International Journal of Emerging Trends in Computer Science and Information Technology
ISSN: 3050-9246 | https://doi.org/10.63282/3050-9246/ICRTCSIT-120
Eureka Vision Publication | ICRTCSIT'25-Conference Proceeding

Original Article

Goal-Oriented Semantic Communication: A Foundational Enabler for Ultra-Reliable Low-Latency Communication in 6G-Enabled Internet of Things

Naresh Kalimuthu
Texas, USA

Abstract - *The sixth-generation (6G) wireless networks aim for a shift towards an AI-driven infrastructure, supporting critical Internet of Things (IoT) applications with extraordinary performance requirements. Ultra-Reliable Low-Latency Communication (URLLC) plays a key role in this vision, but its demanding criteria challenge traditional communication principles. This paper argues that Goal-Oriented Semantic Communication (GOSC), a new approach focused on transmitting only the essential information needed for a specific task, is vital for enabling URLLC in 6G. We examine the primary research challenges associated with processing, knowledge sharing, and resource management. A comprehensive, AI-centered architecture with a dedicated semantic layer is proposed to address these issues. Through case studies in industrial automation, autonomous vehicles, and remote healthcare, we demonstrate GOSC's potential to significantly reduce data loads while improving task success. The paper concludes with future research directions, including standardization, security, and developing a complete theoretical foundation for goal-driven information.*

Keywords - 6G, Semantic Communication, Goal-Oriented Communication, URLLC, IoT, AI-Native Networks, Edge Intelligence, Resource Allocation.

1. Introduction

The journey of wireless communication has been one of constant evolution, with each generation unveiling new capabilities and changing how society interacts with the digital world. The current deployment of fifth-generation (5G) networks introduces three primary service categories: Enhanced Mobile Broadband (eMBB), Massive Machine-Type Communication (mMTC), and Ultra-Reliable Low-Latency Communication (URLLC). While transformative, 5G is merely a stepping stone toward a much more ambitious goal: the development of sixth-generation (6G) networks. Planned for the 2030s, 6G is not just a slight increase in speed and capacity but a fundamental redesign of the network as a distributed, intelligent platform aimed at merging the physical, digital, and human worlds into a seamless cyber-physical continuum. This future will feature applications such as holographic telepresence, large-scale digital twinning, and fully autonomous systems, all of which will require performance guarantees far beyond current technological limits.

At the core of this transition is the challenge of supporting critical Internet of Things (IoT) applications, where communication failures or delays can have serious consequences. URLLC is the service class designed to meet these needs, but the leap from 5G's objectives to 6G's demands is substantial. This reveals the limitations of a communication philosophy that has largely remained unchanged for over 70 years one focused on perfecting the replication of bits, regardless of their meaning or purpose. This paper argues that unlocking the full potential of 6G-enabled critical IoT requires an architectural shift: from the syntax of bits to the semantics of meaning and the pragmatics of goals. This shift is embodied in Goal-Oriented Semantic Communication (GOSC), a groundbreaking approach that aligns communication with the achievement of a specific task.

1.1. The 6G Imperative: Transitioning from Connected Devices to Ubiquitous Intelligence

6G networks will offer more than just connectivity. It is designed as an AI-native system where intelligence is not an afterthought but is integrated into every layer of the network fabric, from the physical air interface to the application layer. The paradigm is "Intelligent Connectivity of Everything." The network can sense, learn, and reason, making it fully autonomous and responsive in real-time. The architecture blueprint for 6G combines terrestrial networks with non-terrestrial elements, such as satellites and unmanned aerial vehicles (UAVs), forming a universal space-air-ground-sea integrated network that offers global coverage.

A set of enabling technologies drives this widespread intelligence. Communications in the Terahertz (THz) and optical frequency bands are expected to deliver unprecedented data rates, while integrated sensing and communication (ISAC) will allow the network to use its own signals to create a high-resolution, real-time map of the physical environment. This combination of communication and sensing turns the network itself into a distributed sensor, capable of providing precise positioning and environmental context as a service.

Nonetheless, this AI-native framework presents a fundamental contradiction in how model communication is understood. If a network has a deep understanding of a context, to what extent do we describe a communication layer as an inefficient "content-blind" pipe dedicated solely to transmitting information? A network with contextual awareness will need an equally advanced mechanism for transferring data. Goal-Oriented Semantic Communication (GOSC) serves as the logical and communicative blueprint for an AI-native 6G ecosystem and beyond. It provides the foundation for a system where autonomous vehicles, industrial robots, and surgical systems are intelligent agents that not only exchange information but also collaborate seamlessly toward shared goals.

1.2. Defining the Extremes: The Role and Requirements of URLLC in Critical IoT

Providing seamless communication during emergencies is vital, as these situations can cause financial, life, equipment, and environmental losses. URLLC is essential for remote healthcare, autonomous driving, and smart grids, focusing on strict latency and reliability requirements. In 5G, requirements are demanding, targeting a 1ms user plane latency and 99.999% reliability. 5G systems use physical and MAC layer enhancements like flexible numerology, mini-slot scheduling, grant-free uplink, and robust channel coding.

