DNA Storage Promotes Long-Term Storage of Internet of Vehicles Data

Guanjin Qu, Huaming Wu, and Ruidong Li

Abstract

With the proliferation of Telematics and autonomous driving, vehicles are generating increasing volumes of data from numerous sensors. To accommodate this influx of in-vehicle data, Data Centers (DCs) must possess ample storage capacity and bandwidth to cater to the rising demand for in-vehicle data uploads. Meanwhile, ensuring the security and integrity of this data is paramount to providing reliable information when needed by car companies or users. The cold nature of in-vehicle data, where most of the data remains unused for long periods of time, poses significant challenges to the storage and operation of DCs. In this paper, we propose a potentially feasible solution for information storage by leveraging DNA molecules. We present the architecture of the new DC and its storage process. Compared to conventional approaches, utilizing DNA molecules for storing in-vehicle data can substantially reduce the operating costs of DCs, while greatly enhancing the storage lifespan and density of information. Finally, we address the practical challenges associated with implementing DNA storage and propose potential solutions to overcome them.

Introduction

The rapid advancements in Artificial Intelligence and cloud-edge collaboration have swiftly reshaped the transportation industry, facilitating the adoption of Telematics and Autonomous Driving technologies. Fig. 1 illustrates the process of vehicles within the Telematics network uploading real-time information about them to Data Centers (DCs) via edge devices within the context of cloud-edge collaboration. However, the realization of autonomous driving necessitates the coordinated operation of various devices such as multiple cameras, millimeter wave sensors, Light Detection and Ranging (LiDAR), and others, resulting in the generation of multidimensional data comprising video streams, point cloud maps, radar waveforms, and more. These sensors generate substantial amounts of data, which must be stored or transmitted to DCs for neural network training and learning, typically ranging from 20 TB to 40 TB per vehicle per day [1].

Current vehicle data logging systems are primarily designed to capture a traditional variety of signals including temperature, braking, throttle settings, engine, speed, etc [2]. However, these systems often struggle to handle sensor data due to the overwhelming volume and complexity of the data. The types of Event Data Recorder (EDR) data currently available often fail to accurately reflect the real-world conditions experienced by vehicles in complex environments. For instance, the US EDR typically records only indicated speed, which may deviate by up to 10% from the vehicle's actual center-of-mass speed during maneuvers involving full braking with ABS control intervention [3]. The rapid advancement of autonomous driving and connected vehicles exacerbates the discrepancy between EDR data and real-world conditions. This disparity can significantly diminish confidence in and utilization of new technologies, particularly in the event of a crash. Consequently, there is a growing need for EDR systems to incorporate additional dimensions and transmit in-vehicle data in real-time via telematics, placing significant pressure on the storage of in-vehicle data.

DCs have become an irreplaceable critical infrastructure for storing the increasing amount of in-vehicle data. Typically, DCs host numerous computing and storage nodes that are interconnected through a dedicated network called Data Center Network (DCN). Serving as a communication backbone, DCNs play a pivotal role in optimizing the operations of DCs. However, traditional DCs suffer from drawbacks such as high energy consumption, extensive physical footprint, and the challenges posed by storing cold data (such as in-vehicle data), which can exacerbate operational costs [4]. Consequently, traditional DCs have struggled to address the storage demands of in-vehicle data adequately.

DNA storage is expected to break this problem, through the use of DNA molecules to store cold data, which can have excellent characteristics such as high storage density, long storage time, and friendly storage environments. Since DNA storage does not need to be electrified and has good antimagnetic properties, data can be stored for a long time and safely. In addition, encryption of the stored data can effectively prevent data leakage in DNA storage. This paper proposes a hybrid storage DC for in-vehicle data, including infrastructure and network framework. By introducing DNA storage facilities in traditional DCs, the deficiencies of existing DCs can be improved to better meet the development of vehicular

Digital Object Identifier: 10.1109/MNET.2024.3481262
Date of Current Version: 15 July 2025
Date of Publication: 15 October 2024

Guanjin Qu and Huaming Wu (corresponding author) are with the Center for Applied Mathematics, Tianjin University, Tianjin 300072, China; Ruidong Li is with the Institute of Science and Engineering, Kanazawa University, Kanazawa 920-1192, Japan.

networks. We present a potentially viable solution for storing in-vehicle data in a DC incorporating DNA storage. The main contributions of our paper are threefold:

- We provide a concise and comprehensive review, including the characteristics of existing DCs, as well as a detailed description of current technical approaches to DNA storage.
- Aiming at the characteristics of in-vehicle data, a new DC architecture is proposed. It introduces DNA storage technology as a means of storing infrequently used data, aiming to reduce long-term storage costs effectively. Simulation evaluation shows that the proposed architecture has the potential to significantly reduce storage energy consumption compared to traditional approaches.
- The paper offers guidance for future research in the field of data storage, especially DNA storage technology. It describes current data trends, evaluates the advantages and limitations of different storage methods, and presents challenges facing DNA storage applications as well as potential solutions.

The remainder of this paper is organized as follows: The section "Challenges for Traditional DCS" describes existing DC models. The section "DNA-Based Data Storage" delves into the approach and attributes of DNA storage. In the section "DCS Incorporating DNA Storage," we propose a new DC architecture integrating DNA storage, followed by a comparison between the new and existing DCs in the section "Performance Evaluation." The section "Conclusion" summarizes our work and highlights the remaining shortcomings of DNA storage.

