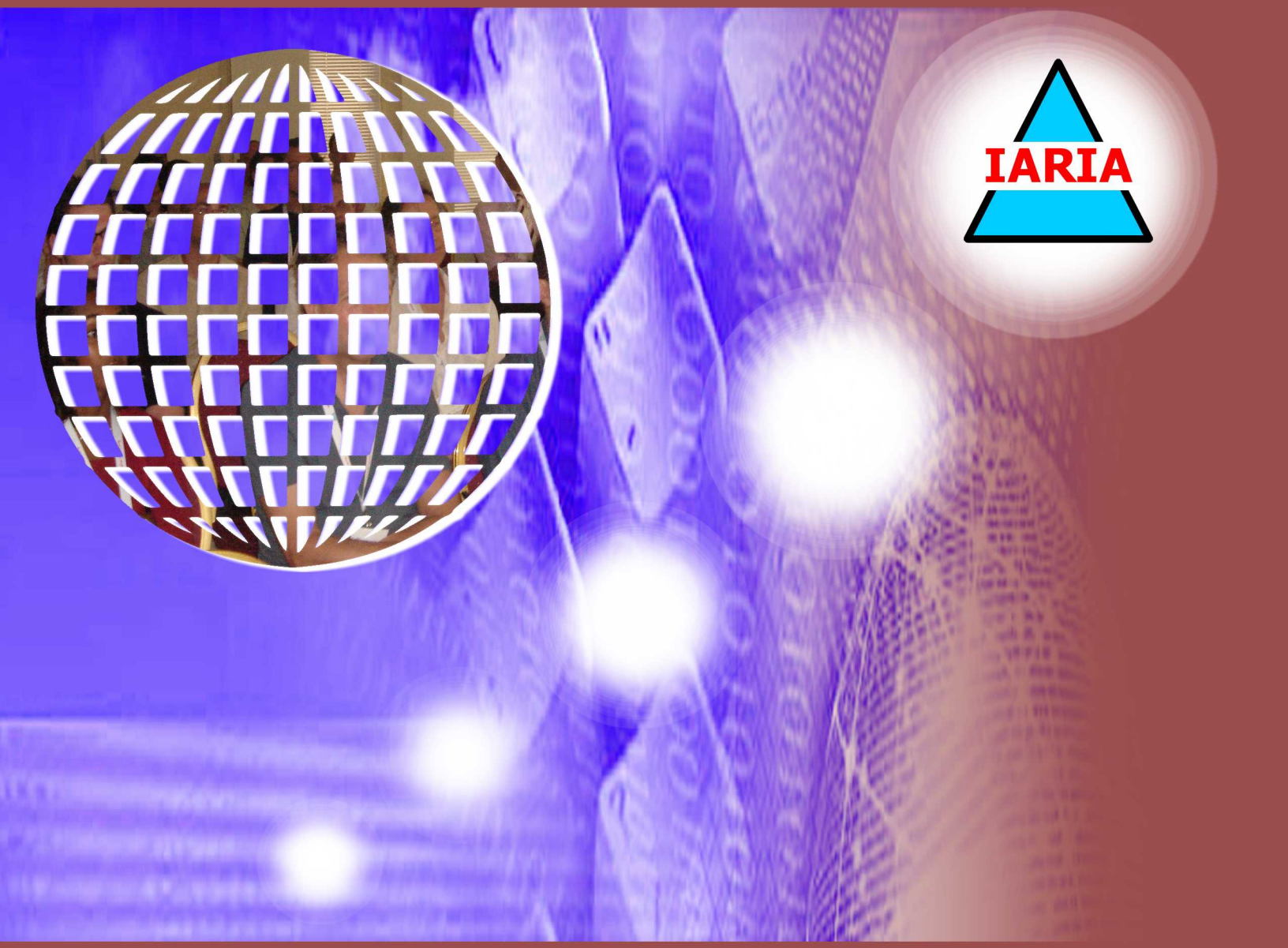
International Journal on Advances in Networks and Services



2025 vol. 18 nr. 3&4

The *International Journal on Advances in Networks and Services* is published by IARIA.
ISSN: 1942-2644
journals site: http://www.iariajournals.org
contact: petre@iaria.org

Sponsored by IARIA
www.iaria.org

International Journal on Advances in Networks and Services
Volume 18, Number 3&4, 2025

International Journal on Advances in Networks and Services
Volume 18, Numbers 3&4, 2025

CONTENTS

pages: 12 - 23
User-Friendly Simulator for Estimating Power Consumption and Communication Availability in Wireless Sensor Networks
Yuta Hosokawa, Railway Technical Research Institute, Japan
Yuki Ogawa, Railway Technical Research Institute, Japan
Ryuta Nakasone, Railway Technical Research Institute, Japan
Minoru Tanaka, Railway Technical Research Institute, Japan
Kazuki Nakamura, Railway Technical Research Institute, Japan

pages: 24 - 35
Beyond Best Effort: Routing with Partially Ordered Requirements
Bradley Smith, University of California Santa Cruz, USA
Paul Tatarsky, Tatarsky.com, USA

pages: 36 - 47
A Proposal of a Sensor Exclusion and Dynamic Cluster Head Selection Algorithm to Improve Energy Efficiency in Cognitive Radio Networks
Alcides Mário Joaquim Tomás, National Institute of Telecommunications - Inatel, Brasil
José Marcos Camara Brito, National Institute of Telecommunications - Inatel, Brasil

pages: 48 - 55
Hourly and Weather-Based Variability in Starlink Internet Performance: A TCP and UDP Throughput Study
Céline Careau, KTH Royal Institute of Technology, Sweden
Emil Fredriksson, KTH Royal Institute of Technology, Sweden
Robert Olsson, KTH Royal Institute of Technology, Sweden
Peter Sjödin, KTH Royal Institute of Technology, Sweden
Claes Beckman, KTH Royal Institute of Technology, Sweden

pages: 56 - 70
Impact of Variable Channel Conditions on Multi-Hop DTN-Based Lunar Communications with Traffic Prioritization
Klara Schaper, Institute of Communication Networks, Hamburg University of Technology, Germany
Teresa Algarra Ulierte, Institute of Communication Networks, Hamburg University of Technology, Germany
Andreas Timm-Giel, Institute of Communication Networks, Hamburg University of Technology, Germany
Felix Flentge, Directorate of Operations, European Space Agency, Germany

pages: 71 - 80
Unveiling Content Traps: A Network Resilience and Topic-Based Study of YouTube's Algorithmic Content Curation
Monoarul Bhuiyan, COSMOS Research Center, UA-Little Rock, USA
Nitin Agarwal, COSMOS Research Center, UA-Little Rock; ICSI, University of California-Berkeley, USA

pages: 81 - 89
Millimeter-Wave Information-Centric Wireless-Sensor-Network Ecosystem: Evaluation under Non-Terrestrial

and Long-Distance Environments in Actual City

Shintaro Mori, Fukuoka University, Japan

User-Friendly Simulator for Estimating Power Consumption and Communication Availability in Wireless Sensor Networks

Yuta Hosokawa, Yuki Ogawa, Ryuta Nakasone, Minoru Tanaka, and Kazuki Nakamura
Information and Communication Technology Division
Railway Technical Research Institute
Kokubunji, Tokyo, 185-8540 Japan
e-mail: {hosokawa.yuta.84, ogawa.yuki.87, nakasone.ryuta.07, tanaka.minoru.96, nakamura.kazuki.26}@rtri.or.jp

***Abstract*—In railway infrastructure monitoring, wireless sensor networks (WSNs) offer considerable potential for collecting sensor data at locations where communication cables and/or power lines cannot be installed. However, for practical adoption, railway operators must design networks that maintain communication reliability while minimizing the maintenance costs, particularly for battery replacement. In this study, we introduced simulation models that predict communication performance and power consumption in WSNs. The proposed simulation model was implemented as a user-friendly, web-based tool. The simulator enables railway operators to compare various network configurations without requiring specialized technical knowledge. To validate the simulation models, we constructed an actual WSN along a test railway line. The experimental results indicated that the simulator provides highly accurate estimations of the data arrival rates and power consumption, demonstrating the effectiveness of the approach. Thus, the proposed simulator supports cost-effective deployment and reduced maintenance effort in railway environments where wire-based sensing of infrastructure is impractical.**

Keywords*—monitoring; power consumption; railway; simulator; wireless sensor network.

I. INTRODUCTION

This study builds upon our preliminary work presented at The Eighteenth International Conference on Sensor Technologies and Applications (SENSORCOMM 2024) [1], where we introduced the concept of a web-based simulator for the deployment of wireless sensor networks (WSNs). We extend the findings of the previous work by obtaining parameters in real-world settings and validating the proposed models using the actual equipment installed in a railway environment.

In recent years, Japanese railway operators have increasingly sought to adopt more efficient and labor-saving technologies for the maintenance of railway equipment. One approach is to use WSNs to remotely monitor equipment distributed along the railway lines. Using WSNs can reduce the frequency of on-site inspections; however, the following elements must be considered when applying this principle: communication range, modulation method, power consumption, and power supply implementation. These elements interact with each other, and the variety of available options renders the design of optimal WSNs challenging for railway operators.

Several types of network simulators have been developed for WSN design. Sharma et al. [2] conducted a comprehensive comparative analysis of the WSN simulation frameworks, while highlighting their merits and demerits based on key characteristics such as scalability, graphical support, and ease of implementation. Idris et al. [3] surveyed the available Internet of Things (IoT) and WSN simulation tools for long-range wide-area network (LoRaWAN) [4] to demonstrate the evolving landscape of network simulation tools. These comparative analyses reveal the trade-offs between simulation flexibility and practical usability, wherein existing simulators typically prioritize accuracy and research depth over operational accessibility.

General-purpose simulators such as Network Simulator-2 (ns-2) [5], Network Simulator-3 (ns-3) [6], and OMNeT++ [7] provide flexible discrete-event simulation platforms for network research. These tools can be used to model various network types, ranging from wired networks to wireless sensor networks. In particular, ns-3 supports power consumption modeling through extensions such as Energy Framework [8]. A WSN-specific module Castalia [9] was developed for OMNeT++. More recently, the INET Framework [10] provided extensions for simulating WSNs, including radio propagation and power consumption models. The ns-2, ns-3, and OMNeT++ are open-source simulators, whereas QualNet [11] is a commercial network simulator that provides high-fidelity models and supports a wide range of wireless technologies. QualNet has been widely used in both academia and industry. Furthermore, it has been integrated into EXata [12], a real-time network emulation platform that inherits the detailed protocol models of QualNet while focusing on real-time network emulation capabilities. However, substantial programming expertise (typically in C++) and a deep understanding of simulation are essential for using these simulators.

Protocol-specific simulators typically focus on particular communication methods and provide detailed analysis for specific protocols. For example, LoRaSim [13] is a Python-based simulator that models large-scale LoRa networks using an analytically derived physical layer model. It focuses on evaluating the scalability and packet collisions in dense deployments, thereby enabling researchers to determine the change in network performance with an increase in the number of LoRa nodes. Another simulator, LoRaWANSim [14], was implemented in MATLAB that offers comprehensive LoRaWAN simulation from the physical layer

to the application layer. MATLAB-based tools present several advantages in terms of rapid prototyping and visualization, making them accessible to engineers familiar with numerical computing. Although these simulators excel at protocol-level performance evaluation, their single-protocol focus limits their utility when comparing multiple communication methods.

Development-integrated platforms such as Contiki-NG [15] provide an operating system for IoT devices combined with the Cooja simulation environment [16], enabling developers to test the same application code on both simulated and real hardware. This framework enables users to simulate the behavior of embedded devices without modifying the code. Although this approach is beneficial for protocol development and debugging, it requires programming expertise in C and embedded systems.

Overall, these simulators typically require advanced technical expertise for command-line operations, custom scripting, or detailed parameter configuration. Conversely, there is extensive demand from non-technical users of railway operators to rapidly evaluate the possible WSN configurations among the various available options. Therefore, conventional, highly detailed network simulators are typically unsuitable for railway operators, as their complexity makes rapid, practical evaluation difficult for non-experts in wireless networking.

To address these issues, we propose a user-friendly simulator to estimate the communication quality and battery consumption of WSNs in railway environments. The proposed simulator supports multi-hop transmission and mesh networking for nodes that extend over long distances along the railways. It features a web-based graphical user interface (GUI), which enables users to set parameters easily, compare node placements and communication methods, and instantly view the results. The simulator satisfies the specific requirements of WSN deployment planning in railway maintenance operations without requiring specialized software or technical expertise.

The remainder of this paper is organized as follows. Section II presents the proposed simulation framework. Section III describes in detail the experimental characterization of the WSN technologies, including the radio propagation properties and power consumption behavior. Section IV presents a detailed explanation of the development of the simulation models. Section V describes the implementation of the web-based simulator. Section VI presents the verification of the simulator using actual WSN deployments along a test railway line. Section VII details a practical case study that depicts the implementation of the simulator in WSN design under realistic constraints. Finally, Section VIII concludes the paper.

II. Proposed Simulation Framework

In railway environments, WSN deployment requires operators to evaluate node placement, communication methods, and battery replacement schedules corresponding to specific monitoring targets and deployment sites. To meet these requirements, we developed a comprehensive simulation framework to evaluate the communication quality and power consumption of wireless sensors designed

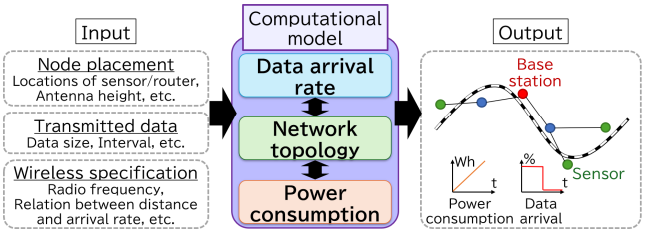


Figure 1. Overview of the proposed simulation framework.

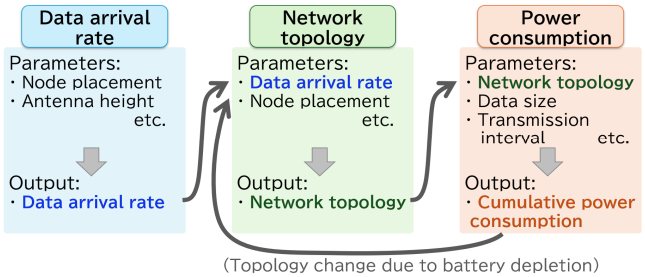


Figure 2. Relationship of the three estimation models.

specifically for the railway infrastructure constraints and operational requirements.

Figure 1 shows the overall architecture of the proposed simulation method. This framework accepts three types of input parameters: node placement information, including wireless sensor locations, and site characteristics; data transmission specifications, such as the payload size and transmission intervals; and wireless specifications. The computational model performs a WSN simulation based on these inputs and outputs power consumption profiles and data arrival rates for each network node.

The computational model comprises three interconnected estimation models: data arrival rate estimation, network topology estimation to determine the data transmission paths, and power consumption estimation. These models operate cooperatively to conduct a comprehensive network simulation (Figure 2).

The proposed framework adopts a feedback mechanism, wherein the output from each estimation model is fed into subsequent models to create an integrated simulation loop that captures dynamic network interactions. This interconnected approach enables the realistic modeling of the network behavior under various operational conditions.

Additionally, modeling the topology reconfiguration based on battery depletion events is a crucial feature of the proposed framework. When the power consumption estimation indicates node battery exhaustion, this information is propagated back to the network topology estimation model, which recalculates optimal routing paths for the modified network. This iterative process continues throughout the simulation, presenting a realistic prediction of long-term network evolution and maintenance requirements.

Furthermore, the proposed framework comprises various WSN architectures, ranging from simple star topologies, as typically employed in LoRaWAN, to complex mesh networks, commonly observed in wireless smart ubiquitous networks

(Wi-SUN) [17] deployments. This flexibility enables the implementation of the proposed framework in various railway monitoring applications. Moreover, the modular architecture facilitates adaptation to different communication protocols by replacing individual estimation models, while preserving the integrity of the overall simulation framework.

Existing network simulators require specialized technical knowledge and command-line expertise. Conversely, the proposed framework prioritizes practical usability for railway operators. The simulation provided immediate feedback on crucial deployment parameters, thereby enabling rapid design iteration and network optimization prior to physical implementation.

III. Experimental Characterization of WSN Technologies

To establish estimation models for communication performance and power consumption, we conducted field experiments using Wi-SUN and LoRa devices.

A. Target Technology Selection and Measurement Approach

The appropriate selection of frequency bands and communication protocols is crucial in WSN deployment, particularly for railway applications with unique operational constraints. We established the following specific selection criteria: communication range capabilities of several kilometers, battery operation sustainability for multiple years, and license-free operation in Japan, enabling railway operators to independently construct private networks. These requirements ensure the practical applicability of railway infrastructure monitoring while maintaining operational autonomy.

Based on these criteria and their widespread implementation in WSN deployments in Japanese railways, we selected Wi-SUN and LoRa as the target protocols for comprehensive characterization. The complementary characteristics of these protocols, i.e., the reliability of Wi-SUN and the exceptional range of LoRa, render them well-suited for railway monitoring scenarios. We performed targeted baseline measurements using the aforementioned technologies. These measurements are not intended to be universally applicable across all wireless systems; instead, they reflect the specific conditions corresponding to railway applications.

1) Frequency Band Selection

The protocol selection adopted in this study focused on specific low-power radio stations operating without radio station licenses in Japan. Among the available frequency bands (e.g., 429 MHz, 920 MHz, and 2.45 GHz), we selected the 920 MHz band for its optimal balance of communication range, data rate, and power efficiency. Although the 920 MHz band presents a shorter range than the 429 MHz band, it provides superior data rates and multi-hop transmission. When compared with the 2.45 GHz band, it sacrifices speed for a significantly longer range and lower power consumption, which are critical factors for battery-operated railway sensors.

2) Protocol Characteristics

a) Wi-SUN

Wi-SUN is an IEEE802.15.4g-compliant protocol that is widely employed in smart meter applications. Despite the higher standby current (tens of mA) when compared with other LPWA technologies, Wi-SUN delivers communication speeds of several hundred kbps over distances up to 1 km. Its field area network (FAN) profile provides robust mesh networking with automatic route reconfiguration during node failures, thereby ensuring high network reliability for critical railway monitoring applications.

b) LoRa

LoRa employs chirp spread spectrum modulation to achieve exceptional power efficiency and transmission range. It excels in remote monitoring scenarios owing to standby currents as low as 0.7 μA and communication distances extending 1–10 km. Although the LoRaWAN typically employs star topologies, private LoRa configurations support multi-hop networking, thereby providing deployment flexibility for the linear railway infrastructure.

The complementary characteristics of these protocols render them essential for different railway monitoring scenarios, justifying their selection for comprehensive characterization.

B. Radio Propagation Characteristics

We systematically measured the RSSI and data arrival rates across various inter-node distances to establish reliable communication feasibility criteria for the proposed simulation framework. For target communication ranges of several kilometers, we conducted experiments on a straight viaduct, which provides the 3 km straight sections required for long-range testing.

The experimental setup involved fixing a base station and moving a remote unit to various distances while measuring the RSSI and data arrival rates at each location. Figure 3 shows the experimental configuration employed in this study. We configured the transmission power at 20 mW, representing the maximum allowable output for the 920 MHz-specified low-power radio stations in Japan, with a standardized payload size

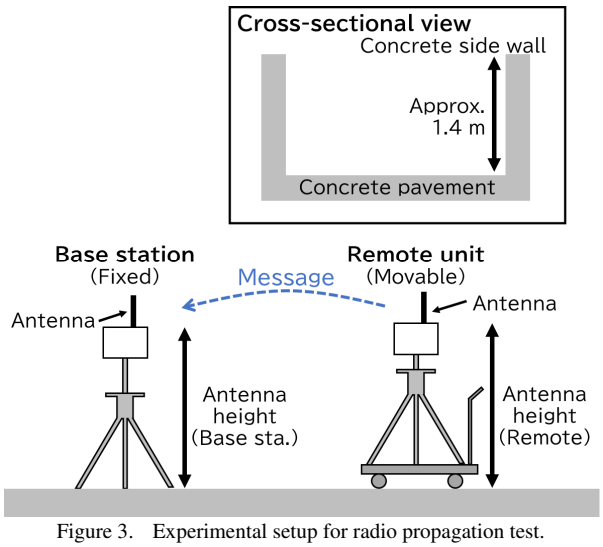


Figure 3. Experimental setup for radio propagation test.

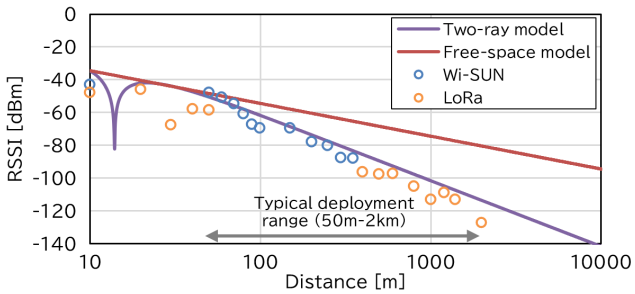
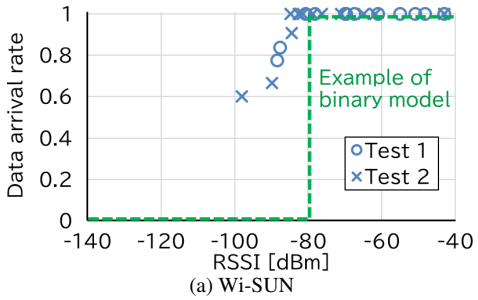
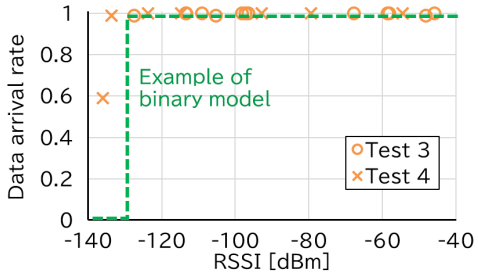


Figure 4. Relationship between distance and RSSI.



(a) Wi-SUN



(b) LoRa

Figure 5. Relationship between RSSI and data arrival rate.

of 16 bytes per transmission. The testing conditions included multiple antenna height configurations and both Wi-SUN and LoRa protocols to ensure comprehensive characterization.

Measurement Results: Figure 4 shows the relationship between antenna distance and RSSI. The measurement results indicated that both the Wi-SUN and LoRa protocols exhibit similar distance-dependent RSSI characteristics under identical antenna and power configurations. In particular, in the range 50 m to 2 km, which represents that of typical WSN deployment scenarios, the measured values closely concurred with the two-ray ground-reflection model [18] predictions rather than free-space propagation models.

Consequently, we adopted the two-ray ground-reflection model as the foundation for our simulation framework. This model assumes a flat ground surface with a clear line-of-sight path between the nodes. It characterizes radio propagation by considering the interference between the direct path and the ground-reflected path. It is particularly advantageous for railway environments, where communication nodes are typically arranged along linear tracks on flat terrain or elevated viaducts. However, it does not account for the additional attenuation or interference caused by railway

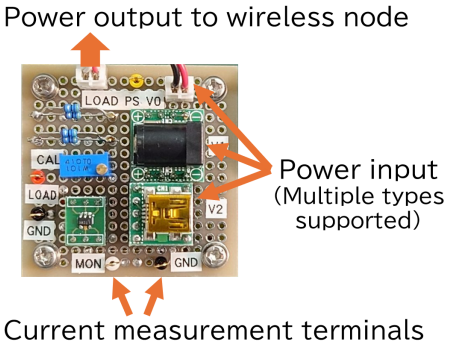


Figure 6. Custom circuit for current detection.

Figure 7. Commercial USB tester.

infrastructure, such as utility poles, overhead power lines, tunnels, or station structures.

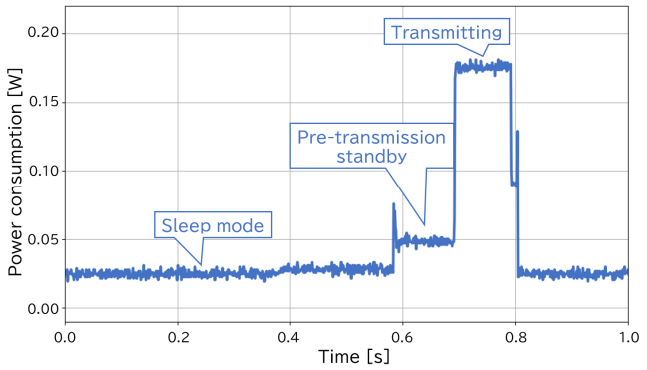
Figure 5 shows the relationship between the RSSI and data arrival rates. The results demonstrated a characteristic threshold behavior for both protocols. Above certain RSSI thresholds, both the Wi-SUN and LoRa maintained arrival rates exceeding 99%. However, the arrival rates decreased rapidly when the RSSI fell below these thresholds, indicating that a binary threshold-based model (binary model) was appropriate for simulation.

C. Power Consumption Characterization

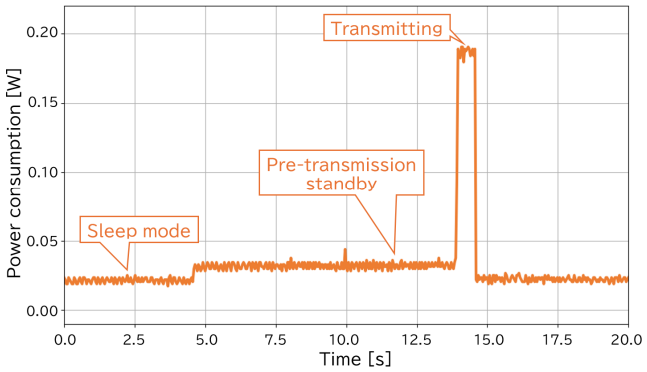
Measurement of power consumption presents unique challenges for the characterization of wireless sensors. Voltage monitoring requires a simple parallel connection, whereas current measurement requires series insertion of measurement equipment into the power path. This requirement was addressed using two complementary approaches: a custom current detection circuit for module-level measurement (Figure 6) and a commercial USB tester for system-level characterization (Figure 7).

The current detection circuit developed in this study enables insertion between the wireless sensor modules and power supplies, presenting a voltage output proportional to the consumed current. This approach enables a detailed oscilloscope-based analysis of the instantaneous power consumption patterns. Additionally, we selected and validated the AVHzY CT-3 USB tester [19], which demonstrated a measurement accuracy with an error within 1% at 1000 samples per second while providing continuous data logging capabilities.

Measurement Results: The characterization of power consumption revealed distinct operational states for both the Wi-SUN and LoRa modules, with consumption varying significantly between the sleep, standby, and transmission



(a) Wi-SUN



(b) LoRa

Figure 8. Example results of power consumption.

phases. Figure 8 shows the representative power consumption waveforms for both the protocols, demonstrating the power consumption profiles during different operational phases. As expected, the transmission duration and the associated power consumption scaled with the data payload size, thereby requiring multiple measurement scenarios to establish the representative power models.

These measurement data present the parameters required for the power consumption estimation model of the proposed simulation framework. An analysis of the power consumption profiles indicates that the node power consumption can be effectively modeled using two representative states. This includes the waiting state, which dominates the overall operation time when the node is not actively transmitting data, and the transmitting state, which exhibits the highest instantaneous power consumption. In this study, we refer to the non-transmitting periods, including sleep or reception standby, as the waiting state. This two-state approximation ensures that the essential power consumption behavior is captured while maintaining the computational efficiency for simulation. However, these results represent specific hardware configurations used in validation testing. Practical deployment requires the characterization of the actual target hardware to ensure accurate power consumption prediction.

IV. Simulation Model Development

The simulation framework employed the three interconnected models highlighted in Section II: data arrival rate estimation, network topology determination, and power consumption prediction models. These models operate cooperatively to capture the real-world network dynamics including topology reconfiguration due to battery depletion.

A. Data Arrival Rate Estimation Model

The data arrival rate estimation model determines the communication feasibility between the node pairs based on their positioning and transmission parameters. For the Wi-SUN and LoRa protocols operating in the 920 MHz band under clear line-of-sight conditions with inter-node distances within 2 km, the relationship between communication distance and RSSI closely follows the two-ray ground-reflection model predictions, as described in Section III. B.

Our measurements demonstrated a clear threshold behavior in the RSSI-arrival rate relationship: communication was highly reliable (>99% arrival rate) above certain RSSI values, but deteriorated rapidly below these thresholds.

Therefore, we employed the two-ray ground-reflection model to calculate the RSSI at each node pair for simulation and then compared these values against the predetermined RSSI thresholds to determine the communication feasibility. This approach presents a practical binary decision criterion: communication was considered to be feasible (arrival rate = 100%) when the calculated RSSI exceeded the threshold and infeasible (arrival rate = 0%) otherwise.

The two-ray ground-reflection model estimation formula based on the inter-node distance was expressed as follows:

$$P_{rx} = P_{tx} - L_{tx} + G_{tx} + G_{rx} - L_{rx} + 20\log_{10}\frac{\lambda\times\left|\sin\frac{2\pi h_{tx}h_{rx}}{\lambda d}\right|}{4\pi d}. \quad (1)$$

$$\begin{pmatrix} d\text{: Distance between the nodes[m]} \\ \lambda\text{: Wavelength [m]} \\ P_{tx}\text{: Transmission power [dBm]} \\ P_{rx}\text{: Received power [dBm]} \\ G_{tx}\text{: Transmitting node antenna gain [dBi]} \\ G_{rx}\text{: Receiving node antenna gain [dBi]} \\ L_{tx}\text{: Cable and filter loss of transmitting node [dB]} \\ L_{rx}\text{: Cable and filter loss of receiving node [dB]} \\ h_{tx}\text{: Transmitting node antenna height [m]} \\ h_{rx}\text{: Receiving node antenna height [m]} \end{pmatrix}$$

Figure 9 shows an example wherein seven nodes were placed, and the communication availability between the nodes was estimated using the aforementioned procedure. The yellow dashed lines represent the communication-available links, and the numbers represent the estimated RSSI values.

B. Network Topology Estimation Model

The network topology estimation model determined the optimal transmission paths when there were multiple routing options between the sensor nodes and the base station. We employed the routing protocol for low-power and lossy networks (RPL), which was widely adopted in Wi-SUN and similar multi-hop networks, to autonomously establish these transmission paths. Our implementation follows the standard RPL procedure, which is given as follows:

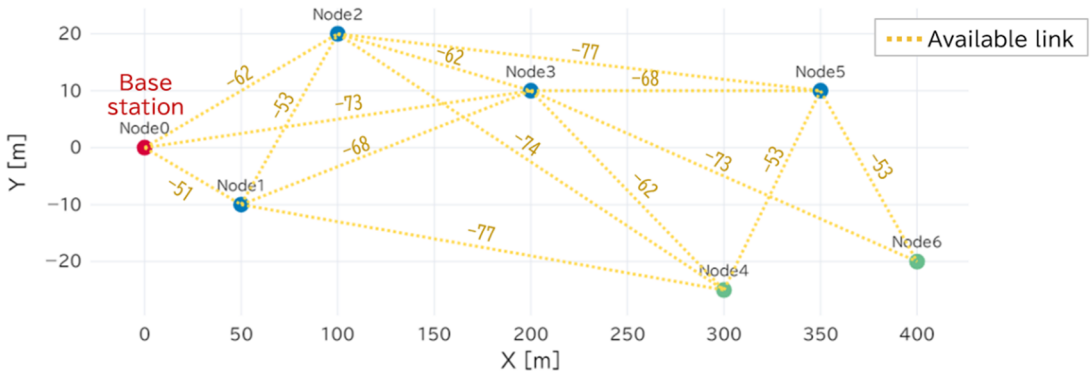


Figure 9. Example of communication availability estimation.

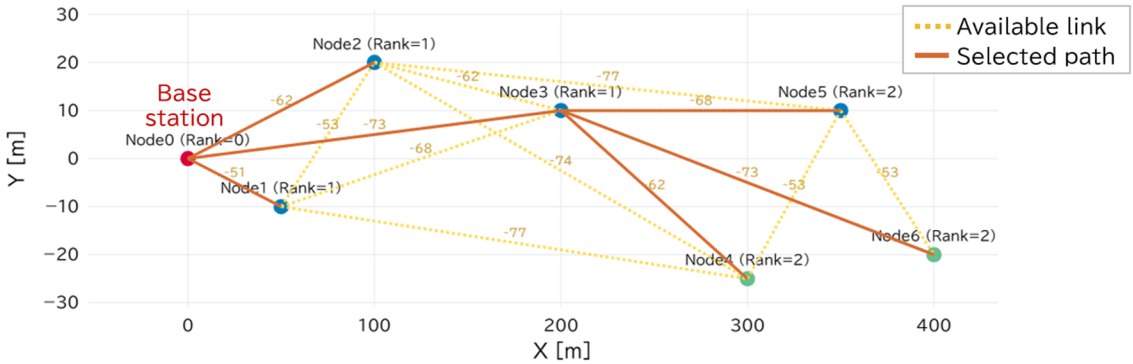


Figure 10. Example of network topology estimation.

1) We assigned a Rank value of 0 to the base station, where Rank represents the hop count required to reach the base station.
2) We assigned a Rank of 1 to the nodes capable of direct communication with the base station, with direct routes established to the base station.
3) We assigned a Rank of 2 to nodes that can communicate with nodes of Rank 1, but lack established routes. When there are multiple parent candidates, the algorithm selects the route with the highest estimated RSSI.
4) This process was repeated iteratively, assigning a Rank of k+1 to the nodes communicating with nodes with Rank k until all the nodes have established routes.

Figure 10 shows example results of estimating the transmission paths from each node to the base station, assuming the node configuration shown in Figure 9. This implementation ensured optimal parent selection based on signal strength, while minimizing hop counts to the base station. The modular design enabled the substitution of alternative routing algorithms to support different multi-hop protocols beyond Wi-SUN.

C. Power Consumption Estimation Model

The power consumption model estimated the battery consumption for each battery-powered node using a simplified two-state approach based on the measurement results obtained in this study. The wireless sensor power consumption varied significantly across operational phases, with the largest differences occurring between the waiting and active transmission periods, as described in Section III. C. Figure 11 shows the employed power consumption model.

The proposed model simplified this behavior into two primary states: a waiting state consuming P_{wait} [W] and a transmitting state consuming P_{tx} [W]. These parameters were configured based on the node functionality. Sensor nodes (leaf nodes), which only collect and transmit the sensor data, can utilize sleep mode while waiting to minimize the power consumption. Conversely, the router nodes must maintain a constant standby operation to receive and relay data from the

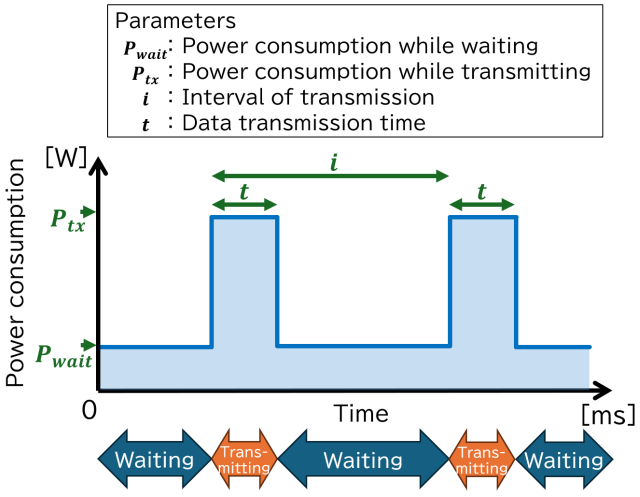


Figure 11. Power consumption model [1].

other nodes, thereby presenting higher power consumption during the waiting state (P_{wait}). The transmission duration scaled linearly with the data payload size. In the proposed model, this relationship was expressed as $t = ax + b$, where x [Bytes] denotes the size of the transmitted data. Each node transmits data at its configured transmission interval i [s]. In multi-hop networks, router nodes must relay the data received from child nodes in addition to transmitting their own sensor data, presenting higher power consumption.

The cumulative power consumption was calculated by integrating these power states over time and generating battery consumption predictions for each network node. This enables the evaluation of battery replacement schedules and identification of nodes requiring higher-capacity batteries or an external power supply. The proposed modeling approach captures the essential WSN operational dynamics while maintaining the computational efficiency required for real-time network design evaluation.