The 6G vision introduces applications like real-time immersive experiences and high-stakes control systems that push requirements to their limits, leading to the concept of "extreme URLLC" (xURLLC). As shown in Table 1, 6G xURLLC aims for latencies of 0.1 to 1ms and reliability of 99.9999% to 99.9999999% (seven nines) or higher. This is a significant leap that challenges current communication design, not just a quantitative tightening.

Traditional methods to improve URLLC face a "performance wall." Increasing reliability via redundancy, like HARQ retransmissions or packet duplication, raises latency and resource use. As latency drops to sub-millisecond levels, there's no time for multiple retransmissions, creating a trade-off: lower latency often reduces reliability, and vice versa. Pushing both to extremes with conventional methods is unfeasible, requiring a disruptive approach that optimizes the relevance of information itself.

Table 1. URLLC Key Performance Indicators (KPIs) across 5 G and 6 G)

KPI	5G URLLC Target	6G (xURLLC) Vision	Relevant Applications
User Plane Latency	~1 ms	0.1−1 ms	Industrial Control, Haptic Feedback, V2X
Reliability	99.999% ($1-10^{-5}$)	99.9999% to 99.9999999% (10^{-6} to 10^{-9})	Remote Surgery, Autonomous Driving
Connection Density	~ 10^5devices/km²	~10^7 devices/km²	Massive Industrial IoT, Smart Cities
Jitter	Not strictly defined	Bounded, microsecond-level	Real-time Robotics, Immersive XR
Data Rate	Up to 100 Mbps	= 1 Gbps	Holographic Telepresence, Digital Twinning

1.3. A Paradigm Shift in Communication Theory: From Syntax to Semantics and Goals

Claude Shannon laid the groundwork for modern digital communication by focusing on the "technical" challenge of transmitting messages accurately and efficiently. This approach, supporting all systems from 1G to 5G, emphasizes syntaxthe correct order of bitsand treats information as a black box of data aiming for perfect bit fidelity. It has been highly successful, driving communication systems toward Shannon capacity limits.

In 1949, Warren Weaver introduced two higher communication levels: the semantic level, which conveys meaning, and the effectiveness level, which impacts recipient actions. Previously, philosophers and linguists were mainly focused on these aspects, as the primary concern was the technical challenges of the syntactic level. With the advent of AI-driven communications in 6G, these higher levels have become pressing engineering challenges.

Semantic Communication (SemCom) emphasizes the 'what' over the 'how,' using AI/ML to convey core meaning instead of raw data. The receiver interprets semantics with shared context. Goal-Oriented Semantic Communication (GOSC) advances this by transmitting only essential information for specific tasks, such as sending key details for collision avoidance rather than full video frames. This approach is crucial for URLLC, enabling efficient, mission-critical data transmission.

Table 2. Comparison Of Communication Standards

Feature	Traditional (Shannon) Communication	Semantic Communication (SemCom)	Goal-Oriented Semantic Communication (GOSC)
Primary Objective	Accurate reconstruction of transmitted bits.	Accurate reconstruction of intended meaning.	Successful completion of a specific task at the receiver.
Key Metric	Bit Error Rate (BER), Throughput.	Semantic Similarity, Semantic Accuracy.	Task Success Rate, Goal-Effectiveness.
Unit of Information	Bit, Symbol.	Semantic feature, concept.	Task-relevant information, decision.
Role of Context / KB	Irrelevant to the communication process.	Essential for encoding/decoding meaning.	Essential for identifying and transmitting task-critical data.
Example	Transmitting a full-resolution image file.	Transmitting a description: A cat is on a mat.	Transmitting an alert: Obstacle detected.

1.4. Thesis and Paper Organization

The rise of AI-native 6G, URLLC demands, and critical IoT needs creates a challenge that traditional communication can't meet. The conflict between latency and reliability requires a new approach to reduce communication load without losing integrity. The paper argues that Goal-Oriented Semantic Communication is essential for meeting URLLC needs in AI 6G, shifting focus from bit accuracy to goal effectiveness. GOSC can surpass traditional limits, enabling intelligent, autonomous, and reliable applications.

2. Core Research Challenges in Realizing Goal-Oriented URLLC

Transforming Goal-Oriented Semantic Communication in URLLC from a theoretical concept to a practical, deployable system presents significant technical challenges. Although transmitting only goal-relevant information offers clear advantages, achieving this requires addressing complex issues across device hardware, network protocols, and AI model management. These challenges are the key research areas essential for realizing GOSC's potential in mission-critical 6G applications.

2.1. Challenge 1: The Processing Bottleneck: Real-Time Semantic Inference on Resource-Constrained Devices

A GOSC transmitter's primary role is to perform semantic extraction using Deep Neural Networks (DNNs) to analyze sensor data, such as images, audio, and LiDAR, and condense it into a compact form. This process is resource-intensive, requiring significant processing power, memory, and energy, which many IoT devices lack. This creates a bottleneck: the time to run these models causes delays, even in URLLC systems where latency must be around one millisecond. A few hundred microseconds of delay can negate the benefits of increased communication speed. Additionally, the unpredictable performance of deep learning models, influenced by input data and device temperature, introduces jitter, complicating deterministic, low-latency operations essential for industrial applications. Addressing these issues requires tackling both communication and computation challenges on constrained hardware.

