TOWARDS DIGITAL TWINS FOR OPTIMIZING METRICS IN DISTRIBUTED STORAGE SYSTEMS - A REVIEW

May Itani

Layal Abu Daher

Ahmad Hammoud

1. INTRODUCTION

Data generated exclusively by Internet of Things (IoT) devices is predicted to double from 15.1 billion in 2020 to more than 29 billion IoT devices in 2030. With this data explosion and the advent of 6G in addition to the increased demand for data-intensive applications including social networks, artificial intelligence, and the Internet of Things there is a crucial need for highly available and reliable distributed storage systems. To ensure availability, scalability, reliability, and cost-effective modern distributed cloud storage systems, such as Amazon Simple Storage Service (Amazon S3), Hadoop Distributed File System (HDFS), and Microsoft Azure, Google Cloud use data replication and erasure codes (Liu et al.,2022; Statista, 2023). These systems offer cloud services to users and need efficient monitoring and management techniques since they rely on a pool of computing and communication resources that can be accessed based on demand. Different performance and energy efficiency metrics are employed to assess cloud computing systems. Those metrics include: (i) Communication Network Energy Efficiency which is defined as the energy to deliver a single bit of information, (ii) Network power usage effectiveness ratio between total IT power and power consumed by network equipment, (iii) Energy proportionality coefficient which constitutes the degree of energy proportionality of a device or a system performance, (iv) Communication latency between data center gateway and computing servers, (v) Uplink/Downlink hop distance between data center gateway and computing servers, (vi) Inter-Server communication latency between computing servers, (vii) Inter-Server hop distance between computing servers, (viii) Average latency of accessing database from computing servers, (ix) Actual bandwidth servers can exploit under full load, (x) Uplink/Downlink error rate of the paths between data center gateway and servers, (xi) Inter-Server error rate of the network paths between computing servers, (xii) Average link utilization ratio, (xiii) Average server degree connectivity (Fiandrino et al., 2015).

From the above, we focus on selected performance metrics that are introduced and optimized in different research works during failure recovery of DSSs. These constitute: Repair bandwidth, load imbalance, latency, disk failure, and reliability.

- ***Repair Bandwidth*** is the transmission efficiency during the recovery process in DSSs, is defined as the average amount of data symbols transferred from the surviving nodes to the new node in the recovery process
- ***Load Imbalance*** is an imbalance in the distribution of recovery load among various storage nodes in the system which can lead to congestion in certain nodes and prolonged recovery times
- ***Latency*** is the time taken to process a data request, measured relative to the time at which it enters the DSS
- ***Disk Failure*** is act when information can’t be accessed due to software or hardware errors (example hard disk malfunction)
- ***Reliability*** is the ability of the system to remain available over a period of time

In this paper, we provide an up-to-date overview of recent research work on distributed storage systems that have used erasure coding for the last four years (2020+). Then we highlight a new trend in using digital twins for monitoring data centers. After we introduce different related performance metrics, in section 2, we present some related surveys to highlight the contribution of our work and show how it differs from existing published surveys. In section 3, we review papers related to adopting erasure codes in DSSs with a table of comparisons and classifications emphasizing the performance metrics that the different authors focused on. In section 4 we survey recent literature where Digital Twins (DTs) are involved in monitoring DSSs and other systems for failure diagnosis and for assisting managers in intelligent decision-making. To conclude, we identify important open issues to inspire future studies for fully efficient DSSs.