V. IMPLEMENTATION

We developed a simulator for WSNs that considers both data transmission and power consumption. The simulator was implemented in Python 3.11 using an open-source Python library called the Streamlit [20] library to enable web-based GUI controls. It was developed on a Linux server in our local network, enabling users to interact with it via a web browser on any PC connected to the network. Figure 12 shows a sample screen of the interface of the proposed simulator. The simulator comprised the following three components:

A. *Parameter input interface*

The simulator provides a web-based interface for users to input all the required parameters, including the node placement, characteristics, and battery capacity. The interface features real-time updates, where variations in the input values are instantly reflected in the corresponding graphs, facilitating rapid iteration of the simulation parameters.

B. *Network and Communication Path Visualization*

The simulator generates a visual representation based on the input parameters, which depicts the

- Node placement
- Inter-node communication availability
- Data transmission routes from each node to the base station

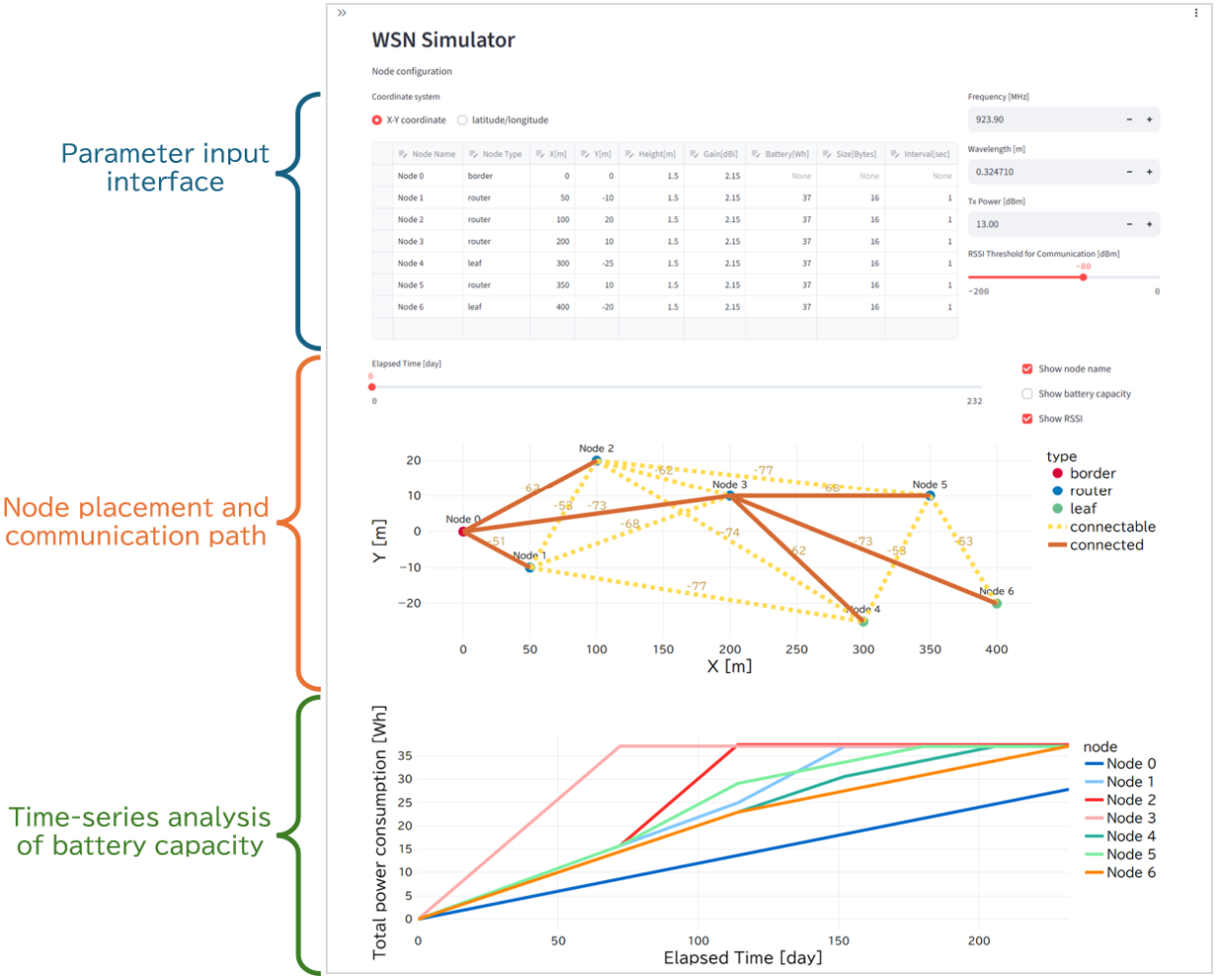


Figure 12. Screenshot of the developed simulator.

A dynamic timeline feature enables users to simulate the battery consumption over time. This includes dynamic re-routing when the nodes deplete their batteries.

C. Time-series analysis of battery capacity

The simulator presents a graph that depicts the relationship between the elapsed time and cumulative power consumption for each node. The key features are as follows:

- Simulation is continued until the entire battery capacity is depleted.
- Time resolution can be selected as one day, one hour, or one minute, allowing users to optimize the simulation for planning battery replacement strategies in WSN deployments.
- Visualization of long-term network behavior in scenarios without battery replacement is feasible.

VI. VERIFICATION

To demonstrate that the proposed simulator accurately reflects the behavior of a WSN implemented with actual devices, we constructed a Wi-SUN-based WSN along a test railway line at the Railway Technical Research Institute and compared the results with those obtained from the simulator. The experiments were conducted to demonstrate the following: (1) reproduction of the network topology and reflection of changes in topology when the battery capacity of a node runs out; (2) accuracy of data arrival rate estimation, and (3) accuracy of power consumption estimation.

A. Experimental Setup

Figure 13 shows the node placement in the verification experiments. We installed wireless sensor nodes based on the 920 MHz Wi-SUN standards at these locations (Figure 14). Each router and sensor node periodically collected the environmental data. In particular, the temperature, humidity, and atmospheric pressure data were obtained and transmitted wirelessly to the base station. The sensor node was placed far away and could not communicate directly with the base station. Consequently, the data transmission paths were established using the RPL, thereby enabling multi-hop communication via intermediate router nodes. We employed ROHM's Wi-SUN communication modules (model: BP35C5-T01).

We artificially increased the size of the transmitted data to accelerate battery consumption and complete the verification experiments within a reasonable time frame. Although real-world sensor data would only contain a few bytes, we appended random data to enlarge each packet to 256 bytes. Furthermore, we configured a short transmission interval of one second, which was considerably more frequent than in the typical use cases, to emphasize the power consumption model.

B. Verification of Network Topology and Arrival Rate Simulation

Figure 15 presents the simulation results of the data transmission path obtained by executing the simulator with the node placement and corresponding parameters depicted in Figure 13. The orange lines in the figure represent the selected transmission path, which was determined using the network

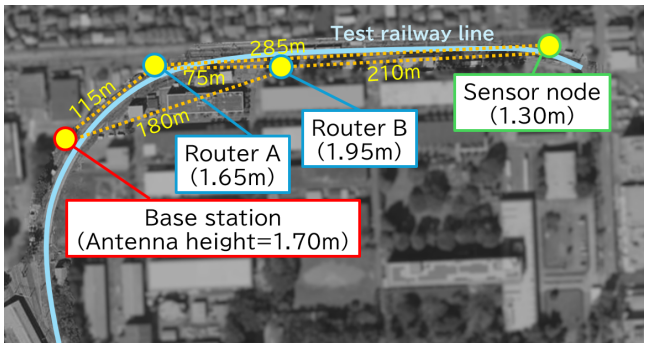


Figure 13. Node placement in the experiments (modified from an aerial photograph provided by the Geospatial Information Authority of Japan [21]).

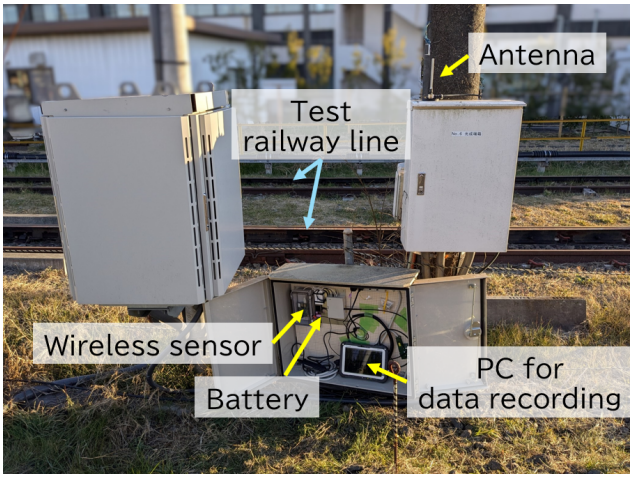
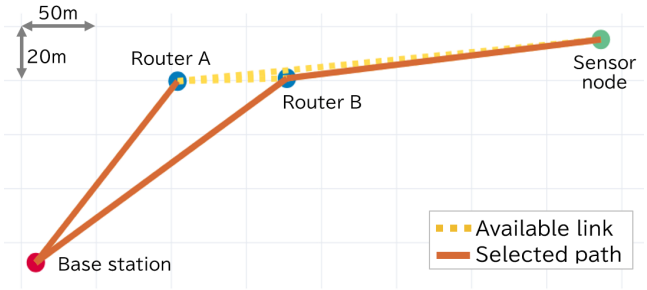
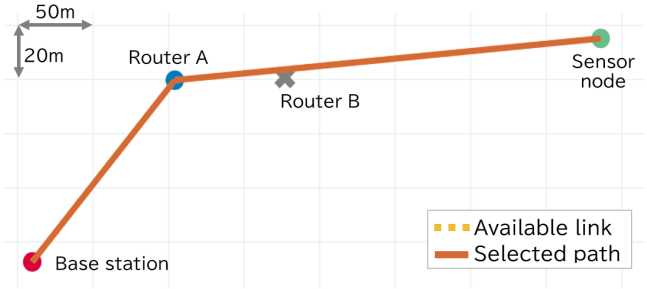


Figure 14. Wireless sensor node deployed along the test railway line.



(a) Initial transmission path



(b) Transmission path after router B became unavailable

Figure 15. Simulation results of the transmission path.

TABLE I. MEASURED RESULTS OF RSSI AND DATA ARRIVAL RATE

Link	RSSI [dBm] (Estimated)	RSSI [dBm] (Measured Avg.)	Arrival Rate (Estimated)	Arrival Rate (Measured)
Sensor node → Router B	-73	-76.82	100%	99.97% (286657/286738)
Router B → Base station	-67	-73.62	100%	100.00% (295314/295318)
Router A → Base station	-61	-66.63	100%	100.00% (295145/295154)

topology estimation model described in Section IV. B. Direct transmission could not be achieved under the test conditions owing to the long distance between the terminal sensor node and the base station. Therefore, the sensor data were initially transmitted via router B. Subsequently, router B experienced battery depletion owing to its relatively high power consumption for relaying operations, presenting a change in the network topology, as shown in the simulation outcome.

We deployed a WSN along the test railway line using the Wi-SUN modules to validate the simulation results. Consequently, the initial network topology matched the simulation output, with the sensor data being transmitted through router B. Furthermore, router B eventually exhausted its battery capacity as time progressed owing to its relaying operations, and the transmission path was shifted to the route via router A (Figure 15 (b)).

Table I presents the measurement results of the RSSI and data arrival rate. The experiment involved continuous data transmission at 1-s intervals over approximately 88 h, totaling approximately 300,000 transmissions. This table lists the average RSSI value and arrival rate for each communication link. The measured RSSI values were up to 6 dB lower than the estimated values. This was attributed to the presence of physical obstacles in the actual environment, such as railway structures and wayside buildings, which were not considered in the simulation that assumed an unobstructed terrain. In particular, the link between router B and the base station likely experienced a deteriorated RSSI owing to the obstruction by wayside buildings (see Figure 13).

Despite these environmental differences, the actual data arrival rates for all the communication links, estimated to be 100% in the simulation, were 99.97% or higher. These results indicated that the estimation error of the simulation model was sufficiently low, thereby demonstrating the practical reliability of the model.

C. Verification of Power Consumption Simulation

We measured the power consumption of routers A and B, and the sensor node in a WSN deployed along the test railway line to evaluate the accuracy of the simulator in estimating the power consumption. Routers A and B were powered using mobile batteries with a rated capacity of 11.84 Wh. The experiment was continued until the data transmission ceased due to the depletion of the battery capacity.

Figure 16 shows the cumulative power consumption over time for each node. The dashed lines represent the simulation results, whereas the solid lines represent the actual experimental measurements. At the beginning of the

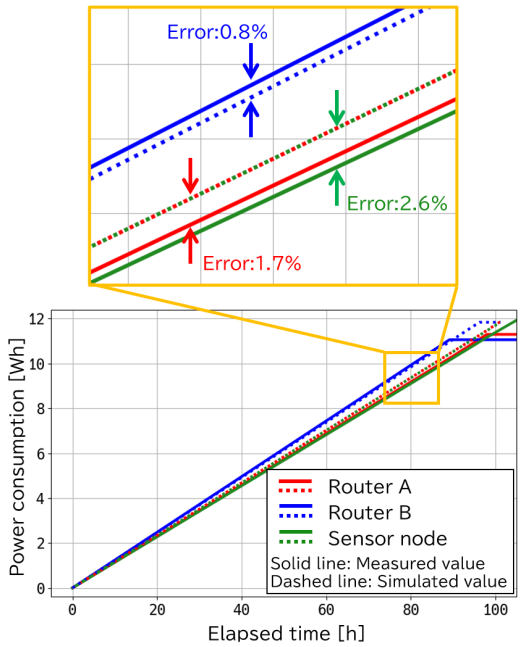


Figure 16. Time-series comparison of simulated and measured cumulative power consumption.

experiment, the data from the sensor node were transmitted via router B (Figure 15(a)). Consequently, router B has to transmit both its own sensor data and that obtained from the sensor node, presenting higher power consumption when compared with router A. As highlighted in the magnified view presented in Figure 16, the estimation error appeared as a difference in the slope of the cumulative power consumption over time, representing the difference in the average power consumption. The operating time until the depletion of the battery capacity was determined by the total energy extracted from the battery, corresponding to the consumption rate. Therefore, we calculated the total energy consumed by each router in the experiment. Table II presents the detailed measurement results.

The maximum error in the estimated power consumption per unit time was 2.6%, demonstrating that the simulator accurately estimated the power consumption of the wireless sensor nodes (Table II). Conversely, the actual-to-rated capacity ratios indicated that the batteries delivered 93.4% to 95.4% of their rated capacity under the test conditions. This

TABLE II. DETAILED MEASUREMENT RESULTS OF POWER CONSUMPTION

Node	Average power consumption [W]		
	Estimated value	Measured value	Error rate
Router A	0.117	0.115	1.7%
Router B	0.123	0.124	0.8%
Sensor node	0.117	0.114	2.6%

Node	Battery capacity [Wh]		
	Rated	Measured capacity	Capacity ratio (Actual/Rated)
Router A	11.84	11.30	95.4%
Router B	11.84	11.06	93.4%
Sensor node	—	—	—

variation was attributed to factors such as the battery type, degradation level, ambient temperature, and humidity.

D. Model Limitations and Practical Considerations

Although the proposed simulation models aligned closely with the measurement results, they still face several limitations.

First, the current network topology estimation assumes the use of RPL routing as defined in Wi-SUN FAN. However, the modular design of the simulator enables this component to be replaced with alternative routing algorithms. This flexibility enables future extensions to various communication standards based on the user requirements.

Second, the radio propagation model is based on the two-ray ground-reflection assumption, which assumes a flat ground surface and clear line-of-sight conditions. This model is well-suited for railway environments where nodes are typically deployed along flat, straight railway tracks or elevated viaducts. However, more detailed propagation modeling is required for environments with complex geometries or significant obstructions.

Lastly, the battery consumption model does not consider temperature-dependent characteristics or long-term degradation of battery cells. Although these factors were outside the scope of this work, incorporating them can improve the lifetime estimation accuracy in real deployments.

VII. Case Study: Applying the Simulator to Real World WSN Design

We conducted a case study on the design process of a WSN to demonstrate the practical utility of the proposed simulator, focusing on how it supports the selection of optimal node configurations and communication methods under realistic constraints. This section describes the procedure, input parameters, and representative results obtained using the proposed simulator.

A. Simulation Setup and Input Parameters

The simulator requires several key input parameters to accurately model WSN behavior:

1) Node Placement

The geographical positions of the nodes were specified in either latitude/longitude or relative distance coordinates. These positions must be determined based on the actual deployment scenarios in railway environments. The locations of the sensor nodes (where the sensors were to be installed) and the base station (where the sensor data were aggregated) were defined. Additionally, the candidate router locations must also be set as parameters.

2) Wireless Communication Parameters

Wireless characteristics, such as the frequency band and receiver sensitivity (e.g., reception threshold) were configured. These parameters can be derived from actual measurements or datasheets of commercially available devices.

3) Power Consumption Parameters

The power consumption values for both the transmission and waiting states, along with the battery capacity, were specified based on the empirical measurements of the candidate hardware. These values enabled the simulator to estimate the power consumption and battery replacement cycles under different traffic patterns and network topologies.

B. Example of WSN Design using the Simulator

In an example scenario, four sensor nodes were placed along a hypothetical railway section; they communicated with the central base station via an intermediate router. The initial configuration is shown in Figure 17(a). In this scenario, the router must forward data from all four sensor nodes, resulting in high power consumption and early battery depletion. Figure 17(b) shows the simulation results of power consumption using a 100 Wh battery. The results indicated that the battery lasted for approximately 61 days.

Based on the simulation results, railway operators aim to extend the battery replacement cycles to reduce the cost of maintaining these networks. We analyzed several redesign strategies for extending the battery replacement cycle.

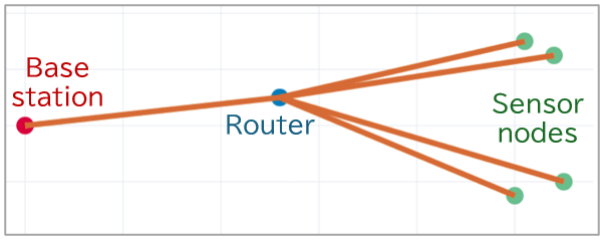
1) Adding an Additional Router Node

The first scenario involved adding another router to reduce the amount of data transmitted by each router node. Figure 18 shows the node placement and simulation results obtained in this case. This approach extended the battery life and increased the battery replacement cycle to 74 days.

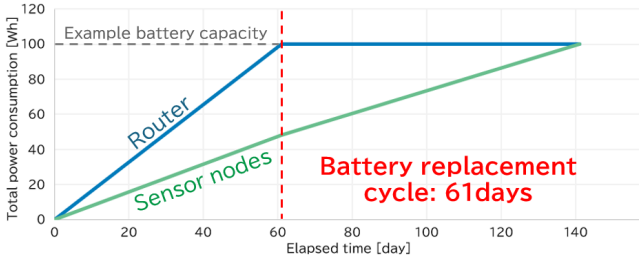
2) Changing the Communication Method

The second scenario involved switching from a short-range protocol to a long-range one (e.g., from Wi-SUN to LoRa), which enabled the sensor nodes to bypass the intermediate relays, presenting simpler topologies and reduced energy consumption. Figure 19 shows the node placement and simulation results. This configuration extended the battery replacement cycle to approximately 140 days.

Based on these illustrative scenarios, users can evaluate multiple network design options using the proposed simulator prior to the actual WSN deployment in railway environments. As observed in these examples, the proposed web-based simulator enabled non-specialist users of railway operators to effectively analyze various network design strategies. By

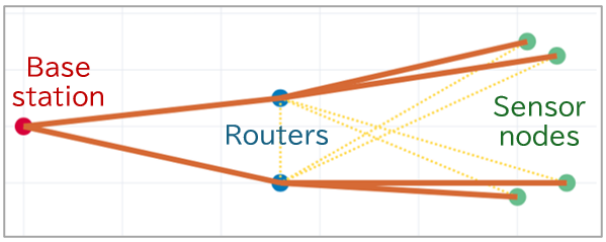


(a) Node placement

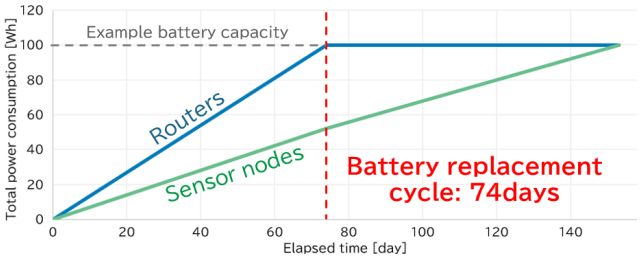


(b) Power consumption simulation results

Figure 17. Example of initial configuration.

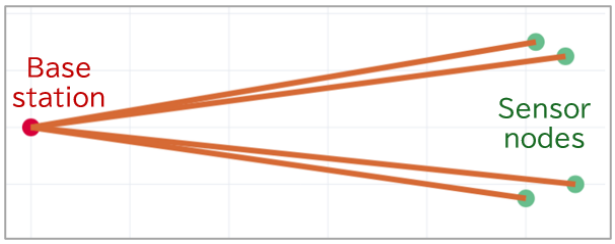


(a) Node placement

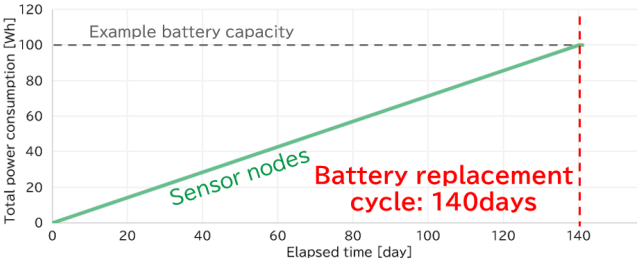


(b) Power consumption simulation results

Figure 18. Adding an additional router node.



(a) Node placement



(b) Power consumption simulation results

Figure 19. Changing the communication method.

configuring and comparing different scenarios, such as those presented here, users with minimal technical expertise can make informed decisions regarding hardware selection, node placement, and maintenance plans. The ability to simulate both the communication quality and power consumption significantly reduced the uncertainty and costs associated with WSN deployment in railway environments. Furthermore, the proposed simulator framework enables easy adaptation to other wireless standards and application domains beyond railways.

VIII. Conclusion

In this study, we proposed a simulation-based approach to support railway operators in designing WSNs. We aimed to provide a practical tool to facilitate network planning without requiring advanced technical expertise. Communication experiments conducted with Wi-SUN and LoRa were used to characterize key parameters—data arrival rate and power consumption—which formed the basis for the modeling. These models were integrated into a user-friendly, web-based simulator capable of estimating the communication availability and power consumption under various deployment conditions. That enabled users to easily compare different node placements and other parameters to satisfy the specific requirements of railway applications.

We conducted validation experiments by constructing a small-scale WSN along the test railway line. The experimental results indicated that the proposed simulator achieves high estimation accuracy for both communication performance and battery usage. The simulator provides a practical and accessible tool for evaluating the WSN configurations in railway environments. It enables non-specialist users to evaluate the network performance under realistic conditions and supports the design of reliable, low-maintenance monitoring systems.

References

[1] Y. Hosokawa, M. Tanaka, and K. Nakamura, “Easy-to-use wireless sensor network simulator for estimating power consumption and communication availability,” The Eighteenth International Conference on Sensor Technologies and Applications (SENSORCOMM 2024), IARIA, Nov. 2024, pp. 4–7.

[2] R. Sharma, V. Vashisht, and U. Singh, “Modelling and simulation frameworks for wireless sensor networks: a comparative study,” IET Wireless Sensor Systems, vol. 10, no. 5, pp. 181–197, Oct. 2020.

[3] S. Idris, T. Karunathilake, and A. Förster, “Survey and comparative study of LoRa-enabled simulators for internet of things and wireless sensor networks,” Sensors, vol. 22, no. 15, p. 5546, Jul. 2022.

[4] LoRa Alliance: About LoRaWAN. [Online]. Available from: https://lora-alliance.org/about-lorawan/ 2025.12.01

[5] ns-2: The Network Simulator—ns-2. [Online]. Available from: https://www.isi.edu/websites/nsnam/ns/ 2025.12.01

[6] ns-3: A discrete-event network simulator for internet systems. [Online]. Available from: https://www.nsnam.org/ 2025.12.01

[7] A. Varga and R. Hornig, “An overview of the OMNeT++ simulation environment,” Proceedings of the 1st International Conference on Simulation Tools and Techniques for Communications, Networks and Systems & Workshops (Simutools ’08), ICST, Mar. 2008, pp. 1–10.

[8] H. Wu, S. Nabar, and R. Poovendran, “An energy framework for the network simulator 3 (ns-3),” Proceedings of the 4th International ICST Conference on Simulation Tools and Techniques (Simutools ’11), ICST, Mar. 2011, pp. 222–230.

[9] A. Boulis: Castalia: an OMNeT-based simulator for low-power wireless networks such as wireless sensor networks and body area networks. [Online]. Available from: https://github.com/boulis/Castalia/ 2025.12.01

[10] INET Framework. [Online]. Available from: https://inet.omnetpp.org/ 2025.12.01

[11] Keysight Technologies: QualNet Network Simulator. [Online]. Available from: https://www.keysight.com/us/en/assets/3122-1395/technical-overviews/QualNet-Network-Simulator.pdf 2025.12.01

[12] Keysight Technologies: EXata Network Modeling. [Online]. Available from: https://www.keysight.com/us/en/assets/3122-1406/technical-overviews/EXata-Network-Modeling.pdf 2025.12.01

[13] M. C. Bor, U. Roedig, T. Voigt, and J. M. Alonso, "Do LoRa low-power wide-area networks scale?," Proceedings of the 19th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM '16), ACM, Nov. 2016, pp. 59–67.

[14] R. Marini, K. Mikhaylov, G. Pasolini, and C. Buratti, "LoRaWANSim: a flexible simulator for LoRaWAN networks," Sensors, vol. 21, no. 3, p. 695, Jan. 2021.

[15] G. Oikonomou, S. Duquennoy, A. Elsts, J. Eriksson, Y. Tanaka, and N. Tsiftes, "The Contiki-NG open source operating system for next generation IoT devices," SoftwareX, vol. 18, p. 101089, Jun. 2022.

[16] F. Osterlind, A. Dunkels, J. Eriksson, N. Finne, and T. Voigt, "Cross-level sensor network simulation with cooja," Proceedings. 2006 31st IEEE conference on local computer networks (LCN), IEEE, Nov. 2006, pp. 641–648.

[17] R. Hirakawa, K. Mizutani, and H. Harada, "Specification and performance analysis of Wi-SUN FAN," IEEE Open Journal of Vehicular Technology, vol. 4, pp. 849–866, 2023.

[18] W. C. Jakes, Ed., Microwave mobile communications. New York: John Wiley & Sons, Inc., pp. 80–83, 1974.

[19] AVHzY: CT-3 USB 3.1 power meter tester digital multimeter current tester voltage detector lua interpreter integrated DC 26V 6A. [Online]. Available from: https://www.avhzy.com/html/product-detail/ct3/ 2025.12.01

[20] Snowflake: Streamlit, a faster way to build and share data apps. [Online]. Available from: https://streamlit.io/ 2025.12.01

[21] Geospatial Information Authority of Japan (GSI): Map and aerial photo browsing service. [Online]. Available from: https://service.gsi.go.jp/map-photos/app/map?search=photo/ 2025.12.01

Beyond Best Effort: Routing with Partially Ordered Requirements

Bradley R. Smith
Department of Computer Science & Engineering
University of California Santa Cruz
Santa Cruz, California, USA
e-mail: brad@soe.ucsc.edu

Paul S. Tatarsky
Tatarsky.com
Washington, D.C., USA
e-mail: paul@tatarsky.com

***Abstract*—Originally designed for the exchange of best effort traffic (email, web, etc.), the Internet had the modest requirements of best-effort service and global reachability. The resulting architecture provides robust and scalable networking, however it is insecure, does not support the performance and policy requirements of modern applications, and makes inefficient use of network resources. While mechanisms have been developed that attempt to address these limitations (firewalls, Policy-Routing, Traffic Engineering with Multi-Protocol Label Switching, Segment Routing, etc.), they are expensive (requiring additional devices and expensive expertise), complicated to configure, and fragile in the context of a changing network. We have developed a new routing architecture based on partially ordered routing metrics. The existing architecture assumes totally ordered metrics where, for any pair of metrics either one is better than the other or they are equal, resulting in a single "*best*" metric for any source/destination pair. Partial orders introduce the possibility of two metrics being *incomparable*. With partial orders, a *best set of (incomparable) paths* is computed between a source and destination that supports the full range of performance and policy available in the network. Traffic is forwarded over the path in this set that meets the flow's requirements and minimizes congestion. This forwarding model is compatible with the Internet's infrastructure. We have developed a prototype and submitted it to an independent testing lab that has verified the functionality and quantified the increase in performance in their testbed network.**

Keywords-Network Routing; Partial Orders; Routing Requirements; Quality-of-Service; Traffic Engineering.

I. Introduction

In this paper we enhance [1] by, in addition to presenting a new routing architecture and results from an independent lab evaluation of a prototype, also exploring *how* this new architecture is implemented using partial orders. The Internet is based on a best-effort communication model where "the network makes no specific commitments about transfer characteristics, such as speed, delays, jitter, or loss. It is assumed that end-system software, both transport layer protocols and applications, would (and must) take this unpredictability into account" [2]. Combined with reliable delivery provided by the Transmission Control Protocol (TCP) transport protocol, best-effort services provide flow-rate fairness, which is defined by the goal of equal flow rates for different flows over the same path. Flow rate fairness is an appropriate goal for best effort traffic (file transfer, email, web, etc.) [3]. As a result, the best-effort service model was a good match for the traffic the Internet was originally designed to carry. "The best-effort paradigm was very powerful - it meant that a wide range of communication technologies could be incorporated into the Internet, technologies with a wide range of basic characteristics. One factor that made the Internet protocols a success was that they could work over 'anything'" [2].

In addition, the Internet adopted a model of universal connectivity. "The original design of the Internet has been described as *transparent*: what goes in comes out. The net does not observe, filter, or transform the data it carries; it is oblivious to the content of packets. This transparency may have been the single most important factor in the success of the Internet, because transparency makes it possible to deploy a new application without having to change the core of the network. On the other hand, transparency also facilitates the delivery of security attacks, viruses, and other unwelcome data." [2]. For the original environment where the network was small and there was a high degree of trust and shared context among the users, the power of universal connectivity outweighed its risk.

In the Internet, packet forwarding is implemented on a hop-by-hop basis where forwarding tables are computed independently at each router, and the forwarding decision is done on a per-packet basis. Paraphrasing [4], packets in a flow traverse a set of interconnected networks (an internet) by, at each hop, forwarding the packet to the next hop router on the path to the packet's destination, where the next hop router *is derived from the packet's destination*. This derivation of the next hop router was initially based on the single best path in terms of a distance metric, and Internet forwarding state was composed of a *single entry for each destination* in the internet giving the next-hop router on the best path to the destination. As a result, only one path is supported to any given destination, and that path is computed to optimize a single metric.

The use of single-path routing significantly compromises the ability of a network to meet the *ordered* Quality-of-Service (QoS) and *categorical* Traffic Engineering (TE) requirements of diverse applications. Single-path routing has a similarly detrimental effect on the utilization of network resources. As the load in a network increases, sending all traffic between a given source and destination over a single path tends to result in links on that path becoming congested.

The hop-by-hop style of packet forwarding used in the Internet exacerbates this problem. With destination-based forwarding each router forwards packets by matching each packet's destination address with a single entry in the router's forwarding table. This leads to the constraint that all traffic

forwarded through an intermediate router to a destination must follow the same path used by traffic sent from that router to the destination. This aggressive tendency to concentrate traffic on a subset of a network's topology causes traffic to experience congestion while usable network resources are left idle, resulting in poor utilization of network resources.

This shortest-path model has been expanded to support Equal-Cost Multi-Path (ECMP) forwarding state composed of the *set of paths* with the same (shortest) distance metric. However, ECMP is not widely utilized, and the result is still limited to the single best path *cost* to a destination. ECMP does not address the QoS or TE requirements of a flow, and only partially addresses the poor utilization of network resources.

However, as the Internet has transitioned to the role of global communication infrastructure, with paying users of more diverse and demanding applications managing increasingly sensitive information, there is a growing need to provide QoS, trust and TE control of network resources as a basic part of the architecture.

These new requirements come from the growth of two new traffic classes called real time and policy-constrained (our term). *Real time* traffic has ordered, time-based constraints for its delivery (delay, jitter, etc.). Examples include voice, video, telemetry (e.g., computer gaming) and real time trading. *Policy-constrained* traffic has categorical constraints for its delivery. Examples include disclosure requirements (sensitive traffic must be carried on eavesdrop-resistant network infrastructure [5]), jurisdictional constraints (restrict genomics data to networks operated in a specific jurisdiction), multitenant networks (a network environment shared by multiple customers), and zero-trust environments where the network is considered untrustworthy, traffic is encrypted and strict access control enforced on what traffic can be shared between which endpoints.

In the early 2000s, the Defense Advanced Research Projects Agency (DARPA) funded the "Future Generation Internet Architecture" project (aka NewArch) to answer the question "if we could now design the Internet from scratch, knowing what we know today, how would we make the basic design decisions?" [2]. The project addressed many issues with the Internet architecture and made many intriguing recommendations. Of particular interest to this paper were two recommendations; one to transition from the Internet's traditional best-effort delivery model to a model they called *trust-modulated transparency*, and another to adopt a generalized version of the routing concept of *regions* as a first-class object in the architecture.

Trust-modulated transparency generalizes the best-effort concept to empower the network to "offer a range of behavior when two (or more) nodes communicate, based on the declared wishes of those nodes. If all the endpoints request, the flow of data among them should be as transparent and unconstrained as the Internet of today. But either end should be able to require that the packets being received be checked, filtered, or constrained in ways that limit the risk of damage and limit the range of unexpected behavior."

This paper presents our solution to this single-path routing model by combining trust-modulated transparency and regions into a unified mechanism for resource allocation in the form of a routing architecture based on computing paths subject to requirements defined by users, applications, and network administrators. This architecture makes it possible to address the problems of scaling and heterogeneity in a wide range of domains including trust, and the articulation, administration, and enforcement of resource allocation policies involving QoS and other policy-constraints. The ultimate goal being to tame the challenges of scale and heterogeneity to maximize trust, user empowerment, and the effective use of network resources.

The spirit of this trust-enhanced region abstraction is not to replace the best-effort model, but to augment it. The resulting Internet will still "work over anything," however it will also allow applications to exploit special functionality when it is available on some paths, thereby ensuring the best experience, in terms of trust, QoS, and policy compliance that is possible in a network.

The existing Internet routing architecture assumes totally ordered metrics where, for any pair of metrics, either one is better than the other or they are equal, resulting in a single "*best*" metric for any source/destination pair. Partial orders introduce the possibility of two metrics being *incomparable*. With partial orders, a *best set of (incomparable) paths* is computed between a source and destination that supports the full range of performance and policy available in the network. Traffic is forwarded over the path in this set that meets the flow's requirements and minimizes congestion.

Recent works [6], [7] have explored the related issue of routing protocols that work with partial ordered metrics. Both explore distributed routing protocols for what we have called here routing over the *best set of paths*. These two works have focused on implementing this approach in Bellman-Ford routing protocols (where paths are computed from destination back to source; see solution to "Problem B" in [8]).

This paper is organized as follows. Section II reviews current solutions that have been developed and deployed in an attempt to address the problems discussed above. Section III reviews the algebraic abstraction of partial orders and describes how we apply this to network routing using both *ordered* and *categorical* (unordered) metrics. Section IV presents the challenges presented by forwarding with partial orders, and shows how these challenges can be met by various mechanisms available in existing Internet infrastructure services. Section V provides a concise overview of our requirements-based routing approach and presents a series of scenarios that illustrate the approach. Scenarios encompass Quality of Service (QoS) management, traffic engineering for multitenant networks, zero-trust networking, the utilization of Boolean variables to reflect network state evaluated at runtime, and finally, the programmatic control of Boolean variables by external systems. Section VI outlines the challenges and opportunities we have identified for this architecture. Section VII presents the outcomes obtained from an independent testing laboratory evaluation of a prototype of this model that

we implemented. Section VIII concludes by summarizing the results and drawing conclusions.

II. Current Solutions

As described in the Introduction, the Internet's best effort communications model is limited in its ability to satisfy the QoS and TE requirements of modern network applications. A number of solutions have been developed to address these limitations under the rubric of *Traffic Engineering*.