Challenges for Traditional DCs

DCs serve as indispensable critical infrastructures, particularly in the current era of the Internet of Everything and information proliferation. Typically, DCs accommodate numerous storage nodes and some compute nodes interconnected by a specially designed network fabric. To handle the immense volumes of data, DCs must expand their storage capacity, enhance computing capabilities, and simultaneously optimize the DCN to support higher bandwidth links.

Storage media, as the basic unit of DC storage, determines the basic performance of the DC as well as the difficulty of subsequent maintenance. Currently, common DC storage media include:

- **Hard Disk Drives (HDD):** HDDs are one of the most prevalent storage devices, utilizing spinning disks and moving read/write heads to access data. They typically offer larger capacities and lower costs, making them suitable for storing vast amounts of data.
- **Solid-State Drives (SSD):** SSDs leverage flash memory chips rather than spinning disks to store data, which is faster, more durable, and requires no mechanical parts. With its superior performance, SSDs are gradually replacing HDDs as the preferred storage medium in DCs.
- **CD-ROMs/Blu-Ray Discs:** While less commonly used, some DCs may still rely on CD-ROMs or Blu-ray Discs for long-term storage. These media provide decent capacity and cost advantages, but tend to have slower access speeds than HDDs and SSDs.
- **Tape Libraries:** Tape remains a traditional yet effective storage medium for long-term data archiving and backup purposes. Tape libraries, automated systems capable of housing numerous tapes, offer high-capacity and long-term storage solutions.

> DNA storage is expected to break this problem, through the use of DNA molecules to store cold data, which can have excellent characteristics such as high storage density, long storage time, and friendly storage environments.

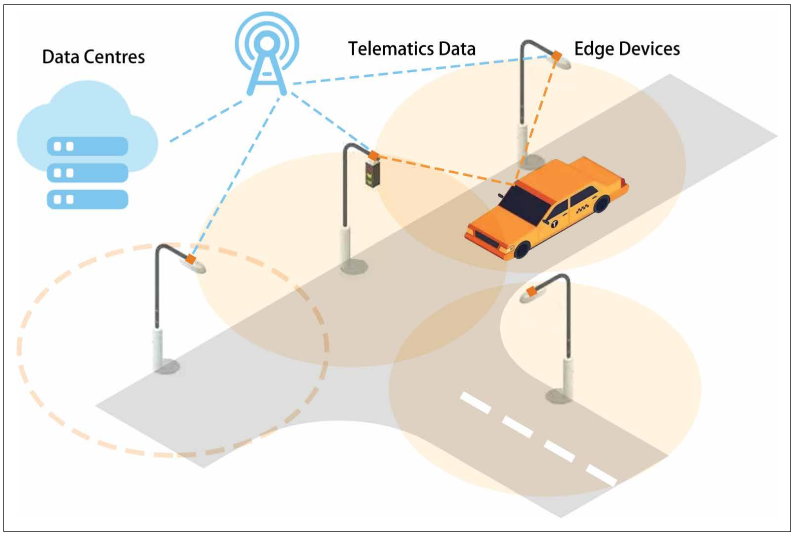

FIGURE 1. The process of uploading in-vehicle information under cloud-edge collaboration.

The structure of the DCN also greatly affects the performance and operating costs of the DC. Currently, common DCNs include leaf-trunk architecture, fat-tree architecture, hybrid fat-tree architecture, and Facebook 4-Post architecture. Different network architectures have different unit costs for various deployment sizes and performance requirements. Architecture selection for DCNs depends heavily on the unique challenges and requirements that DCNs need to address compared to traditional networks such as LANs and WANs. The exponential growth of data has forced DCs to grow larger and larger, for example, Microsoft hosts more than 1 million servers in more than 100 DCs around the world [5]. Such a large system scale further increases the challenges of interconnection, cost, and robustness in network design.

In addition, for in-vehicle data supporting self-driving cars, the requirements for traditional DCs are even higher, including:

- **Data Volume and Velocity:** In-vehicle data can be generated at a relatively high rate, particularly within large-scale fleet or vehicle monitoring systems. As a result, DCs tasked with storing in-vehicle data must possess the capability to handle both high-speed data input and the storage of vast volumes of data.

- **Storage Requirements:** In-vehicle data frequently necessitates long-term storage for activities like follow-up investigations, accident reconstruction, and legal proceedings. Thus, DCs responsible for storing in-vehicle data must offer persistence, scalability, and stringent security measures to guarantee data integrity and confidentiality.
- **Data Privacy and Compliance:** In-vehicle data raises concerns regarding personal privacy and regulatory compliance. Therefore, DCs tasked with storing such data must ensure adherence to pertinent data privacy laws and regulations. Additionally, they should implement appropriate security measures to safeguard this sensitive information.

In the face of these problems and challenges, a new data center architecture is urgently needed to store in-vehicle data. DNA storage, as an emerging storage medium, holds promise as a solution for big data storage in DCs. Its adoption is expected to enhance the efficiency of DCs while reducing storage and operational costs.

