Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Correspondence:

A unified data collection framework based on the data plane for 6G

Yannan YUAN[†‡], Fei QIN, Jiankang LIU, Yuanyuan WANG, Jianan CAI, Xiang PAN, Dajie JIANG

vivo Mobile Communication Co., Ltd., Beijing 100083, China

[†]E-mail: yannan.yuan@vivo.com

Received Apr. 1, 2024; Revision accepted Dec. 1, 2024; Crosschecked Dec. 25, 2024

https://doi.org/10.1631/FITEE.2400247

Fifth-generation wireless communication (5G) offers different data collection methods for different use cases, some of which have faced challenges due to a lack of data providers. The fragmented solutions for individual use cases have high standard overheads and can result in duplicate data collection. Moreover, existing data collection methods are not suitable for collecting a large amount of data to provide new capabilities for sixth-generation wireless communication (6G) systems, such as native artificial intelligence (AI) and sensing. In this paper we propose a unified data collection framework based on the data plane (DP) for 6G. Through the protocol stack of DP and a two-sided data collection mode, the proposed unified data collection framework is more suitable for collecting a large amount of data. The gain of DP in uplink (UL) processing delay is validated by test results based on a 6G user equipment (UE) prototype. As the data packet length increases, the advantage of the DP protocol stack on the UE UL processing delay becomes more significant. By way of rewards or data exchange, the proposed two-sided data collection mode increases the willingness of data providers to provide data. The two-sided data collection mode enables the possibility of digital twin (DT) UE. DT UE can reuse a large amount of UE in the physical network to reduce the infrastructure resource overhead incurred by the DT network and improve performance. Therefore, the unified data collection framework provides extensive data functionalities to realize the full potential of the 6G system.

1 Introduction of 5G data collection

The 5G standards provide various data collection methods to meet the needs of different use cases (Fig. 1). Based on the definitions of physical layer measurements (3GPP, 2024), layer 2 (L2) measurements (3GPP, 2023d), and performance measurements (3GPP, 2022), to support experience optimization, radio resource management, and network operation and maintenance (OAM), data collection solutions for quality of experience (QoE) (3GPP, 2021), minimization of drive tests (MDT) (3GPP, 2023f), and self-optimizing networks (SON) (3GPP, 2019) are independently standardized. 3GPP (2023e) and 3GPP (2023c) designed positioning protocols for the location management function (LMF). LMF, base station, and OAM collect data from UE independently, resulting in duplicate data collection, such as channel status information reference signal received power (CSI-RSRP). In 3GPP (2023g), the data collection coordination function (DCCF) is introduced to support the data collection and exposure of analytics results for the network data

---

[‡] Corresponding author

ORCID: Yannan YUAN, https://orcid.org/0000-0003-1892-612X

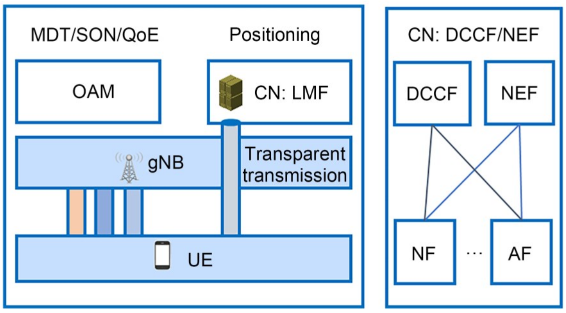

Fig. 1 Data collection methods for 5G

MDT: minimization of drive tests; SON: self-optimizing network; QoE: quality of experience; OAM: operation and maintenance; CN: core network; LMF: location management function; UE: user equipment; DCCF: data collection coordination function; NEF: network exposure function; NF: network function; AF: application function

analytics function (NWDAF) in the 5G core network (CN). The network exposure function (NEF) in 3GPP (2023b) can collect data from network functions (NFs) of the 5G CN via a service-based interface and provide data to application functions (AFs) via northbound application programming interfaces (APIs). DCCF and NEF can collect duplicate data from NFs.