Fundamentally, TE is the ability to route traffic over paths that differ from the lowest cost paths used by best-effort routing [5]. TE mechanisms were originally developed primarily to manage network bandwidth with the goal of minimizing congestion [9]. Since their introduction, these mechanisms have been generalized to address a broader set of requirements, such as meeting QoS requirements (specifically bandwidth and delay), restricting specific classes of traffic to topological regions of a network (i.e., multitenant capabilities), enforcing flow priorities (in the sense of preemption), and meeting administrative goals (e.g., restricting sensitive traffic to paths composed of eavesdrop-resistant media such as fiber).

This section reviews the two generations of TE technology developed to date: MPLS-TE and Segment Routing.

A. MPLS TE

The first comprehensive solution for these issues was called *MPLS TE* (Traffic Engineering with Multi-Protocol Label Switching). On its own, MPLS provides the capability to forward traffic over multiple paths, including paths that are different from the lowest cost paths used by the default best-effort routing, as required for traffic engineering. Using MPLS TE, real-time and policy-constrained traffic can be forwarded over paths that better meet their requirements and, sometimes as a specific goal and sometimes as a side-effect, distribute traffic more broadly over a network, resulting in a reduction in congestion and more efficient use of network resource.

MPLS TE accomplishes this by including additional link attributes in the routing computation, using an enhanced routing algorithm called *Constrained Shortest Path First* (CSPF) [10], and using MPLS forwarding state to forward traffic over diverse paths. In addition to the cost used in best-effort routing, MPLS TE includes additional link information such as a TE metric (distinct from the standard link cost), bandwidth, and administrative "*color*" attributes [5], [11].

For QoS requirements, CSPF computes a single path that minimizes a specified, additive metric (the traditional cost metric and an additional *TE metric*, which enables "engineering" the routing computation). For policy requirements, CSPF assigns "colors" to links and interfaces in the network. The set of colors is represented by a 32 bit color bitmap. Each color represents some attribute of a link; e.g., encryption, jurisdiction, maintenance status, link media (optical, copper, wireless), service-level agreement (Gold, Silver, Bronze), etc. Given a set of constraints (expressed in terms of link colors to be included and excluded), a traditional SPF routing algorithm is run on the subset of the topology that satisfies the constraints using the specified QoS metric.

CSPF is limited in a number of ways. Limiting QoS support to one *least cost* path is painfully restrictive. For example, the requirements for video streaming (high bandwidth and delay tolerant; watching a movie online, a few seconds to startup is tolerated while pauses during the movie are not) and network-based telephony (low bandwidth and low delay) are almost in conflict (a high bandwidth, low delay path would satisfy both, but at a premium price when their individual needs are not that demanding).

Similarly, the color-based abstraction for TE requirements of a network flow is limiting. The number of attributes used for defining a policy is limited to the 32 bits in the color bitmap. The attributes available for defining policies are all related to properties of links and interfaces on a path. Policies are statically defined as a part of the network configuration.

MPLS-TE implements point-to-point (P2P) forwarding state specific to each flow, resulting in very poor utilization of label-swap resources and poor scalability. Lastly, MPLS-TE implements on-demand route computation and path signaling, adding significant overhead to the forwarding process.

These limitations led to the development of the improved Segment Routing architecture.

B. Segment Routing

A more recent solution for the original Internet architecture's limitations involves a combination of network technologies based on Segment Routing (SR) [12]. SR computes and builds paths similar to MPLS-TE that better meet the QoS and TE needs of network applications. When TE is not required, SR is able to implement ECMP paths.

SR improves on MPLS-TE in a number of ways. SR integrates the label distribution, TE path signaling, and routing functions that are implemented separately in MPLS-TE into a single protocol. SR builds any-to-one, "multi-point to point" (MP2P) label-swap forwarding state. SR implements a forwarding model that still includes an on-demand routing computation, but makes use of pre-computed forwarding state. The resulting solution is dramatically simpler to configure and operate than MPLS-TE, much more efficient in its use of label-swap resources, and improves on the MPLS-TE forwarding process.

While SR improves on MPLS-TE in the ways listed above, it inherits some of MPLS-TE's limitations including only supporting least-cost paths, its use of the limited abstraction of colors for TE requirements, and it still requires a routing computation for each new flow.

III. Partial Orders

With its use of single-path routing the Internet adopted a routing architecture based on a total ordering of paths available in the network between a given source and destination. Based on this ordering, a single path is used to send traffic to the destination. As discussed in the Introduction, for a small network where there was a high degree of trust, a set of

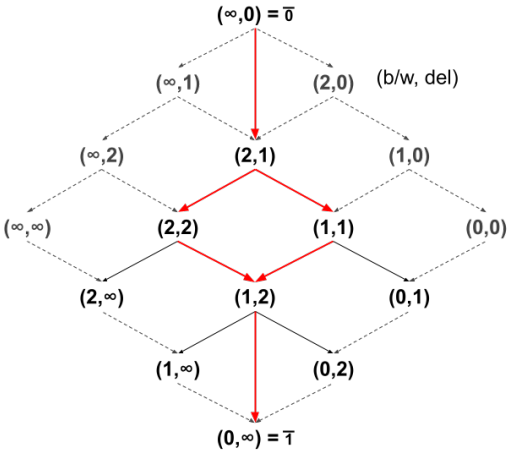


Figure 1. Hasse Diagram for Shortest-and-Widest.

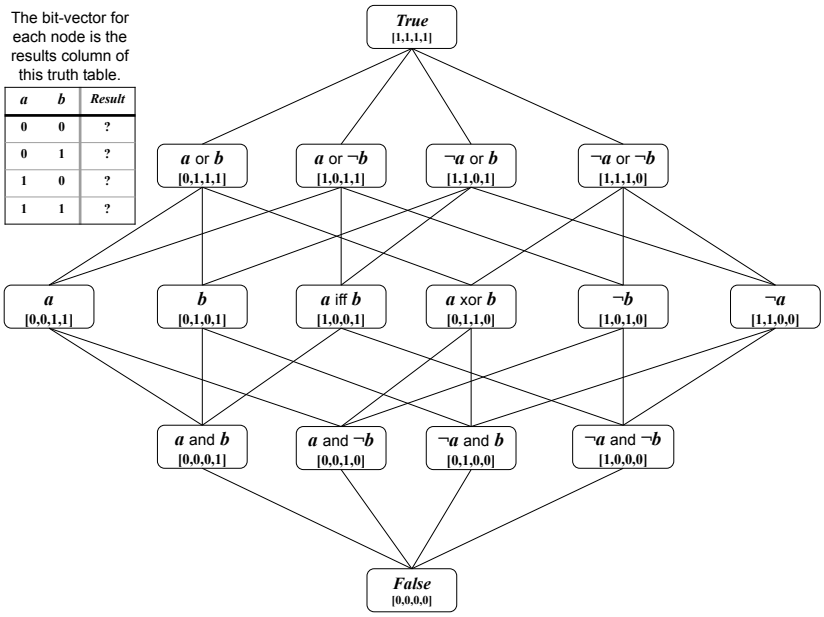


Figure 2. Hasse Diagram of Boolean Constraints on 2 Variables

applications with similar requirements of the network, and minimal requirements for efficient use of network resources, this single-path model was an appropriate choice. However, for the current and future Internet, with the highly diverse requirements and high priority for efficient use of network resources that come with being the converged network infrastructure for the twenty-first century, the single-path limitation is no longer acceptable. Our solution generalizes this architecture to implement a model based on *partial orders* that natively supports the use of diverse paths to a destination that provide the full range of services available in a network and are used based on the requirements of the application generating a given network flow.

A *partial order* $(S, \preceq)$ is a set S with an *order relation* $\preceq$ that is reflexive ($a \preceq a$), transitive ($a \preceq b, b \preceq c \Rightarrow a \preceq c$), and antisymmetric ($a \preceq b, b \preceq a \Rightarrow a = b$). For network routing S is composed of the weights of links and paths in a network, and $\preceq$ defines an order on these paths that is used in the routing computation to identify the set of paths described above.

In general, such an ordering is *partial* in the sense that not all pairs of elements in S are related ($\exists x, y \in S : x \not\preceq y,\ x \not\succeq y$); therefore $(S, \preceq)$ is called a *partially-ordered set* (or *poset*). In *partially-ordered constraint optimization* [13], the relation $x \preceq y$ among constraints is also called $dominates$, and the $dominating$ subset of S (i.e., the set of elements that are not dominated by any other elements in S) is called the *Pareto frontier*. The special case of an ordered set where $\forall x, y \in S$ either $x \preceq y$ or $x \succeq y$ is called a *total order*.

Network routing uses a special kind of partial order called a *lattice*. A lattice is a partial order where every pair of elements x, y in S have both a shared ancestor ($z \in S : z \preceq x, z \preceq y$) and descendant ($w \in S : x \preceq w, y \preceq w$) [14]. These properties are required for network routing as every pair of path weights are guaranteed to have this property. Specifically, all path weights have a guaranteed shared ancestor with the weight of an ideal path (typically used to denote the path from a node to itself), shown as $\overline{0}$ in Figure 1, and shared descendant with the weight of a non-existent path (typically used for the weight between unconnected nodes), shown as $\overline{1}$.

NOTE: order relations are normally defined with the $\succeq$ relation, where $a \succeq b$ is interpreted as a is *better than or equal to* b; however, in network routing, *smaller* (i.e., *shorter*) is considered better, so in this paper we use the network routing convention of $a \preceq b$, and label the top and bottom nodes $\overline{0}$ and $\overline{1}$, respectively (vs the standard notation of $\overline{1}$ and $\overline{0}$).

To help visualize lattices we use *Hasse diagrams*. The rules for drawing the Hasse diagram of a lattice are if $x \preceq y$ is in the poset then y appears below x in the diagram (so a path with weight x is *better* than, or *dominates*, a path with weight y), and a line is drawn between y and x if there is no intermediate value z such that $x \preceq z \preceq y$.

Figure 1 illustrates the use of lattices and Hasse diagrams for routing with the example of a partially ordered version of the Shortest-Widest path algebra, which we'll call *Shortest-and-Widest* to distinguish it from the totally ordered version (discussed in [15]). Weights in Shortest-and-Widest are of the form (bandwidth, delay), and $(b_1, d_1) \preceq (b_2, d_2)$ is defined as $(b_1 \geq b_2)\ and\ (d_1 \leq d_2)$, $(b_1, d_1) \oplus (b_2, d_2)$ ($\oplus$ is used here to distinguish it from the $+$ used in the following) is defined as $(Min(b_1, b_2), d_1 + d_2)$, $\overline{0}$ (i.e., self, or perfect) connectivity by $(\infty, 0)$, and $\overline{1}$ (i.e., no connectivity) is denoted by $(0, \infty)$.

Figure 2 illustrates the partial ordered nature of Boolean constraints with a Hasse diagram for expressions of two

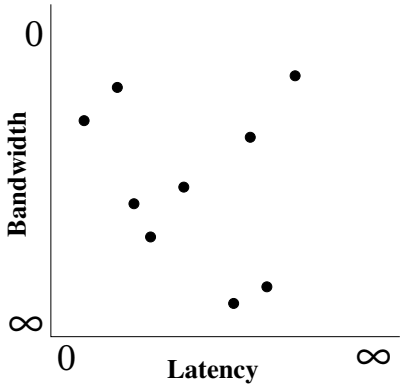


Figure 3. Path Weights

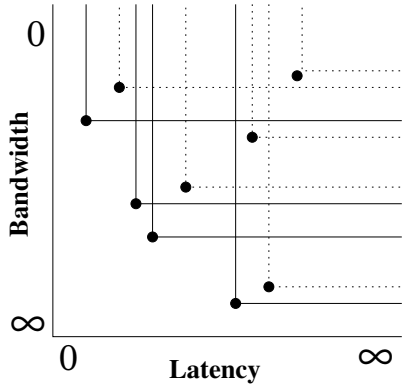


Figure 4. QoS Regions

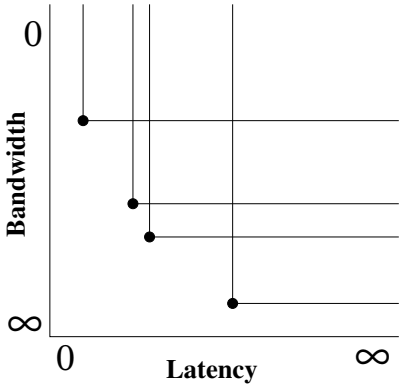


Figure 5. Best Set

Boolean variables. Boolean constraints provide a powerful abstraction for expressing *categorical* (i.e., unordered) resource utilization policies. Path comparison ($\preceq$) is defined as the partial order *dominates* relation where $\varepsilon_1 \preceq \varepsilon_2$ (i.e., ε_1 *dominates* ε_2) when ε_1 is $True$ for every truth assignment where ε_2 is $True$. Formally, this is true when $\varepsilon_2 \Rightarrow \varepsilon_1$ is a tautology (i.e., is always $True$). Path construction ($\oplus$) is defined as the conjunction (and'ing) of two expressions, resulting in values that are generally lower in the diagram (e.g., $(a \circ\mathtt{r}\ b) \oplus (a \circ\mathtt{r}\ \neg b) = a$).

Figures 3 through 5 give an intuitive sense of the partial-ordered nature of Shortest-and-Widest. Figure 3 plots the weights of 9 paths between a specific source and destination in an example network where the metrics composing the weights are bottleneck bandwidth and latency. "Better" values of these metrics are towards the origin of the graph (i.e., a perfect path would have infinite bandwidth and 0 latency).

These points can be interpreted as representing a region, up and to the right (away from the origin) of QoS values that each weight *satisfies* in the sense that the path represented by the weight would satisfy any QoS requirement in that region of the graph. Figure 4 depicts the regions satisfied by each path. Note that regions satisfied by some of the paths are fully contained in the regions of other paths. In the figure these *dominated* regions are represented with dashed lines.

A *best set* of paths to the destination is defined as the dominating set of the weights of paths to the destination. This set of paths is *best* in the sense that any bandwidth and delay requirements that are satisfiable by an existing path between the source and destination, are satisfiable by a path in this set. Figure 5 shows the best set of routes for the example network.

The red-connected subgraph in Figure 1 is the Hasse diagram for Shortest-and-Widest where the bandwidth and delay values range over the set of values $\{0, 1, 2, \infty\}$. Note that the partial ordered nature of this definition of Shortest-and-Widest is evident here in that $(1, 1)$ and $(2, 2)$ are not comparable ($(1, 1)$ has better delay, but $(2, 2)$ has better bandwidth).

Lastly, the *best* metrics (from a routing sense) are towards the top of the Hasse diagram (e.g., $(\infty, 0)$, ∞ bandwidth and 0 delay, is the value for a self-loop in Shortest-and-Widest) and the *worse* metrics are towards the bottom (e.g., $(0, \infty)$, 0 bandwidth and ∞ delay, is the value for unreachable in Shortest-and-Widest), reflecting the partial ordering semantics for the $x \preceq y$ relation being that x is better than y.

Figure 1 also illustrates an implementation issue with partial orders. Specifically, the black or gray-connected subset of the lattice is composed of weights that are false values in the sense that they have (exactly) one component with value 0 or ∞. These values do not reflect real world paths (any value with delay of ∞ or bandwidth of 0 is a synonym for $(0, \infty)$, and any value with the reverse, bandwidth of ∞ or delay of 0, is a variant of $(\infty, 0)$, but is not really a valid weight). However, given the constraint that real link/path weights should have neither component with a 0 or ∞ value, only some of these "false" weights can occur in the normal course of a routing algorithm. Specifically, in this Shortest-and-Widest example, the false values connected to the graph by solid arrows can occur in a real computation. E.g $(2, 2) \oplus (2, 1) = (Min(2, 2), 2 + 1) = (2, \infty)$. The solution for these false values is to identify the ones that can result from the summing of valid weights, and have the path algebra implementation translate those values to $\overline{0}$ or $\overline{1}$, whichever is appropriate. In the figure this means translating $(2, \infty)$, $(1, \infty)$, $(0, 1)$, and $(0, 2)$ to $(0, \infty)$ (i.e., $\overline{1}$).

IV. Forwarding with Partial Orders

In a hop-by-hop routing environment we make a distinction between a path and a *forwarding path*. In hop-by-hop routing, the shortest paths to each destination are computed for each node, and a forwarding table composed of the next hops on these paths is installed at the node. As a packet travels through the network the node at each hop independently selects the next hop based on the forwarding table computed at that node. Forwarding paths are an emergent property of the collective routing tables of all nodes, are not known by any single node, and may be different from paths computed at any given node.

All previous work on path algebras has focused on identifying the properties needed to ensure forwarding paths resulting from use of a path algebra are loop-free and best paths. These properties were appropriate for the implicit focus on totally

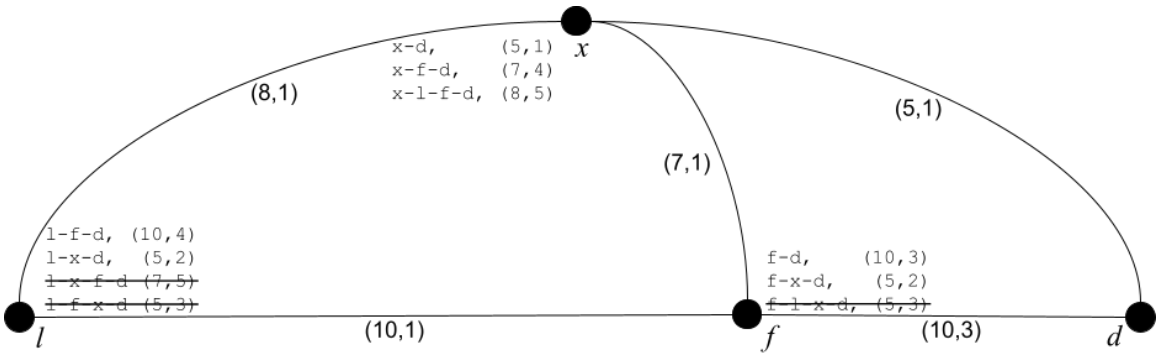


Figure 6. Shortest-and-Widest Scenario.

ordered path algebras where there was guaranteed to be a *single best* path weight (of possibly multiple equal cost paths) for each destination.

However, with the focus here on partially ordered path algebras, this guarantee is no longer valid. In general there will be a set of paths with different weights to each destination, and the concept of best will not apply. The generalized definition we use for *well-behaved* routing and forwarding with partially ordered path algebras is that *forwarding paths are loop free and have the same weight as the weight of the path selected at the source*. Note, this definition works equally well for the totally ordered case, which is appropriate.

Similarly, previous work on totally ordered path algebras has assumed a simplified forwarding model where, at each hop, traffic is forwarded to the next hop on the best path computed by the routing function. Again, in a multi-path environment, this no longer makes sense. With multiple paths to each destination having different path weights, each hop will have to make a choice of the available paths to use in forwarding traffic. Therefore we need to define a generalized forwarding model to be used with the best set of routes computed using a partially ordered path algebra. First we present a simple generalization of the single-path forwarding model and show that it does not work for a path algebra (Shortest-and-Widest) that, intuitively, seems like it should be well-behaved.

With traditional routing, decisions are made at each hop to forward traffic along the best path to the destination from the current node. Given the multiple incomparable paths available with partial orders, this use of a *best* path no longer makes sense and a process along the lines of the following must be used:

> Given a destination and a performance constraint (PC), at each hop choose a path from the set of best paths computed for the destination where the weight of the forwarding path from source to destination satisfies PC. Specifically, given the weight of the path taken to reach this hop (call this the preceding path weight, or PPW), choose a path from the current node's dominant route set where the path's weight (succeeding path weight, or SPW) is such that $PPW \oplus SPW \succeq PC$.

This process is most like traditional forwarding in that decisions are made at each hop on a packet-by-packet basis. The problem with this decision process is it can loop in the context of the Shortest-and-Widest algebra, as shown in Figure 6. In the figure the dominant route set for destination d is shown at each router, and includes the path and path weight (given as *(bandwidth, delay)*). The figure shows the full list of loop-free paths available at each router, with those routes not in the dominating set ~~crossed out~~ to show how the dominant set is selected. Specifically, at node l, $l-f-d \prec l-x-f-d$ (i.e., $(10,4) \prec (7,5)$) and $l-x-d \prec l-f-x-d$ (i.e., $(5,2) \prec (5,3)$.

Using this topology with traditional forwarding there are many opportunities for loops. For example, consider node l forwarding a flow with a PC of $(4,8)$; l could forward it to either node x ($(5,2) \preceq (4,8)$) or node f ($(10,4 \preceq (4,8)$). Both could forward to the other, resulting in a possible loop; e.g., x could forward to f ($(8,1) \oplus (7,4) \preceq (4,8)$), and f could forward back to x ($(8,1) \oplus (7,1) \oplus (5,2) \preceq (4,8)$) until eventually the delay of the looping path gets too great and one of x or f is forced to either forward to d directly or drop the flow (specifically, after 4 trips over the $x-f$ link node x would see that none of its routes satisfy the constraint described above (e.g., $(8,1) \oplus (7,1) \oplus (7,1) \oplus (7,1) \oplus (7,1) \oplus (7,4) = (7,9) \not\preceq (4,8)$), and it would be forced to forward directly ($(8,1) \oplus (7,1) \oplus (7,1) \oplus (7,1) \oplus (7,1) \oplus (5,1) = (5,6) \preceq (4,8)$).

The result is traffic being forwarded over a path that wastes network resource by traversing segment $x-f$ four times and a path that is worse than the options available at node l. This problem comes from the extra degree of freedom offered by the *best set* of routes available at each node. There needs to be some mechanism for communicating the decision made at the ingress router to subsequent routers to ensure these loops are avoided along the lines of the following:

> Given a destination and the full path weight (FPW) selected at the source (where FPW satisfies the flow's performance constraint), at each hop choose a path from the set of paths computed for the destination such that the weight of the forwarding path from

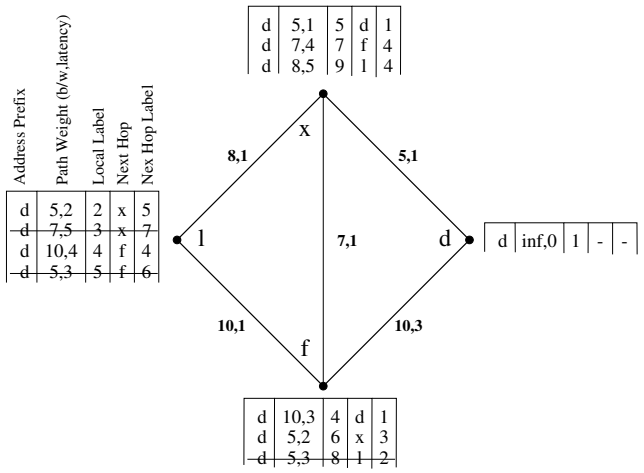


Figure 7. Shortest-and-Widest with Labels for Destination d.

source to destination equals FPW. Specifically, given PPW and SPW described above, choose a path from the current node's dominant route set such that $PPW \otimes SPW = FPW$.

This eliminates the problems experienced by the process described above. Specifically, referring to Figure 6 again, if l forwards the traffic to x, x is forced to forward the traffic directly to d because only route $x - d$ at x satisfies the constraint that $(8,1) \otimes (5,1) = (5,2)$.

While this may initially appear to be an expensive forwarding model (due to the need to communicate FPW and PPW to each hop), it can actually be efficiently implemented in current networks using existing technology that causes the ingress router's forwarding decision to fix the forwarding decisions of each hop along the path. The conceptually simplest solution, in the context of a full-topology routing environment (e.g., link-state routing where each router has a full picture of the network topology) would be the use of source routing (see [16] for a description of source routing in the context of specific requirements).

More recently, segment routing [17] offers a similar solution. Both of these solutions work with the full path to the destination, and therefore the existence of a forwarding path is guaranteed (barring topology changes, which would trigger a re-computation of paths and an update of source/segment routes).

MPLS [18] provides a more general solution, that would support routing environments involving less topology information including, in the extreme, distance-vector routing (where only next hops are known by each router). The MPLS solution works by neighboring routers exchanging label information following route re-computations. Given this mechanism, MPLS follows a path equivalent to the path selected by the ingress router (where there are multiple equal-cost paths, the path selected by the ingress router and the forwarding path may be different paths, but will both be members of this set of equal cost paths).

With label-swap forwarding only the first router that handles a packet has to classify the traffic and select a path for the flow before forwarding it; all subsequent routers perform a simple label-swap lookup. Forwarding state is enhanced to include local and next hop label information, and path weight information to be used in selecting a path.

Figure 7 gives an example network configuration that reflects Figure 6. Of specific interest are the routes between nodes l and d, which include two routes for each path with weights $(5,2)$ (going through x) and $(10,4)$ (going through f) where the path through x has better latency but worse bandwidth than the path through f, illustrating the ability of MPLS to support multiple incomparable paths between two nodes. All of these solutions, which provide control of the path to be used for forwarded traffic to the source router, support the efficient use of multiple paths between a given source/destination pair.

V. Beyond Best Effort

As described in the Introduction, our requirements-based routing architecture implements the *trust-modulated transparency* and routing *region* capabilities identified by the DARPA NewArch project as needed to address the requirements of modern network applications. Specifically, routing based on partially-ordered requirements computes and forwards traffic over paths that satisfy requirements articulated by users, applications and network administrators for each flow carried in a network. As a result traffic carried in a given routing domain ("region") complies with the QoS and TE requirements defined for that domain. The result is an augmented best-effort architecture where the Internet protocols are still able to work over "anything," but now are able to exploit special functionality in the network when it is available, ensuring the best experience in terms of trust, QoS, and policy-compliance that is possible in a given region.

The rest of this section illustrates the mechanics and power of this approach with a number of scenarios. Each scenario is defined by a set of requirements for how traffic in a given class of flows is to be handled. As described in the Introduction, there are two types of requirements: QoS and TE.

QoS requirements of a network application address the *ordered*, performance requirements needed for an application to perform well, typically expressed in terms of bandwidth, latency, jitter (variation in latency), reliability, etc. *TE* requirements specify the *categorical*, non-performance related characteristics of network links such as security (e.g., encryption), jurisdictional issues (for example, restricting private health information to networks within the jurisdiction of a given country), network maintenance status, etc.

A. Quality of Service

As an example, consider a network being used by both an interactive voice application implementing an Internet-based telephony service (commonly called Voice over IP, or VoIP), and a video streaming service such as Netflix.

TABLE I. VOICE/VIDEO QOS REQUIREMENTS

Flow Type	Perf Rqmts	
VoIP	$\leq$ 40ms	$\geq$ 100Kbps
Video Streaming	$\leq$ 10sec	$\geq$ 3Mbps

TABLE II. MULTI TENANT TE REQUIREMENTS

Flow Type	Boolean Variable	Path Expressions
Tenant A	TA	TA
Tenant B	TB	TB
		(TA **or** TB)
		(TA **and** TB)
		True
		False

Interactive voice communication has relatively modest bandwidth requirements (100Kbps provides a high quality voice encoding) but fairly stringent delay requirements (interactive communications is awkward with delays much above 150ms[19]). So, VoIP service requires low delay and can can live with relatively low bandwidth. In contrast, video streaming has very modest delay requirements, but relatively high bandwidth requirements (i.e., even many seconds delay in starting a video is tolerable as long as once it starts there is adequate bandwidth for it to smoothly run to completion). So, a video streaming service requires high bandwidth and can live with high delay. Table I shows these requirements.

Given these performance requirements defined in terms of delay and bandwidth, the routing computation collects topology information that includes QoS metrics for each link. It then runs a modified shortest-path first routing algorithm that computes the set of paths in the network that are not comparable to each other, and forwards traffic over one of these paths that satisfies the flow's performance requirements; in the event there are more than one it uses the least congested.

Using the video and voice example from above, the low and high bandwidth and delay paths can be see as *incomparable*. Specifically, low delay is better than high delay however high bandwidth is better than low. This incomparability can be restated as *it depends on the needs of the flow*, resulting in the opportunity to compute a *best set of routes* as those paths in the network where some application might prefer one path over the others. Further, with potentially a choice of satisfying paths, it is possible to distribute traffic more widely over a network, thereby reducing congestion and increasing utilization.

B. Multitenant

Multi-tenancy is when several network customers are sharing a set of network resources, such as when several different small business are using the same network resources to communicate within their offices in a building and to reach the Internet. Despite the fact that they share resources, these network customers are not aware of each other, and their data is typically kept separate.

TABLE III. ZERO TRUST

Flow Types	Zones	End-to-End Requirements
$\texttt{WEB}_F$	$\texttt{USER}_Z$	($\texttt{WEB}_F$ **and** $\texttt{USER}_Z$ **and** $\texttt{WEB}_Z$)
$\texttt{APP}_F$	$\texttt{WEB}_Z$	($\texttt{APP}_F$ **and** $\texttt{WEB}_Z$ **and** $\texttt{APP}_Z$)
$\texttt{DB}_F$	$\texttt{APP}_Z$	($\texttt{DB}_F$ **and** $\texttt{APP}_Z$ **and** $\texttt{DB}_Z$)
	$\texttt{DB}_Z$	

To implement such a set of requirements we define a set of Boolean variables that reflect policy-relevant attributes of network traffic, the network itself, or of the network's environment. TE requirements are articulated as Boolean expressions composed of these variables, and are used in the routing computation to compute policy-compliant paths for the flow to use.

A subset of these expressions can be used to label links in the network to express the TE constraints each link imposes on traffic that traverses the link. Path expressions are constructed as a part of the routing computation (by *and*'ing together the link expressions), to express the constraints imposed on traffic that traverses the path. Expressions that are not assigned to links define what we will call *end-to-end* requirements that are used to define requirements of traffic in terms of its content, source, and destination. We will see examples of all of these in the following.

Table II shows the Boolean variables that could be defined to support two tenants, and some likely path expressions that would be used to control traffic on a multitenant network. The Boolean expressions extracted from a flow are used to determine if a flow can use a path by determining if the conjunction (and'ing) of the flow expression with the path's expression is *satisfiable* (meaning there is a truth assignment to the variables that results in a $True$ value for the combined expression).

The True and False path expressions indicate any or no flows may use a link, respectively (these expressions can be used for any path expression and are not included in the remaining scenarios). TA or TB represent traffic sent or received by tenant A or B (perhaps set based on a flow's source or destination address). (TA **or** TB) allows tenants A and B to share a link, and (TA **and** TB) indicates a link only for use for flows between tenant A and B.

C. Zero Trust

This scenario illustrates support for Zero Trust security applied to the traditional three layer web application architecture using TE requirements. The general Zero Trust architecture, based on the assumption that networks cannot be trusted, adopts a least privilege strategy by encrypting all traffic and strictly enforcing access control expressed as an access matrix specifying what combination of users, applications, and security zones can access other security zones. Security zones are logical containers for physical interfaces, VLANs, and IP address ranges (i.e., a region of the network) [20].

In the three layer web application architecture, applications are organized into three logical tiers: web, application, and

TABLE IV. CONTROL BACKUPS OVER CORE

Flow Types	Time Periods	Path Requirements
BKP	NT	(**not** BKP **or** (NT **and** BKP))

data. The web (or presentation) tier is the user interface to the application, responsible for collecting data from the user and displaying data from the application to the user. The application (or logic) tier is where data collected from the user is processed, sometimes using information from the data tier, and results are presented to the user or saved in the data tier. The database tier is where information produced by the application is stored and managed. The benefits of this architecture include faster development, and improved scalability, reliability, and security. For security purposes, firewalls are commonly deployed between tiers.

Table III illustrates a three tier architecture implemented on a single subnet using TE requirements. Boolean variables are defined for flow types ($\texttt{WEB}_F$, $\texttt{APP}_F$, $\texttt{DB}_F$) and network zones ($\texttt{USER}_Z$, $\texttt{WEB}_Z$, and $\texttt{DB}_Z$). The zone variables could be set based on the IP prefix of servers in each zone, and TCP ports or application detection technology could be used for setting the flow variables. In this scenario the links have no TE requirements, but end-to-end TE requirements limit traffic between zones to the appropriate classes of flows (e.g., $\texttt{WEB}_F$ traffic is only allowed between the $\texttt{USER}_Z$ and $\texttt{WEB}_Z$ zones, etc.). Note that, with this solution, the integrity of the three tier architecture does not depend on the location of servers. Servers from different tiers could be connected to the same layer 2 switch and the integrity of the tiers would still be maintained.

The two previous scenarios represent static TE requirements in the sense that how a Boolean variables is set is specified as part of configuring TE requirements for the network. So zones in the Zero Trust scenario could be defined by an IP prefix, etc. The remaining two scenarios illustrate an important capability of Boolean expression-based configurations to dynamically define the value of variables based on attributes of the network's state or environment.

D. Dynamic Variables

Table IV illustrates a simple scenario where backup traffic is only allowed to flow over a core portion of the network at night. The idea being that during the day the core portions of an organization's network are reserved for operational data and backups are only allowed to traverse peripheral networks, or be delayed to run at night.

Two Boolean variables are defined including BKP, which is set to true for flows that carry backup traffic, and NT, which is set to true when it is currently nighttime. The link expression (**not** BKP **or** (NT **and** BKP)) is defined for all core network links specifying that BKP traffic can only traverse core links at night.

The Boolean variable NT is a *dynamic* variable whose value is determined by the network at the time the flow is processed.

TABLE V. BOOLEAN SATISFIABILITY AND ONEHOT()

DY	NT	BKP	Path Req	OH(DY,NT)	Result
$False$	$False$	$False$	$True$	$False$	$False$
$False$	$False$	$True$	$False$	$False$	$False$
$False$	$True$	$False$	$True$	$True$	$True$
$False$	$True$	$True$	$True$	$True$	$True$
$True$	$False$	$False$	$True$	$True$	$True$
$True$	$False$	$True$	$False$	$True$	$False$
$True$	$True$	$False$	$True$	$False$	$False$
$True$	$True$	$True$	$True$	$False$	$False$

While the time period to define as night would be configured statically as part of the network configuration, the value of the variable is determined dynamically. This capability introduces a bit of autonomic control into the network configuration, and leads to the more general solution presented next. The primary limitation to the dynamic nature of Boolean variables like NT is they only support state directly available to the network device implementing the routing function (a router, switch, or controller).

There are some subtleties to satisfiability that need explanation. We illustrate this by adding a variable DY that is $True$ for a flow occurring during they day (added for illustration since DY can be expressed as (**not** NT)). The first four columns of Table V show the truth table for the path expression (**not** BKP **or** (NT **and** BKP)) given these three variables. This shows that a flow sent during the day, with DY set to True and NT not set (i.e., in a "don't care" state), would be allowed because the path requirements would be satisfied in the last two rows, which is a mistake. This mistake comes from the fact that we have not expressed the requirement that a flow can only occur either *during the day or night but not both*, which is why the last two rows of the fourth column (where both DY and NT are $True$) show as $True$. To fix this we need a Boolean expression of the DY and NT variables that is $True$ only for truth assignments where only one variable is $True$. We represent such a function as $OH(variables...)$ (short for $OneHot(...)$) in the fifth column of Table V, and use it to complete the satisfiability test.

Applying this to our problem, the "Result" column shows the conjunction of the path requirements and OneHot(DY,NT) columns, where only rows three through six are valid, and show the desired truth table (the only blocked flows are backup flows not sent at night). So whenever we have a set of variables where only one can be $True$ for a given flow, we must include the $OneHot(...)$ function of those variables in the combined flow and path expression to avoid false positives. This is assumed in the examples in the paper. Note, a similar set of constraints is needed for the Zero Trust scenario.

E. Programmatically-Controlled Variables

The final example illustrated in Table VI implements functionality that can demonstrate a fully dynamic Boolean variable. This scenario has two components, DEFCON threat levels and MultiLevel Security (MLS). MLS provides support

TABLE VI. DEFCON WITH MULTILEVEL SECURITY

Flow Types	Threat Levels	Path Requirements
TS_f	D_1	((D_1 **and** (U_f **or** S_f **or** TS_f)) **or** (D_3 **and** (S_f **or** TS_f)))
S_f	D_3	((D_1 **and** (U_f **or** S_f)) **or** (D_3 **and** S_f))
U_f		(U_f)

for multitenant use of networks in the form of the traditional, military-style multilevel security using TE requirements. Traffic is classified at unsecured, secret, or top secret security levels and is routed over infrastructure certified at the traffic's level or above. The Boolean variables U_f, S_f, TS_f are defined for a flow's security level. An unspecified mechanism determines the security level for a new flow, and the flow is assigned to the least congested path that satisfies the MLS routing requirement (e.g., unclassified traffic can be forwarded over paths of any security level, but top secret traffic can only traverse strongly secured paths) as specified by the TE Boolean expressions assigned to each link.