2. RECENT RELATED SURVEYS

Several surveys in literature have addressed different aspects of distributed storage systems. We select the most recent surveys available in literature and present a brief overview of the objective of each survey. Authors in (Chiniah et al., 2022) presented a review on the implementation of erasure codes in DSSs and a thorough understanding of the benefits as opposed to replication. Namely Reed-Solomon, hierarchical, self-repairing, regenerating and locally repairable codes are reviewed. Work in (Aatish et al., 2020) stated that replication has significant drawbacks and conducted a survey showing how erasure codes are the most promising alternative for achieving redundancy in data centers. Authors considered six major DSSs: Amazon's Dynamo, Facebook's Cassandra, Facebook's Haystack, Google's Bigtable, Yahoo's pNuts & Microsoft's Azure and compared them in terms of the design principle used in implementing these systems, query and data models, failure detection and recovery, consistency & security. The review also tackled existing erasure code families to make sure how codes could be adopted in such systems. The survey in (Xiao et al., 2020) summarizes a variety of data update schemes introduced in literature that are based on erasure coding. Authors categorized the surveyed literature in two different ways: based on the metric they optimize and based on a 5-tier model that may help in joint optimization of multiple metrics. This five-tier model was inspired by Zhou's idea of operation tiers that constitute: Coding tier, scheduling tier, network tier, I/O tier and hardware tier. Techniques to improve response to failures were suggested. Authors in (Al-Shami et al., 2020) surveyed literature for the five-year span from 2015 to 2020. Authors classified the literature and provided comparisons of related papers based on paper objectives, tools, and codes used. All the above surveys provided summaries for researchers for the development of more future work. To the best of our knowledge, no survey has comprehensively reviewed recent literature for the past three years that is related to failure of DSSs and/or tackled failure recovery aspects specifically.

3. FAILURE RECOVERY OPTIMIZATION LITERATURE REVIEW

Different research work related to failure recovery optimization in distributed storage systems is present in literature. We selected recent papers that use various erasure codes and target specific metrics in the recovery process. Table 1 classifies the data into two main groups: Studies that target failure prevention only and others that target failure recovery.

Authors in (Chen et al., 2020; Liang et al., 2020; Gao et al. 2021; Dhore et al., 2022) exploited different erasure codes to help in adding redundancy to DSSs to specifically ensure reliability. Chen et al. targeted the decoding process of XOR-based erasure codes as per speed. The decoding process can be improved by enhancing decoding computational complexity and cache behavior. A new decoding algorithm that utilizes CPU cache more efficiently is designed and verified to achieve considerable decoding speed gain on different platforms. Liang et al. 2021 addressed the usage of non-volatile memory and presented a new erasure code with high computation performance eliminating diagonal/row parity dependencies. This code, EaR (Endurance-aware RAID-6), helps enhance system endurance and can reduce read/write overhead by around 30% averagely under real workloads compared to existing XOR-based erasure codes Row Diagonal Parity (RDP) and EVEN-ODD. Ye et al. introduced the Minimum Parity Reconstruction (MPR) Codes that minimize parity block recovery cost and help in reliability maintenance. Authors then proposed hybrid regenerating codes that use the MPR method to reduce repair cost, disk I/O and network traffic during the repair process to provide quick recovery. (Dhore et al., 2022) studied the benefits of cloud computing addressing three main storage tiers: cloud, fog and computer or main memory layers. Authors applied a Hash-Solomon algorithm that splits data into three parts before it encrypts/decrypts and demonstrated how this three-tier architecture enhances DSS performance under attack failure scenarios.

Other papers did not only focus on designing and implementing codes for failure prevention, they considered the failure recovery process (Xu et al., 2020; Zhang et al., 2022; Mu. et al.,2022 and others as specified in Table 1). To mention some, Xu et al. utilized orthogonal arrays and proposed a Deterministic Data Distribution (D3) approach that distributes data/parity blocks uniformly among DSS nodes to provide an efficient failure recovery approach that minimizes the cross-rack repair traffic against a single node failure. The proposed algorithm was implemented in Hadoop Distributed File System (HDFS) with a cluster of 28 machines and proved to enhance the failure recovery speed up to 2.49 times relative to RS codes and 1.38 times relative to Longitudinal Redundancy Check (LRC). Zhang et al. presented a random search failure recovery technique that enhances single-node failure recovery of XOR-based erasure codes. The technique is simulated for optimized recovery by processing minimum amount of data completed in a polynomial time. The proposed algorithm was demonstrated on a real storage system and was shown to reduce the amount of data in the failure recovery process up to 30.0% compared to traditional recovery methods. (Mu et al., 2022) worked on enhancing cross-node and cross-rack bandwidth loss that

https://digitalcommons.bau.edu.lb/stjournal
DOI: 10.54729/2959-331X.1118

results upon failure and degrades the efficiency of failure recovery wasting additional resources. Authors worked on improving the erasure code storage strategy in Hadoop 3.x, by introducing an H-V code (Horizontal/Vertical Code) that adds RS parity check inside the data nodes to minimize cross-node and cross-rack data transmission during recovery, reduce the occupation of cross-rack bandwidth, and improve recovery efficiency. Authors proved that H-V can improve the cross-node/cross-rack bandwidth of RS by at least 25% and the storage redundancy by around 19%.