From the perspective of CN, we take DCCF as an example. NWDAF can collect data from CN NFs and AFs via DCCF, but not from the UE or radio access network (RAN). The analytics results of NWDAF can be delivered to NFs by DCCF. Since DCCF is an optional feature introduced in a later release, NWDAF can also directly collect data based on data subscription from the access and mobility management function, session management function, policy control function, unified data management, OAM, etc. DCCF in 5G CN has an embryonic form of unified data collection framework to collect data from different NFs. However, NFs other than NWDAF are unable to collect data or provide data through DCCF. Neither UE nor RAN can collect data through DCCF.

From the perspective of RAN, we take QoE, MDT, and SON as examples. Although both QoE and signaling-based MDT are initiated by OAM, OAM generates only the QoE configurations of UE. The UE QoE configurations are transparent to the RAN node. The configuration of MDT sent to UE is generated by the RAN node. Finally, the RAN node sends the collected data to the trace collection entity/measurement collection entity (TCE/MCE). Based on the data analysis of TCE/MCE, the configuration of RAN nodes can be optimized through SON. Since the QoE and MDT data are carried by radio resource control (RRC) signaling of the control plane (CP), the granularity of the data collected by existing methods is not fine enough. For example, the time interval for data collection is typically 15 min and the data collected are at the cell level rather than the UE level. This is because CP is designed for small size, high priority signaling rather than bulk data collection.

In addition to the data collection requirements for communication optimization, there are new data collection requirements arising from 6G. ITU-R (2023) proposed six usage scenarios: immersive communication, massive communication, hyper reliable and low-latency communication, AI and communication, integrated sensing and communication (ISAC), and ubiquitous connectivity. Liu et al. (2020) emphasized the importance of native AI and native security in the design of a 6G network architecture. It is foreseeable that sensing data (e.g., sensing measurement data) and AI data (e.g., AI model) will be two new types of data of the 6G system. The collection of these two types of data is characterized by a large data volume and flexible termination node (e.g., UE, RAN nodes, and CN NFs). In addition, the digital twin (DT) network approach in Deng et al. (2023) can realize a high level of autonomy for a 6G wireless network. The key element of the DT network is real-time data collection.

In summary, data collection is a pervasive requirement of the 6G system. However, 5G data collection methods have issues with fragmented solutions for individual use cases, leading to high standardization overheads and UE/NF duplicate data collection. The data collection methods based on CP are not suitable for collecting a large amount of data, such as those needed by 6G native AI, ISAC, and DT networks. Data collection methods based on CP also suffer from low efficiency of data transmission. Moreover, 5G data collection methods adopt mainly the one-sided data collection mode, which results in a lack of data providers.

To solve the above problems, we propose a unified data collection framework. The architecture of the 6G DP is described as a unified data collection framework to solve the problem of duplicate data collection and reduce the standardization overhead. The protocol stack of DP is described which supports a large amount of data. To improve efficiency, a DT network solution based on a two-sided data collection mode provided by the unified data collection framework is described.

2 Architecture of the data plane for 6G

We propose to use DP to provide a unified data collection framework for the 6G system. The logical functions of 6G DP include a data control function (DCF), data privacy and security function (DPSF), data repository function (DRF), data transmission function (DTF), data processing function (DProcF), data provider (DPro), and data consumer (DCons) (Fig. 2). The details are as follows:

DCF is used to support data collection coordination and to configure data service, data transmission (e.g., establishing, modifying, and releasing the bearer of DP), and data processing.

DPSF is used to support the mechanisms of privacy and security, such as authentication, authorization, and access control.

DRF is used to support the long-term storage and retrieval of data collected by DP.

DTF is used to support the transmission and forwarding of DP data based on the configuration from DCF. Depending on the termination node of data transmission, DP transmission can use several data transmission protocols, such as the protocols between UE and RAN, RAN and CN, and UE and CN, as well as the protocol of intra RAN or intra CN functions.

DProcF is an optional function used to preprocess data or provide data analysis according to the requirements of data service.

DPro provides the required data according to the configuration from DCF.

DCons sends data requests and receives data responses.