DNA-Based Data Storage

DNA-based data storage technology is an emerging storage method, which utilizes DNA molecules as a storage medium to store information. It is expected to overcome the shortcomings of existing storage media and become the next generation of data storage technology. DNA storage involves relevant knowledge in the fields of biology, computers, communications, etc. As [6] proved the feasibility of DNA storage by storing information through DNA molecules, more and more research is driving the gradual scaling and practicality of DNA storage. Fig. 2 shows the whole process of DNA storage.

DNA-Based Data Storage Process

1) Data Encoding: The initial stages of DNA-based data storage involve converting digital data into DNA sequences. Digital data, typically represented in binary form, is converted into a DNA sequence composed of four bases: adenine (A), thymine (T), guanine (G), and cytosine (C). Given that the DNA synthesis and sequencing processes can introduce deletion, substitution, and addition errors, data encoding often integrates error-correcting codes capable of managing errors to ensure effective transmission of information.

2) Data Writing: Data writing in DNA storage refers to synthesizing the encoded digital data into real DNA molecules. The synthesis of DNA molecules is mostly based on the principle of phosphoramidite-based chemical synthesis, including technologies such as column synthesis and chemical synthesis. Recently, advancements in chip synthesis have provided potential benefits for encoding data into DNA molecules, such as reduced synthesis costs and increased writing speeds [7].

3) Data Preservation: DNA molecules can be preserved in various forms, including liquid, dry powder, and encapsulation. By isolating DNA molecules from water and oxygen, they can be stored at room temperature for a long time, potentially up to 500 years. This preservation method ensures the longevity and stability of the stored data encoded in the DNA molecules.

4) Data Reading: By employing DNA sequencing technology, the data encoded within DNA molecules can be read out. Before sequencing, Polymerase Chain Reaction (PCR) and library construction are usually required. After sequencing, the sequence information can be restored to the original data through error correction, decoding, etc. The most common sequencing technology at present is Next Generation Sequencing (NGS), which provides an acceptable bandwidth for reading data.

DNA Storage Features

DNA storage, as a storage method combined with biotechnology, possesses distinctive characteristics when compared to traditional storage semiconductor and magnetic storage media:

- **High Storage Density:** DNA storage achieves an exceptionally high storage density by encoding information within DNA molecules.

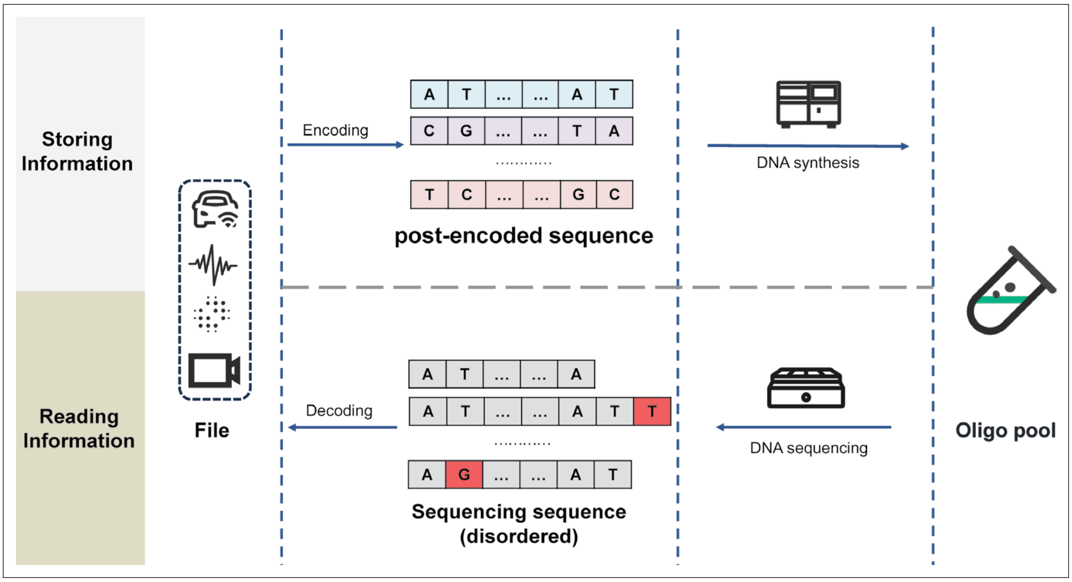

FIGURE 2. Flowchart of DNA storage.

Theoretically, 1 gram of DNA can store up to 455 exabytes of information. Despite the incorporation of channel coding, such as error correction codes, the storage density remains significantly higher, by five orders of magnitude, compared to current storage media.

- **Low Energy Consumption:** Unlike conventional storage mediums, DNA molecules do not require energy to maintain their storage state. As a result, the energy consumption for DNA storage is exceedingly low. Additionally, DNA molecules can retain their biological properties for extended periods, even at room temperature.
- **Longevity:** DNA molecules boast remarkable longevity, enabling storage for exceptionally long durations. In a dry powder state, DNA can be stored at room temperature for at least 500 years.
- **Environmentally Friendly:** The use of DNA molecules for storage is environmentally friendly, as DNA is a naturally occurring substance. This means that the utilization of DNA molecules for storage purposes does not inflict harm on the environment.