Table 1. Classification of selected recent related work

Reference	Erasure Codes/Algorithms Used	Problem	Proposed Solution	Objective
Group A. Failure Prevention				
Chen et al., 2020	XOR-based erasure codes	Data protection for large-scale storage systems	-Improving the decoding speed for XOR-based erasure codes. -Utilizing CPU cache more efficiently	Reliability
Gao et al., 2021	Reed-Solomon codes (RS)	Single-event upsets (SEUs) that can causing failures due to implementation of RS-EC on FPGAs	- Proposing Fault detection and location scheme based on partial re-encoding for the faults in the user memory of the RS-EC decoder. - Improving fault location performance by adding check bits to generator matrix	Reliability
Dhore et al., 2022	Hash-Solomon codes	Security, reliability, and secrecy of cloud data.	- Analyzing usage of Hash-Solomon codes for a more stable and safe cloud storage -Utilizing different tiers from cloud to fog to local server layer	Security & Reliability
Liang et al., 2022	Liberation codes Thou codes	Data protection against co-occurring failure events	-Constructing Thou codes from liberation codes for an efficient searching approach over a cluster of matrices	Reliability
Group B. Failure Recovery				
Ye et al., 2020	Minimum Storage Regenerating (MSR) codes Hybrid Regenerating Codes (Hybrid-RC)	- Disk I/Os and network traffic for recovering unavailable data. - Repair bandwidth under the minimum storage during recovery in real systems	Utilizing the superiority of MSR codes to compute a subset of data blocks and using parity blocks for reliability maintenance	Disk I/O
Xu et al., 2020	Reed-Solomon codes (RCS) Locally Repairable Codes (LCS) Deterministic Data Distribution (D 3)	- Heavy cross-rack traffic, load imbalance, and random access highly affecting failure recovery.	- Minimizing the cross-rack repair traffic against a single node failure.	Load Imbalance
Qiu et al., 2020	Minimum Storage Regenerating (MSR) codes Reed-Solomon codes (RS) Erasure Codes Fusion (EC-Fusion)	- Triple disk failure - Load imbalance during recovery process	- Decreasing the computational overhead and storage cost concurrently for write-intensive workloads.	Load Imbalance

Reference	Erasure Codes/Algorithms Used	Problem	Proposed Solution	Objective
Wang et al., 2020	Local Reconstruction Code (LRC) Hitchhiker (HH) code Combination between algorithms	Increase consumption of network traffic and disk I/O under data recovery process	- Providing an efficient switching algorithm between LRC and HH code with low network and computation costs - Low degraded read latency preserving a low storage overhead	Delay or Latency
Yang et al., 2022	Parity chunks Content Delivery Networks (C2DNs)	- Performance degradation under server unavailability leading to miss ratios	- Achieving a low miss ratio, high availability, high resource efficiency, and close-to-perfect write load balancing by utilizing CDN. - Introducing erasure coding into the CDN architecture and use the parity chunks for re-balancing write load across servers.	Load Imbalance
Zhang et al., 2022	XOR-based erasure codes (SA-RSR) random search recovery algorithm	- Speed up of data recovery process	- Implementing simulated annealing algorithm to search for an optimized recovery solution that reads and transmits a minimum amount of data. - Speeding up single-node failure recovery of XOR-based erasure codes.	Delay or Latency
Mu et al., 2022	Reed-Solomon codes (RS) HRS(n,k) - VRS(n′,k′) abbreviated as H-V	- severe cross-node/cross-rack bandwidth loss during recovery process wasting additional resources.	- Improving the erasure code storage strategy in Hadoop 3.x -Proposing HRS(n,k) - VRS(n′,k′) abbreviated as H-Vcode by adding RS parity check inside the data nodes -Effectively reducing cross-node/cross-rack data transmission during recovery - Reducing the occupation of cross-rack bandwidth and improving recovery efficiency.	Bandwidth Loss

4. DIGITAL TWINS FOR FAILURE RECOVERY

Digital twins are digital replicas of a physical system to help in accurately monitoring and predicting system behavior. Digital Twins help in the process of decision-making by providing a virtual clone of a complex system.

