F. Baldimtsi, J. Camenisch, M. Dubovitskaya, A. Lysyanskaya, L. Reyzin, K. Samelin, and S. Yakoubov, “Accumulators with applications to anonymity-preserving revocation,” in *Proc. of EuroS&P*, 2017, pp. 301–315.

[20] Q. Ye, H. Hu, X. Meng, and H. Zheng, “Privkv: Key-value data collection with local differential privacy,” in *Proc. of IEEE S&P*, 2019, pp. 317–331.

[21] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, “Geo-indistinguishability: Differential privacy for location-based systems,” in *Proc. of ACM CCS*, 2013, pp. 901–914.

[22] Q. Zhao, C. Zuo, B. Dolan-Gavitt, G. Pellegrino, and Z. Lin, “Automatic uncovering of hidden behaviors from input validation in mobile apps,” in *Proc. of IEEE S&P*, 2020, pp. 1106–1120.

[23] S. Kumar, Y. Hu, M. P. Andersen, R. A. Popa, and D. E. Culler, “{JEDI}: Many-to-many end-to-end encryption and key delegation for iot,” in *Proc. of USENIX Security*, 2019, pp. 1519–1536.

[24] C. Kuhn, M. Beck, and T. Strufe, “Breaking and (partially) fixing provably secure onion routing,” in *Proc. of IEEE S&P*, 2020, pp. 168–185.

[25] Z. Zhang, K. Wang, W. Lin, A. W.-C. Fu, and R. C.-W. Wong, “Practical access pattern privacy by combining pir and oblivious shuffle,” in *Proc. of CIKM*, 2019, pp. 1331–1340.

[26] A. Yang, J. Xu, J. Weng, J. Zhou, and D. S. Wong, “Lightweight and privacy-preserving delegatable proofs of storage with data dynam-

ics in cloud storage," *IEEE Transactions on Cloud Computing*, DOI: 10.1109/TCC.2018.2851256, Early Access, 2018.

[27] Z. Tu, K. Zhao, F. Xu, Y. Li, L. Su, and D. Jin, "Protecting trajectory from semantic attack considering k-anonymity, l-diversity, and t-closeness," *IEEE Transactions on Network and Service Management*, vol. 16, no. 1, pp. 264–278, 2018.

[28] Z. Wang, X. Pang, Y. Chen, H. Shao, Q. Wang, L. Wu, H. Chen, and H. Qi, "Privacy-preserving crowd-sourced statistical data publishing with an untrusted server," *IEEE Transactions on Mobile Computing*, vol. 18, no. 6, pp. 1356–1367, 2018.

[29] V. Bindschaedler, R. Shokri, and C. A. Gunter, "Plausible deniability for privacy-preserving data synthesis," *Proc. VLDB Endow.*, vol. 10, no. 5, p. 481492, 2017.

[30] Y. Hu, S. Kumar, and R. A. Popa, "Ghostor: Toward a secure data-sharing system from decentralized trust," in *Proc. of NSDI*, 2020, pp. 851–877.

[31] J. Hao, J. Liu, W. Wu, F. Tang, and M. Xian, "Secure and fine-grained self-controlled outsourced data deletion in cloud-based iot," *IEEE Internet of Things Journal*, vol. 7, no. 2, pp. 1140–1153, 2019.

[32] C. Gentry and D. Boneh, *A fully homomorphic encryption scheme*. Stanford University, 2009, vol. 20, no. 9.

[33] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 10, no. 2, pp. 1–19, 2019.

[34] X. Zhang, F. Li, Z. Zhang, Q. Li, C. Wang, and J. Wu, "Enabling execution assurance of federated learning at untrusted participants," in *Proc. of IEEE INFOCOM*, 2020, pp. 1877–1886.

[35] M. Hastings, B. Hemenway, D. Noble, and S. Zdancewic, "Sok: General purpose compilers for secure multi-party computation," in *Proc. of IEEE S&P*, 2019, pp. 1220–1237.

[36] S. Sun, M. H. Au, J. K. Liu, and T. H. Yuen, "Ringct 2.0: A compact accumulator-based (linkable ring signature) protocol for blockchain cryptocurrency monero," in *Proc. of ESORICS*, 2017, pp. 456–474.

[37] S. Hu, C. Cai, Q. Wang, C. Wang, X. Luo, and K. Ren, "Searching an encrypted cloud meets blockchain: A decentralized, reliable and fair realization," in *Proc. of IEEE INFOCOM*, 2018, pp. 792–800.