2.2. Challenge 2: The Coherency Dilemma: Dynamic Adaptation and Knowledge Base Synchronization

Semantic communication depends on shared understanding via a common Knowledge Base (KB), which can be a shared AI model, ontology, or database. The transmitter encodes messages based on this KB, and the receiver decodes them using its copy. Maintaining consistent KBs across devices, especially in dynamic environments like autonomous vehicles or factories, is challenging. Changes in environment can lead to misinterpretations, risking failures in perception or commands. KB mismatches cause subtle communication failures, even with perfect data transmission, risking catastrophic outcomes in URLLC. Updating KBs frequently ensures coherence but increases resource use and latency, posing a critical reliability challenge.

2.3. Challenge 3: The Resource Allocation Conundrum: Moving Beyond Semantic-Agnostic Networking

Current wireless resource allocation primarily utilizes QoS parameters without considering packet content, resulting in the equal treatment of similar packets regardless of their importance. The goal is to develop 'semantic-aware' protocols that consider the significance, relevance, and urgency of information, requiring a shift from traditional layered architecture by enabling cross-layer data flow and context access. This creates the 'URLLC-GOSC Trilemma,' balancing processing, bandwidth, and computation, demanding a holistic approach. Incorporating semantics extends 'reliability' to include semantic and goal correctness, introducing new security concerns and requiring guarantees across syntax, semantics, and latency.

3. Architectural Frameworks and Mitigation Strategies

Addressing challenges in Goal-Oriented Semantic Communication for URLLC demands a new architecture and techniques to optimize computation, knowledge, and network resources. The proposed AI-driven framework integrates semantic intelligence into 6G, transforming the network from passive to an active participant in communication goals.

3.1. An AI-Native Framework for GOSC: Integrating the Semantic Plane into 6G Architecture

The core solution introduces a "Semantic Plane" to the 6G network, operating in parallel and enhancing the traditional user and control planes. It's a distributed system of smart functions for semantic communication, best implemented following O-RAN principles, which decouple the base station into interoperable parts and include a programmable RAN Intelligent Controller (RIC). In this setup, the Semantic Plane would be managed by a specialized "Semantic RIC" (S-RIC) or a "Semantic Engine," which hosts applications such as O-RAN's xApps and rApps that perform key semantic tasks.

- **Semantic Model Management:** Managing training, deployment, and lifecycle of AI models used for semantic encoding and decoding across all network devices.
- **Knowledge Base (KB) Orchestration:** Keeping shared knowledge bases synchronized, managing versions, and efficiently distributing updates.
- **Semantic-Aware Resource Allocation:** Performing advanced scheduling and resource management based on the semantic importance and relevance of data.
- **Interoperability:** Offering gateways and translation functions to ensure backward compatibility and coexistence with legacy, semantic-agnostic systems.

This AI-driven architecture provides the programmability and intelligence necessary to implement targeted strategies that address the primary research challenges.

3.2. Mitigating the Processing Bottleneck: Lightweight Models, Hashing, and Edge Intelligence

To enable real-time semantic inference on limited devices, a comprehensive strategy is needed to reduce computational demands without losing performance.

- **Lightweight and Adaptive AI Models:** This approach centers on optimizing the AI models themselves. Techniques such as model compression through pruning and quantization can significantly reduce the size and complexity of DNNs. Additionally, developing innovative and efficient neural network architectures, such as transformers tailored for mobile and edge devices, is crucial. Dynamic neural networks that adjust their inference complexity based on available resources or input importance offer a promising route for efficient on-device processing.
- **Hashing-based Semantic Extraction:** For ultra-low latency applications, complex DNNs can be replaced with faster, cheaper methods. Hashing-based semantic extraction uses supervised learning to produce compact binary "semantic signatures" from source data, allowing quicker generation and efficient comparison with simple operations like Hamming distance. This approach balances some semantic loss with reduced latency and energy benefits consumption.
- **Semantic Edge Computing (SEC):** The most effective strategy is shifting computation from the device to the network edge, where resource-limited IoT devices send raw or partial data to a nearby edge server via low-latency 6G links. The server then performs heavy semantic encoding and forwards concise data, leveraging network resources to resolve processing bottlenecks and ensure efficient resource use allocation.

3.3. Ensuring Semantic Coherency: Adaptive Learning and Knowledge Base Orchestration

Maintaining synchronized Knowledge Bases in dynamic environments requires mechanisms that are both adaptive and efficient.

- **Distributed and Adaptive Learning:** In large-scale IoT environments, centralized semantic model training is challenging. Distributed machine learning, particularly Federated Learning (FL), enables devices to collaboratively update a shared AI model without sharing private data, thereby maintaining privacy and reducing communication by sending only small training datasets to a central server.
- **Reinforcement Learning (RL),** particularly multi-agent RL, enables devices to learn and adapt their communication and semantic extraction strategies based on real-time feedback and task success, allowing autonomous development of effective communication policies without explicit programming.
- **Knowledge Base Management and Orchestration (KB-MANO):** To manage knowledge bases effectively, a dedicated orchestration layer, like MANO in network function virtualization, is needed. The KB-MANO system would oversee all knowledge assets, including creating KBs for new services, monitoring status and versions, distributing updates efficiently, and ensuring participants in semantic communication sessions use compatible KB versions. This role would likely be part of the Semantic RIC, serving as a central control point for a distributed knowledge system.