In (Jain et al., 2019) authors discussed the growing importance of distributed photovoltaic (PV) systems integrated into buildings and roofs for smart building applications. This study offered a new way of using digital twins to diagnose faults in these systems. The paper describes the design method, numerical analysis, simulation study, and experimental validation of this method. The digital twin developed in this study aimed to provide real-time estimation of the measurable components of a PV power conversion unit (PVECU) including both PV source and source level power converter. The fault diagnosis was performed by generating and evaluating an error residual vector, which is the difference between the estimated and measured outputs. To validate this approach, the researchers modeled a PV panel-level power converter with sensing, processing, and operation capabilities for real-time fault diagnosis. The results of this study showed that ten difficult faults have been observed in the PVECU. Notably, the fault detection time of the power converter and electrical sensors was less than 290 with an identification time of less than 4 *ms*. Moreover, the time to fault detection is less than 80 *ms* whereas, the time for identification in the PV panel was less than 1.2 *s*. Overall, this method demonstrates higher sensitivity in terms of fault detection than other methods, being able to detect 20% drift in the electrical sensor gain and 20% shading of the PV panel solar cells.

https://digitalcommons.bau.edu.lb/stjournal
DOI: 10.54729/2959-331X.1118

Authors in (Zohdi et al., 2022) highlighted the importance of data center services and subsequent energy use. The significant increase in data centers has drawn significant attention and regulatory inspection regarding the handling of waste heat and its efficient utilization. The objective of this study was to develop an integrated framework that combines Digital Twin technology and Machine Learning to optimize these systems. The primary emphasis lies in the management of ventilation and cooling for the data unit/processor bases within the data center to achieve a desired temperature while minimizing the amount of energy consumed. A data-center model problem was created, with the flow rates, air-cooling at various ventilation ports, and ground-level conduction-based and processor base cooling were considered as the design variables. Using the first law of thermodynamics and based on the Navier-Stokes equations, a thermo-fluid model was built, and based on this model, a quick, stencil-based, iterative solution method was developed. The built model was then coupled with a genomic-based machine-learning algorithm to create a Digital Twin that can operate in real-time or even at a faster rate than the actual physical system making it suitable as an adaptive controller or as a design tool. The authors in this study provided numerical examples to demonstrate the framework and the model built.

Authors in (Hong et al., 2021) presented NetGraph, a digital twin platform that was proposed to enable automatic and intelligent data center network management. NetGraph aims to produce a virtual representation of the physical network with all its elements and to simplify the network flow service management processes. Four functional elements were incorporated in the platform's new operational model for network management, including maintenance: unified digital twin model management, automatic configuration deployment and translation, inventory search, and network element validation. With the implementation of NetGraph in Huawei's data center networks, it has effectively served over fifty thousand devices with millions of network connectivity information such as links and termination points. NetGraph provided comprehensive coverage of network models and supported many critical functions in real-life network scenarios.

An overview of digital twin technology and its application domains is provided in (Mashaly et al, 2021). Thinking of digital twin technology brings to mind that it is similar to an innovation accelerator. Digital twins offer many benefits, including advanced business processes, increased productivity, and faster innovation at a lower cost, by offering a live copy of physical systems. Because of these benefits, digital twins are the perfect answer to a number of issues in industries like Industry 4.0, healthcare, education, and smart cities. However, the network that connects the physical and digital twins must meet several requirements, including low latency for real-time communication, data security, and quality, to ensure that the digital twin contributes to these domains in an effective manner by serving as a synchronized real-time copy of the physical system.

Authors in (Wang et al., 2022) pointed out that the global adoption of the digital economy has led to a flow in competition for battery management solutions which has been growing strongly. The Digital Twin technology presents a new view of managing and servicing lithium-ion batteries. In this work, a digital twin technology was presented along with a cloud-side-end for building future system management of battery. In this work, a four-layer network structure of cloud-side-end collaboration was studied, breaking down the high capacity and storage space which are the limitations in conventional battery management. The battery's digital twin model was built which guarantees refined and safety measures in the process of battery management during its life cycle. In addition to the model built, key technologies such as state estimation and cloud-assisted equalization of the batteries were presented. The result of the study emphasized the importance of Digital Twin on the efficiency of battery management and the full life cycle data which in turn is useful to build the upgrade route of the battery.