Although DNA storage also suffers from defects such as high error rates, expensive synthesis costs, and limited bandwidth for writing and reading data, ongoing advancements in related technologies are anticipated to address these issues. Furthermore, the advantages offered by DNA storage make it well-suited to address the shortcomings of existing storage media and meet the growing demand for data storage in modern DCs.

DCs Incorporating DNA Storage

Based on the unique characteristics of DNA storage and the limitations of current traditional DCs, this paper proposes a novel DC framework that integrates DNA storage technology. By leveraging DNA storage to store cold data and allocating a portion of the computing nodes within the DC to facilitate the reading of DNA data, this framework aims to effectively reduce the scale and energy consumption of DCs. Moreover, for in-vehicle data, DCs equipped with DNA storage offer higher reliability and security for data storage. The inherent stability and longevity of DNA molecules make them an ideal medium for preserving critical in-vehicle data over extended periods. Additionally, the integration of DNA storage technology enhances data security by providing an additional layer of protection against data corruption or loss.

DC Structure With DNA Storage

Since in-vehicle data encompasses various types of information including vehicle sensing data (video, point cloud, and radar waves) and vehicle state information (speed, acceleration, and GPS positioning), it is characterized by diverse data types, large data scales, and low usage frequency. Therefore, compared to traditional DCs, we have integrated DNA storage-related facilities into the improved DC, including DNA synthesis, DNA amplification, DNA sequencing, and other instruments. This enhanced DC leverages DNA storage technology instead of conventional storage media for cold data, such as magnetic tape and hard disk storage. Fig. 3 provides details of the DC incorporating DNA storage.

> The use of DNA molecules for storage is environmentally friendly, as DNA is a naturally occurring substance.

1) Physical Layer: The DC with DNA storage is divided into two distinct modules: a computing equipment module and a biological equipment module.

- **Computing Equipment Module:** This module is similar to traditional DCs, which can be divided into the following three units:
  - **Storage Unit:** This unit, which includes solid-state hard disks, mechanical hard disks, and other storage media, is responsible for storing data.
  - **Communication Unit:** Comprising high-speed switches, routers, optical fibers, and other communication equipment, this unit facilitates communication between different storage units, computing units, and also between the traditional DC and the DNA storage module.
  - **Computing Unit:** Consisting of GPUs, CPUs, and computing servers, this unit handles the computing tasks of the DC as well as the coding and decoding processes of the DNA storage module.
- **Biological Equipment Module:** This module is divided into the following three units:
  - **DNA Synthesis Unit:** This unit includes a DNA synthesizer and related synthesis reagent processing equipment, responsible for synthesizing DNA molecules.
  - **DNA Storage Area:** This area houses the synthesized DNA molecules, serving as the storage medium for DNA data.
  - **DNA Reading Unit:** Comprising a PCR amplification device, a library-building device, and a DNA high-throughput sequencing device, this unit is responsible for reading DNA molecules and translating them into sequenced files.

These two modules are interconnected through optical fibers and other devices to facilitate high-bandwidth data transfer between the computing and biological components of the DC. This integrated approach enables efficient storage, processing, and retrieval of data using both traditional computing resources and DNA storage technology.

2) Network Layer: Since the DC with DNA storage will have a higher data storage capacity compared to traditional centers, it will consequently necessitate increased bandwidth. To address this demand, we propose a combination of optical and traditional electrical switching at the network level. In the computing device module, we employ a hybrid fat-tree structure that incorporates optical switches into the Point of Delivery (PoD) nodes. This approach ensures high-speed connectivity by combining the benefits of the Fat Tree and Thin Tree architectures. The hybrid fat-tree structure leverages the high bandwidth of the Fat Tree Architecture while maintaining the low latency of the Thin Tree Architecture. Additionally, optical switches will be used to link the computing