Based on the 5G architecture (3GPP, 2023h), we propose a 6G system architecture that introduces the sensing function and the aforementioned data functions (Fig. 3). For the communication functions in the green and blue blocks, we have assumed the adoption of current 5G standards without considering the further enhancement that will be brought by 6G. The existing service-based interfaces are used mainly among the CP functions of CN. The service-based interface uses the HTTP/2 protocol with JavaScript Object Notation (JSON) as the serialization protocol (3GPP, 2023a). Since JSON is more suitable for signaling, we propose to enhance the existing protocols of the service-based interface or build a messaging framework to support the transmission of a large amount of data within CN to improve efficiency. Correspondingly, RAN and UE need to introduce peer-to-peer data

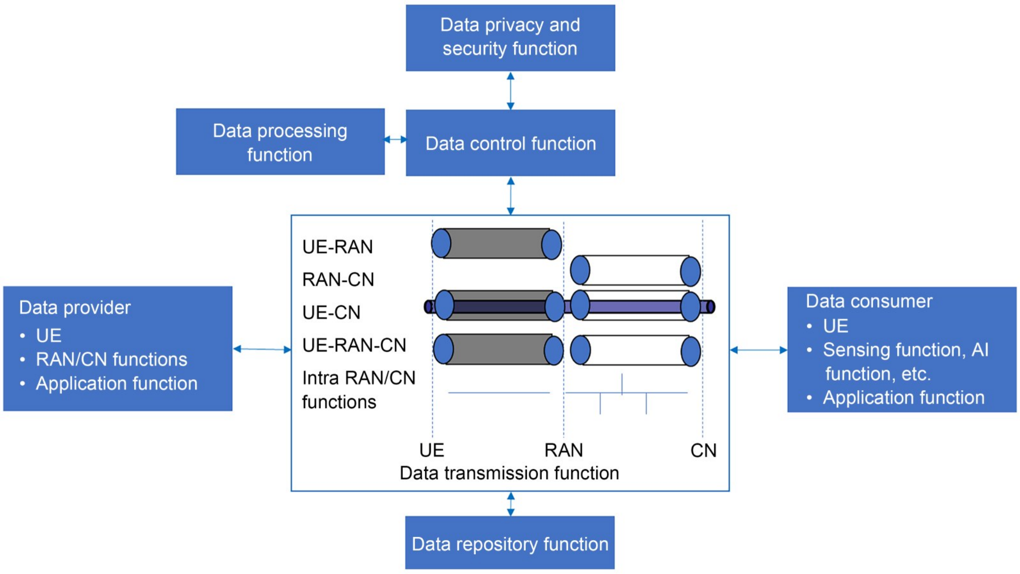

Fig. 2 Logical architecture of the data plane for 6G

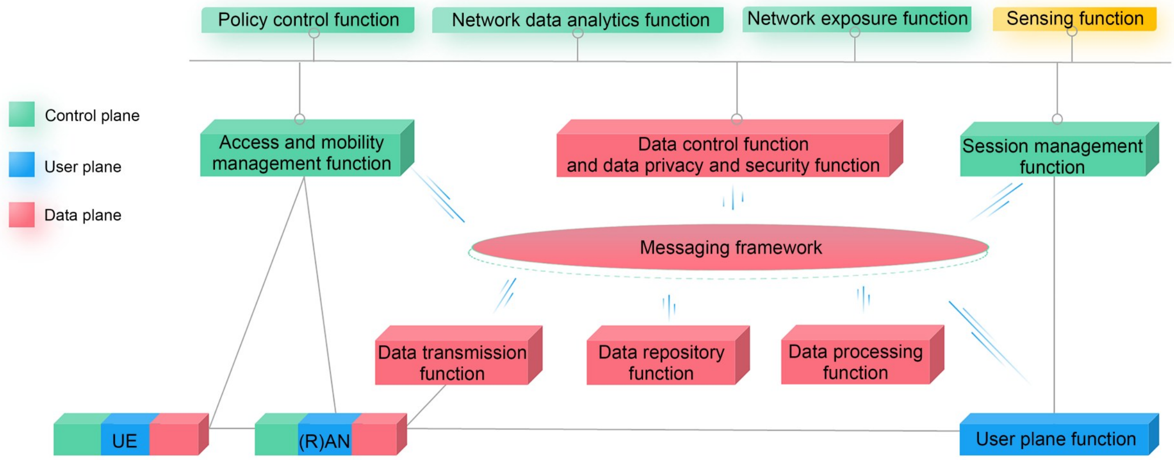

Fig. 3 Six-generation wireless communication system architecture with the data plane (References to color refer to the online version of this figure)

functions, etc., to support the new features of 6G. RAN and UE also need to introduce peer-to-peer functions to support the enhancements of communication and new features of 6G.