DEFCON builds on MLS by adding Boolean variables (D_1 and D_3) reflecting the military *defense readiness condition* (DEFCON) levels used to characterize the current threat level. Higher threat levels are indicated by lower DEFCON numbers (DEFCON1 being the highest threat level). In this scenario link expressions include MLS and DEFCON variables. In these expressions, D_3 enables traffic handling equivalent to standard MLS policies described above (traffic is routed over infrastructure certified at the traffic's level or above), while D_1 enables policies that drop unclassified (U_f) traffic from links rated at S_f and TS_f levels. The logic being that, in a time of heightened threat, secured network resources should be reserved for important traffic.

The dynamic nature of this scenario comes from the ability to implement programmatic control of the DEFCON variables. In our prototype, implemented as a Software-Defined Network (SDN) controller with a web user interface, we implemented programmatic control as a Representational State Transfer (REST) service for setting the values of Boolean variables, which support the remote invocation of functions on the Web server using HTTPS messages. Using such programmatic control mechanisms, Boolean variables can be defined to reflect any state in the network or its environment that has policy significance for the network's configuration. With such variables, the policy enforced in a network can be changed immediately, without the need for reconfiguration of network devices or reprogramming of SDN-based systems.

This capability has profound implications for network management. Imagine a scenario where Boolean variables are defined to reflect workstation configuration acquired using network access control technology (e.g., operating system version and patch levels) combined with variables defined to represent information from threat feeds reflecting the severity of vulnerabilities discovered in operating system versions and patch levels. TE requirements could be defined that only allowed systems to access sensitive parts of a network if they are at patch levels with no known vulnerabilities and traffic from vulnerable systems can be routed to sites that facilitate upgrades of vulnerable systems), with new vulnerabilities being integrated into network behavior as soon as they are discovered.

VI. CHALLENGES AND OPPORTUNITIES

A fundamental challenge of requirements-based routing is the need to determine the *satisfiability* of Boolean expressions used to express categorical requirements [21]. Satisfiability, which is the test of whether there is a truth assignment of the variables in a Boolean expression that cause the expression to evaluate to *True*, is the prototypical NP-Complete problem [22]. The essential meaning of this is there is no known way to determine satisfiability "efficiently".

One possible approach to containing the cost of the satisfiability test is to restrict the syntax of these expressions to forms with efficient algorithms for satisfiability. Significant work has been done along this line, culminating in Schaefer's Dichotomy theorem [23]. Schaefer's theorem comprehensively defines the boundary between expressions for which satisfiability can be determined efficiently and those for which no efficient solutions are known. The theorem shows that efficient solutions exist for six classes of expressions, and any expressions not in these classes are NP-complete.

Unfortunately for the work here, Schaefer also showed that none of these classes support negation, which is required for routing with requirements. However, fortunately, driven by the needs of integrated circuit design testing, there has been dramatic progress in the optimization of satisfiability algorithms such that, in spite of the inherent challenges of the general problem (e.g., current algorithms can determine satisfiability of expressions with millions of variables and clauses in minutes [24]).

These results, and the likely size and characteristics of requirements-based routing problems, give hope that the cost of satisfiability will not be a problem. Experience with our (un-tuned and research-grade) prototype, where path selection based on Boolean requirements are made once per flow, is that the time required for these decisions is consistent with normal switching speeds (single-digit milliseconds). Additionally, we have not implemented the use of "assumptions" [25], which should significantly speed up determining satisfiability in the path selection process.

At a more engineering-level, there are a number of other challenges/opportunities that need to be addressed. Architectures for forwarding traffic over multiple paths to the same destination (currently include OpenFlow [26], P4 [27], and MPLS [5])) are in constant flux. Assessing the scalability and performance of solutions requires attention, and possibly impacts the architecture for a comprehensive solution.

Regarding opportunities, developing and assessing distributed implementations of this technology, along the lines of traditional routing protocols, needs to be evaluated as an approach to addressing scalability and performance issues.

TABLE VII. TCP PERFORMANCE RESULTS

TCP TEST RESULTS			RSTP IAT (sec) 3	
RPO IAT (sec)	RPO Gbps	RSTP Gbps	Throughput Gain	Load Factor
0.25	3.25	1.7	-4.4%	12.0
0.5	3.78	2.32	11.2%	6.0
1	3.87	3.09	13.8%	3.0
1.25	3.76	3.15	10.6%	2.4
1.5	3.79	3.29	11.5%	2.0
2.5	3.79	3.39	11.5%	1.2
3	3.77	3.4	10.9%	1.0

TABLE VIII. UDP PERFORMANCE RESULTS

UDP TEST RESULTS			RSTP IAT (sec) 1.5	
RPO IAT (sec)	RPO loss rate	RSTP loss rate	Relative Loss	Load Factor
0.25	24.9%	42.3%	1.03	6.0
0.5	15.0%	39.1%	0.62	3.0
1	9.3%	31.8%	0.39	1.5
1.25	8.7%	27.9%	0.36	1.2
1.5	9.1%	24.1%	0.38	1.0
2.5	2.7%	14.8%		
3	1.9%	11.2%		
	RPO Gput (Gbps)	RSTP Gput (Gbps)	Goodput Gain	Load Factor
0.25	2.40	1.84	-0.8%	6.0
0.5	2.71	1.95	12.0%	3.0
1	2.87	2.18	18.6%	1.5
1.25	2.91	2.30	20.2%	1.2
1.5	2.90	2.42	19.8%	1.0
2.5	3.10	2.71		
3	3.13	2.83		

As mentioned earlier, recent work along these lines [6], [7] has explored related approaches to routing using distributed Bellman-Ford routing protocols.

VII. PROTOTYPE

To validate this architecture, we developed a prototype that implements policy-based (Layer 2) switching in a software-defined networking (SDN) environment using the OpenFlow protocol, the Ryu open-source controller, and Linux-based Open vSwitch [28] software switches. The prototype includes a web interface that allows users to define the supported traffic classes for a network and the TE and QoS requirements for these classes.

Implementation in Layer 2 was done for both convenience and functionality. A centralized, controller-based implementation made configuration significantly easier by centralizing the definition and implementation of policy in one place. Additionally, implementation of the requirements-based routing model at Layer 2 provides fine-grained control of network traffic down to the switch port level, enabling the full power of this architecture to be displayed. However, with some loss of granularity (working at the subnet vs switching level), this architecture can support a Layer 3 implementation equally well.

We engaged an independent third-party test lab to evaluate the prototype in terms of functionality and performance. Functionality testing involved evaluation of three scenarios: QoS with VoIP and video traffic (see Section V-A), a Zero Trust network environment (Section V-C), and a network segmentation environment (similar to the scenario covered in Section V-B) emulated in a small enterprise network environment.

The findings of these tests verified the expected results for the three scenarios. In the QoS configuration, routing with requirements effectively selects an appropriate path for different types of network traffic that have distinct bandwidth and delay requirements. In the Zero-Trust scenario, routing with requirements implements access control for users, applications, and network zones. And for network segmentation, routing with requirements enforces access control for different zones in the network. Furthermore, across all three scenarios, if there are multiple paths that satisfy the flow's requirements, routing with requirements selects the path with the least amount of traffic (thereby minimizing congestion).

For performance testing, they measured the performance of TCP and UDP traffic among hosts attached to switches connected in a 4x4 torus topology. In the tests, they compared the performance of routing with partial orders (RPO) with the standard Rapid Spanning Tree Protocol (RSTP) in identical topologies using the same Linux-based software switches. For TCP tests they compared throughput bandwidth, and for UDP tests they compared loss rates.

Focusing on performance evaluation, they deployed the system as a 4x4 torus, with two hosts per switch, in a VMware-based virtual environment. Each test involved 10 traffic flows for each host between random nodes in the graph with restrictions on the distribution of hops traversed (2 flows traversed 1 hop, 3 flows 2 hops, 4 flows 3 hops, and 1 flow 4 hops). Tests were run for a range of flow Inter-Arrival Times (IATs) between hosts (0.25, 0.5, 1, 1.25, 1.5, 2.5, and 3 seconds). TCP performance was characterized by the cumulative throughput of all 320 flows and UDP by the average loss rate and cumulative good-put of the flows.

The relevant results are presented in Tables VII and VIII. For TCP, RPO at 0.5sec IAT provides 11.2% better throughput $((3.78Gbps - 3.4Gbps)/3.4Gbps)$ at six times the load $(3sec/0.5sec)$ of RSTP at 3sec IAT. For UDP, RPO at 0.25sec IAT provides roughly the same loss rate and good-put at six times the load of RSTP at 1.5sec IAT.

VIII. CONCLUSION

We have given an overview of routing with partially ordered requirements-based and presented a number of scenarios that demonstrate the power of this paradigm. Specifically, partially ordered QoS and TE requirements enhance network routing to compute a *best set of routes* that satisfy the full range of QoS and TE requirements supported by a given network environment.

Articulating and enforcing the QoS and TE requirements enhances the Internet's original default-allow security model to default-deny, where only requirement-compliant flows are allowed. Security is further enhanced by a dramatic reduction in the network's attack surface as it is limited to network devices whose access is typically tightly controlled (compared to the attack surface of all connected devices).

The use of partially-ordered requirements optimizes the user's experience, ensuring that traffic is forwarded over paths customized to the application's QoS and TE requirements and is compliant with network administration's policies. By working with a *set* of candidate paths, traffic can be forwarded over the least congested requirement-compliant path, dramatically improving network utilization. Simulations predicted a ten-fold increase with a somewhat "meshy" (average node degree of four) network topology [29]; these results have been verified by an independent testing lab using an untuned *prototype* implementation.

Network services can be safely reconfigured with programmatic control of TE Boolean variables as they do not require reconfiguration of network equipment or re-programming of software-defined networking functions. Many functions currently implemented by expensive devices external to the core network, such as firewalls, load balancers and zero-trust network equipment, can be replaced by a software upgrade. Furthermore, implementing these functions using requirements-based routing results in significantly more robust services as they are implemented in the network layer where they have knowledge of the network's topology as it evolves.

Most importantly, for many environments, requirement-based routing provides a more intuitive, high-level network configuration paradigm based on specifying *what* the requirements of the network are, allowing the network to solve the problem of *how* to enforce the requirements rather than depending on highly trained network engineers. This enables the support of significantly more sophisticated network services by available engineers.

References

[1] B. R. Smith and P. S. Tatarsky, "Beyond Best Effort: Routing with Requirements," in *The Seventeenth International Conference on Evolving Internet INTERNET 2025*, Mar. 2025.

[2] D. Clark *et al.*, *New Arch: Future Generation Internet Architecture*, Dec. 2003.

[3] S. Floyd and M. Allman, *RFC 5290: Comments on the usefulness of simple best-effort traffic*, Request For Comments, 2008.

[4] V. Cerf and R. E. Kahn, "A Protocol for Packet Network Intercommunication," *Communications, IEEE Transactions on*, vol. 22, no. 5, pp. 637–648, Jan. 1974.

[5] A. Sanchez-Monge and K. G. Szarkowicz, *MPLS in the SDN Era*. Sebastopol, CA: O'Reilly Media, Dec. 2015.

[6] J. J. Garcia-Luna-Aceves, B. R. Smith, and J. T. Samson, "QoS routing using dominant-distance vectors," in *Proceeding IEEE/ACM International Symposium on Quality of Service (IWQoS 2022)*, Jun. 2022.

[7] J. L. Sobrinho and M. A. Ferreira, "From non-optimal routing protocols to routing on multiple optimality criteria," *IEEE/ACM Trans. Netw.*, vol. 31, no. 1, pp. 294–307, Feb. 2023.

[8] L. R. Ford, "Network flow theory," RAND, Tech. Rep. P-923, Aug. 1956.

[9] D. O. Awduche, J. Malcolm, J. Agogbua, M. O'Dell, and J. McManus, *RFC 2702: Requirements for traffic engineering over mpls*, Request For Comments, Sep. 2008.

[10] I. Minei and J. Lucek, *MPLS-Enables Applications: Emerging Developments and New Technologies*. Wiley, Jun. 2010.

[11] D. Katz, D. M. Yeung, and K. Kompella, *Traffic Engineering (TE) Extensions to OSPF Version 2*, Request For Comments, Sep. 2003.

[12] S. F. Hassan, A. Orel, and K. Islam, *A Network Architect's Guide to 5G*, M. Taub, N. Davis, S. Schroeder, S. Schroeder, and B. Reed, Eds. Addison-Wesley Professional, Jun. 2022.

[13] M. Gavanelli, "Partially ordered constraint optimization problems," *Principles and Practice of Constraint Programming — CP Lecture Notes in Computer Science*, vol. 2239, T. Walsh, Ed., pp. 763–768, Nov. 2001.

[14] C. Jongsma, *Introduction to Discrete Mathematics via Logic and Proof*, 1st. Springer, Nov. 2019.

[15] B. R. Smith and J. T. Samson, "Herding Packets: Properties Needed of Metrics for Loop-Free & Best Forwarding Paths," in *Proceedings International Conference on Computing, Networking and Communications (ICNC)*, Jan. 2017.

[16] S. Previdi, C. Filsfils, B. Decraene, M. Horneffer, and R. Shakir, *Source Packet Routing in Networking (SPRING) Problem Statement and Requirements*, IETF Request for Comments, May 2016.

[17] C. Filsfils *et al.*, *Segment Routing Architecture*, IETF Request for Comments, Jul. 2018.

[18] B. Davie and Y. Rekhter, *MPLS: Technology and Applications*. Morgan Kaufmann, 2000.

[19] *Itu g.114 : One-way transmission time*, ITU-T Rcommendation G.114, 2003.

[20] J. Kindervag, *No more chewy centers: The zero trust model of information security*, Forrester Research Technical Report, Mar. 2016.

[21] B. R. Smith, "Efficient Policy-Based routing in the internet," Ph.D. dissertation, University of California, Santa Cruz, Aug. 2003.

[22] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W.H. Freeman & Co., 1979.

[23] T. J. Schaefer, "The Complexity of Satisfiability Problems," in *10th ACM Symposium on the Theory of Computing*, 1978, pp. 216–226.

[24] S. Garfinkel, J. M. Abowd, and C. Martindale, "Understanding database reconstruction attacks on public data," en, *Commun. ACM*, vol. 62, no. 3, pp. 46–53, Feb. 2019.

[25] A. Nadel and V. Ryvchin, "Efficient SAT solving under assumptions," in *Theory and Applications of Satisfiability Testing – SAT 2012*, ser. Lecture notes in computer science, Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 242–255.

[26] N. McKeown *et al.*, "OpenFlow: Enabling innovation in campus networks," *SIGCOMM Computer Communication Review*, vol. 38, no. 2, Mar. 2008.

[27] P. Bosshart *et al.*, "P4: Programming protocol-independent packet processors," en, *Comput. Commun. Rev.*, vol. 44, no. 3, pp. 87–95, Jul. 2014.

[28] B. Pfaff *et al.*, "The design and implementation of open vSwitch," in *12th USENIX Symposium on Networked Systems Design and Implementation (NSDI 15)*, Oakland, CA: USENIX Association, May 2015, pp. 117–130, ISBN: 978-1-931971-218.

[29] B. R. Smith and L. Thurlow, "Practical multipath load balancing with QoS," in *Proceedings International Conference on Computing, Networking and Communications*, Dec. 2013, pp. 937–943.

A Proposal of a Sensor Exclusion and Dynamic Cluster Head Selection Algorithm to Improve Energy Efficiency in Cognitive Radio Networks

Alcides Mário Joaquim Tomás
National Institute of Telecommunications - Inatel
Minas Gerais, Brasil
e-mail: alcides.tomas@mtel.inatel.br

José Marcos Camara Brito
National Institute of Telecommunications - Inatel
Minas Gerais, Brasil
e-mail: brito@inatel.br

***Abstract*—Cognitive Radio (CR) emerges as a solution for efficient spectrum utilization in response to the growing demand for connected devices driven by 5G and beyond. In this context, numerous devices share network resources, leading to high energy consumption. It is therefore essential to develop strategies that reduce this consumption and extend the operational lifetime of Cognitive Radio Networks (CRNs). This article proposes an algorithm that combines sensor exclusion with the dynamic selection of Cluster Heads (CHs), aiming to reduce energy consumption while balancing detection capability and network longevity. Simulation results for different numbers of Secondary Users (SUs) show that the proposed algorithm maintains a high detection probability with negligible false-alarm probability, while significantly increasing the network lifetime when compared with classical and cluster-based cooperative sensing schemes. In the evaluated scenarios, the proposed solution increases the CRN lifetime from approximately 3×10^4 to about 6×10^4 sensing cycles, which corresponds to an average network lifetime gain close to 100% relative to the classical method, without compromising sensing performance.**

Keywords-Clustering; K-Means; energy efficiency; cognitive radio network; cooperative spectrum sensing; energy detector; probability of detection.

I. Introduction

In recent years, the rapid growth of wireless communication systems has increased the scarcity of the Radio Frequency (RF) spectrum. This occurs mainly due to the fixed allocation policy, which grants exclusive spectrum usage rights to a network of licensed users, known as Primary Users (PUs). The demand for new telecommunication services drives research and technologies such as the Fifth Generation of mobile communication networks (5G), the Internet of Things (IoT) and, in the future, the Sixth Generation of mobile communication networks (6G). However, to enable most wireless communication services, it is essential to overcome spectrum limitations, since multiple frequency bands are required to support the growing number of transmitters and receivers expected in 5G, 6G and IoT networks. In this context, the concept of spectrum sensing through CR emerges as a promising alternative to provide more efficient spectrum access [1].

From a practical standpoint, CRNs must simultaneously satisfy regulatory constraints on detection probability (P_d) while operating with battery-powered Secondary Users (SUs) [2]. High reporting overhead, unbalanced energy consumption among SUs and heterogeneous propagation conditions (e.g., shadowing and fading) often lead to premature network death and performance instability, especially in dense IoT and 5G/6G scenarios [3] [4]. These aspects make the joint design of Cooperative Spectrum Sensing (CSS) and energy-management mechanisms a difficult task. Therefore, there is a clear need for CSS strategies that extend network lifetime without violating minimum detection requirements, thus improving the reliability and operational cost of spectrum-sharing services [5].

From a scientific and engineering perspective, the main problem addressed in this proposal is how to jointly design sensor-selection and CH management policies in CRNs so as to minimize energy consumption while preserving cooperative detection performance. Prolonging the lifetime of the secondary network is particularly relevant for large-scale applications such as environmental monitoring, smart-grid supervision, and massive IoT connectivity, where battery replacement is costly or even infeasible. In this context, this proposal investigates a cluster-based CSS scheme that introduces temporary and permanent exclusion of low-performance sensors together with dynamic CH selection based on spatial position [6] [7].

The central research questions are:

- (i) To what extent can sensor exclusion and dynamic CH selection increase the lifetime of a secondary network when compared with classical and cluster-based CSS schemes?
- (ii) What is the impact of these mechanisms on the probability of detection and on the false-alarm probability for different numbers of SUs?

Accordingly, the purpose of this article is to quantify the energy–reliability trade-off provided by the proposed algorithm and to compare its performance with established CSS approaches through numerical simulations.

A. Contributions and Structure of the Article

This article proposes a sensor exclusion algorithm combined with dynamic CH selection, aiming to reduce the energy consumption of the secondary network and, thus, increase its lifetime without compromising the system's detection capability. The main contributions of this proposal are as follows:

- Temporary exclusion of low-performance sensors, which can be reintegrated if their performance improves, and permanent exclusion applied after a defined number of consecutive inactivity periods.
- Dynamic CH selection, defined as the sensor closest to the Center of Mass (CM).

The proposed algorithm is evaluated through simulations in *MATLAB*, considering a CRN with a static set of SUs and having the secondary network lifetime as the main focus. An important limitation of this study is that SU mobility was not considered in the analysis and is therefore left as a perspective for future work.

The remainder of this article is organized as follows. Section II presents the related work. Section III describes the system model. Section IV discusses the cluster-based cooperative spectrum sensing scheme. Section V details the signal and channel modeling. Section VII introduces the proposed algorithm. Section VIII presents the analysis and results. Finally, Section IX concludes the paper.

II. Related work

Cluster-based CSS has been widely explored as a strategy to reduce energy consumption in CRNs. Due to the various contributions presented in this proposal, we conducted a literature review on works related to sensor selection techniques, CH election, fusion algorithms, and communication schemes, each presenting specific benefits and limitations. Below is a review highlighting the main contribution of each work. Table I summarizes the analyzed studies.

In [8], a hybrid information fusion scheme is proposed. This model employs the Pietra–Ricci (PRIDe) detector for data fusion at the CH level and decision fusion at the Fusion Center (FC) level. This technique improves the robustness of decision-making and can minimize the impact of CH energy depletion. However, simulations indicate that, in dense networks, this approach may not yield significant energy savings, as the high cost of intra-cluster communication and between CHs and the FC compromises some of the gains achieved.

In [9], a model for sensor alternation between active and idle modes was developed, aiming to distribute energy consumption more evenly. As an alternative to minimize overhead and better balance consumption, the use of pairs of sensors operating synchronously, alternating between active and idle states, is proposed. Although this approach offers this advantage, the constant alternation between states may introduce latency, particularly in scenarios that require a high detection rate and fast response.

In [10], a distributed CH election method based on the residual energy of the nodes is proposed. In this approach, in each round, the node with the highest energy within the cluster is selected as the CH, responsible for transmitting the data to the FC. This method reduces the need for frequent alternation between sensors and improves the energy efficiency of communication. However, the continuous selection of the same CH nodes tends to cause premature depletion of these nodes, compromising the network's stability over time.

In [11], a weighted linear fusion scheme in CRSNs is proposed, assigning weights to the nodes based on the Signal-to-Noise Ratio (SNR) and historical detection accuracy. The technique has proven effective in increasing the P_d and reducing the error rate by better exploiting the differences between the nodes.

In [12], a selection method based on the remaining energy of sensors is presented, prioritizing those with higher energy levels for spectrum sensing. This model allows for more dynamic energy balancing between sensors, preventing the premature depletion of a specific subset. However, this approach tends to cause overload in dense networks, as sensors with more energy may be frequently activated, increasing communication consumption and reducing detection accuracy.

In [13], the use of multiple sequential reporting channels is proposed to reduce delays and improve decision accuracy at the FC. This strategy optimizes the sensing time of nodes and reduces transmission latency, achieving superior performance compared to conventional methods in terms of both accuracy and delay. In [14], a centralized routing protocol with clustering for mobile nodes in Wireless Sensor Networks (WSNs) is proposed, optimizing cluster formation based on energy and mobility.

In [15], the selection of sensors with higher SNR and residual energy is proposed for CSS. This approach significantly reduces energy consumption by enabling only an optimized subset of sensors to participate in sensing, while others remain in energy-saving mode. However, the lack of a sensor rotation mechanism may result in uneven wear of selected nodes, thus reducing the network's lifespan.

In [16], a comprehensive review of energy-efficient CSS techniques is presented, classifying methods according to various criteria and highlighting the potential for integrating dynamic sensor selection, adaptive clustering, and intelligent information fusion. In [17], a dynamic clustering algorithm for large-scale Mobile Sensor Networks (MSNs) is developed, considering residual energy and transmission delays, making the system more adaptable to mobility and uneven energy consumption.

In all the studies analyzed, clustering is used as the initial technique for energy savings. However, in [18], clustering is not applied. Instead, a CSS scheme without clustering is proposed, specifically adapted for CRSNs. The distinguishing feature of the model is the use of real data correlations collected by the sensors (e.g., temperature and humidity) to more efficiently select nodes participating in the detection process. This approach reduces the number of active nodes in sensing, allowing others to remain in idle mode, thus leading to greater energy savings and lower latency. However, the effectiveness of this solution depends on the accuracy of correlation estimation and the SNR scenario, which may require additional adjustments in dense networks.

Thus, the literature encompasses multiple strategies — from node selection and CH rotation to hybrid fusion schemes and mobile protocols — but still presents gaps in integrative approaches that reconcile energy efficiency, detection robustness, and support for dynamic environments. This proposal contributes by suggesting an algorithm that combines temporary and permanent sensor exclusion with dynamic CH selection, thereby extending the network's lifespan without compromising the reliability of CSS.

TABLE I. ANALYSIS OF RELATED WORK

Reference	Year	Energy	Cluster	CSS	Fusion	Main Contribution
[8]	2023	✓	✓	✓	✓	Hybrid fusion scheme using Pietra–Ricci (PRIDe) detector at the CHs and decision at the FC.
[9]	2021	✓	✓	✓		Clustering model with active/idle alternation for paired nodes.
[10]	2021	✓	✓	✓		Distributed clustering algorithm for CH election based on residual energy.
[11]	2021	✓	✓	✓	✓	Weighted linear fusion with weights based on SNR and historical accuracy.
[12]	2020	✓		✓		Sensor selection method based on residual energy.
[13]	2020	✓	✓	✓	✓	Proposal for multiple sequential reporting channels to reduce latency and increase accuracy at the FC.
[14]	2019	✓	✓			Centralized routing protocol with efficient clustering for mobile nodes in WSNs.
[15]	2019	✓	✓	✓		Sensor node selection for CSS based on energy and SNR.
[16]	2016	✓	✓	✓	✓	Review and classification of energy-efficient CSS techniques.
[17]	2013	✓	✓			Dynamic clustering algorithm in large-scale mobile sensor networks based on residual energy and transmission delays.
[18]	2013	✓		✓		Explores the correlation of real data collected by sensors to select participating nodes in the sensing process.

III. SYSTEM MODEL

This section describes the cognitive radio network considered in this work, detailing the spatial distribution of SUs, the formation of clusters, and the role of CHs and the FC. We first present how SUs and the primary transmitter are positioned in the coverage area, including the definition of shadowed regions and their impact on sensing performance. Then, we introduce the clustering model adopted to organize the SUs into groups and to support cooperative spectrum sensing and energy-management mechanisms.

A. Spatial distribution of SUs

The SUs are devices that operate in CRNs, using radio spectrum dynamically and without a license. In other words, SUs take advantage of spectrum gaps that are not used by PUs, without interfering with the operations of these licensed users. The positions of the SUs are determined by two-dimensional coordinates (x, y), randomly distributed within the coverage area of the primary network. This coverage area has a radius r around the FC, located at coordinate $(0, 0)$. The Primary Network Transmitter (PU_{tx}) is positioned at coordinates $(-r, r)$, with the coverage area radius defined as $r = 1000\,\text{m}$. Additionally, shadowed areas are defined to evaluate the impact of the received signal on SUs located in these regions. Due to attenuation caused by obstacles, these sensors are more susceptible to detection failures and are consequently penalized more frequently.

This penalty results in periods of inactivity during which the sensors cease to participate in the sensing process and enter energy-saving mode. However, they can still transmit data if the channel is detected as free and, if their performance improves, they may be reintegrated into the network. On the other hand, sensors with successive cycles of inactivity are permanently excluded to prevent compromising the decisions of the cluster and, consequently, the FC.

Figure 1 illustrates the execution of the k-means clustering process. As shown in Fig. 1, the process was carried out with $m_T = 20$ and $c_{\max} = 3$, generated in MATLAB. The colors of the points distinguish the clusters and their respective SUs, while the shadowed areas are represented by dashed circles. The centroids resulting from the clustering are marked with crosses. The PU_{tx}, FC, and CHs are also highlighted, along with the circular coverage area of the primary network.

The secondary network consists of three main components: the FC, which makes global decisions; the SUs, which perform spectrum sensing; and a subset of SUs that act as CHs, coordinating the cluster decisions.

B. Clustering model

For cluster formation, the K-Means clustering algorithm was used due to its simplicity, ease of implementation, and computational efficiency. This algorithm aims to partition the SUs into k clusters, where each SU corresponds to the cluster whose centroid is the closest [19].

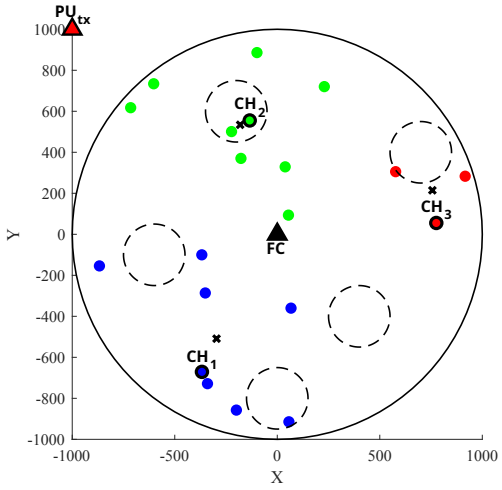


Figure 1. Spatial distribution of SUs, shadowed regions, and cluster formation using the K-Means algorithm.

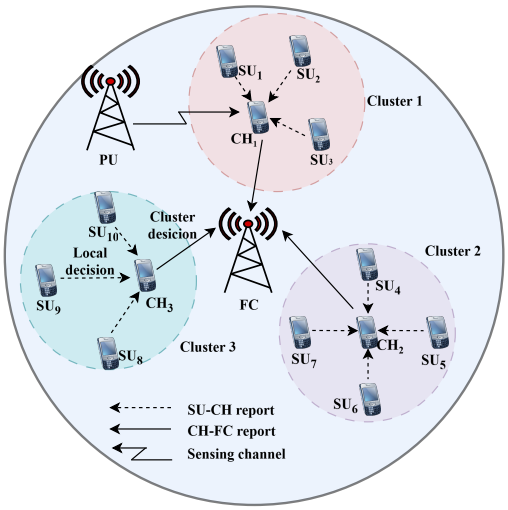


Figure 2. Architecture of a CSS based on cluster.

The process begins with the random selection of k initial centroids, which are iteratively adjusted until the cluster positions stabilize or no longer change significantly. This method allows the partitioning of SUs based on their two-dimensional positions (x, y). The necessary information for the clustering process includes the total number of SUs (m_T) and the maximum number of clusters ($c_{\max}$).

IV. Cluster-Based Cooperative Spectrum Sensing

In cluster-based CSS, all SUs equipped with spectrum sensing capability actively participate in the spectrum detection process by collecting samples of the primary signal within their respective operating regions. Moreover, each SU is capable of processing this information and making a local decision regarding the presence or absence of the PU. Instead of reporting directly to the FC, the SUs forward their local decisions to the CH, which is responsible for aggregating the detection decisions within its cluster.

The CH considers the decisions of the nodes belonging to its cluster and its own decision to determine the final decision of the cluster and transmits this decision to the FC. The FC processes the decisions of multiple CHs and makes the final decision on spectrum occupancy, determining whether the channel is available or if there is an active primary user in the network. All sensors have direct communication capability with the FC and, for this reason, can assume the role of CH when selected.

Figure 2 shows the architecture of a cluster-based CSS, highlighting the different decision reporting channels. The SUs transmit their local decisions to the corresponding CH via reporting channels, represented by dashed lines. Then, the CHs forward the cluster decisions to the FC using channels indicated by solid arrows, ensuring data consolidation for the final decision-making. Additionally, the sensing channel, represented by a zigzag arrow, illustrates the reception of the primary signal by the SUs.

A. Binary hyphotesis test

Spectrum sensing can be mathematically modeled as a binary hypothesis test, where the objective is to decide between two possible conditions of the radio frequency (RF) spectrum:

- H_0 (Null Hypothesis): The spectrum does not contain a primary signal, e.g., the band is unoccupied. In this case, the received signal consists only of noise.
- H_1 (Alternative Hypothesis): The spectrum contains an active primary signal, indicating that the band is occupied. Thus, the received signal is composed of the sum of the PU signal and noise.

This decision is made by comparing a test statistic T with a pre-established decision threshold γ. The criterion for choosing the hypothesis follows the following rule:

$$T > \gamma \Rightarrow H_1, \tag{1}$$

$$T \leq \gamma \Rightarrow H_0, \tag{2}$$

If the statistic T exceeds the threshold γ, it is concluded that a primary signal is present in the sensed band (H_1). Otherwise, it is decided that there is no active transmission in the band (H_0). Mathematically, the hypotheses can be expressed as:

$$y(t) = \begin{cases} n(t), & \text{under } H_0, \\ h(t)x(t) + n(t), & \text{under } H_1, \end{cases} \tag{3}$$

where:

- $y(t)$: Signal received by the SU.
- $n(t)$: Additive white Gaussian thermal noise (AWGN).
- $h(t)$: Gain or attenuation factor of the transmission channel.
- $x(t)$: Signal transmitted by the Primary Transmitter.

The test statistic T is derived from the processing of the signal $y(t)$ received by the SU, and its construction varies according to the chosen sensing technique. Different methods

for creating this statistic give rise to various types of detectors, with Energy Detection (ED) being the most widely used due to its low implementation complexity.

B. Energy Detection

ED differentiates the presence or absence of the primary signal based on the energy of the samples collected during a sensing interval. This technique is widely used due to its simplicity and effectiveness, especially when there is no prior knowledge of the transmission signal characteristics [10].

The test statistic of the ED for the j-th SU is:

$$T_j = \frac{1}{n} \sum_{i=1}^{n} |y_{j,i}|^2 , \tag{4}$$

where n is the number of samples at the j-th SU, and $|y_{j,i}|$ represents the magnitude of the i-th sample at the j-th SU. In CSS with decision fusion, local decisions (at the SUs) are made by comparing T_j with the corresponding decision threshold.

The performance of spectral sensing is commonly measured using two main metrics: the probability of false alarm (P_{fa}) and the P_d, mathematically defined by (5) and (6):

$$P_d = \Pr\left[T > \gamma \mid H_1\right], \tag{5}$$

and P_{fa} is defined as:

$$P_{\text{fa}} = \Pr\left[T > \gamma \mid H_0\right], \tag{6}$$

where H_1 and H_0 represent the hypotheses of the presence and absence of the signal transmitted by PUs, respectively.

A low P_{fa} is desirable, as it maximizes efficient spectrum utilization, allowing the secondary network to exploit communication opportunities when the spectrum is truly unoccupied. Conversely, a high P_d is crucial to ensure the protection of the primary network, minimizing the risk of interference caused by the secondary network.

V. Signal and Channel Modeling

Consider a CSS with m_T SUs, each collecting n complex samples of the signal transmitted by the PU in each sensing interval. The samples collected by the SUs can be organized as in [8].

The Gaussian distribution was adopted for the transmitted signal x, as it simplifies mathematical analysis. Moreover, the choice of the Gaussian distribution facilitates the evaluation of detection performance, proving to be an effective and reasonable approach for signal modeling in various practical communication scenarios.

The elements of h are complex samples of zero-mean Gaussian variables, independent and identically distributed (i.i.d.), modeling a flat Rayleigh fading channel between the PU and each SU. The Rayleigh channel was chosen to represent a scenario without a direct line of sight between the transmitter and receiver, characterizing an environment dominated by multipath propagation and reflections.

The samples in x are complex Gaussian random variables with zero mean and variance $P_{\text{tx,PU}}$, where $P_{\text{tx,PU}}$ represents the transmission power of the PU. Thus, the power of the primary signal received by the j-th SU ($P_{\text{rx,SU}_j}$) is given by the log-normal prediction model, as described by (7):

$$P_{\text{rx,SU}_j}(\text{dBm}) = P_{\text{tx,PU}}(\text{dBm}) - 10\alpha \log_{10}\left(d_{\text{PU},j}\right), \tag{7}$$

where $d_{\text{PU},j}$ is the distance from the PU to the j-th SU, and α is the path loss exponent. Higher values of α indicate greater signal attenuation. The shadowing effect was considered by adjusting the value of α, resulting in greater attenuations in the power received by the SUs in these regions.

VI. Energy Consumption Modeling

The energy of the SUs decreases as they participate in the sensing process. SUs that reach energy levels below the threshold are deactivated to preserve the reliability of the network and maintain the balance of energy load.

Initially, the SUs perform sensing simultaneously during the period τ_s. Subsequently, they transmit their decisions to the corresponding CH within the period $\tau_{r,\text{SU}}$. The CH then makes the cluster decision and forwards it to the FC within the period $\tau_{r,\text{CH}}$.