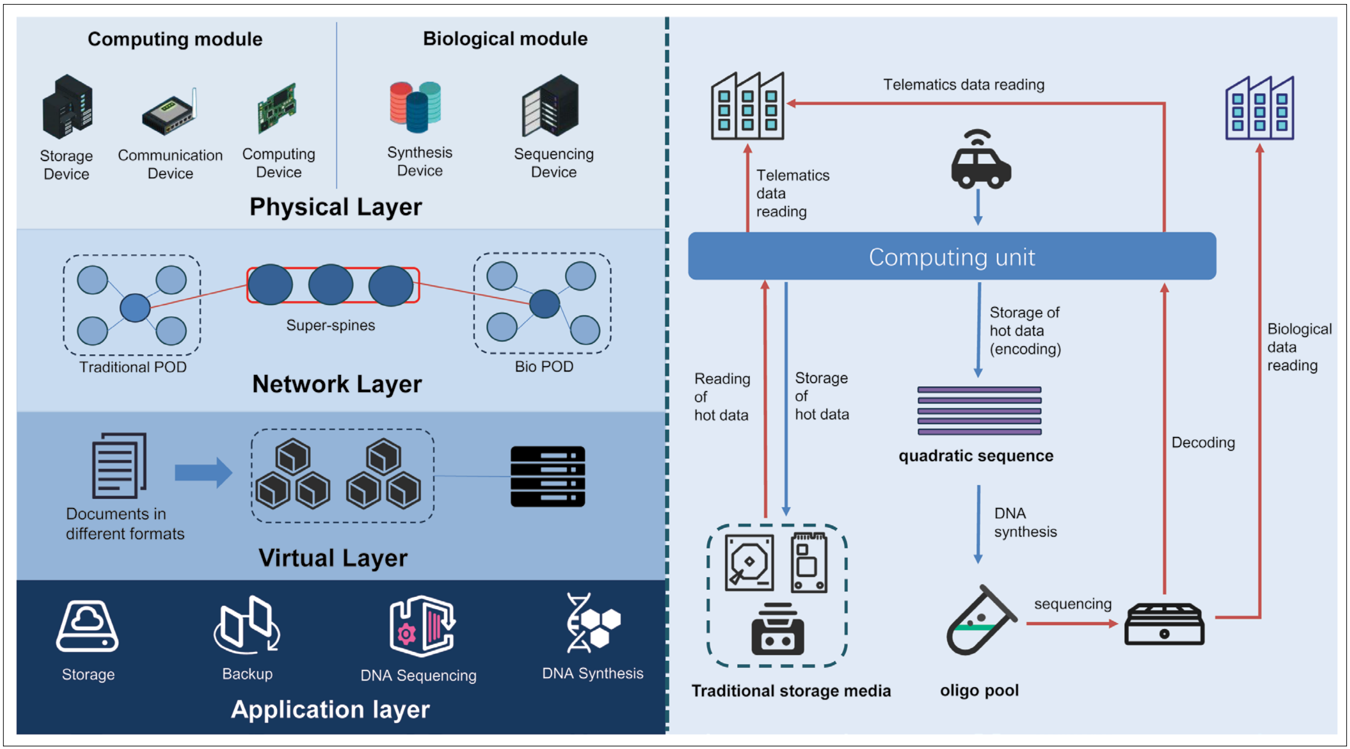

FIGURE 3. The DC framework incorporating DNA storage.

and biological device modules, effectively meeting the bandwidth requirements throughout the DC.

3) Virtual Layer: To enhance the data center's compatibility with various vehicle EDR storage formats and its deployment capabilities, we introduce Software Defined Networking (SDN) and Network Functions Virtualization (NFV). With SDN and NFV, multiple virtual networks are established on top of a shared physical network underlay, enabling independent implementation and management of each virtual network. By decoupling the virtual network from the underlying physical network, customized network protocols and management policies can be introduced to facilitate performance isolation and ensure application Quality of Service (QoS). Control and management systems can be set up to customize the control and management systems under different virtual networks for the needs of different automotive companies and different data format requirements. Furthermore, the network virtualization environment provides isolation, thereby minimizing security threats to in-vehicle data. This ensures robust protection and confidentiality of sensitive automotive data stored within the DC.

4) Application Layer: The primary focus of the DC's application lies in the storage of in-vehicle data captured by the vehicle, including both EDR information and sensory data. In contrast to traditional DCs, those integrating DNA storage services can extend their capabilities to encompass bioinformatics-related services. These may include DNA synthesis services, gene design, protein prediction, and more, which can be executed during idle windows, such as late at night or during periods of low vehicle data transmission. The physical proximity of computing units to biological units within the new DC facilitates a more seamless utilization of computing resources for processing biological services. Consequently, this enhances the efficiency of experimental processing in related fields compared to traditional biological service providers. The integration of biological and computational services is anticipated to advance the development of these fields, fostering innovation and accelerating scientific discovery.

Data Writing and Reading

The primary objective of a DC handling in-vehicle data is to securely store vast amounts of this data while ensuring its integrity and privacy. In critical scenarios, such as accidents or autonomous driving model training, the data must be swiftly retrievable within specified timeframes, with its integrity guaranteed. These requirements necessitate a DC with substantial storage capacity, adequate bandwidth, and robust security measures. DCs incorporating DNA storage effectively address these requirements by leveraging a combination of storage media, including DNA storage and flash storage, along with advanced transmission technologies such as optical and electrical switching. Below, we outline the methods for data writing and reading within DCs equipped with DNA storage.

1) Data Writing: The information collected by the vehicle first needs to be offloaded to the DC through a discriminator, which can exist at the roadside or at the vehicle side. The data offloaded to the server will first pass through a dedicated computing unit, for the low frequency of cold data will be encoded with error correction codes, and the encoded data will be transmitted to the DNA storage module, for the high frequency of hot data will be encoded with compressed sources, and the encoded data will be transmitted to the storage unit in the computing module. Due to the volatility of the amount of information data collected by the vehicle and DNA synthesis with the current bandwidth compared to the traditional storage media is low, there will be a certain buffer in the computing module, when the peak data acceptance rate exceeds the speed of DNA synthesis will be temporarily stored with the buffer, and in the low-peak by the buffer to transfer out. After DNA synthesis, the data will be stored in the form of dry powder in a special DNA storage room after dehydration and other operations. In contrast to other storage media that require electricity, dust-free, and heat dissipation, the DNA storage room only needs to maintain a light-proof and dry environment to ensure the integrity of the information.