3 Protocol stack of the data plane for 6G

Oriented to different transmitters and receivers, DP protocols include protocols among UE, RAN, and CN, such as those between UE and RAN, RAN and CN, UE and CN, RAN functions, or CN functions. The design of the protocol stack of DP is the key factor that determines its efficiency, especially the stack between UE and RAN. If the protocol of DP is called the data service application protocol (DSAP), we propose a DP protocol stack of the air interface between UE and RAN (Fig. 4).

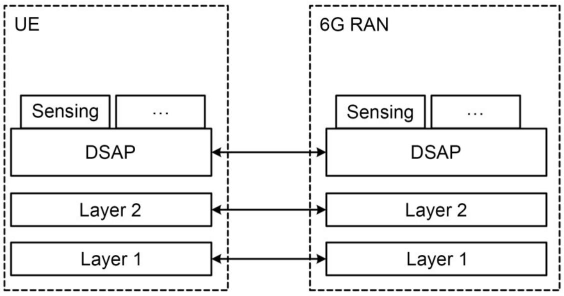

Fig. 4 A data plane protocol stack of the air interface (DSAP: data service application protocol)

3.1 DP protocol stack of the air interface

DSAP includes DCF, DPSF, DRF, DTF, and DProcF. Based on the requirements of data collection, DSAP establishes appropriate DP radio bearers (DPRBs) through RRC. If the collected data are visible to the RAN node, the type of DPRB is a bearer terminated at the RAN node. If the collected data are invisible to the RAN node, the type of DPRB is a bearer transparently transmitted by the RAN node. There is also a type of DPRB in which the data carried are processed by the RAN node before being forwarded to the next appropriate node. DPRB has the characteristics of flexible priority. The priority of DPRB can be set according to the data volume, quality of data transmission, and traffic load of the user plane (UP). Therefore, compared to the signaling radio bearer with the highest priority, DPRB is more suitable for collecting a large amount of data. Additionally, DPRB can meet the requirements of flexible data termination nodes and on-demand processing.

Next, we describe how the air interface DP protocol stack can achieve efficient data transmission from the perspectives of DSAP, L2, and layer 1 (L1). DProcF in the DSAP of data providers supports data compression, including both lossy algorithms and lossless algorithms. The volume of data can be reduced by data compression. When the collected data are ready for transmission, data are not subject to Abstract Syntax Notation One (ASN. 1) processing in DSAP.

ASN. 1 is used in RRC to express the complex data structures of signaling and to serialize data into bit streams. The pure data provided by data providers do not have a complex structure, and thus the data representation or serialization can be optional or optimized. L2 of the DP protocol stack uses simplified protocol functions. For example, unified data storage can reduce the overhead of L2 headers and data copying; encryption/decryption and segmentation/re-segmentation can be configured on demand. Finally, L1 of the DP protocol stack can adopt more efficient technologies of the physical layer, such as joint source–channel coding (Xu et al., 2023) or semantic communication (Zhang et al., 2022; Dai et al., 2023), to improve the spectral efficiency of data collection.

3.2 Test and performance analysis

The efficiency of the DP protocol stack between UE and RAN is verified via a 6G UE prototype developed by vivo, the parameters of which are shown in Table 1. The test case is that the UE collects data and sends the collected data to the network according to the configuration.

Table 1 Parameters of the UE prototype

Parameter	Value/Description
CPU (@frequency)	Arm A53 (@1 GHz)
Memory	4 GB
ASN.1 tool/version	Objective/V773
ProtoBuf tool/version	protoc/3.6.1
Encryption algorithm	AES128

The UL processing delay is tested under different lengths and value ranges of the collected data, as well as under different UP traffic loads. The test results (Fig. 5) show that the proposed unified data collection framework based on DP is more suitable for collecting a large amount of data. The statistical average UL processing delay does not include the L1 processing delay because the L1 of the existing CP-based collection method in the prototype is the same as that of the DP stack. Analyzed from the UE side, the average UL processing delay of the CP-based solution is 63.2090 μs when the length of data is 19 bytes. With the same length of data, the average UL processing delay of DP is 13.7305 μs, which is about 22% of the processing delay of the existing CP-based solution. When the encryption is disabled via the network configuration, the average UL processing delay of the DP-based solution is 3.9210 μs. As the length of data increases, the efficiency improvement of the proposed DP protocol stack increases.