3.4. Re-engineering the Air Interface: Semantic-Driven Scheduling and Resource Management

To advance beyond semantic-agnostic networking, the air interface and resource management protocols need a fundamental redesign to include semantic awareness.

- **Semantic-Aware MAC Protocols:** The MAC layer scheduler optimizes by prioritizing data flows based on semantic metrics, such as Value of Information (VoI) or Age of Information (AoI), rather than just packet-level QoS parameters. This allows dynamic resource allocation to maximize overall value to applications' throughput.

- **Deep Reinforcement Learning (DRL) for Resource Allocation:** Given the complexity of allocating multiple resource types (like time, frequency, power, and computational resources) in a dynamic environment, DRL offers a strong solution. A DRL agent at the base station or RIC learns resource policies by observing the network state and semantic requirements, and receives rewards for achieving communication goals. It adapts to changing conditions and optimizes for semantic performance metrics, such as maximizing "semantic spectral efficiency" (S-SE)the amount of meaningful information transmitted per unit of spectrum.
- **Coexistence with Legacy Systems:** A 6G network won't be a greenfield deployment; it must coexist with legacy BitCom services. Techniques such as Non-Orthogonal Multiple Access (NOMA) can be utilized to enable GOSC-based URLLC and eMBB traffic to share spectrum efficiently by serving users simultaneously at different power levels, with interference cancellation at the receivers. This enables high-priority semantic packets to overlay lower-priority bit traffic, boosting spectral efficiency.

These strategies turn the network into a "network-as-a-cognitive-system,' making it active and intelligent with edge computing, adaptive learning, and semantic-aware scheduling. The O-RAN architecture with a Semantic RIC is its central nervous system.

This shift highlights a trade-off between "semantic compression" and "semantic robustness." GOSC aims for high compression by removing redundancy, but in traditional systems, redundancy through channel coding ensures robustness against noise. A highly compressed semantic representation can be fragile; a single bit error may cause significant meaning loss. Therefore, the semantic encoder and channel coder should be designed together. Future research should focus on Joint Source-Channel Coding (JSCC) for semantics to develop compact, meaningful, and resilient representations, enabling graceful degradation when errors occur instead of abrupt communication failure entirely.

4. Performance Analysis and Validation through Case Studies

The theoretical benefits of Goal-Oriented Semantic Communication require testing through practical cases and performance analysis. This section relates earlier points to real-world examples, demonstrating how GOSC supports URLLC in crucial IoT sectors. A key initial step is recognizing that traditional communication metrics are inadequate for this approach.

4.1. Evolving Performance Metrics: From Bit Error Rate to Goal-Effectiveness

Traditional communication metrics, such as Bit Error Rate (BER), Packet Error Rate (PER), and throughput, focus on syntactic details but can be misleading for GOSC systems. A system with a non-zero BER may still succeed if errors are semantically irrelevant. Conversely, a zero BER doesn't guarantee success if important semantic info is missing misinterpreted.

Therefore, a new set of performance metrics, operating at the semantic and effectiveness levels, is required:

- **Semantic Accuracy/Similarity:** This metric assesses how well meaning is preserved. For different data types, specific tools are used: BLEU (Bilingual Evaluation Understudy) scores for text, Structural Similarity Index Measure (SSIM or Peak Signal-to-Noise Ratio (PSNR for images, and F1-score or Intersection over Union (IoU for classification and object detection tasks.
- **Task/Goal Success Rate:** This is often the most critical metric for URLLC, measuring the effectiveness of communication and the likelihood of completing the intended task within the required time budget.
- **Effectiveness-Aware Metrics:** These metrics measure the practical value of information. The Value of Information (VoI) quantifies the usefulness of information for decision-making, while the Age of Information (AoI) measures the freshness of data at the receiver, which is vital for real-time control systems. Prioritizing these metrics over raw throughput becomes the main optimization goal.

Using this new evaluation framework, we can analyze the performance of GOSC in several key URLLC domains.

4.2. Case Study: Precision and Safety in Industrial Automation (IIoT)

- **Scenario**: A modern "Industry 4.0" smart factory uses autonomous mobile robots (AMRs) and collaborative robotic arms on a reconfigurable assembly line. These systems require real-time monitoring to ensure efficiency and safety. Key tasks include predictive maintenance, such as tracking wear on high-speed cutting tools to prevent catastrophic failure.
- **GOSC Application:** In traditional setups, high-resolution video or raw vibration data are sent to a central server, consuming bandwidth and causing latency. GOSC, an intelligent sensor with edge processing, performs local semantic analysis, transmitting only relevant info. Usually, it sends nothing or a simple 'status normal.' If early signs of wear are detected, it sends a compact alert, e.g., 'Tool T 4 on CNC Mill 7: Flute wear exceeds 90% at (X,Y,Z). Predicted failure in 15 Cycles.".
- **URLLC Enablement:** The alert payload is just a few dozen bytes, much smaller than raw data megabytes, enabling fast, reliable wireless transmission with low latency. This allows the control system to quickly stop or reroute

production, preventing damage and costs downtime.
- **Performance Analysis:** The key metric isn't bandwidth but "goal-effectiveness," measured by the average time to prevent failure or reduce unscheduled downtime. Research shows that formal ontological models in automated production systems help maintain consistency and assess change impact, boosting system reliability. GOSC's main benefit is supporting stable, real-time control over wireless, a challenge for traditional communication due to latency jitter.