Energy consumption in the sensing process is directly related to the fusion method adopted. In decision fusion, energy consumption is higher than in data fusion because the processing of the received signal involves additional steps beyond simply collecting samples. However, in the reporting step, the local decision can be represented by just 1 bit, significantly reducing energy consumption compared to transmitting multiple bits per sample in data fusion. Given this advantage in the reporting step, the decision fusion approach was adopted in this work. It is assumed that the reporting channel is error-free, ensuring accurate communication between the SUs and the CH, as well as between the CH and the FC, for sensing decisions.

SUs whose energy falls below a certain threshold are permanently excluded from sensing, being classified as inactive or "dead." Since any SU can assume the role of CH, the energy threshold (λ) was calculated as the minimum energy required to perform sensing and report the decision to the FC. This value considers the distance between the SU and the FC to be equal to the coverage area radius (r), as shown in the equation below:

$$\lambda = \left(P_s + P_{\text{rx,FC}}\, r^{\alpha}\right), \tag{8}$$

where P_s is the power used for spectrum sensing, and $P_{\text{rx,FC}}$ corresponds to the receiver sensitivity of the FC, representing the minimum acceptable received power level.

The residual energy, in joules, of the j-th SU in the secondary network during a sensing cycle can be calculated by:

$$E_r^{(j)} = E^{(j)} - \left(P_s\, \tau_s + P_{\text{tx,SU}}^{(i,j)}\, \tau_{r,\text{SU}} + P_{\text{tx,CH}}^{(j)}\, \tau_{r,\text{CH}}\right), \tag{9}$$

where $E_r^{(j)}$ represents the residual energy of the j-th SU during the sensing process, and $E^{(j)}$ refers to the energy available

in the j-th SU, which is initially uniform across all SUs. P_s is the power for spectral sensing, $P_{\text{tx,SU}}^{(i,j)}$ is the power used by the j-th SU to transmit its decision to the i-th CH, and $P_{\text{tx,CH}}^{(j)}$ is the power of the signal transmitted by the j-th CH to communicate the cluster decision to the FC, calculated in (10) and (11). If the reporting occurs from SU to CH, the term $P_{\text{tx,CH}}^{(j)}\tau_{r,\text{CH}} = 0$. On the other hand, if the reporting is from CH to FC, the term $P_{\text{tx,SU}}^{(i,j)}\tau_{r,\text{SU}} = 0$.

Applying the distance path loss model, the values of $P_{\text{tx,SU}}^{(i,j)}$ and $P_{\text{tx,CH}}^{(j)}$ are calculated as follows:

$$P_{\text{tx,SU}}^{(i,j)} = P_{\text{rx,CH}}\, d_{\text{CH}_{ij}}^{\alpha}, \tag{10}$$

and

$$P_{\text{tx,CH}}^{(j)} = P_{\text{rx,FC}}\, d_{\text{FC}_j}^{\alpha}, \tag{11}$$

where $d_{\text{CH}_{ij}}$ represents the distance from the j-th SU to the i-th CH, and d_{FC_j} is the distance from the j-th CH to the FC. $P_{\text{rx,FC}}$ and $P_{\text{rx,CH}}$ are the sensitivities of the FC's and CHs' receivers (minimum admissible levels of received power), and α is the path loss exponent.

At each sensing cycle, the energy of the active SUs is reduced as described in [12]. Active SUs are those that have not been excluded, either temporarily or permanently, and have sufficient energy to continue participating in the sensing process.

VII. Proposed Algorithm

An algorithm is presented that integrates temporary and permanent sensor exclusion with the dynamic selection of CHs. The components of the algorithm are detailed individually, as described below:

A. SUs Exclusion Algorithm

The exclusion of SUs is determined based on the individual performance of each one, continuously evaluated during the sensing cycles. This performance is monitored by the system's global decision, made by the FC, which penalizes SUs that make detection errors. This approach ensures that only SUs with satisfactory performance remain active in the sensing process.

In the considered model, each SU makes a local decision about the channel status (free or occupied) and sends it to the CH, which in turn aggregates the decisions of its SUs and forwards the cluster's final decision to the FC. The FC then makes the final decision that prevails for the network, referred to here as the global decision. Since the true spectrum status is unknown, the SUs do not have access to their actual sensing performance. However, the FC is able to compare each SU's local decision with its global decision. If the spectrum is considered idle, the SUs compete for spectrum access using appropriate multiple access techniques; otherwise, a new sensing cycle is initiated.

To better understand how decisions are made, the following terms are used:

- **TX (Transmission Medium State)**: Represents the true state of the spectrum. $\text{TX} = 0$ indicates that the medium is free (idle, not occupied by the PU), while $\text{TX} = 1$ indicates that the medium is occupied by the PU.
- **GD (Global Decision)**: Decision made by the FC based on the local decisions reported by the CHs. $\text{GD} = 0$ indicates that the system decided under hypothesis H_0 (spectrum idle), while $\text{GD} = 1$ indicates hypothesis H_1 (spectrum occupied).
- **LD (Local Decision)**: Individual decision of each SU about the spectrum state, based on its local sensing. $\text{LD} = 0$ indicates that the SU judged the spectrum as idle, and $\text{LD} = 1$ indicates that the SU considered the spectrum to be occupied.

An SU's failure in the sensing process is referred to as a local failure, while the FC's failure is referred to as a global failure. Furthermore, failures are classified as either verifiable or non-verifiable. If there is a transmission by the SU, which occurs when $\text{GD} = 0$, the success or failure of the decision can be verified, since it is assumed that an acknowledgment message from the receiver will indicate whether the message was correctly received, characterizing a successful decision, or if it was received incorrectly or not at all, resulting in a negative acknowledgment or absence of acknowledgment, which characterizes a failed decision.

If there is no transmission by the SU, which occurs when $\text{GD} = 1$, both the SU and the FC are unable to determine whether their decisions were actually correct or incorrect. Thus, if $\text{LD} = 1$ and $\text{GD} = 1$, the SU's decision matches the FC's decision, and the outcome of the decision process is undefined. Similarly, if $\text{LD} = 0$ and $\text{GD} = 1$, since the true channel state is unknown, this situation is again considered undefined. If the SU fails in its decision, whether verifiable or non-verifiable, a penalty is assigned, which will be used in the SU exclusion algorithm. Sensors with verifiable failures receive more severe penalties, with a weight of 2, while those with non-verifiable failures are penalized less severely, with a weight of 1. Table II summarizes all possible scenarios.

The penalty weight of the j-th SU, denoted by f_j ($f_j = 1$ or 2), is reset at each cycle and varies according to the type of failure committed in the current cycle, as described above. If $f_j \neq 0$, a backoff time is assigned to the SU, given by (12), which determines the number of cycles during which the sensor will be temporarily excluded from the sensing process.

The backoff is defined to prevent unnecessary energy consumption by SUs with low detection performance. However, these inactive SUs can still participate in the data transmission process if the channel is detected as idle ($\text{GD} = 0$). The type of failure determines the penalty value, which directly affects the backoff time. The backoff time is calculated as follows:

$$\text{backoff}_j = \left(2^{f_j} - 1\right), \tag{12}$$

where backoff_j is the number of sensing cycles during which the sensor will not participate in the sensing process.

The proposed algorithm also considers the possibility of permanently excluding the SU from the decision process. For

TABLE II. POSSIBLE SCENARIOS OF SUCCESS AND FAILURE IN THE SENSING PROCESS

TX	LD	GD	Result	Action taken
0	0	0	Verified Local and Global Success	None
0	0	1	Undefined	None
0	1	0	Verified Local Failure	SU is penalized with weight 2
0	1	1	Undefined	None
1	0	0	Verified Local and Global Failure	SU is penalized with weight 2
1	0	1	Unverified Local Failure	SU is penalized with weight 1
1	1	0	Verified Local Success and Global Failure	None
1	1	1	Undefined	None

this purpose, a consecutive backoff occurrences counter (β) is defined, which is incremented by 1 each time the SU enters a backoff. After the backoff, when the sensor is reintegrated into the sensing process, if the SU successfully detects the signal, β is reset to zero. Otherwise, β is incremented by 1 again. When β reaches a predefined limit, the SU is considered dead or permanently excluded from the spectrum sensing process.

B. Dynamic CH selection algorithm

After grouping the SUs, k clusters are formed, each with a centroid representing the cluster's CM. The CH is selected as the SU closest to the CM, considering that all SUs have the same initial energy. However, after a certain number of cycles, the CM is recalculated, now weighted by the residual energy of the SUs in the cluster. This new centroid reflects the current energy distribution of the SUs, and the SU closest to the centroid, with the highest available residual energy, is selected as the new CH. The CM is calculated as follows:

$$X_{\text{cm}} = \frac{\sum_{j=1}^{m} E_r^{(j)} x_j}{\sum_{j=1}^{m} E_r^{(j)}}, \tag{13}$$

$$Y_{\text{cm}} = \frac{\sum_{j=1}^{m} E_r^{(j)} y_j}{\sum_{j=1}^{m} E_r^{(j)}}, \tag{14}$$

where m is the total number of sensors in the cluster, $E_r^{(j)}$ represents the residual energy of the j-th sensor, and x_j and y_j represent the x and y coordinates of the j-th sensor, respectively.

The energy consumption of SUs directly impacts the position of the CM, as after a certain number of cycles, some sensors may be deactivated (considered dead) when their residual energy falls below the desired threshold. To maintain an efficient energy balance within the cluster, the current CH is replaced by an eligible SU that meets the minimum energy requirements. This process is continuously repeated until no qualified SUs remain to take on the role of CH, ensuring that sensors with available energy continue to actively participate in the system. The following pseudocode outlines the main steps of this dynamic CH selection process.

TABLE III. PSEUDO-CODE 1: DYNAMIC CH SELECTION ALGORITHM

1	**If** $cycle == cycle_{CH}$ **then**
2	**For** the i-th cluster in the set, up to $i = c_{\max}$ **do**
3	Obtain the m SUs of cluster i
4	Identify the live SUs among the m SUs in cluster i
5	Obtain the coordinates of the live SUs
6	Obtain the residual energy of the live SUs
7	Compute the CM weighted, based on 13 and 14 (new centroid)
8	Select the new CH as the SU closest to the new centroid with the highest residual energy available in the cluster
9	**If** the CH has changed **then**
10	Update the CH coordinates and distances (d_{CH}, d_{FC})
11	**End If**
12	**End For**
13	$cycle_{CH} = cycle_{CH} + 1000$
14	**End If**

In this pseudocode, $cycle$ is the sensing cycle number, $cycle_{CH}$ represents the periodicity of the CH change process, which is initially set to 1000 cycles. Live SUs are the sensors with sufficient energy to participate in the sensing process.

VIII. Analysis and Results

The results presented in this section were obtained through computational simulations performed in MATLAB. Pseudo-code 2 outlines the main steps of the code used to generate these results. Different CSS approaches were compared, with the following being evaluated:

- **Classic CSS (without clustering)**: each SU operates independently, can act as a CH, and reports its detection decision directly to the FC [18].

- **Cluster-based CSS**: the SUs are organized into clusters and send their local decisions to the corresponding CHs, which aggregate the cluster's decisions and forward the consolidated decision to the FC [10].
- **CSS with the proposed algorithm**: combines clustering with a strategy for temporary and permanent SU exclusion and dynamic CH selection, aiming to simultaneously improve detection capability and the lifespan of the secondary network.

The analyses consider a primary network composed of a single primary transmitter and a secondary network operating under different values of m_T and $\max_{\text{back}}$ in the CSS. The parameter $\max_{\text{back}}$ was defined as the maximum value that the consecutive backoff occurrence counter β can reach before the sensor is permanently excluded or considered dead in the sensing process.

The analysis of the systems is primarily conducted based on two metrics: system lifetime and P_d. The system lifetime is defined as the elapsed time until all SUs in the network become inoperative. To evaluate it, two complementary metrics were used: the average number of active SUs, representing the mean number of operational SUs (with sufficient energy for the sensing process) throughout the simulation cycles; and the average drop start, which corresponds to the cycle in which the first SU failure occurs during system operation.

The test statistic for the hypotheses H_0 and H_1 of the SUs was derived using the ED, with $P_{\text{fa}} = 0.1$ adopted in all scenarios, and the corresponding P_d determined from the sensing performance simulations. In all scenarios, decision fusion was implemented using the majority voting (MAJ) logic.

In this pseudocode, $E^{(j)}$ represents the energy available at the j-th SU, which is initially equal for all SUs. It is calculated as the minimum power required to perform sensing and report the decision to the FC, multiplied by the number of cycles the secondary network can operate. The shadowed areas were defined as indicated in Table V, considering regions where signal propagation experiences additional attenuation due to physical obstacles.

The energy parameters of the secondary network were adjusted so that, when implementing the classic method, the network lifetime is approximately 30,000 cycles for each SU draw, with this draw being performed 100 times. During the simulation, the metrics were collected, and the average of the results was then plotted. Table VI presents the system parameters used in the simulations.

In the simulation, the primary transmitter activity was modeled using the function `randi([0,1])`, which follows a Bernoulli distribution, alternating between active and inactive states. The signal attenuation was adjusted according to the path-loss exponent α, defined as:

- $\alpha = 2$ in areas without shadowing (lower attenuation);
- $\alpha = 4$ in areas with shadowing (higher attenuation).

Figure 3 shows the P_d performance of the SUs considering different values of m_T. In Fig. 3(a), for $m_T = 50$, it can be observed that the cluster-based systems and the proposed system exhibit a P_d below 0.9, while the classic system maintains a P_d at or above 0.9. This behavior is related to the fusion technique used at the FC: in the classic system, all SU reports are considered, increasing detection accuracy. In contrast, in the cluster-based and proposed systems, the decision is based only on the reports sent by the CHs, which reduces the accuracy of the decision-making process. This drop compromises the system's efficiency in the later stages.

In Fig. 3(b), corresponding to $m_T = 100$, it can be observed that all systems maintain a P_d above 0.9 during the first sensing cycles, indicating high detection accuracy and reflecting the robustness of the network and the reliability of the adopted fusion technique. In the classic system, performance also increases, approaching 1. In Fig. 3(c), with $m_T = 200$, the

TABLE IV. PSEUDO-CODE 2: STEPS ASSOCIATED IN MATLAB

1	Set the values of the system parameters
2	Define the simulation scenario (PU, FC and r)
3	Define the shading areas
4	Calculate the energy threshold λ
5	**For** each draw, do the following:
6	Generate m_T SUs
7	Run the algorithm *K-Means*
8	Find the CH of each cluster
9	Calculate the distances
10	Calculate the reception power
11	**While** $E^{(j)} > \lambda$, run the detection round:
12	Calculate y_j for each SU
13	Calculate the test statistic
14	Find the local decision of the SUs (LD_j)
15	Find the system's global decision (GD)
16	Update the energy consumed
17	Penalize SUs
18	Find new CHs
19	Calculate the distances
20	Disable SUs
21	Define which SUs are still active
22	Calculate the current value of P_d
23	**End** of sensing round
24	Calculate the average of the metrics
25	**End** of draw round

TABLE V. SHADED AREAS

Area	r	X	Y
1	150	-200	600
2	150	700	400
3	150	400	-400
4	150	0	-800
5	150	-600	-100

TABLE VI. SYSTEM PARAMETERS

Parameters	Values
m_T	50, 100 or 200
$c_{\max}$	5
r	1000 m
FC	$(0, 0)$
P_{txPU}	1 W
n	60
α	$[2, 4]$
P_s	1 μW
P_{ref}	0.1
n_{sensing}	30,000
$cycle_{CH}$	1000
P_{rxCH}	-100 dBm
P_{rxFC}	-100 dBm
$\max_{\text{back}}$	10, 15 or 20

cluster-based and proposed systems exhibit even higher P_d values, close to 1, while the classic system remains stable with P_d equal to 1. These results highlight that increasing the number of SUs significantly contributes to improving system accuracy. The increase in the number of sensors raises the amount of available information, often redundant, strengthening the decision-making process, since the greater the number of reports transmitted, the higher the FC's ability to accurately identify the occupancy status of the band.

The performance in terms of P_{fa} is approximately zero in all analyzed scenarios, due to the criteria adopted in the system design.

The classic system showed higher P_d in all analyzed scenarios due to the fusion technique applied at the FC, as previously described. However, the proposed system demonstrates greater stability, maintaining satisfactory performance even in the final stages, when the SUs start being permanently deactivated. It is worth noting that, in the P_d graph, the curves end at the point where no more SUs are available in the network to perform sensing.

Figure 4 illustrates the lifespan of the secondary network. It can be observed in Fig. 4(a), (b), and (c) that the classic system operates for just over 30,000 cycles. Since the SUs are randomly positioned, some may be closer to the FC, resulting in lower energy consumption when transmitting their decisions. This variation causes certain SUs to take longer to be deactivated, slightly extending the overall network lifespan.

In all the curves of Fig. 4, it can be observed that the proposed system exhibits a longer lifespan compared to the others. However, Fig. 4(a), with $m_T = 50$, shows a more significant relative gain compared to the curves in Fig. 4(b) and (c), with $m_T = 100$ and $m_T = 200$, respectively. This can also be attributed to the higher dispersion of the SUs in scenarios with lower density, resulting in more varied energy consumption patterns. As the number of SUs increases, they tend to be positioned closer to each other, leading to more

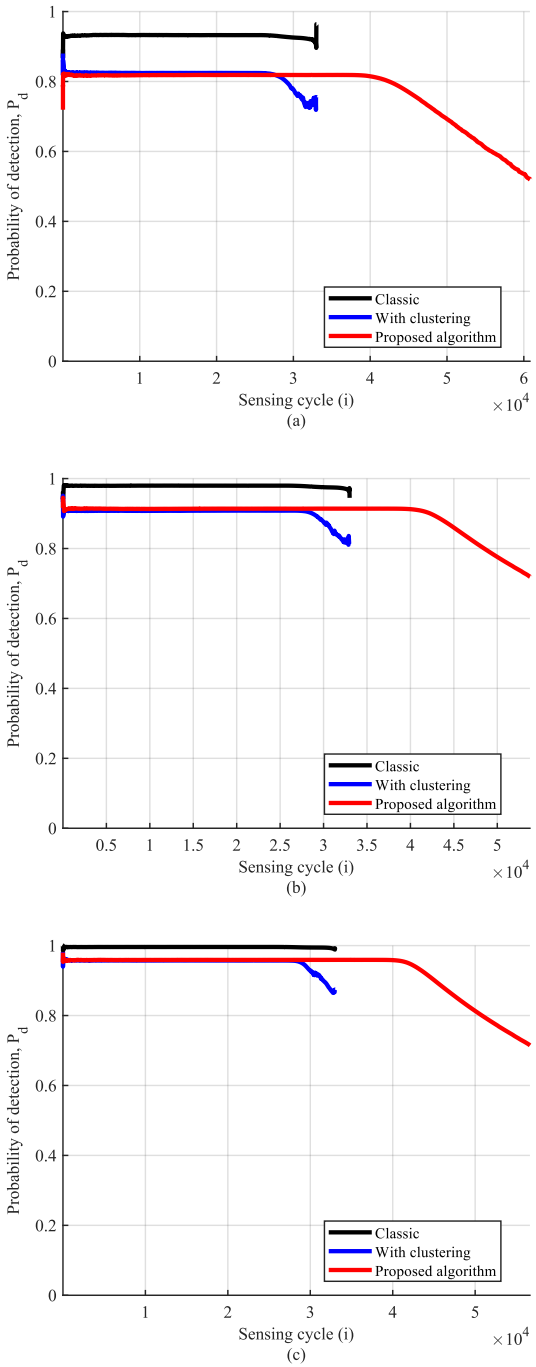


Figure 3. P_d for $c_{\max} = 5$, $n = 60$ samples per SU, with $m_T = 50$, 100, and 200 from top to bottom.

uniform energy consumption and, consequently, reducing the impact of heterogeneity on the network's lifespan.

For this reason, the curves for $m_T = 100$ and $m_T = 200$ exhibit similar behaviors, with a slight reduction in lifespan observed in Fig. 4(c). Furthermore, it is important to highlight that in Fig. 4(b), the CSS is cluster-based, which means the network operation depends directly on the CHs. When a CH ceases to operate, all nodes in the associated cluster

are deactivated. In contrast, in Fig. 4(c), where the proposed algorithm is applied, if a CH stops operating, another CH is assigned to the cluster.

It can be inferred, therefore, that increasing the number of SUs tends to stabilize the network's lifespan, as the SUs begin to operate under similar conditions. To ensure that the addition of more SUs continues to provide significant durability gains, it would be necessary to expand the network's coverage area, keeping the SUs relatively dispersed.

Although the classic system and the clustered system exhibit approximately the same lifespan, in the classic system all SUs send their decisions directly to the FC. In dense networks, this can cause communication overload, compromising system efficiency. On the other hand, the clustered system organizes the SUs into smaller groups, which improves network scalability and allows for the application of more efficient strategies for energy management, thereby contributing to the extension of the network's lifespan.

The oscillation observed in the average number of active SUs during the final cycles of the proposed algorithm results from the dynamics between penalties for detection failures, shutdown conditions due to low energy, and possible temporary reactivations of the SUs. SUs are disabled when their residual energy falls below a threshold λ and an additional parameter β. The system alternates between deactivating SUs with insufficient energy and reintegrating them when detection conditions become favorable. This intermittent behavior, combined with the use of an average-based metric, causes noticeable fluctuations in the curve, especially when the number of active SUs approaches the minimum threshold required for the system's continued operation.

Although the classic system exhibits higher P_d values, as illustrated in Fig. 3, it can be observed in Fig. 4(a)–(c) that the proposed system provides a significantly longer network lifespan.

Figure 5 illustrates the lifespan of the fixed scenario with variations in the maximum number of consecutive backoffs ($\max_{\text{back}}$). This parameter defines the limit for the consecutive backoff occurrence counter β of each sensor; that is, when the value of β_j (the backoff occurrence counter of the j-th sensor) reaches $\max_{\text{back}}$, the respective sensor is permanently disabled from the sensing process.

The highlighted points in Fig. 5 indicate the average onset of SU deaths, i.e., the cycle in which the SUs begin to be deactivated due to poor spectral sensing performance. As this technique is exclusive to the proposed algorithm, the analysis focuses only on the curves corresponding to this system.

The backoff mechanism, described in (12), consists of assigning temporary inactivity periods to SUs with poor detection performance, reducing energy consumption during these intervals. SUs located in shadowed areas experience higher signal attenuation due to the presence of obstacles, which increases the likelihood of errors in spectral sensing. As a result, these SUs tend to fail more frequently, remaining in backoff for several consecutive cycles until, upon reaching the maximum allowed number of backoffs, they are permanently disabled from the sensing process.

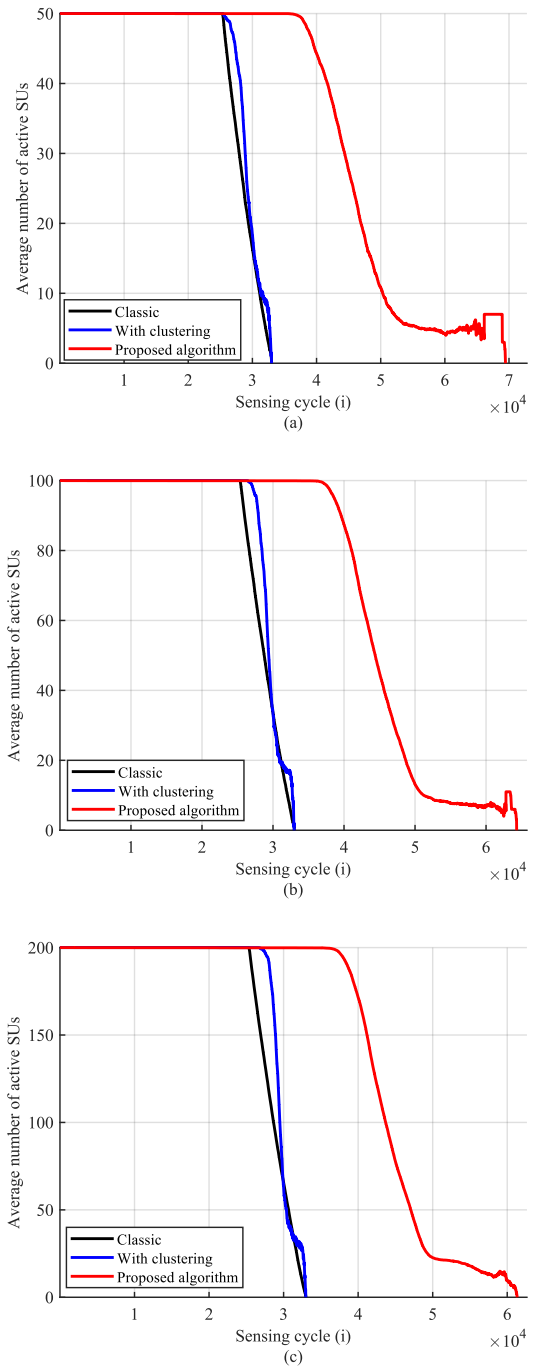


Figure 4. Network lifespan for $c_{\max} = 5$, $n = 60$ samples per SU, with $m_T = 50$ in graph (a), $m_T = 100$ in graph (b), and $m_T = 200$ in graph (c).

The premature deactivation of SUs is directly related to the value of $\max_{\text{back}}$. In Fig. 5, with $\max_{\text{back}} = 10$, SUs in unfavorable regions are disabled more quickly, as they accumulate consecutive backoffs in fewer cycles. When β_j reaches the limit, the SU is considered dead for the sensing process, even if it still has available energy.

In Fig. 5, with $\max_{back} = 15$, the SUs take longer to be permanently disabled. With $\max_{back} = 20$, the SUs remain active for an even longer period. This analysis shows that, in the proposed system, the quality of spectrum sensing is prioritized over residual energy. That is, a sensor may be excluded from the sensing process even if it has sufficient energy, if its detection performance is unsatisfactory.

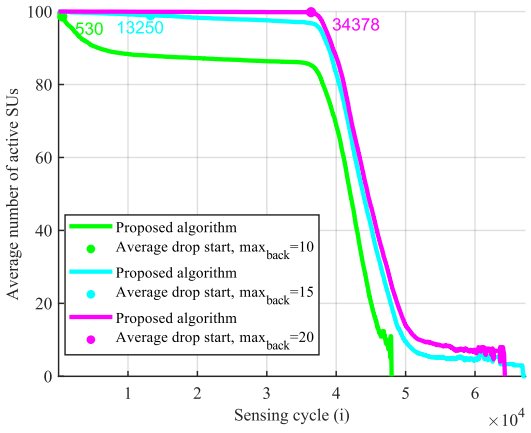


Figure 5. Network lifetime for $m_T = 100$, $c_{\max} = 5$, $n = 60$ samples per SU, with variable $\max_{back}$.

IX. Conclusion

The proposed algorithm, which integrates temporary and permanent sensor exclusion with dynamic CH selection, provides significant gains in energy efficiency and in the longevity of the secondary network in CRNs. In the evaluated scenarios, the proposed solution increases the CRN lifetime from approximately 3×10^4 to about 6×10^4 sensing cycles, which corresponds to an average network lifetime gain close to 100% relative to the classical method, without compromising sensing performance. Simulations demonstrated that, even in scenarios with varying SU densities and backoff configurations, the algorithm maintains a robust P_d while distributing energy consumption evenly among the sensors. Compared to the classical and clustering-based systems, the proposed method stands out for its performance stability, adaptability under adverse conditions, and reduction of the impact of localized failures. Thus, the algorithm contributes not only to the protection of the primary network by ensuring reliable sensing decisions but also to the operational sustainability of the secondary network, making it an effective approach for implementing more resilient and energy-efficient CRNs.

Acknowledgment

This work was partially supported by RNP, with resources from MCTIC, Grant No. 01245.020548/2021–07, under the Brazil 6G project of the Radiocommunication Reference Center (Centro de Referência em Radiocomunicações – CRR) of the National Institute of Telecommunications.

References

[1] A. M. J. Tomás and J. M. C. Brito, "Algorithm for Sensor Exclusion and Dynamic Cluster Head Selection in Cognitive Radio Networks", in *Proc of the Twentieth International Conference on Digital Telecommunications (ICDT 2025)*, Nice, France, May 2025.

[2] *ISO/IEC/IEEE 8802-22.2:2015/Amd.2:2017(E). Part 22, Cognitive Wireless RAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Policies and Procedures for Operation in the TV Bands AMENDMENT 2: Enhancement for broadband services and monitoring applications : ISO/IEC/IEEE International Standard - Information technology – Telecommunications and information exchange between systems – Local and metropolitan area networks – Specific requirements*, eng. New York: IEEE, 2018, ISBN: 1-5044-4761-1.

[3] Z. Quan, S. Cui, and A. H. Sayed, "Optimal linear cooperation for spectrum sensing in cognitive radio networks", *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 1, pp. 28–40, 2008.

[4] L. Wang, J. Wang, G. Ding, F. Song, and Q. Wu, "A survey of cluster-based cooperative spectrum sensing in cognitive radio networks", in *Proceedings of 2011 Cross Strait Quad-Regional Radio Science and Wireless Technology Conference*, IEEE, vol. 1, 2011, pp. 247–251.

[5] W. Xia, S. Wang, W. Liu, and W. Chen, "Cluster-based energy efficient cooperative spectrum sensing in cognitive radios", in *2009 5th International Conference on Wireless Communications, Networking and Mobile Computing*, IEEE, 2009, pp. 1–4.

[6] C. Sun, W. Zhang, and K. B. Letaief, "Cluster-based cooperative spectrum sensing in cognitive radio systems", in *2007 IEEE International Conference on Communications*, IEEE, 2007, pp. 2511–2515.

[7] Z. Bai, L. Wang, H. Zhang, and K. Kwak, "Cluster-based cooperative spectrum sensing for cognitive radio under bandwidth constraints", in *2010 IEEE International Conference on Communication Systems*, IEEE, 2010, pp. 569–573.

[8] D. A. Guimarães, "Hybrid fusion of pietra–ricci index detector information for cooperative spectrum sensing", *Ad Hoc Networks*, vol. 150, p. 103 265, 2023.

[9] A. Rauniyar and S. Y. Shin, "A novel energy-efficient clustering based cooperative spectrum sensing for cognitive radio sensor networks", *International Journal of Distributed Sensor Networks*, vol. 11, no. 6, p. 198 456, 2015.

[10] R. Mustafa, M. J. Khan, T. Kaur, et al., "Novel energy efficient clustering approach for cognitive radio sensor networks", *Turkish Journal of Computer and Mathematics Education*, vol. 12, no. 9, pp. 3109–3116, 2021.

[11] H. Ye and J. Jiang, "Optimal linear weighted cooperative spectrum sensing for clustered-based cognitive radio networks", *EURASIP Journal on Wireless Communications and Networking*, vol. 2021, no. 1, p. 84, 2021.

[12] M. Monemian and M. Mahdavi, "Analysis of a new energy-based sensor selection method for cooperative spectrum sensing in cognitive radio networks", *IEEE Sensors Journal*, vol. 14, no. 9, pp. 3021–3032, 2014.

[13] M. A. Hossain, M. Schukat, and E. Barrett, "Enhancing the spectrum sensing performance of cluster-based cooperative cognitive radio networks via sequential multiple reporting channels", *Wireless Personal Communications*, vol. 116, no. 3, pp. 2411–2433, 2021.

[14] J. Zhang and R. Yan, "Centralized energy-efficient clustering routing protocol for mobile nodes in wireless sensor networks", *IEEE Communications Letters*, vol. 23, no. 7, pp. 1215–1218, 2019.

[15] F. Kong, Z. Jin, and J. Cho, "A novel sensing nodes selection scheme for energy efficiency of cooperative spectrum sensing in cluster-based crsns", in *2015 International Conference on Information and Communication Technology Convergence (ICTC)*, IEEE, 2015, pp. 26–31.

[16] I. F. Akyildiz, B. F. Lo, and R. Balakrishnan, "Cooperative spectrum sensing in cognitive radio networks: A survey", *Physical Communication*, vol. 4, no. 1, pp. 40–62, 2011.

[17] C. Ma, N. Liu, and Y. Ruan, "A dynamic and energy-efficient clustering algorithm in large-scale mobile sensor networks", *International Journal of Distributed Sensor Networks*, vol. 9, no. 11, p. 909 243, 2013.

[18] O. Ergul and O. B. Akan, "Energy-efficient cooperative spectrum sensing for cognitive radio sensor networks", in *2013 IEEE Symposium on Computers and Communications (ISCC)*, IEEE, 2013, pp. 000 465–000 469.

[19] P. Sasikumar and S. Khara, "K-means clustering in wireless sensor networks", in *2012 Fourth International Conference on Computational Intelligence and Communication Networks*, IEEE, 2012, pp. 140–144.

Hourly and Weather-Based Variability in Starlink Internet Performance: A TCP and UDP Throughput Study

Céline Careau, Emil Fredriksson, Robert Olsson, Peter Sjödin and Claes Beckman
School of Electrical Engineering and Computer Science
KTH Royal Institute of Technology
Stockholm, Sweden
e-mail: {careau | emifre | roolss | psj | claesb}@kth.se

***Abstract*—Starlink provides satellite internet connectivity to customers worldwide using Low Earth Orbit (LEO) satellites connecting to ground stations and user equipment. Precipitation, hourly variability, and the use of different transport protocols, all have impact on throughput. The study was conducted in Stockholm, Sweden, at a latitude of 59.3 degrees north, which is well north of the main coverage area of Starlink. Higher latitudes are covered by fewer satellites compared to Central Europe and the main regions of the United States. The study consists of throughput measurements with the network performance measurement tool iPerf3 using two different transport protocols: Transmission Control Protocol (TCP) and User Datagram Protocol (UDP). Precipitation (rainfall) measurements were conducted simultaneously. The results show a notable decrease in the throughput when moderate rainfall (about 1 mm per hour) is present, about 16 percent for UDP and 28 percent for TCP. The data also show that the throughput varies during different hours of the day, with around 21 percent for UDP and 32 percent for TCP. The highest throughput is received at night and early mornings for both transport protocols. The throughput achieved through the Starlink network with the TCP protocol fluctuates more than on 4G mobile networks. In conclusion, our study provides further knowledge about the effects of precipitation and hourly variability with TCP and UDP on Starlink's performance, specifically when operated at latitudes outside of Starlink's main coverage area.**

Keywords-starlink; leo-antennas; network; tcp; udp; weather; precipitation; iperf3; ping; throughput; latency; internet measurements.

I. Introduction

This is an extended version of our research paper "Throughput Analysis of Starlink Satellite Internet: Study on the Effects of Precipitation and Hourly Variability with TCP and UDP", including more figures of the data collected as well as a comparison between Starlink connectivity and 4G [1].

Starlink provides broadband connectivity mainly over Central Europe and the main regions of the United States (within the latitudes of ±55 degrees). Regions at higher latitudes, e.g., Scandinavia, are covered by fewer satellites but still receive good enough service for sparsely populated areas [2]. The satellite distribution, which is seen in Figure 2, is a screenshot of a live Starlink satellite map, where more than 4000 satellites were active at the time of the study [3]. In the northern parts of the world, there is a clear decrease in satellite density orbiting around the globe in comparison with central parts of the world. The data path for the Starlink network is visualised in Figure 1, showing both the ground station and a "Point of Presence" (PoP) that is used to transfer data between the user antenna and the Internet. The Starlink system makes use of ground stations scattered around the world to be able to connect the satellites to the PoPs. The ground stations are connected via leased fibre to the closest PoP [4]. Aggregated data is then able to travel between the ground station and the PoP, where the data enters the Internet as regular traffic [5]. SpaceX has not revealed where Starlink's PoP and ground stations are located.

Due to the LEO satellites' limited coverage and high travelling speed, the satellites quickly move out of range from user antennas and ground stations. To connect a satellite within range, the Starlink system performs a network reconfiguration every 15 seconds [6]. This process introduces a short interruption, but it is necessary to maintain a connection to the satellite constellation.

The effect of precipitation on the Starlink system performance has been investigated in Central Europe (Germany and the Netherlands) [7], but remains unexplored in Scandinavia. Previous papers have provided data on Starlink's performance over "Transmission Control Protocol" (TCP) [8] and "User Datagram Protocol" (UDP) [7]. However, no studies have been found comparing the two protocols over the Starlink network in Scandinavia.

This study examines the throughput performance of the Starlink system, how it is affected by moderate rainfall and how the throughput varies by hourly variability when operated in Stockholm, Sweden. In addition, throughput comparisons are made using two different transport protocols: TCP and UDP.