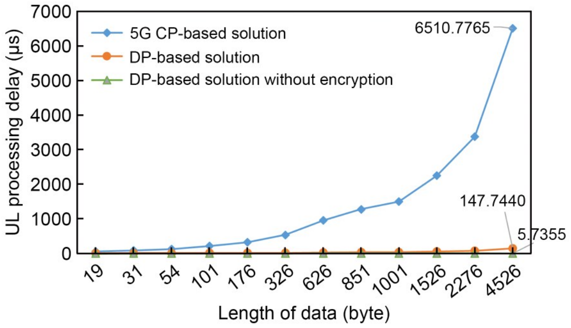

Fig. 5 Average UL processing delay of data collection based on DP and CP

The processing delay of ASN. 1 and ProtoBuf is also tested to evaluate potential optimization methods for data representation and serialization. In the minimum value test case, the output data length of proto3 gradually decreases from 127% to 77% of ASN. 1 as the data length increases. The processing delay of proto3 is shorter than that of ASN. 1 for both the maximum and minimum test cases. In particular, the encoding time of proto3 is about 10% of that of ASN. 1, and the decoding time about 17% of that of ASN. 1. As the data length increases, the processing delay advantage of ProtoBuf becomes more obvious. ASN.1 is therefore more suitable for short messages with a small amount of data. When a large amount of data are collected and transmitted, ProtoBuf performs similarly in terms of output data length but has a shorter processing delay than ASN.1.

4 Two-sided data collection mode enables data applications of 6G

4.1 Two-sided data collection mode

The unified data collection framework supports both one- and two-sided data collection modes. The one-sided data collection mode refers to the process where only one party obtains the required data during data collection. For example, the existing data collection methods are used by network nodes to obtain the required data, which leads to a low willingness of data providers (such as UE) to contribute data. The two-sided data collection mode we propose means that both data collection nodes and data providers obtain the required data or rewards. Potential rewards include the number of times to obtain network data, quality of service (QoS) guarantee and priority promotion, data quota, minutes of voice, etc. When data providers can also obtain the required data or rewards, the number of data providers increases. The two-sided data collection mode can improve the success rate of data collection and enable data applications of 6G, such as sensing, AI, and DT networks.

Taking the data collection between RAN and UE as an example, RAN initiates a data request with the indication of the data collection mode. For the two-sided mode, the data request can also indicate the data that the UE can obtain during data collection. Potential data obtained by UE can be UL measurements collected by the network during the data collection process, such as reference signal received power (RSRP) and resource block utilization rate. Potential data can also include paging failure data of UE already collected by the network, such as the time of paging failure, the tracking area, and the cell identifier involved in the paging failure. Since the data provided by the network are helpful for the UE to detect problems and fine-tune local parameters (e.g., policy parameters and AI model parameters), the UE prefers to provide user consent or accept the data request with the indication of the two-sided data collection mode. Meanwhile, two traditional one-sided data collection procedures are required to accomplish the above two-sided data collection. In addition to improving the success rate of the data request, the two-sided data collection mode reduces nearly 50% of the signaling message overhead and uses only one DPRB.

In addition, existing network-initiated data collection prevents data providers from sharing data in an active manner. No matter whether it is the network node or the UE, the existing data collection methods can lead to data wastage due to factors such as storage space. Thus, the two-sided data collection mode also provides a data-sharing procedure initiated by the data provider. The data provider releases the information about the shared data and the expected rewards. Therefore, the number of data collections or the volume of data collected is expected to become a new billing parameter. Data services will offer new revenue growth for operators. Since the potential data consumers that each UE can communicate with directly are limited, the network can assist UE in sharing data by releasing the information of shared data or by storing shared data.

4.2 A DT network solution enabled by the two-sided data collection mode

A DT network is one of the potential data applications of 6G. A DT network can obtain a large amount of real-time status data from the physical network through DP. In addition, high-precision algorithms (e.g., environment modeling algorithms) are crucial to achieve good performance in terms of similarity and prediction accuracy. However, the DT of complex channel environments and large-scale networks requires substantial computational resources. This creates a dilemma for DT solutions. To achieve a trade-off between resource overhead and performance, we propose a DT network solution that aims to use the existing physical network infrastructure as much as possible.