2) Data Reading: For hot data requiring immediate access, the reading method is similar to that of a traditional DC. However, for cold data stored in DNA molecules, the procedure involves several steps. Initially, the dry powder DNA undergoes PCR, followed by library construction and high-throughput sequencing. Subsequently, the sequenced sequences are transmitted to the computational unit where they are decoded to retrieve the original data. It's worth noting that, leveraging the characteristics of DNA molecules, for large-scale data required by off-site DCs, specific DNA molecules can be PCR amplified in the DC and then mailed directly to the off-site location for sequencing and decoding. Additionally, unlike tape libraries where random reads are impractical, DNA storage allows for the retrieval of specified data using specific primers or alternative methods.

Performance Evaluation

We compare DCs using DNA storage technology with traditional ones across various dimensions such as cost and performance. Although DNA storage currently faces high synthesis costs and low bandwidth, it offers considerable advantages over tape and other media in terms of operational costs, retention time, and storage density, especially for the vast amounts of cold data typical in IoT applications. Furthermore, leveraging the biological characteristics of DNA molecules and their resultant functionalities may foster the emergence of novel application paradigms within DCs. As DNA storage serves primarily as alternative to tape libraries for storing cold data, we provide a performance comparison between DNA storage and tape libraries in Fig. 4(b). Our simulation experiment uses a server with an Intel(R) Xeon(R) Gold 6348H CPU @ 2.30 GHz and 1 TB of RAM, and the simulation code is written in Python.

Storage Cost

Compared to traditional storage media, the storage cost of DNA storage includes DNA synthesis, DNA sequencing, etc., of which the cost of DNA synthesis occupies the major cost. Based on predictable low-cost DNA synthesis methods [9], DNA storage is expected to fall below $10,000 per TB. While the cost of DNA synthesis remains higher than that of existing storage media, it's important to note that the overall storage cost in DCs also encompasses preservation and operational expenses.

In Fig. 4(a), we provide a comparative analysis of operational costs between DCs employing DNA storage and traditional DCs for storing 1 ZB of data. It can be seen that despite the extremely high cost of DNA molecule synthesis, the subsequent maintenance cost is very low as the DNA molecules require little operation and maintenance after encapsulation. The total cost will be lower than HDD-based DCs after 72 years of operation and lower than tape-based DCs after 122 years of operation. This is because tapes need to be rewritten and read every 15-30 years, while HDDs have drawbacks such as high energy consumption and the need to dissipate heat. For in-vehicle data, DCs combined with DNA

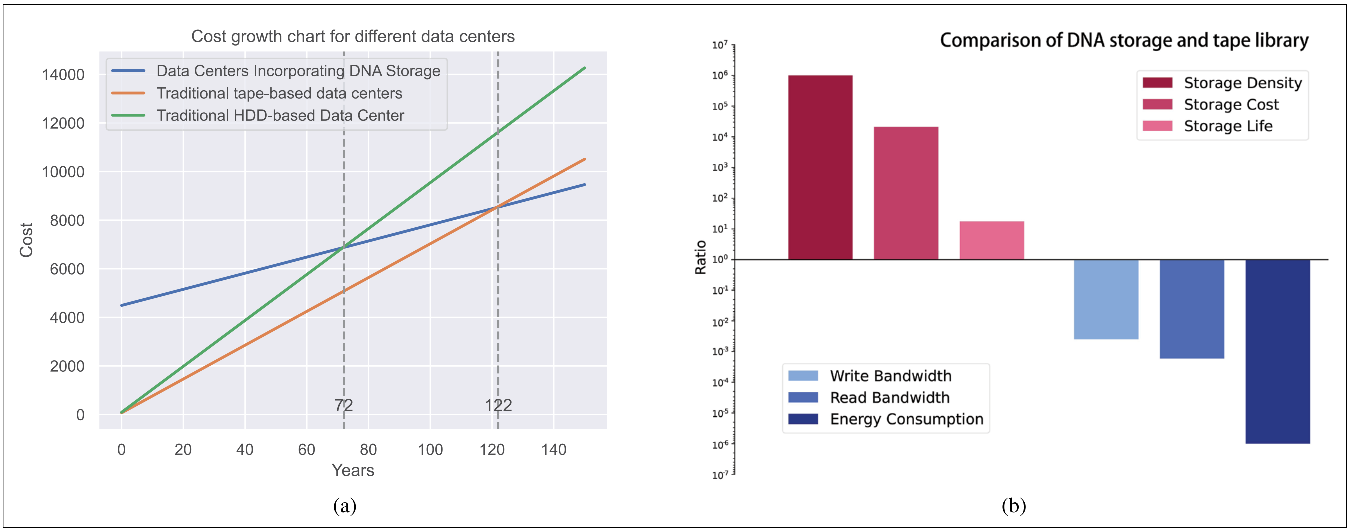

FIGURE 4. Effectiveness of DCs incorporating DNA storage vs. traditional DCs. a) shows the trend of operating costs for storing 1ZB of data under different DCs, where the storage ratio of DCs incorporating DNA storage is 70% DNA storage, 15% HHD storage, and 15% SSD storage, and the storage ratio of storage centers based on tape libraries is 60% tape storage, 25% HHD storage, and 15% SSD storage; and the storage ratio of HDD-based storage centers is 85% HHD and 15% SSD storage. The operating cost per unit time is estimated based on the adjustment of [8]. b) shows the performance comparison between DNA storage and its main alternative storage medium, the tape library.