This paper is structured as follows: Section II gives insight into TCP and UDP measurements on the Starlink system. Section III describes the measurement setup and Section IV presents an analysis of the results obtained from the experiments. The results are then further discussed in Section V. The paper is concluded, and future work is explored, in Section VI.

II. Related work

This section covers studies of Starlink's performance related to TCP and UDP measurements. TCP is a connection-oriented Internet transport protocol. This implies that before data is sent, the connection between the sender and receiver has to be confirmed. The acknowledgement between the two points is referred to as a TCP handshake. The TCP protocol re-transmits

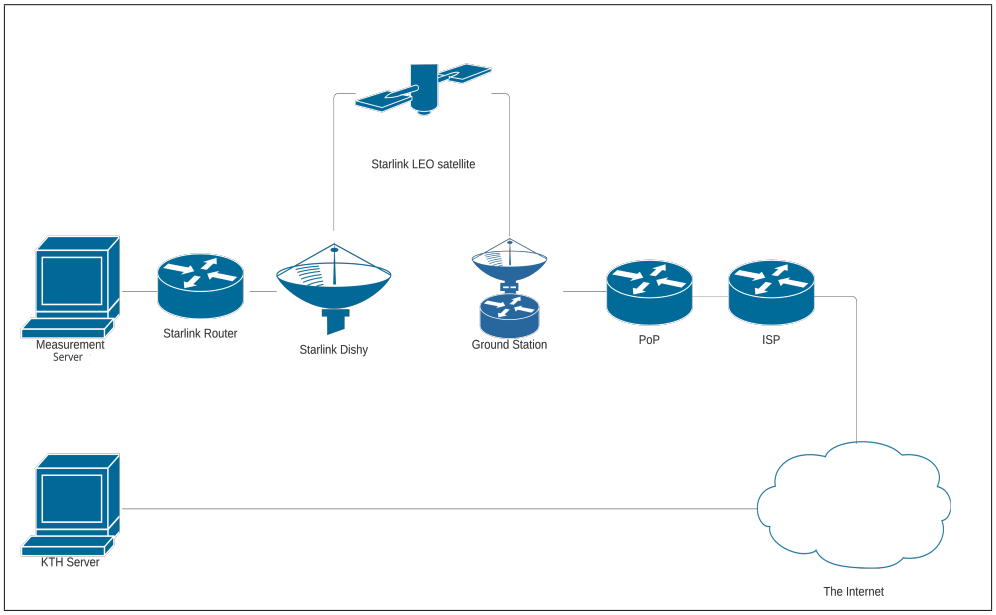


Figure 1. Data path for throughput measurements.

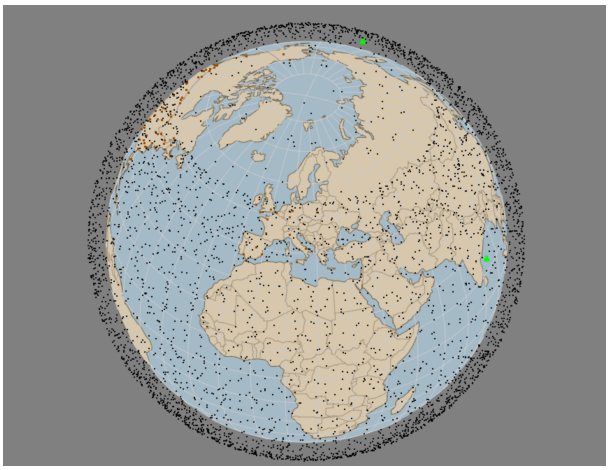


Figure 2. Screenshot taken of live map during data collection period showing over 4000 active Starlink satellites [3].

data if an error occurs. An error could, for example, occur due to packet loss, corrupt data or data being transmitted in the wrong order. The re-transmission process ensures data integrity at the cost of throughput due to an increase in total transmission time [9]. Parallel TCP connections can be used to achieve higher bandwidth. The number of parallel connections for maximum bandwidth utilisation depends on the network bandwidth available.

UDP is a transport protocol that, unlike TCP, does not require an established connection before data transmission. The transmission rate when using the UDP protocol is decided by the sender. This allows high transfer speed by using the entire link bandwidth. However, a static transmission rate can cause packet loss if the receiver is incapable of receiving data at the same rate as it is sent.

A. *Previous studies of UDP*

The "WetLinks" paper by Laniewski et al. [7] presents a large dataset of Starlink performance measurements gathered through experiments conducted in Germany and the Netherlands. This data set allowed the authors to analyse the correlation between Starlink's performance and weather conditions. The authors collected weather data both independently and from national weather services in their respective countries. In the paper, UDP was used to measure the throughput of Starlink during different weather conditions. The two measurement locations give a somewhat better view of Starlink's performance than from just one location. However, both places are located at latitudes with dense concentrations of Starlink satellites. In contrast, our paper reports measurements made at a location with much fewer Starlink satellites in nearby orbits [2]. The "WetLinks" paper reports UDP throughput in the range of 170-250 Mbps (median 210 Mbps) during days without precipitation. The paper also includes an analysis of how performance varies over the hours of the day. The time-of-day analysis can contribute to a better understanding of how the Starlink network is affected by user traffic. The paper reports that the minimum average throughput throughout a day is approximately 20% lower than the maximum. The median UDP throughput decreased by 17% when it was raining, highlighting the impact of moderate rain showers on Starlink's performance.

B. *Previous studies of TCP*

The majority of internet traffic is sent with TCP [10]. High levels of packet loss, which can be caused by, for example, interruptions in the satellite connection, are expected to negatively affect TCP throughput and have a large impact on end-user performance. Michel et al. [8] measured TCP throughput (using Speedtest by Ookla [11]) at the UCLouvain campus in Louvain-la-Neuve, Belgium. The reported TCP

throughput range was 100-250 Mbps (median 178 Mbps), which is considerably lower than the UDP throughput reported in the "WetLinks" paper [7].

III. Method

In our study, the throughput data is collected using a Starlink "Dishy McFlatface" antenna [12] located on the roof of the Electrum building in Kista, Stockholm (Figure 3). The Starlink device is directly connected to a server from which all measurements are conducted (Figure 1). The weather data is collected using a "Davis Rain Collector" [13] (rain bucket) (Figure 4) located next to the antenna.

The measurements are designed to give a real-life estimate of the system performance expected from Starlink Internet connectivity in Scandinavia.

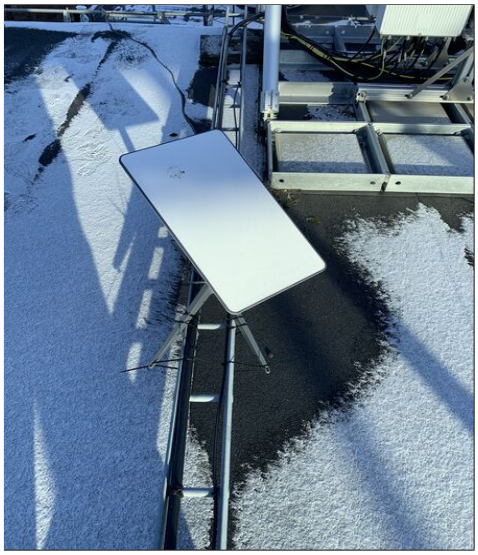


Figure 3. The Starlink user antenna "Dishy McFlatface" [12].

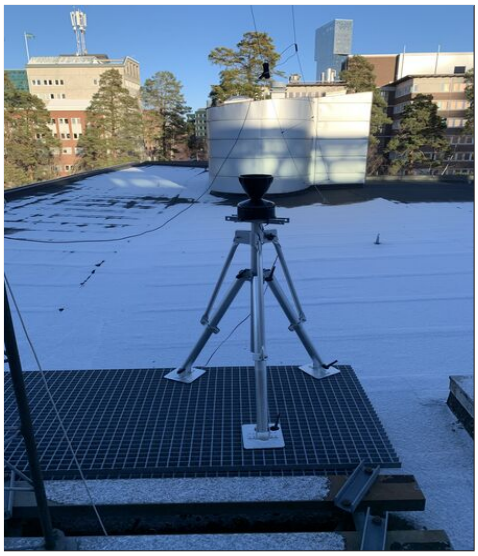


Figure 4. Davis Rain Collector [13].

TABLE I. Measurement hardware specifications for the Starlink Dishy[12]

Hardware specifications for the Starlink Dishy	
Antenna	Electronic phased array
Orientation	Motorised self orienting
Environmental rating	IP54
Snow melt capability	Up to 40 mm/hour
Operating temperature	-30°C to 50°C
Field of vision	110°
Average power consumption	50–75 W

TABLE II. Specifications Davis 6464 Rain Collector[13]

Specifications Davis 6464 Rain Collector	
Sensor type	Tipping spoon with magnetic switch
Collection area	214 square cm
Range daily rainfall	0.0 mm to 999.8 mm
Range total rainfall	0.0 mm to 6553 mm
Accuracy	For rain rates up to 50 mm/hr ±4% of total or +0.2mm (one tip of the spoon) whichever is greater
Update interval	20–24 seconds

The throughput data collection consists of four different iPerf3 measurements for TCP and UDP, scheduled to run in series. Since the Starlink network undergoes a complete reconfiguration every 15 seconds, each measurement runs for 40 seconds. This duration ensures that at least two reconfigurations occur and allows the TCP connection to readjust its speed, providing more realistic real-world performance results. The iPerf3 measurement for UDP is limited to a bitrate of 250 Mbps to prevent unnecessary network load, alongside a configured buffer size of 1400 bytes to reduce packet loss [14]. For TCP connections, the iPerf3 command is set to use 8 parallel streams with a buffer size of 128 kB.

IV. Results and Analysis

This section presents and analyses the results from the study, categorised into three sections based on the findings.

A. Precipitation

The following section contains a comparison between three consecutive days with rain being present on the third day. The day with rain is referred to as "the rainy day". Figure 6 and Figure 7 show the throughput measured in the interval 8:00-22:00. To get a clearer view of the data, a rolling average is applied to the data and can be seen in Figure 8 for TCP and Figure 9 for UDP. Figure 10 (TCP) and Figure 11 (UDP) illustrate the days using Kernel Density Estimation (KDE) [15], providing a clearer picture of how throughput is distributed across the different days. Over the three days, the median throughput for TCP was 120 Mbit/s on the first day, 118 Mbit/s on the second, and 86 Mbit/s on the third, rainy day. For UDP, the throughput was 194 Mbit/s, 202 Mbit/s, and 169 Mbit/s, respectively. This corresponds to an approximate 28% decrease in TCP throughput and a ~16% decrease in UDP

throughput on the rainy day. Figure 5 presents the measured throughput for TCP and UDP on the rainy day. The blue dots represent the amount of rainfall in one-minute intervals, with a total measured rainfall of 15 mm during April 14th, 2024. This is well in agreement with the data provided by The Swedish Meteorological and Hydrological Institute (SMHI) for that date [16]. The TCP throughput is significantly affected, even though the rainfall is classified as moderate (less than 4 mm/hour) [17].

TABLE III. COMPARISON OF MEDIAN THROUGHPUT (MBIT/S) OF THE TWO DAYS WITHOUT PRECIPITATION (DAY 1 & DAY 2) AND THE DAY WITH PRECIPITATION (DAY 3).

Median throughput [Mbit/s]	Day 1	Day 2	Day 3
TCP	120	118	86
UDP	194	202	169

TABLE IV. SUMMARY OF THE DATA ON THE DAY WITH PRECIPITATION.

Description	Data
TCP decrease on rainy day	~ 28%
UDP decrease on rainy day	~ 16%
Total rainfall [mm]	15 mm (April 14, 2024)
Rainfall intensity	Moderate <(4 mm/hour)

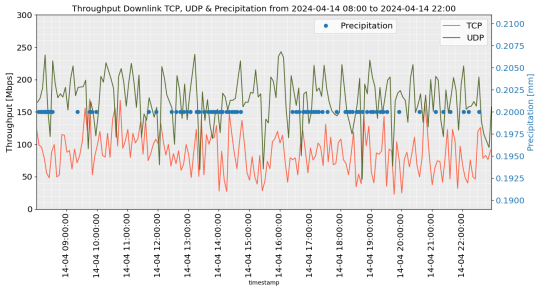


Figure 5. Measured TCP and UDP throughput during the rainy day. The blue dots represent the amount of rainfall for one minute.

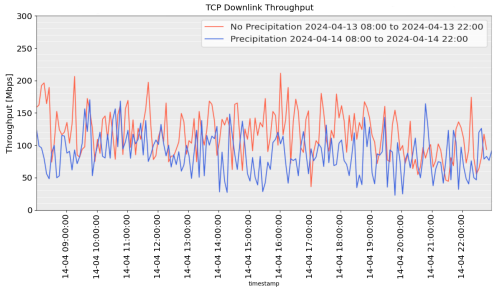


Figure 6. TCP: Throughput on the rainy day (blue) vs a day without rain (red).

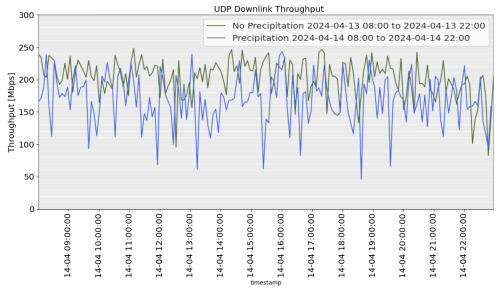


Figure 7. UDP: Throughput on the rainy day (blue) vs a day without rain (green).

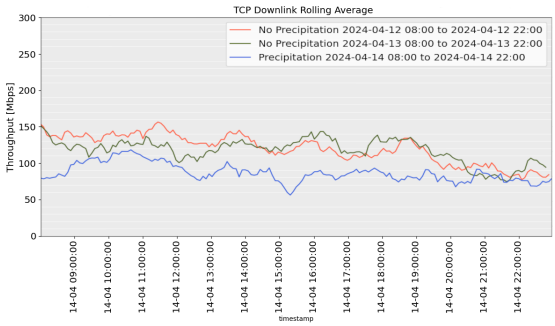


Figure 8. TCP: Throughput the rainy day (blue) vs two days without rain (red and green) using a rolling average on every 10th measurement.

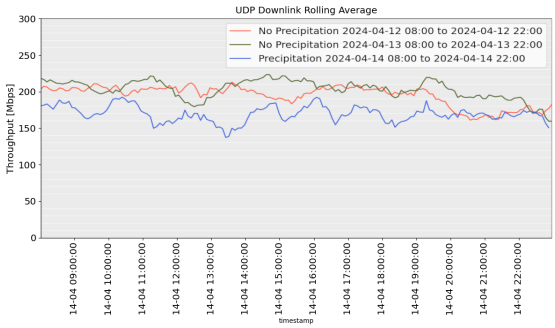


Figure 9. UDP: Throughput the rainy day (blue) vs two days without rain (red and green) using a rolling average on every 10th measurement.

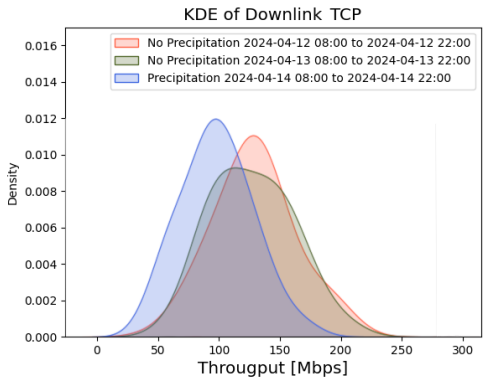


Figure 10. TCP: Probability density function for throughput measurements on the rainy day (blue) vs two days without rain (red and green), using *Seaborn* KDE-plots, with a bandwidth of 0.5 [18].

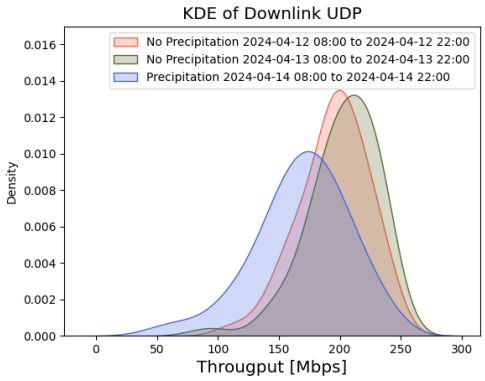


Figure 11. UDP: Throughput density measurement on the rainy day (blue) vs two days without rain (red and green), using *Seaborn* KDE-plots, with a bandwidth of 0.5 [18].

B. Hourly variations

Figure 12 shows the throughput data for TCP (red) and UDP (green) during 72 hours without rain. As can be seen, there is a significant decrease in Starlink throughput during the daytime compared to the night. The highest throughput was measured during the nights and early mornings, while the lowest throughput was observed in the late afternoon and evenings.

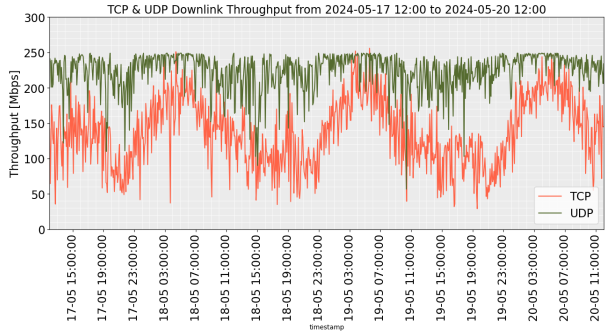


Figure 12. Throughput data from TCP (red) and UDP (green) for a 72-hour rain-free period.

C. Internet protocol

Figure 14 shows the hourly variations in throughput over seven days with and without precipitation. As seen in Figure 14, the mean throughput for UDP peaks in the early morning with a mean throughput 243 Mbit/s at 04:00-05:00. The lowest throughput occurs at around 21:00 with a mean of 185 Mbit/s. In contrast, the TCP measurements in Figure 14 show much lower mean throughput. The highest mean for TCP throughput is observed at around 05.00 with 188 Mbit/s, while the lowest mean is found at around 21.00 with 66 Mbit/s. By calculating the difference between the 75th and 25th quantile (IQR), we find an average difference of 46.05 Mbit/s for UDP and 41.88 Mbit/s for TCP. The average mean throughput for UDP is 208 Mbit/s, while for TCP, it is 132 Mbit/s, resulting in a ~57% higher mean for UDP compared to TCP. To get a sense of the variability, we express the average interquartile range as a percentage of the average median. This analysis shows that for UDP, the throughput varied by ~21%, and for TCP, it varied by ~32%.

TABLE V. Summary of TCP and UDP Throughput Measurements

Throughput [Mbit/s]	UDP	TCP
Highest mean	243	188
Lowest mean	185	66
Average mean	208	132
IQR (75th–25th percentile)	46.05	41.88
IQR as % of median	~ 21%	~ 32%

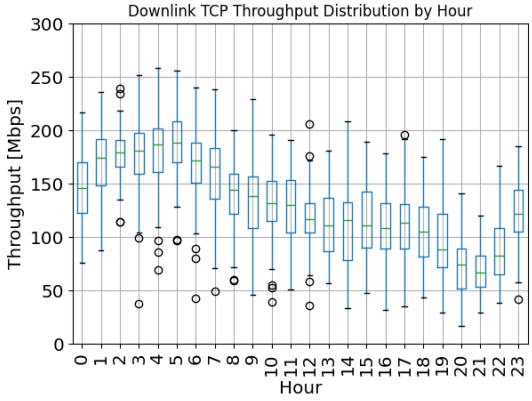


Figure 13. Boxplot of TCP throughput data from 7 consecutive days.

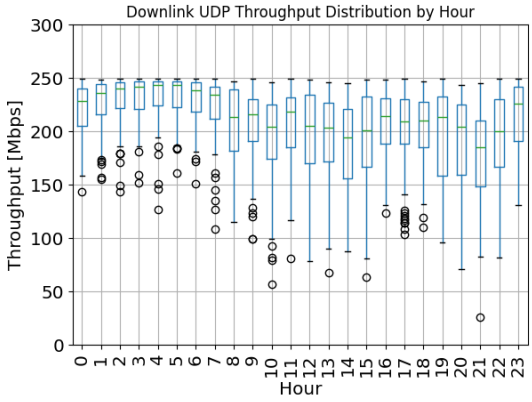


Figure 14. Boxplot of UDP throughput data from 7 consecutive days.

D. Validation analysis

According to the Starlink website, a throughput between 111 Mbps and 212 Mbps is expected at the location of the Starlink antenna within the 20th and the 80th percentile [19]. Figure 15 and Figure 16 show the TCP and UDP throughput within the 20th and 80th percentile over 6 days. With the TCP protocol, the distribution is similar to the numbers given by Starlink. As for the UDP protocol, the results show a higher average throughput.

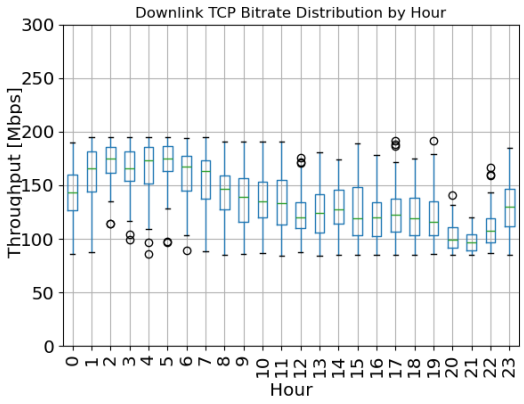


Figure 15. TCP: Throughput distribution by the hour for the 20th to 80th percentile, using box plots. The data is taken from 6 days without rain.

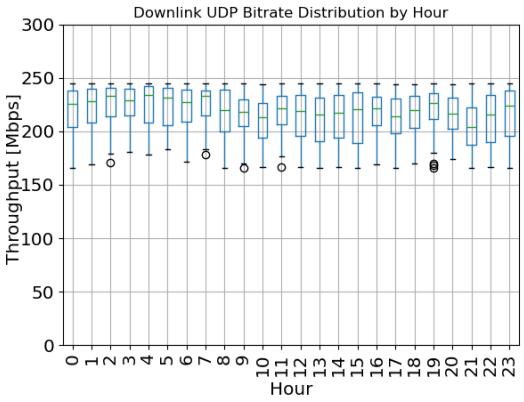


Figure 16. UDP: Throughput distribution by the hour for the 20th to 80th percentile, using box plots. The data is taken from 6 days without rain.

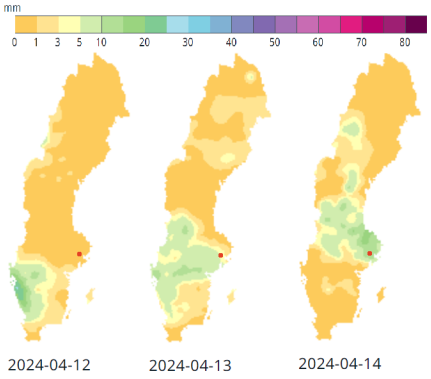


Figure 17. Data of precipitation (rainfall) from SMHI [16] during the testing dates. The red dot indicates the position of the test equipment.

The protocol used by Starlink to set the estimated throughput is not specified. However, the throughput results for both the TCP and UDP measurements are within the stated throughput range.

To validate the rain collector measurements, the data were compared with SMHI data. Figure 17 presents data from SMHI, which show precipitation at the test location during the same date as the measurements in Subsection IV-A [16]. It also verifies the two days without rainfall, which corresponds with our precipitation measurements.

V. Discussion

Our results show that Starlink's downlink throughput is affected by rain. This is expected, as terrestrial antennas generally struggle to transmit and receive signals during precipitation [20], especially at higher frequencies. Starlink operates in three bands above 10 GHz [21], where rain attenuation is more significant [22]. These include the Ku-band (10.7–14.5 GHz), Ka-band (17.3–30.0 GHz), and E-band (71–76 GHz and 81–86 GHz) [21].

The Ku-band, used for both uplink and downlink communication with Starlink user terminals, is the focus of our study, as rain measurements were collected at the user terminal location. The higher-frequency Ka and E bands are used for communication between satellites and ground stations [4]. Since these bands are more susceptible to rain attenuation [22], further investigation is needed to analyse the throughput impact of precipitation at ground stations.

The Starlink system shows a distinct variation in throughput depending on the time of day. The throughput is higher during the night and early mornings than during the day and evenings. The cause of this pattern could be that the data traffic is higher during the day, implying a higher load on the network. Hence, areas with a lower density of Starlink satellites may be more affected by network load, since more users need to share the same capacity.

Laniewski et al. [7] concluded that the throughput for UDP varies by ±10% during the day. This is similar to our results. For TCP, our data shows that the throughput varies by ±30%. The variation is expected because of the inherent properties of the two transport protocols.

The Starlink system's TCP variation can be compared to 4G cellular networks to give perspective on the results. Garcia et al. conducted a study that included a comparison of network throughput, depending on the hour of the day. The study used the TCP protocol over different 4G cellular network operators [23]. The operator with the most fluctuations in throughput during the day varied between approximately 35-45 Mbit/s, resulting in a daily variation of ±12.5%. The throughput for 4G shows a significantly lower fluctuation than Starlink's throughput using the TCP protocol. This illustrates how the TCP protocol can have less variable throughput within a network under certain circumstances. Since TCP is a connection-oriented protocol, latency also impacts the throughput [24]. The performance impact due to latency depends on the congestion control algorithm.

There are different types of congestion control algorithms. When data is sent with a high transfer rate and/or high latency, congestion control will have larger impact on throughput. Congestion control algorithms are designed for network environments with different characteristics than LEO satellite networks [25]. The LEO satellite network is a new kind

of network environment, with no specialised algorithm for the network's congestion. The Starlink network environment includes a reconfiguration of the entire network every 15 seconds, and the current congestion control algorithms are not adapted to such reconfigurations. Barbosa et al. published a comparison of congestion control algorithms on a simulated LEO satellite network. The results of the papers show that there is a latency and throughput difference when using different algorithms, with BBR being able to adapt to the LEO network the best. According to the authors [26], BBR adapted best to the simulated LEO satellite network. According to the paper by Barbosa et al., Linux uses the CUBIC congestion control as a standard, and therefore, it is assumed to be the congestion control algorithm used in the research experiments in that paper. Our study did not have latency and congestion control algorithm analysis within its scope, although it would be an interesting topic for future studies. Expanding on the testing done by Barbosa et al. is also an area worth exploring. By doing a similar test on a real-world LEO satellite constellation, there would be a better understanding to gain of how different TCP congestion control algorithms affect satellite internet throughput and latency.

VI. Conclusion and Future Work

This study shows that for a Starlink satellite terminal in Stockholm, Sweden, the throughput varies dramatically with precipitation, time of day, and choice of transport protocol.

For future Starlink users and researchers, it is important to understand the limitations and variations in throughput depending on these factors. However, Starlink is constantly being updated and changed, which will have an effect on future performance.

There are still issues that deserve examination within the Starlink system. Possible future work includes:

- Measuring packet loss, latency and jitter.
- Testing different methods to measure throughput.
- Testing throughput and other parameters with the Starlink API. [27].
- Examining Starlink's performance in relation to satellite alignments.
- Examining how throughput via the Starlink network is affected by rain at the ground station.

Acknowledgment

This study was supported by the Sustainable Mobile Autonomous and Resilient 6G SatCom research centre (SMART 6GSAT), funded by the Swedish Foundation for Strategic Research.

References

[1] C. Careau, E. Fredriksson, R. Olsson, P. Sjödin and C. Beckman, 'Throughput Analysis of Starlink Satellite Internet: A Study on the Effects of Precipitation and Hourly Variability with TCP and UDP', presented at the SPACOMM 2025, The Seventeenth International Conference on Advances in Satellite and Space Communications, 18th May 2025, pp. 1–4, ISBN: 978-1-68558-276-0. [Online]. Available: https://www.thinkmind.org/library/SPACOMM/SPACOMM_2025/spacomm_2025_1_10_20007.html (visited on 10/12/2025).

[2] C. Beckman, J. Garcia, H. Mikkelsen and P. Persson, 'Starlink and Cellular Connectivity under Mobility: Drive Testing Across the Arctic Circle', in *2024 Wireless Telecommunications Symposium (WTS)*, 2024, pp. 1–9. DOI: 10.1109/WTS60164.2024.10536679.

[3] 'Satellite tracker', Starlink satellite tracker, [Online]. Available: https://satellitemap.space (visited on 08/03/2024).

[4] 'FCC Clears SpaceX to Use E-Band for Starlink Capacity Improvement', PCMag UK, Section: Networking, 11th Mar. 2024, [Online]. Available: https://uk.pcmag.com/networking/151385/fcc-clears-spacex-to-use-e-band-for-starlink-capacity-improvement (visited on 16/06/2024).

[5] *Case No. 2021-00002. 2021. Application of Starlink Services.* [Online]. Available: https://psc.ky.gov/pscecf/2021-00002/kerry.ingle%40dinslaw.com/01042021010318/Application_-_Designation_as_ETC.PDF (visited on 18/04/2024).

[6] N. Mohan *et al.*, 'A Multifaceted Look at Starlink Performance', 13th Oct. 2023, p. 15. (visited on 17/04/2024).

[7] D. Laniewski, E. Lanfer, B. Meijerink, R. van Rijswijk-Deij and N. Aschenbruck, *WetLinks: a Large-Scale Longitudinal Starlink Dataset with Contiguous Weather Data*, 13th Mar. 2024. DOI: 10.48550/arXiv.2402.16448. arXiv: 2402.16448[cs]. [Online]. Available: http://arxiv.org/abs/2402.16448 (visited on 10/05/2024).

[8] F. Michel, M. Trevisan, D. Giordano and O. Bonaventure, 'A first look at starlink performance', in *Proceedings of the 22nd ACM Internet Measurement Conference*, ser. IMC '22, event-place: Nice, France, New York, NY, USA: Association for Computing Machinery, 2022, pp. 130–136, ISBN: 978-1-4503-9259-4. DOI: 10.1145/3517745.3561416. [Online]. Available: https://doi.org/10.1145/3517745.3561416.

[9] J. F. Kurose and K. W. Ross, *Computer networking: a top-down approach*, 7. edition. Boston Munich: Pearson Education, 2017, 824 pp., p.123-127, ISBN: 978-0-13-359414-0.

[10] L. Qian and B. E. Carpenter, 'A flow-based performance analysis of TCP and TCP applications', in *2012 18th IEEE International Conference on Networks (ICON)*, Singapore, Singapore: IEEE, Dec. 2012, pp. 41–45, ISBN: 978-1-4673-4523-1. DOI: 10.1109/ICON.2012.6506531. [Online]. Available: http://ieeexplore.ieee.org/document/6506531/ (visited on 06/05/2024).

[11] 'Speedtest by Ookla - The Global Broadband Speed Test', Speedtest.net, [Online]. Available: https://www.speedtest.net/ (visited on 09/05/2024).

[12] 'Starlink | Specifikationer', Starlink, [Online]. Available: https://www.starlink.com/specifications (visited on 24/05/2024).

[13] 'AeroCone Rain Collector with Flat Base for Vantage Pro2 and EnviroMonitor (tipping spoon) - SKU 6464, 6464M', Davis Instruments, [Online]. Available: https://www.davisinstruments.com/products/aerocone-rain-collector-with-flat-base-for-vantage-pro2 (visited on 24/05/2024).

[14] S. H. Ali, S. A. Nasir and S. Qazi, 'Impact of router buffer size on TCP/UDP performance', in *2013 3rd IEEE International Conference on Computer, Control and Communication (IC4)*, Sep. 2013, pp. 1–6. DOI: 10.1109/IC4.2013.6653751. [Online].

Available: https://ieeexplore.ieee.org/document/6653751/?arnumber=6653751 (visited on 14/02/2025).

[15] M. Waskom, 'seaborn: statistical data visualization', *Journal of Open Source Software*, vol. 6, no. 60, p. 3021, 6th Apr. 2021, ISSN: 2475-9066. DOI: 10.21105/joss.03021. [Online]. Available: https://joss.theoj.org/papers/10.21105/joss.03021 (visited on 14/06/2024).

[16] 'Dygnskartor | SMHI', [Online]. Available: https://www.smhi.se/data/meteorologi/dygnskartor/nederbord/2024/april/ (visited on 27/05/2024).

[17] 'Rainfall calculator, metric-How much water falls during a storm? USGS Water Science School', [Online]. Available: https://water.usgs.gov/edu/activity-howmuchrain-metric.html (visited on 13/02/2025).

[18] 'seaborn.objects.KDE — seaborn 0.13.2 documentation', [Online]. Available: https://seaborn.pydata.org/generated/seaborn.objects.KDE.html#seaborn.objects.KDE (visited on 14/06/2024).

[19] 'Starlink | Availability Map', Starlink, [Online]. Available: https://www.starlink.com/map?view=download (visited on 27/05/2024).

[20] T. J. Smyth and A. J. Illingworth, 'Correction for attenuation of radar reflectivity using polarization data', *Quarterly Journal of the Royal Meteorological Society*, vol. 124, no. 551, pp. 2393–2415, Oct. 1998, ISSN: 0035-9009, 1477-870X. DOI: 10.1002/qj.49712455111. [Online]. Available: https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.49712455111 (visited on 09/05/2024).

[21] *NTIA Docket No. 230308-0068*. [Online]. Available: https://www.ntia.gov/sites/default/files/publications/spacex.pdf.

[22] 'Comparing the Ka-Band vs. the Ku- Band', 27th Feb. 2023, [Online]. Available: https://resources.pcb.cadence.com/blog/2023-comparing-the-ka-band-vs-the-ku-band (visited on 16/06/2024).

[23] J. Garcia, S. Alfredsson and A. Brunstrom, 'Examining TCP Short Flow Performance in Cellular Networks Through Active and Passive Measurements', in *Proceedings of the 5th Workshop on All Things Cellular: Operations, Applications and Challenges*, ser. AllThingsCellular '15, event-place: London, United Kingdom, New York, NY, USA: Association for Computing Machinery, 2015, pp. 7–12, ISBN: 978-1-4503-3538-6. DOI: 10.1145/2785971.2785974. [Online]. Available: https://doi.org/10.1145/2785971.2785974.

[24] B. B. N. S. www.blackbox.be, '3678 - TCP and Latency', Black Box, [Online]. Available: https://www.blackbox.be/en-be/page/28414/Resources/Technical-Resources/Black-Box-Explains/lan/tcp-and-latency (visited on 13/06/2024).

[25] I. G. T. B. Team, 'Understanding Congestion Control for Better Big Data Optimization', Intel Granulate, 19th Apr. 2023, [Online]. Available: https://medium.com/intel-granulate/understanding-congestion-control-7b0890657710 (visited on 13/06/2024).

[26] G. Barbosa, S. Theeranantachai, B. Zhang and L. Zhang, 'A Comparative Evaluation of TCP Congestion Control Schemes over Low-Earth-Orbit (LEO) Satellite Networks', in *Proceedings of the 18th Asian Internet Engineering Conference*, Hanoi Vietnam: ACM, 12th Dec. 2023, pp. 105–112, ISBN: 9798400709395. DOI: 10.1145/3630590.3630603. [Online]. Available: https://dl.acm.org/doi/10.1145/3630590.3630603 (visited on 14/06/2024).

[27] sparky8512, *sparky8512/starlink-grpc-tools*, original-date: 2020-12-22T22:40:55Z, 23rd May 2024. [Online]. Available: https://github.com/sparky8512/starlink-grpc-tools (visited on 27/05/2024).