4.3. Case Study: Enabling Cooperative Perception in Autonomous Driving (V2X)
- **Scenario:** A dense urban environment with connected autonomous vehicles (CAVs) and roadside infrastructure. To ensure safety at blind intersections or in bad weather, vehicles share sensor data to build a shared view, enabling them to "see" around corners and through obstacles.
- **GOSC Application:** Transmitting raw sensor data like full LiDAR point clouds or 4K videos from each vehicle to neighbors is impractical due to bandwidth and latency needs of V2X safety apps. Instead, vehicles use GOSC. Onboard systems (e.g., DNNs) process raw data to identify objects, extracting key info such as class, position, size, velocity, and predicted path of others, then broadcast these concise "semantic object lists" to nearby vehicles.
- **URLLC Enablement:** A semantic object can be represented with a few hundred bytes, while a LiDAR scan may reach several megabytes. This data reduction enables real-time cooperative perception, allowing safety-critical information to be shared among dozens of vehicles and infrastructure nodes within sub-millisecond latency, essential for decisions like triggering emergency brakes for pedestrians obscured by parked vehicles truck.
- **Performance Analysis:** Significant quantitative gains are evident in the literature. One study on urban traffic surveillance using a semantic communication framework showed a 99.9% reduction in data transmission size by sending compact embedding vectors instead of raw image data, with only a slight decrease in task accuracy (traffic condition description by a Large Language Model) from 93% to 89%. Another study on vehicular image segmentation for autonomous driving achieved a 70% reduction in data volume and a nearly 6 dB coding gain at 60% mean Intersection over Union, compared to traditional image transmission. The key goal is faster hazard reaction and better environmental modeling, improving road safety.

4.4. Case Study: The Future of Telemedicine and Remote Robotic Surgery (H-IoT)
- **Scenario:** A surgeon at a city hospital performs complex minimally invasive surgery on a patient in a rural clinic using an advanced robotic system. The system transmits multiple data streams: HD 3D video, real-time control commands, and haptic feedback signals.
- **GOSC Application:** For remote surgery, closed-loop control and haptic feedback require near-instant, flawless communication. Latency or data loss risks incorrect instrument movement. GOSC manages data streams by importance; critical signals get top priority despite low bandwidth. The high-bandwidth 3D video is semantically compressed: AI identifies focus areas, assigning high resolution to the surgical site and compressing peripheral regions more video.
- **URLLC Enablement:** GOSC reduces bandwidth for video, freeing network resources to meet URLLC needssub-millisecond latency and over 99.999% reliability. This ensures stable master-slave control and real-time tactile feedback for precise manipulation. Consequently, procedures impossible on standard networks become safe and feasible effective.
- **Performance Analysis:** The primary measure is end-to-end latency of the haptic and control loop. Early telesurgery experiments over dedicated networks showed round-trip latencies of 264-280ms, acceptable for some procedures but too high for others. Tests with 5G have demonstrated the potential to reduce this significantly. Using GOSC is essential for lowering latency to the sub-millisecond range needed for precise operations, ensuring less urgent data never delays critical, time-sensitive tasks data.

These case studies reveal that GOSC redefines "network efficiency" by shifting focus from spectral efficiency to "goal efficiency," measuring successful tasks per second per Hertz. For example, the V2X case shows a 99.9% reduction in data traffic, which, despite reducing revenue from data volume, enables life-saving features. This calls for models that sell guaranteed task execution, measuring value by outcomes, not bits. Achieving GOSC requires close collaboration between communication systems and application logic, with applications needing to be "semantic-native." Developers and engineers must integrate data structures and semantic transmission formats, likely via standards, marking a significant shift in designing future critical systems.

5. Conclusion and Future Trajectories
The sixth-generation wireless network marks a key milestone, evolving from simple data transfer to supporting pervasive, intelligent systems. In this landscape, Ultra-Reliable Low-Latency Communication (URLLC) is vital for autonomous industries, intelligent transportation, and remote healthcare. However, 6G URLLC's demanding standards challenge the traditional Shannon model, revealing a conflict between ultra-high reliability and ultra-low latency time.