DCs employing DNA storage offer distinct advantages, including low energy consumption, extended storage lifespan, and reduced operation and maintenance costs, making them more suitable for storing cold data compared to traditional DCs.

storage are expected to reduce the operational cost of long-term storage. Moreover, the costs associated with DNA data storage and synthesis are on a declining trend. In 2016, the Genome Project-write (GP-write) program was announced, which aims to dramatically reduce the cost of DNA synthesis [10].

Energy Consumption

DCs incorporating DNA storage have a huge energy consumption advantage over traditional DCs. Recent energy statistics show that the DC industry accounts for 1.3% of the world's electricity consumption [11]. The annual energy consumption of global DCs is estimated to be more than 20.5 billion kWh in 2018 [12], and storage devices in some DCs consume 20-30% of the energy. The power consumption of DNA storage will be reduced by 5 orders of magnitude compared to traditional storage media 4b, which will significantly reduce the energy consumption of DCs. On one hand, this is because DNA storage can be stored at room temperature, which requires almost no power consumption; on the other hand, DNA storage does not require the use of a cooling system, however, in traditional DCs, about 38% of the power is used for cooling systems.

1) Storage Lifetime: DNA has an extremely long shelf life, compared to tape libraries where information is stored on two-dimensional (2D) substrates that need to be rewritten every 15 to 30 years to prevent data loss. DNA can be encapsulated and effectively stored for more than 500 years. This is at least 100 times the storage time of current tape libraries. In addition, a new storage scheme has been proposed, ensuring that DNA stored at 9.4 degrees Celsius remains readable for up to 20,000 years [13].

2) Storage Density: State-of-the-art tape libraries have an information density of about 1.8×10^{-4} PB/cm^3. In contrast, DNA storage boasts an impressive information density of up to 295 PB/g [13]. This indicates that DNA storage can achieve information densities approximately 100,000 times greater than those of tape libraries.

3) Read-Write Speed: The read and write speeds of DNA storage are currently significant drawbacks. Compared to tape libraries, which are often used for cold data storage, the write speed of DNA storage is merely around 1 MB/s [7], making it roughly 100 orders of magnitude slower than magnetic tapes.

This situation is expected to improve with the further development of chip synthesis, which is currently state-of-the-art at $25 * 10^6$ sequences per square centimeter [7]. Benefiting from the development of high-throughput sequencing technology, currently based on the latest high-throughput sequencing technology has been able to do 72 Tb single run throughput, i.e., theoretically read 0.29 MB/s, compared with the 400 MB/s of magnetic tape there is still a big gap. It should be noted that DNA storage can also be based on some of the characteristics of the DNA molecule to operate, and then realizing the function of the traditional storage media is difficult to do. For example, PCR amplification of DNA molecules can quickly store multiple copies of the information. The DNA molecule after copying has a very high bandwidth for data transmission based on the form of transportation due to its extremely small size and easy transportation. In addition, the design of primers based on two different segments of the DNA sequence allows for random reading of stored information, which is an unparalleled advantage over tape libraries, the most likely alternative medium for DNA storage.

Challenges and Solutions

Although DNA storage is expected to be integrated into DCs and for storing in-vehicle data to address existing limitations, however, it still faces the following difficulties and challenges before it can be widely implemented:

- **Difficulties in DNA Synthesis:** DNA synthesis remains a significant bottleneck hindering the widespread adoption of DNA storage. On the one hand, the synthesis flux of DNA is low, and the length of each sequence is limited, resulting in lower information storage speeds compared to existing storage media. On the other hand, the cost of DNA synthesis is prohibitively high compared to reusable mechanical hard disks and other storage media. Although DNA synthesis technology based on array synthesis offers higher synthesis density, advancements are needed to enhance DNA synthesis throughput and reduce costs.
- **Lack of Source Coding Solutions for DNA Storage:** Unlike traditional storage media, which mainly face erasure errors, DNA storage can suffer from synchronization errors such as deletions and additions, significantly affecting information retrieval. While error correction algorithms exist for synchronization issues in DNA storage, there is a notable lack of source coding algorithms specifically designed for it. Consequently, in current storage systems, a single unrecovered synchronization error in a file can lead to the failure of the entire file retrieval process, which is unacceptable, especially for onboard data applications. Thus, there is an urgent need for source coding algorithms capable of mitigating synchronization errors, particularly for onboard data formats specifically designed for DNA storage.
- **Lack of Uniform Standards:** The integration of DNA storage into DCs faces challenges due to the absence of standardized regulations. Specifically, the phosphoramidite-based chemical synthesis used in DNA synthesis involves toxic reagents, raising significant safety and environmental concerns. Consequently, establishing unified standards is crucial for the safe and effective implementation of DNA storage in DCs.
- **Difficulties in Random Reading:** Another challenge for DCs incorporating DNA storage technology is the random reading of files. Since current DNA sequencing based on next-generation sequencing

technologies cannot sequence in real time, it is not possible to read the specified files. Although molecular sequencing technology can achieve real-time reads, the reads still need to be read in the order of the DNA sequence. It is still difficult to read specific regions of the DNA sequence. Achieving random reads for DNA storage will facilitate the introduction of DNA storage into the DC. Currently, there are some algorithm-based approaches to achieve random reads for DNA storage [14], but it is still a challenge to achieve data reads under large-scale DNA storage.