Impact of Variable Channel Conditions on Multi-Hop DTN-Based Lunar Communications with Traffic Prioritization

Klara Schaper, Teresa Algarra Ulierte, Andreas Timm-Giel
Institute of Communication Networks, Hamburg University of Technology, Hamburg (Germany)
e-mail: research@klaraschaper.de, {teresa.algarra.ulierte | timm-giel}@tuhh.de

Felix Flentge
Directorate of Operations, European Space Agency, Darmstadt (Germany)
e-mail: felix.flentge@esa.int

***Abstract*—The Moon is the closest object in space to Earth, and therefore of high value for space research. This can be seen in the numerous planned missions involving the Moon, which will require a stable and robust infrastructure for Earth-to-Moon communication. The existing architectures and protocols — Delay- and Disruption-Tolerant Network (DTN) and Bundle Protocol (BP) — were designed for the challenges of space communication. Nevertheless, they currently lack Quality of Service (QoS) assessments to provide a stable link under high transmission rates and limited bandwidth. One promising solution is the use of traffic prioritization, which has been shown in previous work to improve the performance for a limited amount of critical information bundles at the expense of bulk bundles. This research builds on the aforementioned previous work to model a multi-hop Earth-to-Moon communication path, accounting for differences in space and atmospheric link characteristics and error sources, and extends the work previously published [1]. Through a set of simulations, traffic prioritization was evaluated for one-hop, two-hop and three-hop paths together with optimal, average and worst conditions of the link. An impactful performance improvement was found for high priority bundles throughout all the experiments. In addition, it was found that the efficiency of the high priority packets decreased as the number of hops increased, while the exact opposite was observed for the low priority packets. The study recommends traffic prioritization for time-sensitive data and highlights the need to incorporate priority in routing decisions to enhance QoS.**

Keywords-Solar System Internet (SSI); Space Communications; Delay- and Disruption Tolerant Network (DTN); Bundle Protocol (BP); Quality of Service (QoS); Traffic Prioritization

I. Introduction

The Moon, being the nearest celestial body to Earth, plays a vital role in the future of space exploration [2]. This is particularly evident through the growing number of planned missions and satellite deployments by European Space Agency (ESA), National Aeronautics and Space Administration (NASA), Japan Aerospace Exploration Agency (JAXA) and Indian Space Research Organisation (ISRO) among others. The presence of a reliable communication network is essential for the success of these missions. This is an especially challenging task in the context of space communication, since the characteristics of the communication link in space are marked by intermittent connectivity, long or variable delays, asymmetric data rates and high error rates.

To overcome these obstacles, the concepts of Delay- and Disruption-Tolerant Network (DTN) and Bundle Protocol (BP) were created. Their main feature is their packet-switching approach, this solving the problems of no end-to-end paths, intermittent connectivity and the long delays [3]. However, certain challenges remain unresolved, particularly the bandwidth limitations caused by the high demand from the numerous aforementioned missions. A prominent example of this challenge is the Lunar Gateway — the first space station to orbit beyond Earth — which is intended to function as a communication relay between Earth and the Moon. The European System Providing Refueling, Infrastructure and Telecommunications (ESPRIT) module of the Lunar Gateway is expected to provide a maximum bandwidth of $25\,\mathrm{Mbit\,s^{-1}}$ [4], yet an average scientific mission will presumably consume up to $20\,\mathrm{Mbit\,s^{-1}}$ [5]. When combined with telemetry, tracking and command (TT&C), essential communications, and multiple science missions, this demand is likely to exceed the available bandwidth, resulting in significant queuing delays and potential data loss.

To maximize the utilization of such limited bandwidth, traffic prioritization, a Quality of Service (QoS) mechanism, can be used as a mitigation strategy by prioritizing urgent data. Such a system is however not included in the BP specification, but a theoretical approach exists to include this and several other QoS parameters into an extension block, researched by Algarra et al. [6]. Especially with traffic prioritization they showed a significant enhancement in the delay of high priority bundles [7]. The issue with this research is that a direct Earth-to-Moon transmission path is assumed. Many of the missions will not be equipped with a Direct-to-Earth communication link [8], but will depend on relay satellites to increase the frequency of communication windows, or to have it at all in the case of missions targeting the far side of the Moon. For such missions, deploying and utilizing relay satellites is essential to establish connectivity. As a result, a more realistic analysis that incorporates

multi-hop communication paths is necessary.

The purpose of the original paper [1] was to analyze the effect of traffic prioritization on multi-hop Moon-to-Earth communication links. To achieve that the model of Algarra et al. [6] was adapted to allow flexible numbers of hops, enabling representation of diverse multi-hop paths and future use cases. These modifications necessitated re-evaluating traffic prioritization in DTN using the BP. The adapted model was then used in simulations to compare bundle transmission delays across three priority classes with a non-prioritized scenario. Since this work extends the original all these propositions remain accurate with an additional focus on the link through various conditions. This not only enables a deeper evaluation of traffic prioritization but also allows for additional insights into traffic routing and the overall link performance, building on observations from the original work.

The structure of this paper is as follows: Section II provides the necessary background information, covering topics such as DTN, BP, QoS, and Markov chains. Subsequently, the work of Algarra et al. [6] is discussed in Section III, along with other relevant studies. The adapted model and its key components are detailed in Section IV. Section V presents the analysis and evaluation of the experiments conducted using the implemented model. Finally, Section VI summarizes the findings and outlines directions for future research.

II. Background

Space communication networks face unique challenges that distinguish them from terrestrial systems. To address these challenges, specialized protocols and architectures have been developed, forming the foundation for reliable Earth-to-space communication links.

A. DTN

The characteristics of communication links in space are marked by several challenges which are not present in terrestrial communications: lack of end-to-end path, intermittent connectivity, long or variable delays, asymmetric data rates and high error rates. Traditional internet protocols, designed for terrestrial networks with stable end-to-end connectivity, fail to operate effectively under these conditions. The DTN was created to provide a reliable communication link in environments where these challenging conditions occur. Its key innovation is the store-and-forward message switching mechanism, which replaces the end-to-end path assumption of terrestrial internet protocols. With this approach, the entire message or parts of it are moved from one persistent storage node to the next, enabling a more suitable hop-by-hop communication method as illustrated in Figure 1 [3].

Since end-to-end paths are not guaranteed in space environments, messages can be stored at intermediate nodes until a communication link becomes available. This approach ensures that packet delivery is always faster than traditional methods or, in the worst case, equally fast, making it particularly valuable for cases in which connectivity windows are limited and unpredictable [3].

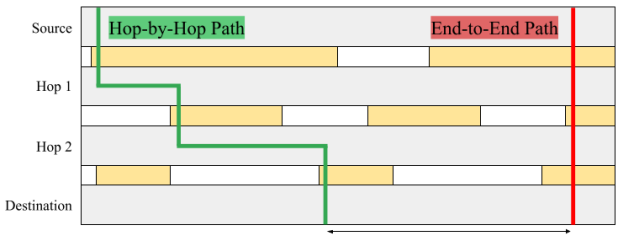


Figure 1. Hop-By-Hop vs End-To-End

B. BP

The store-and-forward message switching capability of DTN is implemented through BP. It operates as an overlay layer positioned below the application layer and above the transport layer, as shown in Figure 2. This architecture allows for terrestrial protocols to still be used even under the circumstances present in space environments. Additionally, this approach preserves compatibility with existing network infrastructures, making BP conveniently interoperable [3].

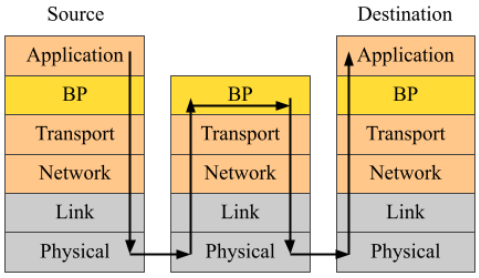


Figure 2. Bundle Protocol (BP)

In BP, messages to be transmitted are encapsulated into a bundle, which contain both the payload data and the metadata required for successful transmission across the network and proper decoding at the destination. Each bundle consists of three types of blocks:

- Primary Bundle Block: contains essential information such as the source node ID, destination node ID, and the bundle's creation time. This block is protected against modification using the Block Integrity Block (BIB) [9].
- Extension Blocks: optional blocks that provide supplementary information, such as the bundle's age or hop count. Since these are not mandatory, some intermediate nodes may be unable to interpret or process specific extension blocks.
- Payload Block: holds the actual data being transmitted. To ensure its integrity, this block is also protected by the BIB.

This bundle structure, combined with the store-and-forward mechanism, enables reliable data transmission across the challenging space communication environment

while maintaining compatibility with existing network protocols and infrastructure.

III. Related work

This section reviews existing approaches to QoS in space communications, their modelling and characterization, and establishes the foundation for our multi-hop analysis.

A. QoS and Channel Modelling for DTN

Space channels are highly stressed communication environments, with periods of extremely low Frame Error Rate (FER) that can abruptly transition into outages lasting from minutes to hours [10]. This high variability and prolonged downtime render conventional terrestrial QoS approaches unsuitable for space communications. Just as DTNs developed to address the unique challenges of space links, tailored QoS mechanisms are required to ensure reliable performance under these extreme conditions.

One of the approaches that been demonstrated as an effective mechanism to mitigate the impacts of space channel volatility is traffic prioritization. It was first implemented in BP with the Class of Service block for DTNs defined in RFC 5050 [11], which includes "expedited", "bulk" and "normal" classes. However, this classification proved to be inadequate, resulting in the formulation of the Extended Class of Service (ECOS) block with which the classes "critical", "streaming", "ordinal", and "reliable" were added. While defined in the Consultative Committee for Space Data Systems (CCSDS) BP Specification [12], ECOS still had significant limitations, mainly the inflexibility caused by the predefined usage of each class, as well as the lack of adaptability due to the inclusion of retransmission schemes within traffic prioritization [12]. Though ECOS did not get standardized, it was implemented in the Interplanetary Overlay Network (ION) [13].

Building on this foundation, Apollonio et al. [14] investigated the integration of ECOS with Contact Graph Routing (CGR), the routing algorithm used in DTN with BP. They tested flags such as "streaming", "bulk", "expedited", and "critical" for a communication path from Earth to the far side of the Moon. Each flag was handled optimally by CGR algorithm: "streaming" bundles followed the best routes, "bulk" bundles made the most of limited bandwidth, and "critical" bundles were sent earlier, postponing lower-priority traffic. This work demonstrates the potential benefits of traffic prioritization in space environments. However, it assumed uniform loss rates, whereas real-world conditions exhibit both uncorrelated (e.g., thermal loss) and correlated losses (e.g., atmospheric or solar effects). This work addresses this limitation by using a multi-hop model with more accurate loss characteristics to validate results under more realistic conditions.

Additional research has explored channel modeling for space communications. Various Markov modeling approaches have been used [15], with the most common being the Gilbert-Elliot model (2-state Markov chain) [16], where the channel alternates between a "good" and a "bad" state, providing a binary state distribution. More complex models have been applied to non-terrestrial networks, such as land mobile satellite (LMS) communications [17][18]. In these, three-state Markov chains are used to classify losses according to their severity, such as moderate or deep attenuation. Chu et al. [19] further extended this approach with a four-state Markov chain, incorporating two error and two good states. This model focused on losses across the entire low-Earth orbit path rather than on burst error sequences.

These studies support the applicability of Markov chains to space communication channels. However, they are primarily tailored for satellites and Earth-orbit scenarios, not for lunar, Martian, or deep-space links. In that regard, Pan et al. [20] provide a broad overview of deep-space channel models, categorizing them into near-Earth, interstellar, and near-planet links. For each scenario, they describe potential losses that can occur, and provide a mathematical model for those losses. While this can serve as a basis for constructing more comprehensive models, it lacks a concrete channel model combining short- and long-term losses in a way suitable for near-Earth and deep-space scenarios, which is the focus of this work.

B. Current Approach

Previous research is based on the assumption that a robust QoS approach must be created considering the space-specific challenges and link characteristics that the transmission will face [10]. To study these challenges, a channel model approach was developed using a Markov chain [6].

While the available literature already uses Markov chains to model channel behavior, such as the Gilbert-Elliott models [16] and the multi-state LMS models [17]–[19] presented in Section III-A, the novelty in our approach lies in the use of a three-state Markov chain specifically tailored for cislunar and deep-space links. The state differentiation allows for more accurate and precise depiction of all the different error sources that can be present in the channel, and makes a separation between long-term or correlated losses, and short-term or uncorrelated losses. The three states are defined as follows:

- Success state: in this state, the transmission is successfully sent by the source and received by the end node without major losses.
- Short-term loss state: here, the transmission failed due to short-term effects such as antenna pointing errors, interference, or light atmospheric phenomena.
- Long-term loss state: lastly, if the transmission fails due to long-term phenomena, the channels will suffer from burst losses that may last from several seconds up to hours. Causes for this include solar storms or severe atmospheric events.

While the three-state Markov chain has proven to model the channel more accurately than other approaches

found in the literature, the values taken in the presented previous work lacked precision. The improved Markov model presented in Section IV takes into account a wider range of error sources and models them in a more accurate and complete manner.

This model enabled the identification of key QoS parameters relevant for space scenarios, which were implemented in an extension block for BP, named the User QoS Extension Block (UQEB) [6][21]. It includes the following QoS mechanisms:

- Traffic prioritization: in order to cope with the long delays and possible bottlenecks, traffic prioritization is critical to enable time-sensitive information to arrive at their destination within the desired latency bounds. Three main classes are described: critical, normal and bulk. Within those, there are sub-priority classes which allow for fine-tuning of the latency requirements.
- Required reliability: due to the high error rates, transmissions often fail to reach their destination. Deciding whether an Acknowledgement (ACK) upon reception is required or not depends on the data that is being transmitted, as well as on the characteristics of the specific link at hand (some links might be uni-directional). This UQEB entry allows for specification of this requirement.
- Latest-only delivery: due to the intermittent connectivity and bottlenecks, it is not uncommon for bundles to accumulate at a node. For some data types, only the latest information might be relevant, such as the case of sensor networks. Should several of these bundles meet at a node, only the latest one will be forwarded and the rest will be dropped.
- Bundle storage: given that the persistent storage in the intermediate nodes is limited, a long period of no connectivity or a surge of bundles might cause the storage to run low. In this case, bundle storage allows for the user to indicate which bundles can be dropped first and which should be kept.

The feasibility studies performed using the UQEB in conjunction with the Markov chain show that the transmissions which included their QoS needs (especially priority classes) had an improved adjustment to their respective requirements: the critical and normal bundles had a shorter delay at the expense of a longer delay for bulk bundles [7][22]. However, this prior work assumed direct Earth-to-Moon visibility in the communication link, which is unrealistic for many future missions that will rely on relay satellites and multi-hop paths.

C. Contribution

As a conclusion, existing work on space channel modelling predominantly relies on simplified or uniform loss models, which do not reflect the heterogeneous nature of the channel. Furthermore, these studies generally lack a channel model capable of representing the impact of both short-term uncorrelated losses and long-term correlated losses in transmissions. To address this gap, both prior work and the present study define a channel model based on a three-state Markov chain that incorporates distinct space and atmospheric error sources (see Section IV). This work expands the previous work by considering varying conditions (optimal, average, and worse-case), enabling evaluation of the full range of potential link performance.

The second critical research gap concerns QoS for DTN and BP. Prior work has primarily evaluated traffic prioritization under the assumption of a direct Earth-to-Moon link, without considering multi-hop communication paths. This is a significant limitation, as a growing number of missions will depend on orbiters and relay satellites to establish connectivity. To address this, the present work introduces a multi-hop Earth-to-Moon channel model that enables assessment of traffic prioritization across multiple hops. This approach provides a more realistic evaluation of priority behavior in lunar communication networks (see Section V).

IV. Model

To assess traffic prioritization under realistic conditions, a comprehensive channel model must be developed which accounts for the diverse loss characteristics encountered in the scenarios at hand. This model distinguishes between different types of losses and communication settings to accurately represent the challenges faced by the communication link.

There are two fundamental types of losses:

- Uncorrelated losses: losses that occur sporadic and shortly. These arise from causes with short duration, such as thermal loss or short interferences.
- Correlated losses: losses that occur repeatedly for extended periods of time, resulting in the loss of multiple packets in a row. Such losses can be caused by atmospheric phenomena (e.g. storms) or space weather (e.g. solar flares).

To analyze the specific losses affecting the communication link, this work models a transmission path from a lunar node (e.g., astronaut, rover, or mission equipment) to an Earth ground station via relay satellites. This path encounters two distinct environments:

- Space setting:
  - Starting at the lunar source, the transmission travels through the exosphere of the Moon. Since this is a very thin layer of gas, and the atmosphere is negligible, no weather can develop. This results in no atmospheric disturbances on the surface of the Moon [23].
  - The next stage of the transmission is the travel through outer space, exposed to the Sun without any protection. While regular solar radiations are uniform and, therefore, calculable and manageable, solar storms are unpredictable.
- Atmospheric setting:
  - As the transmission approaches its destination, it encounters Earth's atmosphere, which is significantly

denser than that of the Moon. This increased density can lead to interference from atmospheric weather phenomena.

Moreover, the communication system must be defined in order to identify the applicable error sources. This can be either Radio Frequency (RF) or Free Space Optical (FSO). Due to its rising popularity, FSO is expected to be used in ground-to-Earth orbit scenarios, and RF is taken as the standard for space communications due to the long distances to be conversed [24].

With this separation of losses and settings set, the Markov model presented in Section III-B can be extended to the following one (Figure 3):

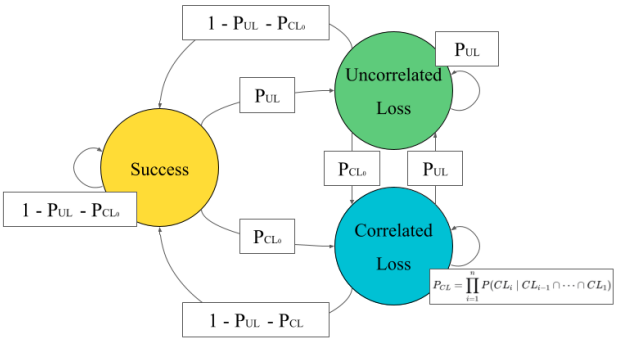


Figure 3. Complete Markov Model

The associated transition matrix to the defined Markov chain is given in Equation 2. It depends on the input parameters P_{UL}, P_{CL_0} and P_{CL}, where P_{CL} is defined in Equation 3.

$$\mathbf{P} = \begin{bmatrix} P_{S\to S} & P_{S\to UL} & P_{S\to CL} \\ P_{UL\to S} & P_{UL\to UL} & P_{UL\to CL} \\ P_{CL\to S} & P_{CL\to UL} & P_{CL\to CL} \end{bmatrix} = \tag{1}$$

$$= \begin{bmatrix} 1-P_{UL}-P_{CL_0} & P_{UL} & P_{CL_0} \\ 1-P_{UL}-P_{CL_0} & P_{UL} & P_{CL_0} \\ 1-P_{UL}-P_{CL} & P_{UL} & P_{CL} \end{bmatrix} \tag{2}$$

$$P_{CL} = \prod_{i=1}^{n} P(CL_i \mid CL_{i-1} \cap \cdots \cap CL_1) \tag{3}$$

To maintain realistic transition probabilities, the simulation employs two separate Markov chains: one representing the space setting and one representing the atmospheric setting. The boundary between these environments is taken as the International Space Station (ISS) for this work, as it is the closest fixed node to Earth, located approximately 400 km above the surface.

The losses depicted in Figure 3 are specific to each environment, and are explained in detail in Section IV-A and Section IV-B. To provide even more accuracy, the system will model three different types of conditions: optimal, average and worst. Optimal conditions feature high throughput because of low influences on the communication link. Average conditions portray more typical loss probabilities, which are higher than those presented under optimal conditions. In case of worst conditions, the probabilities for loss are as high as possible without loss of realism.

A. Space losses

To send data through RF, the digital data is encoded onto a high frequency electromagnetic wave. This output is then amplified and sent through space to the receiver, which decodes the received signal back into digital data. The frequency used depends on the needs. A higher frequency offers a higher data rate, but with the risk of higher loss, whereas with a lower frequency only lower data rates are possible, but with the advantage of a lower error rate. Since the used frequency varies significantly, no specific frequency is modeled [24].

The antenna of the sending satellite has to be aligned for any frequency band used, although the alignment precision needed depends on the frequency. The higher the frequency, the lower the tolerance for pointing errors [24]. Since pointing errors do not happen in long bursts, they can be considered as an uncorrelated loss. Satellites themselves are susceptible to small errors, especially since both solar and cosmic radiation damages the electronics by overcharging them or through high-energy particles penetrating them [25]. Such system errors are considered as uncorrelated losses, since they do not happen in bursts, but as single random losses. These two losses are simulated together in one probability with 0.0014% for optimal, 1.9% for average and 4.35% for worst case conditions. The values are taken from a study of typical satellite-to-ground systems, which models the expected erasure percentage in a transmission under several circumstances [26]. These percentages taken depict the probability of the losses being higher than 6 dB, which is the standard link budget margin used by the missions at hand.

The correlated losses considered in this work do not stem from the satellites themselves but from the environment they are in. The solar storms are caused by explosions on the Sun releasing a large amount of energy [27]. They can be categorized according to their peak brightness in X-ray wavelength and sorted by their level of impact. X-class flares have the highest impact, followed by M-class flares and then C-class flares. All three classes can have consequences on Earth, according to their level of force. The weakest flares are the B-class and A-class flares, with no influence on Earth. The effect of these solar flares on the communication is a complete blockage of the channel, since they are a burst of radiation that spreads across the electromagnetic spectrum, which includes the radio frequencies [25]. This can be depicted as a continuous transmission over all channels, blocking the medium completely for all other sending attempts.

Figure 4 illustrates this process showing how the emitted radio waves can not reach the next satellite because of the solar flares blocking the way.

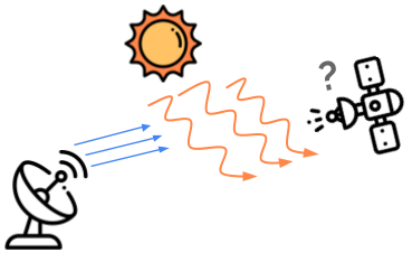


Figure 4. Influence of a Solar Storm on a RF Communication Link

Solar flares can also have a direction, such that even though a solar storm is taking place, the communication might not be interrupted. Nevertheless, this research assumes solar storms to cause a general channel blockage, regardless of their direction. Regarding the level of force, only the classes X, C and M are considered to have enough impact on the communication channel. Solar storms belong to the correlated losses since they cause burst errors due to their duration being anything from seconds to minutes. Moreover, their area of impact is extensive, influencing all satellites in the path.

To determine the probability of entering a solar storm, the work of Nishizuka et al. [28] was used. They developed a forecasting model for solar storms for which they observed the occurrences of the different solar flare types in a specific time period. In the time frame of 2009 days they observed 26 X-class, 383 M-class and 4054 C-class solar flares. These numbers were used for a calculation of percentage per second. For the worst case condition all three types were included, which makes a sum of 4463 solar flares and results in an entry probability of 0.003%. The optimal case leaves out the C-class flares, since they are the weakest ones of the considered classes, and they do not necessarily block the channel, such that the sum only amounts to 818 and a probability of 0.002%. The average condition is set as a middle between those two probabilities with 0.0025%.

For the probability of staying in correlated loss in the space setting, only one curve for all three conditions is used. This curve is taken from the work of Guo et al. [29], which researched the appearances of Quasi-Periodic Pulsation (QPP) in solar flares. QPPs are a common feature of solar flares, and the graphical representation of their occurrence gives a function for the general course of a solar flare. This function can be seen in Figure 5. The x-axis gives the time in seconds, and the y-axis the number of QPPs occurring at that time.

This graph represents the length distribution of solar storms, from which we derive the probability of staying in the solar storm after a certain time. It is important to note that solar storms are always at least 6 s long, such that the probability of staying is 100% until 6 s have passed. The function follows a logarithmic normal distribution and in the simulation an approximation of the given curve is used, because the original data is not available. Since the referenced graph consists of bars, all values were normalized

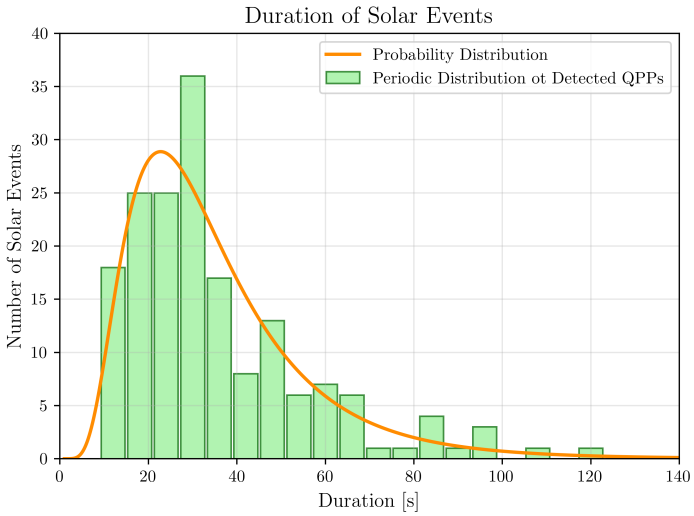


Figure 5. The Periodic Distribution of the Detected QPPs (reproduced from [29])

by multiplying them with the range of one interval, to get a continuous function from the discrete values.

B. Atmospheric Losses

As previously stated, this research has chosen to model the link as FSO, which is a technique based on optical wavelengths of electromagnetic radiation. The fundamental process is the same as in RF based communications, meaning that the data to be sent is modulated into electromagnetic waves and sent to a receiving system, but the frequency used is much higher, generally at near-infrared bands which explains the comparison to lasers. FSO offers much higher data rates than RF, even when using high-frequency bands, and it is much less likely to be jammed. However, these advantages come at the cost of some error-prone properties, from which the losses to be modeled in this link can be drawn [24].

Beginning with the uncorrelated losses, the pointing of the antenna has to be much more accurate than for RF, since the beam is narrower. Additionally, the beam does not only travel through empty space, but through the atmosphere of the Earth, with many more elements to influence a communication. The changes in the atmosphere's density and its composition of particles can scatter the beam, and thus result in loss [24]. For the pointing errors and interferences in the atmospheric setting, the probabilities from the space settings are taken again, since such losses are not dependent on the setting and should therefore be the same.

Aerosols in the atmosphere are another challenge for FSO communication. They are minute particles and can vary from liquid to gaseous. The size, concentration and origin differs and depends on the location, season, time of day, weather conditions and natural events such as volcanic eruptions or desert dust. The impact of aerosols is an attenuation of the transmission, and the amplification of

sky radiance, which is a scattering of the sunlight, such that the noise in transmissions is increased [30]. Because of its high influence, aerosols are modeled separately from atmospheric turbulence in the uncorrelated losses. Their dependence on the time of day, the weather, and the season, would also make them fitting for an inclusion into correlated losses, but the effect on the scenario at hand would be negligible in comparison to the complexity of the analysis needed, as well as the lack of available data on this topic due to said complexity. The probability for aerosols is taken from an informational report on real-time weather and atmospheric characterizations [30]. The report provided two graphs, representing the cumulative probability of loss at 1550 nm and at 1064 nm. Since the frequency was not defined for the simulation, the graph is selected based on higher probability for loss, to rather depict the worst scenario than to provide an overly optimistic view (see Figure 6). For the average case the curve for worldwide is taken at the point of 6dB loss, because it is averaging the transmission loss at 6dB over the whole world. The worst condition depicts the curve of Beijing at 6dB loss, since it is by far the curve with the slowest increase. In the optimal case there are no aerosols influencing the transmission, so the probability is taken as 0%.

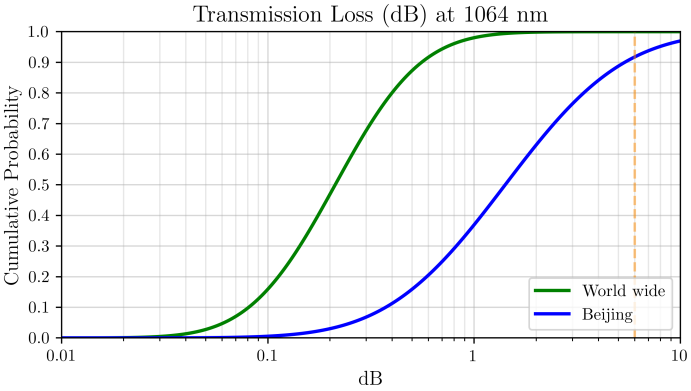


Figure 6. Transmission Loss Due to Aerosols (reproduced from [30]).

The biggest influence on FSO remains the clouds since cloud coverage has a similar effect on FSO as solar storms: Their occurrence results in a complete blockage of the channel. The moisture of the clouds causes such a large attenuation, that communication becomes impossible [24]. The water droplets and/or ice crystals that clouds consist of, scatter and absorb the optical wavelengths. The level of interruption depends on the thickness of the clouds; the thicker the cloud, the higher the disturbance. However, independent of the thickness, clouds always interfere with the communication link to some degree. Additionally, during daytime, clouds have a similar effect as aerosols in terms of scattering Sun irradiance, hence increasing the sky radiance [30]. In order to simplify the complexity that clouds present with regard to their type, size and thickness, the presence of a cloud is taken as a complete blockage of the channel. This still presents a realistic scenario since the blockage would come either from the thickness of the cloud, or from the sky radiance. Another factor to reflect in the model is the duration of clouds. Depending on the wind conditions, clouds can cover the sky from minutes to hours, resulting in a high amount of loss. Therefore, clouds are categorized as correlated loss in this model.

For the probability of entering the correlated loss state, a calculation based on the rainy days at three stations out of the ESA core network was done. Using the number of cloudy days per year as a basis would have been the optimal procedure, but unfortunately no reports on this metric were found. The optimal condition is given at Santa Maria (Portugal), with only 13 rainy days per year [31]. Converting that into a probability per second by dividing it by 365 and then by 86400, we get a probability of 0.000041% to get into the correlated loss state. For the average condition Kiruna (Sweden) was chosen with 92 days of rain per year, which results in a probability of 0.00029% per second [32]. The worst case condition is given at Kourou (French Guiana), with the number of rainy days amounting to 188, which gives a probability of 0.00043% [33].

However, the probability of staying in the correlated loss state in the atmospheric setting depends on time. The curves used can be seen in Figure 7, which depicts the probabilities of the sky staying cloudy over time. In the original work from Iver A. Lund [34] the sky cover is represented in tenths, such that 0 indicates a clear sky and 10 indicates a completely covered sky. Therefore, each curve serves as the respective tenths of the sky staying covered. The worst case would be that the sky does not clear up very much after cloud occurrence, which is characterized by the probability curve of 9 tenths of the sky staying covered. The curve of 10 tenths was intentionally excluded, because it is nearly linear and gives the clouds too high an influence in the Markov model. In the optimal case the sky clears up completely, which is the curve of 0 tenth staying covered. Lastly for the average case the curve of 5 tenth of the sky staying covered was chosen as a middle ground. The functions used in the simulation are approximations of these given curves, since the original data is not available.

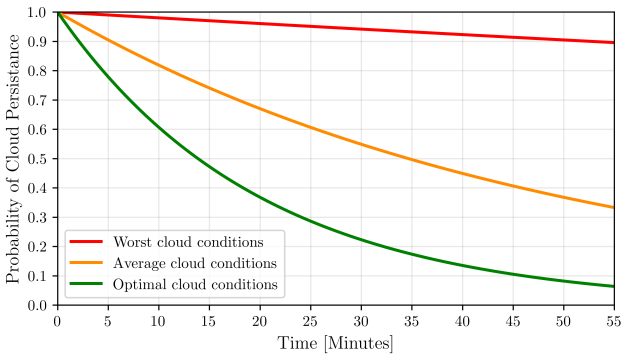


Figure 7. Cloudy Line-Of-Sight Persistence Probabilities (reproduced from [34]).

All probabilities and their affiliations are summarized in

Table I.

TABLE I. Correlated and Uncorrelated Losses for the Space and Atmospheric Settings

Setting	Loss Type		Optimal	Average	Worst
Atmospheric	Uncorrelated	Pointing errors and interferences	0.0014%	1.9%	4.35%
		Aerosols	0%	2%	5%
	Correlated	Clouds	0.000041%	0.00029%	0.00043%
Space	Uncorrelated	Pointing errors and interferences	0.0014%	1.9%	4.35%
	Correlated	Solar storms	0.002%	0.0025%	0.003%

With these values the following simulation was build and studied.

C. Simulation

The code basis of this simulation is from the work of Algarra et al. [22], which was explained in Section III-B and was kindly provided for extension to cover the adapted model.

To simulate several days within a few hours SimPy [35] is used, a process-based discrete-event simulation framework. The first two processes introduced into the SimPy environment are the Markov chains modeling the atmospheric and space conditions. Both share the same abstract structure described in Section IV, differing only in their state-transition probabilities; therefore, they are explained together.

During initialization,

- the probability of transitioning into uncorrelated loss is set based on the selected condition,
- counters for all three states (success, correlated loss, uncorrelated loss) are created, and
- the start state is set to *success*.

Each Markov-chain process then runs in a loop until the set simulation time. In each loop the probability of entering the correlated-loss state is set anew according to the condition, if the Markov-chain is currently not located in the correlated-loss state. Otherwise, this value is instead computed using the corresponding solar-storm or cloud model. The duration spent in this state needed for the computation is tracked by storing the entry time and subtracting it from the current simulation time during each update.

To get into any state, a random number in $[0, 1]$ is drawn. As seen in Figure 3, if it is below $1-P_{CL_0}-P_{UL}$ the process is in the success state; if it exceeds $1 - P_{CL_0}$ it enters the correlated loss state; otherwise, it enters uncorrelated loss. After determining the state, the process waits one second before repeating. Throughout the simulation, other processes query these Markov chains to obtain the current atmospheric or space-link state.

Once the Markov chains are running, the communication link is simulated, transmitting bundles from source to destination. Each priority class has its own traffic generator, which creates and sends bundles with IDs, priorities, and arrival timestamps. A benchmark case with no traffic prioritization (First-In-First-Out (FIFO) queuing) uses the same traffic volume for comparison. Each traffic generator begins after a random initial delay between zero and the set bundle arrival rate, preventing synchronized starts. It then runs in a loop, generating new bundles at uniform distributed intervals.

Before adding the traffic generators, the node storages are initialized to handle incoming bundles. The number of storages equals the number of hops plus one special storage called 'destination'. Initialization proceeds in reverse order so each storage knows its successor (the last one points to destination) and its predecessor (the first one receives bundles from the traffic generator). Figure 8 illustrates this setup, showing that the first node storage functions as the source, even though bundles originate from the traffic generator. Therefore, transfers from the traffic generator to the first storage do not count as hops, whereas transfers into the destination do.

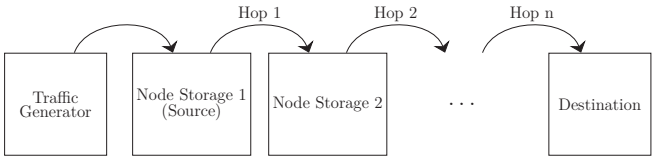


Figure 8. Structure of the Node Storages

Each storage has a 'receive' function that records bundle arrival times and adds bundles to the internal queue. The main function of each node handles forwarding: in an infinite loop, it requests the current state from the appropriate Markov chain according to the hop scenario. In case of a hop bypassing the border, both Markov chains are requested, and the state is only set to success if both chains are in the success state, otherwise resulting in a loss state. On a success, the bundle is sent to the next node after a timeout equal to the bundle service time. On a loss, the timeout is doubled, after which the bundle is resent.

The destination storage differs only in that it lacks a main forwarding function. Its 'receive' function sorts bundles by priority and records delivery statistics, while still appending the arrival time to the bundle, since the transition to the destination counts as a hop as Figure 8 demonstrated.

Throughout the whole simulation data is recorded not only for the evaluation of the results but also for the evaluation of the simulation. The number of occurrence of each state, the consecutive errors for uncorrelated and correlated loss as well as the bundle delivery rates and bundle loss ratios were recorded to check for plausibility. These verification mechanism along with the careful researched parameters ensure the authenticity of the results.

V. Results

This section presents a comprehensive evaluation of traffic prioritization performance across different multi-hop scenarios and channel conditions. Each run consists of

an experiment with critical, normal and bulk classes and an experiment with First-In-First-Out (FIFO) approach to see the impact of traffic prioritization in comparison to the benchmark. All three priority classes have the same amount of traffic and the FIFO experiment has the same amount as all priority classes together. The simulations run for 500 simulated days each to ensure enough data is collected. The parameters to configure for every experiment are the conditions and the scenario.

The condition represents the state of the communication link and changes the probabilities of the events accordingly as seen in Section IV. The modeled conditions are optimal, average and worst conditions. This range allows for a detailed analysis of the impact of traffic prioritization in different link states, and to derive where it performs best.

The experiments have the goal to reevaluate the impact of traffic prioritization on a multi-hop path. Additional focus will lie on the differences occurring between a direct path and a multi-hop path, and the conclusions that can be drawn from that for the future Moon-to-Earth communication. For that, three scenarios were chosen (see Figure 9), all depicting the expected near-future:

- Direct communication (one-hop): the distance of 405 500 km represents the Earth-Moon link at apogee. It was selected for a comparison between the work of Algarra et al. [6] and the modified model.
- Relayed communication with two hops: the distances are 70 000 km from the Moon to the Lunar Gateway, and 335 500 km from the Lunar Gateway to Earth. As highlighted in Section I, the Lunar Gateway will improve bandwidth for critical missions, making a traffic prioritization analysis on such a scenario desirable.
- Relayed communication with three hops: the distances are 70 000 km from the Moon to the Lunar Gateway, 335 100 km from the Lunar Gateway to the ISS, and 400 km from the ISS to the ground station on Earth. Since the ISS was chosen as the boundary between the atmospheric and the space setting in the modeling, this scenario strictly separates the environments, such that each link only corresponds to one setting.

