space
Internet-Draft
Intended status: Informational
Expires: 3 September 2026
J. Wang
China Mobile
P. Zhang
Beihang University
2 March 2026

Consideration for Space-Based Computing Infrastructure Network

draft-wang-space-computing-consideration-00

Abstract

This document presents considerations for a Space-Based Computing Infrastructure Network from use cases and requirements.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

In recent years, the global satellite industry has experienced rapid development. The deployment of low-Earth orbit satellite constellations, advancements in satellite communication technologies, and improved space launch capabilities have propelled global satellite networks towards a more interconnected and intelligent system. These developments have greatly improved the coverage, transmission speeds, system stability, and networking flexibility of satellite networks, allowing for seamless integration across air, land, and space domains.

This increasingly mature global satellite network has broken the traditional constraints of space information transmission, resulting in more efficient inter-satellite and satellite-to-ground data exchange. This has also laid a solid foundation for extending computing power into space. On one hand, the stable and reliable satellite links provide efficient interconnection channels for computing facilities such as in-orbit computing, data processing, and intelligent sensing. On the other hand, the widespread deployment of satellites has created opportunities for the distribution of computing nodes in space.

This has led to the evolution of space computing power from isolated single-satellite operations to multi-satellite coordination, space-ground synergy, and global-scale orchestration. This evolution is crucial in building space computing networks and achieving ubiquitous computing services across all domains.

2. Conventions

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

3. Use Cases

Considering use cases on Space-Based Computing Infrastructure Network.

3.1. Emergency Response and Disaster Monitoring

During natural disasters, such as earthquakes and floods, traditional communication and computing systems are at risk of damage, resulting in delays in the transmission of critical information. However, by utilizing satellite computing networks, emergency communication and computing nodes can be quickly deployed to process disaster imagery in real time. This allows for the creation of precise disaster maps and optimal rescue routes, providing decision support at a minute or even second level.

This greatly improves the efficiency of disaster warning, emergency response, and resource allocation. Additionally, in the event of terrestrial network failures, these satellite networks can seamlessly provide communication and edge computing capabilities to support emergency command, drone search-and-rescue operations, and post-disaster reconstruction data processing.

3.2. Environmental Monitoring and Ecological Management

nder traditional models, large amounts of raw satellite data, such as 0.3-meter high-resolution imagery, must be transmitted back to Earth for processing. However, due to limited satellite-to-ground communication bandwidth, less than one-tenth of the data can be transmitted, resulting in low efficiency.

To address this issue, AI models can be deployed in orbit to perform real-time target detection, classification, change monitoring, and feature extraction on remote sensing imagery. This allows only critical analysis results to be transmitted to the ground, improving efficiency. This technology can accurately identify farmland, forests, water bodies, and glaciers, making it easier to track carbon sinks, monitor water environments, and track vegetation degradation.

As a result, data utilization rates have increased from 10% to nearly 100%, greatly enhancing the timeliness and autonomy of national land resource surveys, environmental monitoring, agricultural assessments, and related fields.

3.3. Deep Space Exploration Mission Support

Deep-space probes experience significant communication delays with Earth, with delays of several minutes being common for missions to Mars. This reliance on ground control can be inefficient.However, by deploying computational nodes in deep-space orbits, these probes can perform in-orbit preprocessing, compression, and intelligent filtering of data.

This allows for coordination through inter-satellite communication networks, resulting in a significant reduction in the volume of raw data that needs to be transmitted back to Earth. This approach not only enhances the autonomous operation capabilities of probes, but also improves their mission response speed. It serves as a critical foundation for future long-term exploration missions to destinations such as the Moon, Mars, and beyond.

3.4. In-orbit Training and Inference for Large AI Models

Training AI models with hundreds of billions of parameters requires immense computational power, which can pose energy and thermal bottlenecks for ground-based data centers. However, by leveraging the distributed computing capabilities and green energy advantages of space computing networks, it is possible to distribute model training and inference.

This approach provides a new "zero-carbon" computing pathway for AI development.

4. Requirements

Considering requirements on Space-Based Computing Infrastructure Network..

4.1. Space-Based Computing Resource Monitoring

Spaceborne equipment faces significant constraints in terms of computational resources, including CPU/GPU processing power, storage capacity, and energy consumption limits. These limitations are due to the size, power consumption, and payload capacity of the equipment. Additionally, the computational configurations of different satellites can vary greatly. Some prioritize edge computing, while others focus on data relay.

Furthermore, the computational load of satellites can fluctuate depending on mission requirements. For example, sudden spikes in remote sensing data processing or IoT terminal access within a specific region can overload local satellites, while satellites in other areas may remain idle.

This highlights the need for a technical solution that can monitor the computational load, available resources, and energy consumption status of each satellite in real-time. This data would then be used to support cross-satellite resource allocation.

4.2. On-demand Traffic Scheduling

Satellite networks support a wide range of service types, each with unique demands for network and computing power. For example, emergency communications require low latency and high reliability, while remote sensing data processing requires significant computing power but is less sensitive to latency. IoT data transmission prioritizes high bandwidth and low power consumption.

However, a unified scheduling strategy may lead to issues such as "computing power mismatch" (e.g. assigning high-latency services to long-range satellites) or "resource wastage" (e.g. using high-performance computing satellites for simple data relay tasks).

Therefore, it is crucial to establish a matching mechanism between service requirements and resource capabilities, including network resources such as link status, in order to enable efficient on-demand scheduling.

5. Conclusion

This document makes some considerations on Space-Based Computing Infrastructure Network.

6. Security Considerations

TBD.

7. IANA Considerations

TBD.

8. Informative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

Authors’ Addresses

Jing Wang
China Mobile
No.32 XuanWuMen West Street
Beijing
100053
China
Email: wangjingjc@chinamobile.com

Pengfei Zhang
Beihang University
No.37 Xueyuan Road, Haidian District
Beijing
100191
China
Email: zhangpengfei@buaa.edu.cn