Conclusion

Traditional DCs face increasing challenges in storing in-vehicle data due to the predominance of cold data and the proliferation of automated driving, which generates an overwhelming amount of data. DCs employing DNA storage offer distinct advantages, including low energy consumption, extended storage lifespan, and reduced operation and maintenance costs, making them more suitable for storing cold data compared to traditional DCs. While the integration of DNA storage into DCs is not yet widespread, it can already fulfill the entire storage process. The primary remaining challenges lie in cost reduction and throughput improvement. It is predicted that the cost of DNA synthesis will halve approximately every 30 months, based on trends observed over the past two decades [15]. This suggests that DNA storage is poised to become a pivotal storage medium for DCs in the future, particularly in storing in-vehicle data for the rapidly advancing field of Telematics.

Acknowledgment

This work was supported in part by the National Key Research and Development Program of China under Grant 2020YFA0712100, in part by the National Natural Science Foundation of China under Grant 62071327, and in part by the Graduate Student Innovation Award Program for Arts and Sciences under Grant B2-2022-001.

References

[1] R. Wang, L. Liu, and W. Shi, "HydraSpace: Computational data storage for autonomous vehicles," in *Proc. IEEE 6th Int. Conf. Collaboration Internet Comput. (CIC)*, Dec. 2020, pp. 70–77.

[2] S. Ilic, J. Katupitiya, and M. Tordon, "In-vehicle data logging system for fatigue analysis of drive shaft," in *Proc. Int. Workshop Robot Sens., ROSE.*, 2004, pp. 30–34.

[3] P. Vertal et al., "Auswertung von CDR-Crashversuchen," *VKU Verkehrsunfall und Fahrzeugtechnik*, vol. 56, no. 10, pp. 1–10, 2018.

[4] W. Xia et al., "A survey on data center networking (DCN): Infrastructure and operations," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 1, pp. 640–656, 1st Quart., 2017.

[5] (2015). *Microsoft's Cloud Infrastructure: Datacenters and Network Fact Sheet*. [Online]. Available: http://download.microsoft.com/download/8/2/9/8297F7C7-AE81-4E99-B1DB-D65A01F7A8EF/Microsoft_Cloud_Infrastructure_Datacenter_and_Network_Fact_Sheet.pdf

[6] G. M. Church, Y. Gao, and S. Kosuri, "Next-generation digital information storage in DNA," *Science*, vol. 337, no. 6102, p. 1628, Sep. 2012.

[7] B. H. Nguyen et al., "Scaling DNA data storage with nanoscale electrode wells," *Sci. Adv.*, vol. 7, no. 48, Nov. 2021, Art. no. eabi6714.

[8] V. S. Prakash, Y. Wen, and W. Shi, "Tape cloud: Scalable and cost efficient big data infrastructure for cloud computing," in *Proc. IEEE 6th Int. Conf. Cloud Comput.*, Jun. 2013, pp. 541–548.

[9] P. L. Antkowiak et al., "Low cost DNA data storage using photolithographic synthesis and advanced information reconstruction and error correction," *Nature Commun.*, vol. 11, no. 1, p. 5345, Oct. 2020.

[10] J. D. Boeke et al., "The genome project-write," *Science*, vol. 353, no. 6295, pp. 126–127, 2016.

[11] *Data Centres and Data Transmission Networks*. Accessed: Oct. 1, 2024. [Online]. Available: https://www.iea.org/energy-system/buildings/data-centres-and-data-transmission-networks

[12] N. Jones, "How to stop data centres from gobbling up the world's electricity," *Nature*, vol. 561, no. 7722, pp. 163–166, Sep. 2018.

[13] L. Song et al., "Robust data storage in DNA by de Bruijn graph-based de novo strand assembly," *Nature Commun.*, vol. 13, no. 1, p. 5361, Sep. 2022.

[14] L. Organick et al., "Random access in large-scale DNA data storage," *Nature Biotechnol.*, vol. 36, no. 3, pp. 242–248, Mar. 2018.

[15] N. C. Seeman and H. F. Sleiman, "DNA nanotechnology," *Nature Rev. Mater.*, vol. 3, no. 1, pp. 1–23, 2017.

Biographies

Guanjin Qu (guanjinqu@tju.edu.cn) is currently pursuing the Ph.D. degree with the Center for Applied Mathematics, Tianjin University. His main research interests include DNA storage, edge computing, reinforcement learning.

Huaming Wu (Senior Member, IEEE) (whmingg@tju.edu.cn) received the B.E. and M.S. degrees in electrical engineering from the Harbin Institute of Technology, China, in 2009 and 2011, respectively, and the Ph.D. degree (Hons.) in computer science from Freie Universität Berlin, Germany, in 2015. He is currently a Professor with the Center for Applied Mathematics, Tianjin University. His research interests include wireless networks, mobile edge computing, the Internet of Things, and DNA storage.

Ruidong Li (Senior Member, IEEE) (liruidong@ieee.org) received the bachelor's degree in engineering from Zhejiang University, China, in 2001, and the Ph.D. degree from the University of Tsukuba in 2008. He is currently an Associate Professor with the College of Science and Engineering, Kanazawa University, Japan. His research interests include big data networking, information-centric networks, network security, and quantum internet.