The study in (Stergiou et al., 2022) explores various issues in the realm of managing and analyzing big datasets in the context of the Industrial Internet of Things (IoT) within cloud environments. It sheds light on challenges arising from machine learning within cloud infrastructures, artificial intelligence methods for analyzing extensive data in cloud settings, and federated learning systems in the cloud. In addition, it explores reinforcement learning, a new approach that enhances the allocation of energy-efficient resources in large cloud-based data centers. Furthermore, the work proposes an architectural framework aimed at combining the capabilities provided by multiple cloud service providers to establish an energy-efficient framework for managing industrial IoT-based large data management framework (referred to as EEIBDM) accessible to all cloud users. This approach incorporates IoT data integration with techniques like reinforcement and federated learning to come up with a digital twin scenario for creating a virtual representation of industrial IoT-based large data. Additionally, an algorithm is presented to verify the energy consumption of the infrastructure by assessing the EEIBDM framework.

Authors in (Almasan et al., 2022) explored that the increasing popularity of emerging network applications such as Augmented Reality or Virtual Reality, telesurgery, and real-time communication is creating a growing challenge for the management of modern communication networks. These applications have strict requirements, especially those that require very low and predictable latency, making it increasingly difficult for network operators to allocate network resources efficiently. The authors in this paper introduced digital twin networks (DTNs) as an important tool to improve the management of modern networks. They described the basic structure of the DTN and argued that recent advances in machine learning (ML) have enabled DTN to be built that can effectively and accurately replicate networks in the real world. In addition, we investigate the key ML technologies that enable the development of different components of the DTN architecture. Finally, the authors highlighted the challenges and barriers to the success of DTN implementation in the real world within the research community.

The authors in (Alaasam et al., 2020) worked on Digital Twin (DT) and stated that the DT of processes and devices use sensors to synchronize their state with the physical entity in the real world. Stream computing is a concept that allows efficient processing of data generated by these sensors. However, since objects must track their state, it is impossible to organize digital twin instances as stateless services. Another characteristic of digital twins is that they need to respond almost in real time to events that arise, and cloud computing is inappropriate due to latency problems. Fog computing addresses this challenge by bringing some computation tasks closer to data sources. The most recent solution for generating loosely connected distributed systems is the microservice approach, which means organizing a system as a coherent and independent service that communicates through a message. Microservices are usually isolated using containers to reduce resource overhead associated with virtual machines. However, microservices and containers are inherently stateless. Container technology is still faced with limitations in supporting the seamless migration of containers between physical hosts without data loss, and it is difficult to ensure uninterrupted service operation in fog computing environments. Thus, an important challenge is to create a state-of-the-art stream processing microservice that supports digital twins in fog computing environments. In this article, we explore the concept of virtual stream processing migration and how to redistribute computation tasks over cloud and fog nodes using Kafka middleware and its stream DSL API.

With the rapid growth of mobile Internet, the Internet of Things (IoT), and cloud computing demand for data services has increased considerably. This increase in the number of data centers (DCs) leads to an increase in energy consumption and is harmful to energy conservation, emission reduction, and sustainable development. To address this challenge, the authors in (Zhang et al., 2022) introduced a data center energy savings solution called "SmartDC" based on the concepts of artificial intelligence (AI) and digital twins. The proposed solution aims to reduce energy consumption in data centers by optimizing air distribution and minimizing cooling redundancy. In this work, the authors used digital twin models to validate and improve AI strategies and address the problem of limited data availability in physical data centers. The Data for AI training and information extraction is often limited, as DC environments are experiencing minor changes. Furthermore, to ensure the safe operation of data centers at a suitable temperature, parameter adjustments must be conservative, leaving space for cooling redundancy. By combining digital twin technology with AI, we explore the temperature rise limits within digital data centers and collect more data pairs. This approach increases the robustness of AI models and results in more effective energy savings. Simulations and experiments have shown that the proposed solution ensures a safe and efficient DC operation, reducing energy consumption by 41.07 percent in cooling systems.