5.1. Synthesis of Findings: GOSC as a Foundational Pillar for 6G URLLC

This paper argues that Goal-Oriented Semantic Communication (GOSC) provides a crucial solution to the deadlock by shifting the goal from perfect bit-fidelity to successful goal achievement, introducing a new optimization dimension. The core findings are threefold:

- **GOSC is a Natural Fit for AI-Native 6G:** The 6G vision of an intelligent network conflicts with a content-blind layer. GOSC offers the "native language" for this architecture, aligning information exchange with the network's understanding of context and user intent.
- **A Holistic Approach is Required to Overcome Key Challenges:** Implementing GOSC for URLLC is complex due to challenges in on-device processing latency, knowledge base synchronization, and semantic-aware resource allocation, which are interconnected. The solution is an AI-driven framework based on O-RAN, with a Semantic Plane and Semantic RIC to coordinate mitigation strategies like edge computing, federated learning, and DRL-based resource management.
- **GOSC Delivers Transformative and Quantifiable Performance Gains:** Case studies in critical IoT domains show GOSC's impact is transformative, not incremental. By transmitting only essential info, GOSC reduces data payloads by up to 99.9%, freeing network resources to meet URLLC latency and reliability needs that would be otherwise infeasible. This shifts network efficiency from bits per second to successful tasks per second.
- In summary, GOSC is a core technology, not just an optional feature or niche optimization. It bridges the gap between 6G ambitions and communication limits, unlocking a future of intelligent, reliable autonomous systems.

5.2. Open Research Problems and Future Directions

While GOSC's potential is clear, the field is still in early stages, with many research challenges unresolved before wide adoption. The following areas highlight future research directions.

- **Standardization of Semantics:** The effectiveness of GOSC depends on shared understanding. For large-scale, multi-vendor ecosystems like V2X or IIoT, this requires developing standardized ontologies and Knowledge Base formats to ensure interoperability. This is a complex, multidisciplinary effort that will need collaboration among communication engineers, AI researchers, and domain experts.
- **Security and Robustness**: Integrating AI models into communication adds security risks. Urgent research is needed to understand and defend against adversarial attacks targeting the semantic layer, which manipulate meaning without altering bits. Developing reliable, verifiable, and trustworthy AI for semantic transceivers is crucial for mission-critical applications systems.
- **Hardware and Algorithm Co-design:** The on-device processing bottleneck is a key obstacle. Future research should focus on co-design ultra-efficient semantic algorithms and hardware accelerators (e.g., AI chipsets for IoT) for real-time inference within strict power and latency limits.
- **A Complete Theoretical Foundation:** AI-driven implementations advance quickly, but the field lacks a comprehensive mathematical theory of semantic and goal-oriented information to complement Shannon's information theory. Developing this would establish limits, design principles, and enhance understanding of trade-offs between rate, reliability, and goal effectiveness.
- **Multi-modal and Multi-task GOSC:** Many advanced applications like remote surgery or augmented reality involve diverse data streams (video, audio, haptic, LiDAR) combined for complex goals. Expanding GOSC frameworks to handle multi-modal inputs and support multiple, possibly conflicting, tasks is a key future focus research.

Addressing these challenges will drive the next decade, leading to a 6G era of faster, more reliable, and smarter communication.

References

[1] Z. Qin, G. Y. Li, and H. V. Poor, "A survey on semantic communications," *arXiv preprint arXiv:2101.01309*, 2021. H. Zhou, Y. Deng, X. Liu, N. Pappas, and A. Nallanathan, "Goal-Oriented Semantic Communications for 6G Networks," *arXiv preprint arXiv:2312.10555*, 2023. [Online]. Available: https://arxiv.org/abs/2312.10555

[2] Gui, Guan & Liu, Miao & Tang, Fengxiao & Kato, Nei & Adachi, Fumiyuki. (2020). 6G: Opening New Horizons for Integration of Comfort, Security and Intelligence. IEEE Wireless Communications. PP. 1-7. 10.1109/MWC.001.1900516.

[3] Wang, Yining & Han, Han & Feng, Yixiao & Zheng, Jianchao & Zhang, Bo. (Jan 2025). Semantic Communication Empowered 6G Networks: Techniques, Applications, and Challenges. IEEE Access. PP. 1-1. 10.1109/ACCESS.2025.3532797.

[4] J. Park, S. Samarakoon, M. Bennis, and M. Debbah, "Extreme URLLC: Vision, challenges, and key enablers," arXiv preprint arXiv:2001.09683, 2020. [Online]. Available: https://arxiv.org/abs/2001.09683

[5] P. Popovski et al., "Wireless Access for Ultra-Reliable Low-Latency Communication: Principles and Building Blocks," in IEEE Network, vol. 32, no. 2, pp. 16-23, March-April 2018, doi: 10.1109/MNET.2018.1700258.

[6] M. Bennis, M. Debbah and H. V. Poor, "Ultrareliable and Low-Latency Wireless Communication: Tail, Risk, and Scale," in Proceedings of the IEEE, vol. 106, no. 10, pp. 1834-1853, Oct. 2018, doi: 10.1109/JPROC.2018.2867029.