[38] Y. Zhang, C. Xu, X. Lin, and X. Shen, "Blockchain-based public integrity verification for cloud storage against procrastinating auditors," *IEEE Transactions on Cloud Computing*, 2019.

[39] H. T. Vo, A. Kundu, and M. K. Mohania, "Research directions in blockchain data management and analytics," in *Proc. of EDBT*, 2018, pp. 445–448.

[40] B. Faber, G. C. Michelet, N. Weidmann, R. R. Mukkamala, and R. Vatrapu, "Bpdims: a blockchain-based personal data and identity management system," in *Proc. of HICSS*, 2019.

[41] N. B. Truong, K. Sun, G. M. Lee, and Y. Guo, "Gdpr-compliant personal data management: A blockchain-based solution," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 1746–1761, 2020.

[42] I. Makhdoom, I. Zhou, M. Abolhasan, J. Lipman, and W. Ni, "Privysharing: A blockchain-based framework for privacy-preserving and secure data sharing in smart cities," *Computers & Security*, vol. 88, p. 101653, 2020.

[43] R. Cheng, F. Zhang, J. Kos, W. He, N. Hynes, N. Johnson, A. Juels, A. Miller, and D. Song, "Ekiden: A platform for confidentiality-preserving, trustworthy, and performant smart contracts," in *Proc. of IEEE EuroS&P*, 2019, pp. 185–200.

Xuemin (Sherman) Shen (M97-SM02-F09) is currently a University Professor with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research focuses on network resource management, wireless network security, social networks, 5G and beyond, and vehicular ad hoc networks. He is a Canadian Academy of Engineering Fellow, a Royal Society of Canada Fellow, and a Chinese Academy of Engineering Foreign Fellow. He received the R.A. Fessenden Award in 2019 from IEEE, Canada; the James Evans Avant Garde Award in 2018 from the IEEE Vehicular Technology Society; and the Joseph LoCicero Award in 2015 and Education Award in 2017 from the IEEE Communications Society.

Cheng Huang (M'20) received his B.Eng and M.Eng in information security from Xidian University, China, in 2013 and 2016 respectively, and was a Project Officer with the INFINITUS laboratory at the School of Electrical and Electronic Engineering, Nanyang Technological University till July 2016. He received the Ph.D. degree in Electrical and Computer Engineering, University of Waterloo, ON, Canada in 2020. His research interests are in the areas of applied cryptography, cyber security and privacy in the mobile network.

Dongxiao Liu (M'20) is a Postdoctoral Research Fellow with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. He received the PhD degree at the Department of Electrical and Computer Engineering, University of Waterloo, Canada in 2020. His research interests include privacy enhancing technologies for blockchain and mobile networks.

Liang Xue is working toward the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. She received the B.S. degree in Information Security and the M.S. degree in Computer Engineering from the School of Computer Science and Engineering, University of Electronic Science and Technology of China, China, in 2015 and 2018, respectively. Her research interests include applied cryptography, cloud computing, and blockchain.

Weihua Zhuang (M93-SM01-F08) has been with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada, since 1993, where she is currently a professor and a Tier I Canada Research Chair in Wireless Communication Networks. Dr. Zhuang is a Fellow of the Royal Society of Canada, the Canadian Academy of Engineering, and the Engineering Institute of Canada. She is also an elected member of the Board of Governors and VP-Publications of the IEEE Vehicular Technology Society. She was a recipient of the 2017 Technical Recognition Award from the IEEE Communications Society Ad Hoc & Sensor Networks Technical Committee.

Rob Sun currently is a Principal Engineer with Huawei Technologies Canada Co. ltd. His work primarily focuses on the advancement of the NG wireless, including 5G/6G and WiFi/IoT security architecture and standardization. Rob was also vice chair of IEEE PMP (Privacy Management Protection) Task Group which was to set out the best practices for protecting personal privacy information, and support efficient, adaptable and innovative approaches for privacy governance. He was regarded as one of the core contributors to the standardizations of a series of NG Wi-Fi security protocols and certifications, including the most recent Wi-Fi WPA3 protocol suites. He also co-authored a few books on wireless security technologies.

Didi Ying (Phd) is currently working in Huawei Technologies Canada Co., LTD as a senior network architecture engineer. Her main research is about security and privacy in wireless network. Before that, she was working in University of Ottawa. During the past 15 years, she has published more than 200 papers in top conferences and reputable journals.