The service management and orchestration (SMO) is responsible for managing functions of a DT network based on the template of a DT network (Fig. 6). We assume that CN functions and some RAN functions (e.g., a central unit of the base station) are deployed on the cloud platform. Through the orchestration function of SMO, it is possible to reuse part of the infrastructure of the physical network to create instances of DT CN functions and instances of DT RAN nodes. The potential reusable infrastructure resources include hot backup nodes and surplus resources based on the tidal effect of network traffic.

For the challenges posed by a large amount of UE and complex channel environments, the two-sided data collection mode enables the possibility of DT UE. According to the data or rewards provided by the network, UE can provide user consent and accept the request of a DT based on the usage habits of UE. For example, when the UE is charging during the early morning hours, users can authorize the network to use their UE as DT UE. Based on the registration information of DT UE, the network selects suitable UE and configures DT UE on the physical UE.

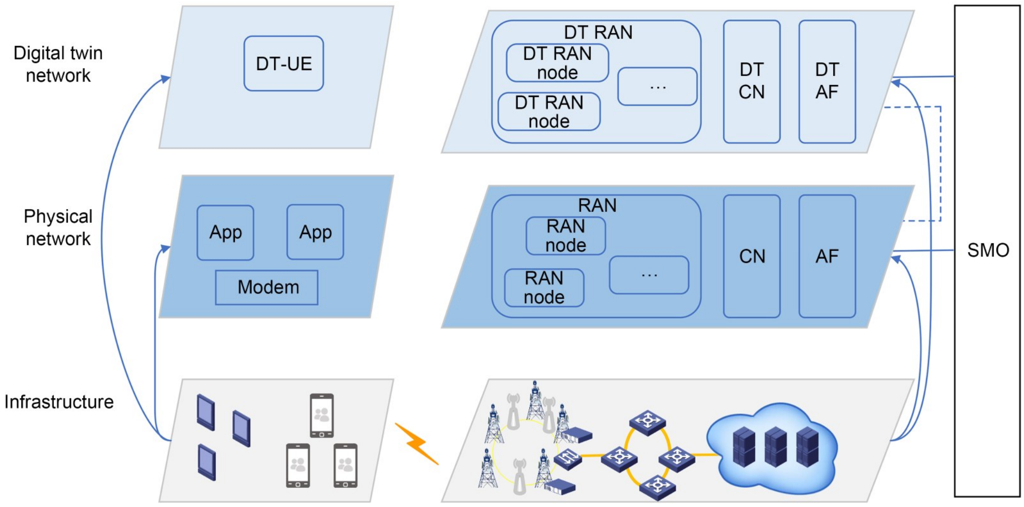

Fig. 6 A digital twin network solution using the existing infrastructure of the physical network

This creates an end-to-end DT network based on the cloud-based network infrastructure and UE.

By frequency or time division multiplexing of spectrum resources, DT UE can achieve a DT for a large number of UE and complex channel environments. Data of DT UE can be collected to train AI models of DT networks. Configurations of DT UE and the collected data can also be used to pre-verify the performance of the network optimization configuration. Moreover, DT UE can provide high performance of DT compared to simulation UE and simulation channels. Furthermore, we will pay more attention to the privacy and security of DT UE. Basically, the isolation between DT UE and physical UE supports the decoupling of DT UE and UE in aspects such as the UE identifier, location information, and context information.

5 Conclusions

We propose a unified data collection framework based on DP to meet the ubiquitous data collection requirements of 6G. The DP protocol stack consists of a DSAP, simplified L2, and optimized L1. The reduction in processing delay by the DP in UL has been validated by test results based on a 6G UE prototype. Moreover, the efficiency improvement of the DP protocol stack increases as the data length increases. For data representation and serialization methods, ASN.1 and ProtoBuf are suitable for small and large amounts of data, respectively. The proposed two-sided data collection mode increases the willingness of data providers to provide data, enabling the possibility of DT UE. DT UE can reuse a large number of UE in the physical network to reduce the infrastructure resource overhead of the DT network. However, some of the rewards (e.g., data or voice quotas) in the two-sided data collection mode may involve billing systems and business models that need to be further validated and refined in commercial networks. We will continue to simulate and analyze the performance of joint source–channel coding and semantic communication for 6G data collection. The DP-based unified data collection framework provides pervasive data functionalities to realize the full potential of the 6G system. Therefore, it is expected to become an important part of the 6G system architecture, providing the data foundation for sensing, AI, and DT networks.