[7] E. Calvanese Strinati et al., "6G-GOALS: Goal-Oriented and Semantic Communications for AI-Native 6G Networks," arXiv preprint arXiv:2402.07573, 2024. [Online]. Available: https://arxiv.org/abs/2402.07573

[8] Y. E. Sagduyu, T. Erpek, A. Yener, and S. Ulukus, "Will 6G be Semantic Communications? Opportunities and Challenges from Task Oriented and Secure Communications to Integrated Sensing," arXiv preprint arXiv:2401.01531, 2024. [Online]. Available: https://arxiv.org/abs/2401.01531

[9] M. Shokrnezhad, H. Mazandarani, T. Taleb, J. Song and R. Li, "Semantic Revolution From Communications to Orchestration for 6G: Challenges, Enablers, and Research Directions," in IEEE Network, vol. 38, no. 6, pp. 63-71, Nov. 2024, doi: 10.1109/MNET.2024.3422348.

[10] W. Jiang, B. Han, M. A. Habibi and H. D. Schotten, "The Road Towards 6G: A Comprehensive Survey," in IEEE Open Journal of the Communications Society, vol. 2, pp. 334-366, 2021, doi: 10.1109/OJCOMS.2021.3057679.

[11] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan and M. Zorzi, "Toward 6G Networks: Use Cases and Technologies," in IEEE Communications Magazine, vol. 58, no. 3, pp. 55-61, March 2020, doi: 10.1109/MCOM.001.1900411.

[12] Yating Liu, Xiaojie Wang, Zhaolong Ning, MengChu Zhou, Lei Guo, Behrouz Jedari, A survey on semantic communications: Technologies, solutions, applications and challenges, Digital Communications and Networks, Volume 10, Issue 3, 2024, Pages 528-545, ISSN 2352-8648, https://doi.org/10.1016/j.dcan.2023.05.010.

[13] Z. Zhang et al., "6G Wireless Networks: Vision, Requirements, Architecture, and Key Technologies," in IEEE Vehicular Technology Magazine, vol. 14, no. 3, pp. 28-41, Sept. 2019, doi: 10.1109/MVT.2019.2921208.

[14] P. Yang, Y. Xiao, M. Xiao and S. Li, "6G Wireless Communications: Vision and Potential Techniques," in IEEE Network, vol. 33, no. 4, pp. 70-75, July/August 2019, doi: 10.1109/MNET.2019.1800418.

[15] Chen, Shanzhi & Liang, Ying-Chang & Sun, Shaohui & Kang, Shaoli & Cheng, Wenchi & Peng, Mugen. (2020). Vision, Requirements, and Technology Trend of 6G: How to Tackle the Challenges of System Coverage, Capacity, User Data-Rate and Movement Speed. IEEE Wireless Communications. PP. 1-11. 10.1109/MWC.001.1900333.

[16] C. Chaccour, M. N. Soorki, W. Saad, M. Bennis, P. Popovski and M. Debbah, "Seven Defining Features of Terahertz (THz) Wireless Systems: A Fellowship of Communication and Sensing," in IEEE Communications Surveys & Tutorials, vol. 24, no. 2, pp. 967-993, Secondquarter 2022, doi: 10.1109/COMST.2022.3143454.

[17] T. S. Rappaport et al., "Wireless Communications and Applications Above 100 GHz: Opportunities and Challenges for 6G and Beyond," in IEEE Access, vol. 7, pp. 78729-78757, 2019, doi: 10.1109/ACCESS.2019.2921522.

[18] Tataria, Harsh & Shafi, Mansoor & Molisch, Andreas & Dohler, Mischa & Sjoland, Henrik & Tufvesson, Fredrik. (2021). 6G Wireless Systems: Vision, Requirements, Challenges, Insights, and Opportunities. Proceedings of the IEEE. PP. 1-34. 10.1109/JPROC.2021.3061701.

[19] W. Saad, M. Bennis and M. Chen, "A Vision of 6G Wireless Systems: Applications, Trends, Technologies, and Open Research Problems," in IEEE Network, vol. 34, no. 3, pp. 134-142, May/June 2020, doi: 10.1109/MNET.001.1900287.

[20] K. B. Letaief, W. Chen, Y. Shi, J. Zhang and Y. -J. A. Zhang, "The Roadmap to 6G: AI Empowered Wireless Networks," in IEEE Communications Magazine, vol. 57, no. 8, pp. 84-90, August 2019, doi: 10.1109/MCOM.2019.1900271.

[21] Luong, N. C., Hoang, D. T., Wang, P., Niyato, D., Kim, D. I., & Han, Z. (2016). Data Collection and Wireless Communication in Internet of Things (IoT) Using Economic Analysis and Pricing Models: A Survey. *ArXiv*. https://arxiv.org/abs/1608.03475

[22] J. A. Stankovic, "Research Directions for the Internet of Things," in IEEE Internet of Things Journal, vol. 1, no. 1, pp. 3-9, Feb. 2014, doi: 10.1109/JIOT.2014.2312291.

[23] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari and M. Ayyash, "Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications," in IEEE Communications Surveys & Tutorials, vol. 17, no. 4, pp. 2347-2376, Fourthquarter 2015, doi: 10.1109/COMST.2015.2444095.

[24] M. Kavre, A. Gadekar and Y. Gadhade, "Internet of Things (IoT): A Survey," 2019 IEEE Pune Section International Conference (PuneCon), Pune, India, 2019, pp. 1-6, doi: 10.1109/PuneCon46936.2019.9105831.