https://digitalcommons.bau.edu.lb/stjournal
DOI: 10.54729/2959-331X.1118

5. CONCLUSION

In this study we surveyed the literature for different failure prevention and recovery approaches in distributed storage systems using erasure coding and moving towards digital twins. It has been a challenge for researchers to optimize different performance metrics while ensuring data availability and reliability. With the advent of digital twin technology, a virtual clone is used for monitoring and fault diagnosis of different systems. Digital twins will be further developed in research and development in the coming years. Researchers can utilize digital twins to monitor the performance of the DDS system and manage different metrics accordingly.

REFRENCES

- Liu, K., Peng, J., Wang, J., Huang, Z., & Pan, J. (2022). Adaptive and scalable caching with erasure codes in distributed cloud-edge storage systems. IEEE Transactions on Cloud Computing.
- Telecommunications, Statista 2023. Number of Internet of Things (IoT) connected devices worldwide from 2019 to 2023, with forecasts from 2022 to 2030 (Accessed 2023). Available at: IoT connected devices worldwide 2019-2030 | Statista.
- Chiniah, A., & Mungur, A. (2022). On the Adoption of Erasure Code for Cloud Storage by Major Distributed Storage Systems. EAI Endorsed Transactions on Cloud Systems, 7(21), e1-e1.
- Li, Z., & Xiao, C. (2021). ER-Store: A Hybrid Storage Mechanism with Erasure Coding and Replication in Distributed Database Systems. Scientific Programming, 2021, 1-13.
- Aatish, C., & Avinash, M. (2020, May). Data management in erasure-coded distributed storage systems. In 2020 20th IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing (CCGRID) (pp. 902-907). IEEE.
- Xiao, Y., Zhou, S., & Zhong, L. (2020). Erasure coding-oriented data update for cloud storage: A survey. IEEE Access, 8, 227982-227998.
- Al-Shami, R., & Al-Mutawkkil, A. (2020). A Survey of Solutions and Future Directions For The Challenges Of Implementing Erasure Codes In Cloud And Fog Computing. International Journal of Advanced Research in Computer Science, 11(5).
- Zhou, T., & Tian, C. (2020). Fast erasure coding for data storage: A comprehensive study of the acceleration techniques. ACM Transactions on Storage (TOS), 16(1), 1-24.
- Liang, L., He, H., Zhao, J., Liu, C., Luo, Q., & Chu, X. (2021). An erasure-coded storage system for edge computing. IEEE Access, 8, 96271-96283.
- Store, D. C., & Cloud-of-Clouds, A. D. B. (2020). Storage Services.
- Fiandrino, C., Kliazovich, D., Bouvry, P., & Zomaya, A. Y. (2015, June). Performance metrics for data center communication systems. In 2015 IEEE 8th International Conference on Cloud Computing (pp. 98-105). IEEE.
- Liang, N., Zhang, X., Chen, H., & Zhang, C. (2022). Thou code: a triple-erasure-correcting horizontal code with optimal update complexity. The Journal of Supercomputing, 78(7), 10088-10117.
- Ye, L., Feng, D., Hu, Y., & Wei, X. (2020). Hybrid codes: flexible erasure codes with optimized recovery performance. ACM Transactions on Storage (TOS), 16(4), 1-26.
- Yang, J., Sabnis, A., Berger, D. S., Rashmi, K. V., & Sitaraman, R. K. (2022). {C2DN}: How to Harness Erasure Codes at the Edge for Efficient Content Delivery. In 19th USENIX Symposium on Networked Systems Design and Implementation (NSDI 22) (pp. 1159-1177).
- Chen, R., & Xu, L. (2020). A new decoding algorithm for XOR-based erasure codes. SN Computer Science, 1, 1-19.
- Xu, L., Lyu, M., Li, Z., Li, Y., & Xu, Y. (2020). Deterministic data distribution for efficient recovery in erasure-coded storage systems. IEEE Transactions on Parallel and Distributed Systems, 31(10), 2248-2262.
- Zhang, X., Liang, N., Liu, Y., Zhang, C., & Li, Y. (2022). SA-RSR: A read-optimal data recovery strategy for XOR-coded distributed storage systems. Frontiers of Information Technology & Electronic Engineering, 23(6), 858-875.
- Mu, T., Song, Y., Yang, M., Wang, B., & Zhao, J. (2022, April). HV: An improved coding layout based on erasure coded storage system. In International Conference on Database Systems for Advanced Applications (pp. 203-213). Cham: Springer International Publishing.
- Qiu, H., Wu, C., Li, J., Guo, M., Liu, T., He, X., ... & Zhao, Y. (2020, May). Ec-fusion: An efficient hybrid erasure coding framework to improve both application and recovery performance in cloud