Contributors

Yannan YUAN and Fei QIN designed the research. Jiankang LIU, Yuanyuan WANG, and Jianan CAI developed a prototype of 6G UE and completed the test of the proposed protocol stack. Yannan YUAN drafted the paper. Xiang PAN helped organize the 5G standard status. Yannan YUAN and Dajie JIANG revised and finalized the paper.

Conflict of interest

All the authors declare that they have no conflict of interest.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

3GPP Technical Specification Group Radio Access Network, 2019. Study on RAN-Centric Data Collection and Utilization for LTE and NR, Release 16, Document 3GPP TR 37.816.

3GPP Technical Specification Group Radio Access Network, 2021. Study on NR QoE (Quality of Experience) Management and Optimizations for Diverse Services, Release 17, Document 3GPP TR 38.890.

3GPP Technical Specification Group Services and System Aspects, 2022. Management and Orchestration, 5G Performance Measurements, Release 17, Document 3GPP TS 28.552.

3GPP Technical Specification Group Core Network and Terminals, 2023a. 5G System, Technical Realization of Service Based Architecture, Stage 3, Release 18, Document 3GPP TS 29.500.

3GPP Technical Specification Group Core Network and Terminals, 2023b. 5G System, Network Exposure Function Northbound APIs, Stage 3, Release 18, Document 3GPP TS 29.522.

3GPP Technical Specification Group Radio Access Network, 2023c. LTE Positioning Protocol (LPP), Release 17, Document 3GPP TS 37.355.

3GPP Technical Specification Group Radio Access Network, 2023d. NR, Layer 2 Measurements, Release 17, Document 3GPP TS 38.314.

3GPP Technical Specification Group Radio Access Network, 2023e. NG-RAN, NR Positioning Protocol A (NRPPa), Release 18, Document 3GPP TS 38.455.

3GPP Technical Specification Group Radio Access Network, 2023f. Radio Measurement Collection for Minimization of Drive Tests (MDT), Overall Description, Stage 2, Release 18, Document 3GPP TS 37.320.

3GPP Technical Specification Group Services and System Aspects, 2023g. Architecture Enhancements for 5G System (5GS) to Support Network Data Analytics Services, Release 18, Document 3GPP TS 23.288.

3GPP Technical Specification Group Services and System Aspects, 2023h. System Architecture for the 5G System (5GS), Stage 2, Release 18, Document 3GPP TS 23.501.

3GPP Technical Specification Group Radio Access Network, 2024. NR, Physical Layer Measurements, Release 18, Document 3GPP TS 38.215.

Dai JC, Zhang P, Niu K, et al., 2023. Communication beyond transmitting bits: semantics-guided source and channel coding. *IEEE Wirel Commun*, 30(4):170-177. https://doi.org/10.1109/MWC.017.2100705

Deng J, Yue LX, Yang HW, et al., 2023. A digital twin network approach for 6G wireless network autonomy. Proc IEEE Int Conf on Communications Workshops, p.415-420. https://doi.org/10.1109/ICCWorkshops57953.2023.10283539

ITU-R, 2023. M.2160: Framework and Overall Objectives of the Future Development of IMT for 2030 and Beyond. Recommendation M.2160-0 (11/2023).

Liu GY, Huang YH, Li N, et al., 2020. Vision, requirements and network architecture of 6G mobile network beyond 2030. *China Commun*, 17(9):92-104. https://doi.org/10.23919/JCC.2020.09.008

Xu JL, Ai B, Wang N, et al., 2023. Deep joint source-channel coding for CSI feedback: an end-to-end approach. *IEEE J Sel Areas Commun*, 41(1):260-273. https://doi.org/10.1109/JSAC.2022.3221963

Zhang P, Xu WJ, Gao H, et al., 2022. Toward wisdom-evolutionary and primitive-concise 6G: a new paradigm of semantic communication networks. *Engineering*, 8(1): 60-73. https://doi.org/10.1016/j.eng.2021.11.003