[25] G. Fettweis and S. Alamouti, "5G: Personal mobile internet beyond what cellular did to telephony," in IEEE Communications Magazine, vol. 52, no. 2, pp. 140-145, February 2014, doi: 10.1109/MCOM.2014.6736754.

[26] J. G. Andrews et al., "What Will 5G Be?," in IEEE Journal on Selected Areas in Communications, vol. 32, no. 6, pp. 1065-1082, June 2014, doi: 10.1109/JSAC.2014.2328098.

[27] A. Osseiran et al., "Scenarios for 5G mobile and wireless communications: the vision of the METIS project," in IEEE Communications Magazine, vol. 52, no. 5, pp. 26-35, May 2014, doi: 10.1109/MCOM.2014.6815890. Ian F. Akyildiz, Shuai Nie, Shih-Chun Lin, Manoj Chandrasekaran, 5G roadmap: 10 key enabling technologies, Computer Networks, Volume 106, 2016, Pages 17-48, ISSN 1389-1286, https://doi.org/10.1016/j.comnet.2016.06.010.

[28] M. Agiwal, A. Roy and N. Saxena, "Next Generation 5G Wireless Networks: A Comprehensive Survey," in IEEE Communications Surveys & Tutorials, vol. 18, no. 3, pp. 1617-1655, thirdquarter 2016, doi: 10.1109/COMST.2016.2532458.

[29] G. Wunder et al., "5GNOW: non-orthogonal, asynchronous waveforms for future mobile applications," in IEEE Communications Magazine, vol. 52, no. 2, pp. 97-105, February 2014, doi: 10.1109/MCOM.2014.6736749.

[30] E. Hossain and M. Hasan, "5G cellular: key enabling technologies and research challenges," in IEEE Instrumentation & Measurement Magazine, vol. 18, no. 3, pp. 11-21, June 2015, doi: 10.1109/MIM.2015.7108393.

[31] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta and P. Popovski, "Five disruptive technology directions for 5G," in IEEE Communications Magazine, vol. 52, no. 2, pp. 74-80, February 2014, doi: 10.1109/MCOM.2014.6736746.
[32] T. S. Rappaport et al., "Millimeter Wave Mobile Communications for 5G Cellular: It Will Work!," in IEEE Access, vol. 1, pp. 335-349, 2013, doi: 10.1109/ACCESS.2013.2260813.
[33] S. M. R. Islam, N. Avazov, O. A. Dobre and K. -s. Kwak, "Power-Domain Non-Orthogonal Multiple Access (NOMA) in 5G Systems: Potentials and Challenges," in IEEE Communications Surveys & Tutorials, vol. 19, no. 2, pp. 721-742, Secondquarter 2017, doi: 10.1109/COMST.2016.2621116.
[34] L. Dai, B. Wang, Y. Yuan, S. Han, I. Chih-lin and Z. Wang, "Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends," in IEEE Communications Magazine, vol. 53, no. 9, pp. 74-81, September 2015, doi: 10.1109/MCOM.2015.7263349.
[35] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan and V. K. Bhargava, "A Survey on Non-Orthogonal Multiple Access for 5G Networks: Research Challenges and Future Trends," in IEEE Journal on Selected Areas in Communications, vol. 35, no. 10, pp. 2181-2195, Oct. 2017, doi: 10.1109/JSAC.2017.2725519.
[36] B. Zhou, I. Howenstine, S. Limprapaipong and L. Cheng, "A Survey on Network Calculus Tools for Network Infrastructure in Real-Time Systems," in IEEE Access, vol. 8, pp. 223588-223605, 2020, doi: 10.1109/ACCESS.2020.3043600.
[37] S. Zhang, "An Overview of Network Slicing for 5G," in IEEE Wireless Communications, vol. 26, no. 3, pp. 111-117, June 2019, doi: 10.1109/MWC.2019.1800234.
[38] Haque, M. E., Tariq, F., Khandaker, M. R., Hossain, M. S., Imran, M. A., & Wong, K. (August 2025). A Comprehensive Survey of 5G URLLC and Challenges in the 6G Era. ArXiv. https://arxiv.org/abs/2508.20205
[39] K. R. Kotte, L. Thammareddi, D. Kodi, V. R. Anumolu, A. K. K and S. Joshi, "Integration of Process Optimization and Automation: A Way to AI Powered Digital Transformation," *2025 First International Conference on Advances in Computer Science, Electrical, Electronics, and Communication Technologies (CE2CT)*, Bhimtal, Nainital, India, 2025, pp. 1133-1138, doi: 10.1109/CE2CT64011.2025.10939966.
[40] Thirunagalingam, A. (2024). Transforming real-time data processing: the impact of AutoML on dynamic data pipelines. Available at SSRN 5047601.
[41] Swathi Chundru, Arunkumar Thirunagalingam, Praveen Maroju, Pushan Kumar Dutta, Harsh Yadav, Pawan Whig, (2024/12/1), Internet of water: quantifying IoT's impact on urban water management and resource optimization in smart cities, 8th IET Smart Cities Symposium (SCS 2024), 2024, 528-533, IET.