storage systems. In 2020 IEEE International Parallel and Distributed Processing Symposium (IPDPS) (pp. 191-201). IEEE.

- Wang, Z., Wang, H., Shao, A., & Wang, D. (2020, August). An adaptive erasure-coded storage scheme with an efficient code-switching algorithm. In Proceedings of the 49th International Conference on Parallel Processing (pp. 1-11).
- Gao, Z., Zhang, L., Cheng, Y., Guo, K., Ullah, A., & Reviriego, P. (2021). Design of FPGA-implemented Reed–Solomon erasure code (RS-EC) decoders with fault detection and location on user memory. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 29(6), 1073-1082.
- Geng, W., Zhang, H., Liu, L., He, R., & Tang, D. (2020, September). Optimization of the Recovery Time of Pyramid Code in Distributed Storage System. In Journal of Physics: Conference Series (Vol. 1634, No. 1, p. 012075). IOP Publishing.
- Liang, N., Zhang, X., Wu, X., Chen, H., & Zhang, C. (2021, September). An Endurance-aware RAID-6 Code with Low Computational Complexity and Write Overhead. In 2021 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom) (pp. 939-946). IEEE.
- Dhore, A. M., & Tiwari, N. (2022). Data Distribution in Reliable and Secure Distributed Cloud Environment Using Hash-Solomon Code. In Machine Intelligence and Smart Systems: Proceedings of MISS 2021 (pp. 537-545). Singapore: Springer Nature Singapore.
- Jain, P., Poon, J., Singh, J. P., Spanos, C., Sanders, S. R., & Panda, S. K. (2019). A digital twin approach for fault diagnosis in distributed photovoltaic systems. IEEE Transactions on Power Electronics, 35(1), 940-956.
- Zohdi, T. I. (2022). A digital-twin and machine-learning framework for precise heat and energy management of data-centers. Computational Mechanics, 69(6), 1501-1516.
- Hong, H., Wu, Q., Dong, F., Song, W., Sun, R., Han, T., ... & Yang, H. (2021, August). NetGraph: An Intelligent Operated Digital Twin Platform for Data Center Networks. In Proceedings of the ACM SIGCOMM 2021 Workshop on Network-Application Integration (pp. 26-32).
- Mashaly, M. (2021). Connecting the twins: A review on digital twin technology & its networking requirements. Procedia Computer Science, 184, 299-305.
- Wang, Y., Xu, R., Zhou, C., Kang, X., & Chen, Z. (2022). Digital twin and cloud-side-end collaboration for intelligent battery management system. Journal of Manufacturing Systems, 62, 124-134.
- Stergiou, C. L., & Psannis, K. E. (2022). Digital twin intelligent system for industrial IoT-based big data management and analysis in cloud. Virtual Reality & Intelligent Hardware, 4(4), 279-291.

  - Almasan, P., Ferriol-Galmés, M., Paillisse, J., Suárez-Varela, J., Perino, D., López, D., ... & Barlet-Ros, P. (2022). Network digital twin: Context, enabling technologies, and opportunities. IEEE Communications Magazine, 60(11), 22-27.
- Alaasam, A. B., Radchenko, G., Tchernykh, A., & González Compeán, J. L. (2020). Analytic study of containerizing stateful stream processing as microservice to support digital twins in fog computing. Programming and Computer Software, 46, 511-525.
- Zhang, Z., Zeng, Y., Liu, H., Zhao, C., Wang, F., & Chen, Y. (2022, April). Smart DC: an AI and digital twin-based energy-saving solution for data centers. In NOMS 2022-2022 IEEE/IFIP Network Operations and Management Symposium (pp. 1-6). IEEE.