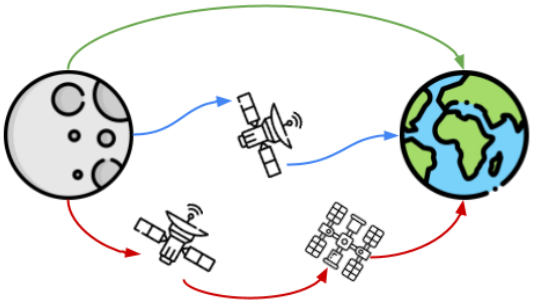


Figure 9. One-Hop Scenario (Green), Two-Hop Scenario (Blue) and Three-Hop Scenario (Red).

The results from all experiments are analyzed in detail in the following subsections, organized by scenario type with comparative analysis across different channel conditions. Focus of the analysis is the end-to-end delay of each bundle generated at a rate sufficient to prevent queuing under ideal conditions. The Cumulative Distribution Function (CDF) graphs illustrate the percentage of bundles that arrive within a given time threshold. Of particular interest are the 2.5 s and 5 s marks, representing the threshold defined by ESA for qualifying as normal communication, and twice that threshold, respectively, for reference [6].

A. One-Hop Scenario

Figure 10 shows the end-to-end delay of the direct communication scenario under optimal conditions. These optimal conditions are clearly visible in the graph, since approximately 83% of the critical bundles arrive at the first mark and within a maximum of 3 s even every critical bundle has arrived. Not far below are the curves of the normal and the FIFO bundles. Their course is rather similar with even occasional overlaps. The FIFO bundles arrive with a chance of 70% within 2.5 s and are only slightly lower than the normal bundles with 73%. At the reach of 100% they overlap at around 4.6 s. The curve of the bulk bundles is below the FIFO and the normal curves. It reaches the limit of 2.5 s with a probability of 62%, and the limit of 5 s with exactly 100%, ensuring that all bulk bundles also arrive within the limits. Overall, although the number of losses is low, traffic prioritization still has a positive impact on the critical bundles. This comes at the expense of the bulk, but especially in such optimal conditions, this does not demand a great trade-off, since all bundles arrive within the limits regardless of their priority.

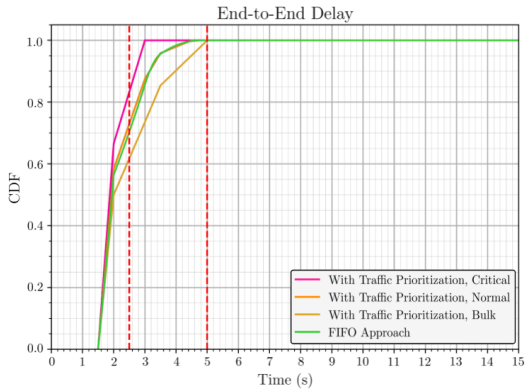


Figure 10. End-to-End Delay in the One-Hop Scenario with Optimal Conditions

With average conditions the probability of critical bundles arriving within 2.5 s decreased compared to optimal conditions and is now 73%. The decrease of normal (60%) and FIFO (48%) bundles is slightly greater. The bulk probability shows a bigger decline achieving a probability of 43% with average conditions. It is also the only curve not reaching 100% before 12 s showing the high impact of the worsened conditions on the bulk bundles, since the interference from high priority bundles increased the

waiting time of the low priority classes. As the losses remain manageable, as evidenced by 98% of critical bundles arriving before the second mark, the impact on the normal bundles is less significant. 92% of the normal bundles still arrive within the limits, which is rather high, especially comparing both the critical and the normal percentages with the FIFO approach, where only 85% arrive before the second mark. Even with this interference from high priority bundles, 74% of bulk arrive before 5 s, which is still a good outcome. The resulting trade-off yields a performance gain, with both the critical and normal showing higher efficiency compared to the FIFO baseline.

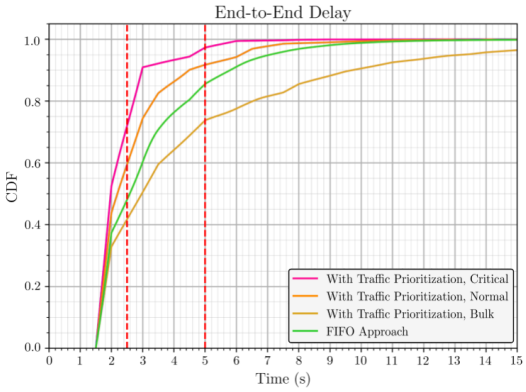


Figure 11. End-to-End Delay in the One-Hop Scenario with Average Conditions

The end-to-end delay displayed in Figure 11 shows high similarities with the results of the experiment with 3 priority classes of Algarra et al. [6]. Therefore, the split of the environments did not change the overall modeling of the direct communication link too much to receive similar results as with the single environment. Furthermore, the results from the two- or three-hop scenarios are applicable to the research of Algarra, providing validation for their findings.

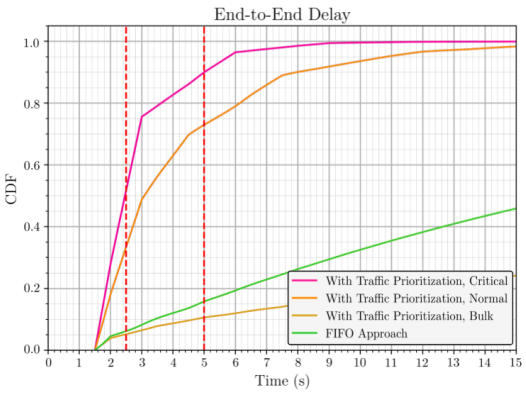


Figure 12. End-to-End Delay in the One-Hop Scenario with Worst Conditions

The higher percentages of loss present in the worst conditions can be seen in the end-to-end delay displayed in Figure 12. Only 52% of the critical bundles arrive at the 2.5 s mark, but 90% arrive before the 5 s mark. While the conditions do dampen the performance of the critical bundles, the resulting arrival rate still represents a satisfactory outcome. The performance of the normal bundles is less favorable compared to the critical bundles, with 53% delivered before the first mark and 73% before the second mark. FIFO and bulk bundles demonstrate an even poorer performance, with their curves showing an almost linear progression and only a very slight increase over time. Only 5% of the bulk bundles arrive before the 2.5 s mark and for FIFO it is only a slightly higher percentage. Even at the second mark, the values did not significantly increase. FIFO achieves 16%, while bulk only gets to 6%. After 15 s both curves did not reach 50% of transmitted bundles.

B. Two-Hop Scenario

The overall delay of the two-hop scenario with optimal conditions can be seen in Figure 13. A high similarity with the end-to-end delay from the optimal one-hop scenario is visible, because the critical bundles dominate the graph, while normal and FIFO overlap and bulk is located below them. Therefore, the results remain good, but critical now reaches the first mark at 76%, normal at 65%, FIFO at 62%, and bulk at 55%, which means a drop of around 8% for all classes. Associated with the similarity to the one-hop scenario are also the conclusions drawn from it, meaning that traffic prioritization keeps its positive impact even under optimal conditions and even in a multi-hop scenario. Additionally, the trade-off did not worsen the overall results, since all bundles still arrive with 100% probability within the limits.

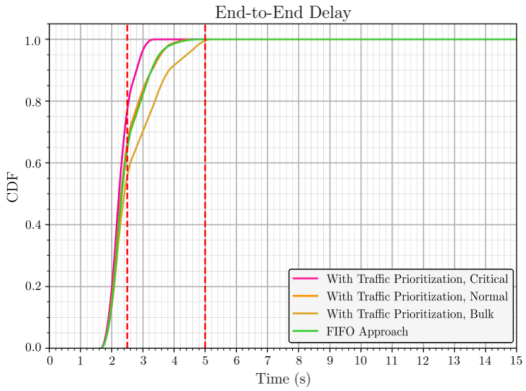


Figure 13. End-to-End Delay in the Two-Hop Scenario with Optimal Conditions

The performance characteristics of the two-hop scenario with average conditions in Figure 14 show that the normal and FIFO curves are much closer now compared to the one-hop scenario. However, the probabilities at the first mark decreased for all curves due to the higher losses. Critical decreased from 76% to 65%, normal from 65% to 55%, FIFO from 62% to 53% and bulk from 55% to 45%. Additionally, the curves do not reach 100% within the limits anymore, but all continue to get above 85%. Even

though the conditions got worse and an additional hop was included, the critical bundles keep outperforming the FIFO approach on the expense of the bulk bundles. The curve of the critical bundles is very steep, showing that even if the total propagation delay is now longer because of the two hops, traffic prioritization remains a positive impact. The higher loss because of multiple hops is a hindrance especially for critical bundles that should be researched further.

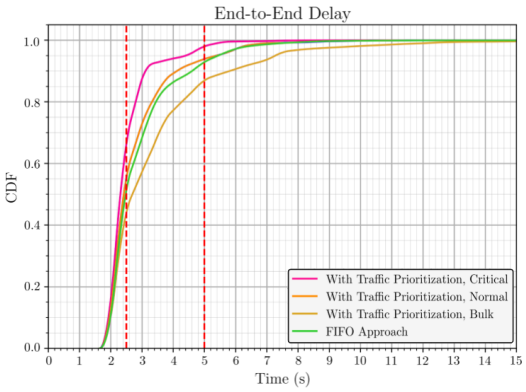


Figure 14. End-to-End Delay in the Two-Hop Scenario with Average Conditions

In Figure 15 the end-to-end delay of the two-hop scenario with worst conditions can be seen. Again similarities to the one-hop worst-case scenario are visible, but with an overall drop in percentages. Critical is now at 50% at the first mark, normal at 38%, FIFO at 30% and bulk at 25%. The pattern already observed from the previous scenarios continue here, supporting the preference of the highest priority class. However, the transmission over two hops appears to have improved the relative performance of the FIFO approach and bulk, since they were far below the other curves in the one-hop scenario.

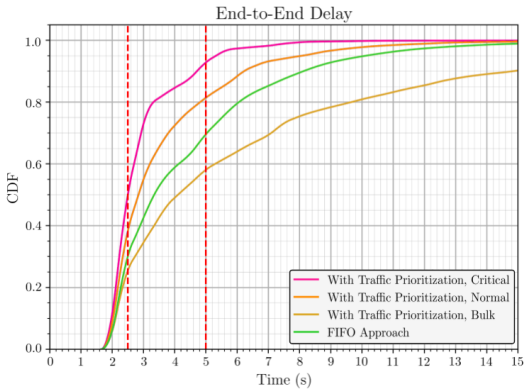


Figure 15. End-to-End Delay in the Two-Hop Scenario with Worst Conditions

Across all conditions, the validation of the better performance of critical bundles at the expense of low priority bundles continues to be demonstrated in the results.

C. Three-Hop Scenario

Looking at Figure 16 the three-hop scenario with optimal conditions shows similarities to the two-hop scenario. The curves have an alike shape, with the greatest parallels being with the FIFO and normal curves. However, the percentages of these curves are still different, since their origin is not at 1.6 s, but at around 2 s, because the higher servicing time shifts the origin in the final graph. Since the curves are all rather close to each other when hitting the first mark, a direct assignment of each class to its probability is hardly possible. The area of impact is from 23% to 33%, which is a significant decrease from the previous 76%, 65%, 62% and 55%. However, before the second mark all curves reached 100% except for the bulk being at 96%. Even though the first percentages are rather low, the positive impact of traffic prioritization continues to be shown, since it outperforms all other classes. Although, the high impact of the service time makes the target mark hard to reach even in optimal conditions.

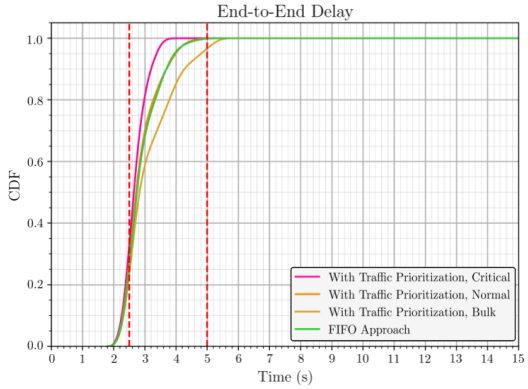


Figure 16. End-to-End Delay in the Three-Hop Scenario with Optimal Conditions

Before analyzing the overall end-to-end delay in the three-hop scenario with average conditions, the delay of each hop is shown first in Figure 17. These separated delays are especially interesting in the three-hop scenario, because each hop only corresponds to one environment, which allows for deductions about the atmospheric and space setting. The similarities of these three graphs to the ones of optimal and worst case are very high, with only small degradations, which is why only for average conditions this split is shown.

For the first hop the origin of all curves is now 0.4 s, since the bundle service time depends on the distance. From there it is a very steep curve to 100%, which is reached by the critical bundles at 2.2 s, for normal at 2.3 s and for bulk at 2.7 s. In the second hop critical arrives at the first mark at 99%, normal at 88% and bulk at 75%. These two plots show the impact of the losses in space on the delay. It becomes clear from the second hop that the impact increases proportionally with the distance. The plot of the third hop then shows the low impact of the atmospheric losses, since all curves are overlapping.

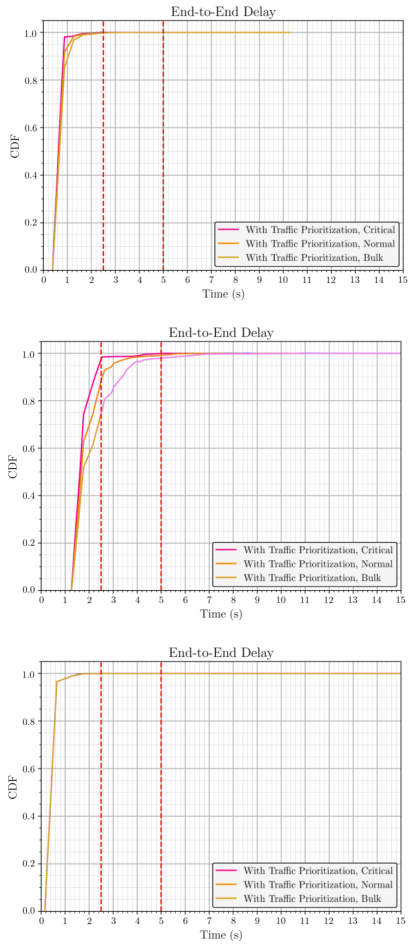


Figure 17. Breakdown of the End-to-End Delay per Hop in the Three-Hop Scenario with Average Conditions

Resulting from these plots, the end-to-end delay displayed in Figure 18 only shows small changes compared to the optimal case. The area of impact with the 2.5 s mark did not change noticeably. However, the small changes in all three-hop delays are obvious in this graph, since no curve reaches 100% before the second mark now. Additionally, all curves are slightly lower in rise such that the probabilities at the 5 s mark are 99% for critical, 98% for FIFO and normal, and 83% for bulk. Therefore, the higher error rates are displayed in the delays even though the overall graph is not very different from the optimal case. This suggests that the influence of the error rates on the delay of each bundle decreases with increasing number of hops. A fitting explanation would be the lower service times, which allow for faster reaction to losses. This shortens the time until retransmission, resulting in a higher throughput, which also benefits the lower priorities. Furthermore, the distribution of the priority classes over all nodes becomes more balanced such that all bundles have a good chance of being sent regardless of their priority class. Based on this analysis, the changes under the worst conditions should be minimal again, following the trend set in the two-hop scenario, which showed minimal differences between the optimal and the average conditions.

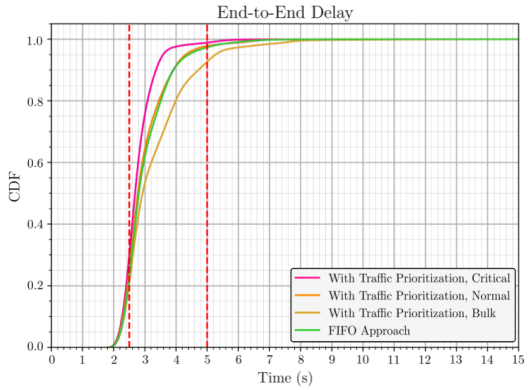


Figure 18. End-to-End Delay in the Three-Hop Scenario with Average Conditions

The end-to-end delay of the worst conditions also only shows small changes to the average conditions, as can be seen in Figure 19. However, the normal and the FIFO approach develop some separation, reestablishing the general order of curves. Because of the higher losses, the area of impact with the first mark is set slightly lower than before at 18% to 33%. The curves reach the second mark at 97% for critical, 94% for normal, 93% for FIFO and 86% for bulk.

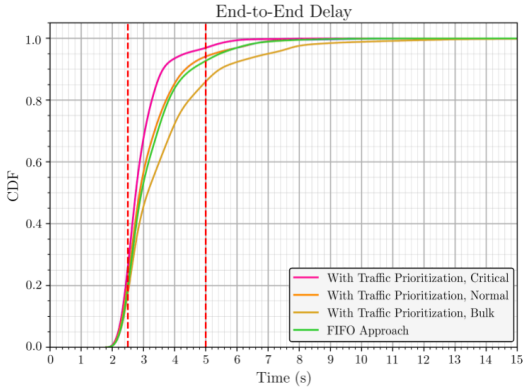


Figure 19. Comparison of the End-to-End Delay in the Three-Hop Scenario with Worst Conditions

These results support the hypothesis that a higher number of hops weakens the influence of losses on the end-to-end delay, because of the higher throughput of each node. Additionally, if a sending attempt of a low priority bundle is already started, it does not get interrupted even if a higher priority bundle arrives, which could contribute to a better throughput for low priority bundles. Nevertheless, the positive impact of traffic prioritization remains visible even with these findings. The critical class arrives even under worst conditions and over multiple hops faster at the destination than the FIFO bundles.

D. Evaluation

The findings from all experiments have been systematically analyzed to provide comprehensive insights into traffic prioritization performance across multi-hop scenarios. Therefore, a comprehensive summary of all insights will confirm or refute the drawn conclusions, and allows for statements that span all experiments.

The main goal of the experiments was to analyze the impact of traffic prioritization not only on a one-hop, but also on a multi-hop path. First, the validation of the results of the work of Algarra et al. [6] had to be performed. In the analysis of the one-hop scenario with average conditions, strong similarities appeared to their work. Therefore, the changes to the Markov chain in this thesis were not so drastic that the results of Algarra's work are not comparable anymore. Thus, their approach, although less detailed in its losses compared to this work, still adequately represented the Earth-to-Moon communication link. These similarities result in an applicability of the findings from the multi-hop scenarios to the work of Algarra.

The overall outcome of the traffic prioritization evaluation is a clear performance improvement for the critical bundles. In all experiments regardless of the conditions or number of hops, the critical class always performed better than all other classes, and especially compared to the FIFO approach. The extent of the improvement depends on the condition and number of hops. It is for example significantly higher for a one-hop case with worse conditions than with optimal conditions, resulting from the overall loss experienced being generally higher, such that the urgency of the critical class proves more useful.

In the comparison of the performance improvements of the critical bundles, it was discovered that the higher the number of hops, the lower the improvement. This is presumably because of lower service time at each hop, resulting in a faster retransmission attempt. Stemming from that is also a higher spread of bulk bundles over multiple hops, such that the lower priorities are serviced due to lack of higher priority bundles. Overall, the delay increased for all classes as more hops were involved, since each hop introduced an additional service time and an additional transmission attempt to pass. This service time decreased the percentage of bundles arriving before the 2.5 s mark especially for the high priority bundles.

Combining those two insights, it becomes apparent that a change in the routing according to the priority class of the bundle might provide improvement. Critical bundles should be sent as directly as possible to their destination to avoid multiple service times and minimize the risk of being blocked by potential loss more than once. However, low priority bundles may be routed through multiple hops to have them distributed widely in the network improving their chances of being serviced next and of not being blocked by a higher priority bundle. If the communication is only possible over multiple hops, it might be beneficial to interrupt the transmission of a low priority bundle in favor of a critical bundle to enhance the transmission time of the high priority bundles.

Apart from findings about traffic prioritization, it also became apparent that the space setting has a much higher influence on the communication link than the atmospheric setting. This is particularly noticeable in the three-hop scenarios, where each hop was assigned to exactly one setting, and therefore the delay plot of the atmospheric setting showed a very low impact. The second hop, in the three-hop scenario, however, displayed the highest delay, since the distance of that hop was the largest, resulting in a bottleneck. In this case interrupting low priority transmissions upon the arrival of high priority bundles might be desirable as well to decrease the delay for critical bundles.

VI. Conclusion and Future Work

A. Conclusion

The communication infrastructure between Earth and Moon needs to develop in order to be able to support all the planned upcoming missions [8]. However, to support this development more robust and reliable communication mechanisms need to be in place. As of now the DTN with BP is envisioned to be the future standard for space communications because of its hop-by-hop approach fitting for the highly stressed communication environment that this setting presents. However, it lacks QoS mechanisms to maximize the utilization of limited bandwidth, provide reliability and be able to achieve the delivery time requirements.

Traffic prioritization is a QoS mechanism fitting for handling limited resources, and so the overall goal of these experiments was to evaluate the impact of traffic prioritization on multi-hop paths building on previous work [6]. All experiments regardless of the scenario or the conditions demonstrated a performance improvement of critical bundles compared to FIFO. The implementation of traffic prioritization into the BP is therefore recommended based on this study. The highest gain is achieved in a one-hop scenario with worst conditions. Throughout the two- and three-hop scenario the gain is reduced whereas the performance of the low priority bundles increased. This is due to the lower service time enabling faster retransmission attempts, which, in turn, lead to a better distribution of bulk bundles, increasing their probability of being serviced. Based on these findings, the suggestion is to additionally adjust the routing of the bundle according to its priority class, which would benefit not only the high priority bundles, but also the low priority ones. The critical bundles should therefore be sent via the most direct route possible to the destination, whereas the bulk bundles can alternatively traverse through multiple nodes on their way. However, it is not uncommon that one node in the path is a critical congestion point, such that all bundles travel through it, creating a bottleneck. The research showed in

the three-hop scenario the significant influence bottlenecks can have on all classes. In such cases the dropping of a low priority class bundle might be desirable in favor of a high priority bundle, such that the effect of the bottleneck is mitigated for the critical bundles.

B. Future Work

The effect of routing according to the bundle's priority as well as the preemptive dropping of a low priority bundle in favor of a high priority bundle requires individual research studies to validate its potential. In addition, the work on prioritization is not yet complete, missing a deployment to a satellite or spacecraft and an analysis of the resulting findings to identify as the actual gains or potential issues. In preparation for deployment, more experiments with the constructed Markov chain could be conducted, testing a different traffic distribution more representative of realism. Additionally, the approach of using weighted queuing instead of strict prioritization as researched in Algarra et al. [6] would provide valuable insights and might further validate the existing results for a multi-hop path.

Moreover, further research on the model would also be beneficial. All data used for the probabilities of the Markov chain were carefully researched. However, in some areas a lack of information made fine-tuning challenging, as for example the entry probability for clouds, which is based on the number of rainy days rather than cloudy days. A purpose-oriented gathering of data would make these values more realistic. Additionally, some loss factors were not included as to reduce the complexity of the model, but might be a valid addition depending on its application area for scenarios in which a highly realistic model is needed.

The Markov model can also be used for the research of additional QoS mechanisms. Especially for reliability, only the implementation was investigated but a study on its actual benefits for the communication link was not provided. The usage of the communication link will increase in the future, such that its improvement is essential.

Beyond testing new features for the link, the Markov chain could also be used for a general study of the Moon-to-Earth communication link. An additional step for that could be the modification to support flexible transmission links instead of a hard coded FSO link through the atmospheric setting and RF link through space. With that, the current state of the link in combination with different scenarios and conditions might provide insights into routing or the types of loss influencing the transmission the most, such that further work can be done on prevention. An alternative to prevention is the correction of errors. By treating the bundles as bits, such error correction algorithms could be tested for their performance in the Moon-to-Earth communication link and optimized with the information gained. The model's realistic depiction can also provide information about actual transmission times. The program could be modified to add contact windows, so that no continuous communication is assumed. Combining the contact windows with the losses would yield the actual success communication availability timeframes.

Overall, the research into lunar communications and therefore the importance of the Moon-to-Earth communication link will continue to grow in the coming years. Therefore, improvements such as traffic prioritization are crucial to ensure the success of the missions. The demonstration of the performance gain should result in an implementation of traffic prioritization into BP, permanently enhancing the communication link. Additionally, the flexible model of a multi-hop transmission path will remain relevant and prove useful in future research.

References

[1] K. Schaper, T. A. Ulierte, A. Timm-Giel, and F. Flentge, "Modeling of multi-hop dtn-based lunar communications for the evaluation of traffic prioritization", in *Proceedings of the Seventeenth International Conference on Advances in Satellite and Space Communications*, Available: https://personales.upv.es/thinkmind/SPACOMM/SPACOMM_2025/spacomm_2025_1_20_20015.html [accessed Nov. 15, 2025], IARIA, May 2025, pp. 5–11.

[2] J. M. Logsdon, *Space exploration*, Available: https://www.britannica.com/science/space-exploration [accessed Dec. 08, 2024], Encyclopedia Britannica, Nov. 2024.

[3] F. Warthman, "Delay- and disruption-tolerant networks (dtns), A tutorial", Tech. Rep., Sep. 2015, Available: https://www.nasa.gov/wp-content/uploads/2023/09/dtn-tutorial-v3.2-0.pdf [accessed Nov. 15, 2025].

[4] European Space Agency, *Gateway: Lunar link*, Available: https://www.esa.int/Science_Exploration/Human_and_Robotic_Exploration/Gateway_Lunar_Link [accessed Dec. 08, 2024].

[5] CDF Lunar Caves Study team, *Lunar caves cdf study - executive summary*, Available: https://esamultimedia.esa.int/docs/preparing_for_the_future/Lunar_Caves_Executive_Summary_1.0.pdf [accessed Nov. 15, 2025], European Space Agency, Mar. 2022.

[6] T. Algarra Ulierte, K. Kuladinithi, A. Timm-Giel, and F. Flentge, "Adding quality of service support to bundle protocol through an extension block", in *2024 IEEE 10th International Conference on Space Mission Challenges for Information Technology (SMC-IT)*, 2024, pp. 115–124. DOI: 10.1109/SMC-IT61443.2024.00020.

[7] T. Algarra Ulierte, K. Kuladinithi, A. Timm-Giel, and F. Flentge, "Lunar communication services: Feasibility study on traffic prioritization of quasi-real time communications over dtns", in *2023 IEEE International Conference on Wireless for Space and Extreme Environments (WiSEE)*, IEEE, Sep. 2023, pp. 35–40. DOI: 10.1109/wisee58383.2023.10289636.

[8] Lunar Communications Architecture Working Group, "The future lunar communications architecture", Interagency Operations AdvisoryGroup, Tech. Rep., Jan. 2022.

[9] E. Birrane III and K. McKeever, *Rfc 9172: Bundle protocol security (bpsec)*, 2022.

[10] T. Algarra Ulierte, F. Flentge, J. Quintanilla, and A. Timm-Giel, "Characterizing space communication channel behavior: An analytical approach with esa mission data validation", in *10th ESA International Workshop on Tracking, Telemetry and Command Systems for Space Applications (TT&C)*, 2025.

[11] K. Scott and S. C. Burleigh, *Bundle Protocol Specification*. Nov. 2007. DOI: 10.17487/rfc5050.

[12] Consultative Committee for Space Data Systems, *Bundle protocol specification (ccsds 734.2-b-1)*, Sep. 2015.

[13] S. Burleigh, "Interplanetary overlay network: An implementation of the dtn bundle protocol", in *2007 4th IEEE Consumer Communications and Networking Conference*, IEEE, Jan. 2007, pp. 222–226. DOI: 10.1109/ccnc.2007.51.

[14] P. Apollonio, C. Caini, and V. Fiore, "From the far side of the moon: Delay/disruption-tolerant networking communications via lunar satellites", *China Communications*, vol. 10, no. 10, pp. 12–25, Oct. 2013, ISSN: 1673-5447. DOI: 10.1109/cc.2013.6650316.

[15] H.-P. Lin, M.-J. Tseng, and D.-B. Lin, "Modeling fading properties for mobile satellite link channels using markov model approaches", vol. 4, 192–195 vol.4, 2003. DOI: 10.1109/APS.2003.1220153.

[16] J. Bitó, "On the markov modeling of digital communication channels", in *2008 IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications*, IEEE, 2008, pp. 1–6.

[17] M. Tropea and F. De Rango, "A comprehensive review of channel modeling for land mobile satellite communications", *Electronics*, vol. 11, no. 5, p. 820, 2022.

[18] F. P. Fontan *et al.*, "Complex envelope three-state markov model based simulator for the narrow-band lms channel", *International journal of satellite communications*, vol. 15, no. 1, pp. 1–15, 1997.

[19] V. Chu, P. Sweeney, J. Paffett, and M. N. Sweeting, "Characterising error sequences of the low earth orbit satellite channel and optimisation with hybrid-arq schemes", in *IEEE GLOBECOM 1998 (Cat. NO. 98CH36250)*, ser. GLOCOM-98, vol. 5, Sydney, NSW, Australia: IEEE, 1998, pp. 2930–2935. DOI: 10.1109/glocom.1998.776610.

[20] X. Pan, Y. Zhan, P. Wan, and J. Lu, "Review of channel models for deep space communications", *Science China Information Sciences*, vol. 61, no. 4, p. 040 304, Mar. 7, 2018, ISSN: 1869-1919. DOI: 10.1007/s11432-017-9345-8.

[21] T. Algarra Ulierte, F. Flentge, and A. Timm-Giel, "Quality of service extension for delay- and disruption-tolerant networking", in *18th International Conference on Space Operations (SpaceOps)*, 2025.

[22] T. Algarra Ulierte, K. Kuladinithi, A. Timm-Giel, and F. Flentge, "Enabling traffic prioritization for space communications over dtns", *IEEE Journal of Radio Frequency Identification*, vol. 8, pp. 748–760, 2024, ISSN: 2469-729X. DOI: 10.1109/jrfid.2024.3415508.

[23] C. Barry, *Weather on the moon*, Blog, Available: https://science.nasa.gov/moon/weather-on-the-moon/ [accessed Nov. 15, 2025], National Aeronautics and Space Administration.

[24] B. Yost and S. Weston, "State-of-the-art of small spacecraft technology", in National Aeronautics and Space Administration, Feb. 2024, ch. 9 Communications, pp. 243–271, Available: https://ntrs.nasa.gov/citations/20240001462 [accessed Nov. 15, 2025].

[25] European Space Agency, *What are solar flares?*, Blog, Available: https://www.esa.int/Science_Exploration/Space_Science/What_are_solar_flares [accessed Nov. 21, 2024].

[26] European Space Agency, *Optimised ccsds protocol stack for high data rate (erasurevectors)*, Login required.

[27] M. Hatfield, *Solar flares faqs*, Available: https://blogs.nasa.gov/solarcycle25/2022/06/10/solar-flares-faqs/ [accessed Dec. 14, 2024], National Aeronautics and Space Administration, Jun. 2022.

[28] N. Nishizuka, Y. Kubo, K. Sugiura, M. Den, and M. Ishii, "Reliable probability forecast of solar flares: Deep flare net-reliable (defn-r)", *The Astrophysical Journal*, vol. 899, no. 2, p. 150, Aug. 2020, ISSN: 1538-4357. DOI: 10.3847/1538-4357/aba2f2.

[29] Y. Guo *et al.*, "Feature identification and statistical characteristics of quasi-periodic pulsation in solar flares using the markov-chain-monte-carlo approach", *The Astrophysical Journal*, vol. 944, no. 1, p. 16, Feb. 2023, ISSN: 1538-4357. DOI: 10.3847/1538-4357/acb34f.

[30] Consultative Committee for Space Data Systems, "Real-time weather and atmospheric characterization data (ccsds 140.1-g-2)", Tech. Rep. Issue 2, Mar. 2024, Available: https://public.ccsds.org/Pubs/140x1g2.pdf [accessed Nov. 15, 2025].

[31] World Weather & Climate Information, *Monthly average rainy days in santa maria*, Available: https://weather-and-climate.com/average-monthly-Rainy-days,santa-maria-cv,Cape-Verde [accessed Dec. 14, 2024].

[32] Climates To Travel, *Climate in kiruna (sweden)*, Available: https://www.climatestotravel.com/climate/sweden/kiruna [accessed Dec. 14, 2024].

[33] World Weather Online, *Kourou annual weather averages*, Available: https://www.worldweatheronline.com/kourou-weather-averages/gf.aspx [accessed Dec. 14, 2024].

[34] I. A. Lund, "Persistence and recurrence probabilities of cloud-free and cloudy lines-of-sight through the atmosphere", *Journal of Applied Meteorology*, vol. 12, no. 7, pp. 1222–1228, Oct. 1973, ISSN: 0021-8952. DOI: 10.1175/1520-0450(1973)012<1222:parpoc>2.0.co;2.

[35] Team SimPy, *Overview - simpy documentation*, Available: https://simpy.readthedocs.io/en/latest/ [accessed Nov. 29, 2024].

Unveiling Content Traps: A Network Resilience and Topic-Based Study of YouTube's Algorithmic Content Curation

Md Monoarul Islam Bhuiyan*, Nitin Agarwal*†
*COSMOS Research Center, University of Arkansas at Little Rock, Arkansas, USA
†International Computer Science Institute, University of California, Berkeley, California, USA
e-mail: {mbhuiyan, nxagarwal}@ualr.edu

***Abstract*—YouTube's recommendation algorithm drives 70% of total video views, playing a pivotal role in shaping user engagement and content consumption. However, this algorithm often contributes to the creation of content traps, or attractor content, where users are nudged towards certain videos. Repeated nudges to users in turn create filter bubbles and echo chambers. This study aims to identify and analyze such content traps within YouTube's recommendation network using Focal Structure Analysis (FSA), a social network analysis approach designed to detect key groups of nodes in a network that together act as powerful attractors due to their position in the network. Such focal structures in a video recommendation network act as strong content traps. By constructing a network based on YouTube's recommendation hops, we identify content traps using the FSA approach. We then confirm the presence of dominant themes and high topical consistency by leveraging topic modeling and information divergence metrics, suggesting less topical diversity in content traps. Engagement analysis of content trap videos with high topic uniformity reveals user interactions reinforce these traps. Our findings provide insight into content trap dynamics that could inform strategies for content consumers to break out of them and help platform owners to develop debiasing techniques in their content curation algorithms. This research highlights the critical role of recommendation algorithms in shaping content exposure and calls for strategic interventions to foster greater content diversity and mitigate the effects of content traps.**

***Keywords*-Content Traps; Network Resiliency; YouTube Recommendation Network; Social Network Analysis; BERTopic**

I. Introduction

With the rise of social media platforms, several activities have gained wide popularity, including content creation and sharing, news consumption, community engagement, societal influence, and many others. The rapid adoption of these has become possible due to massive user engagement with video content that can often lead to the entrapment of content [1]. YouTube is the second-most popular social media platform and the number one video-sharing platform globally, available in over 100 countries and 80 languages [2]. Moreover, YouTube's recommendation algorithm drives 70% [3] of its watch time, making it one of the most influential aspects of the platform. Since this algorithm helps users discover content aligned with their viewing habits, it often also creates content traps or attractor content, where users are repeatedly exposed to similar content. The emergence of content traps is particularly concerning for sensitive or controversial topics, as it can influence public opinion, deepen polarization, and propagate misinformation. Given YouTube's critical role in shaping digital discourse, understanding the mechanisms driving content traps is essential. This phenomenon highlights concerns about the recommendation algorithm's influence on user behavior and its role in reinforcing content loops.

With an emphasis on how YouTube's algorithm directs users toward repeating and thematically limited content loops, we examine in this study the phenomena of content traps inside the platform's recommendation network. We employ Focal Structure Analysis (FSA) [4], a social network analysis (SNA) approach, to detect powerful set of nodes, known as focal structures, that play a significant role in shaping user engagement and reinforcing content traps.

Our analysis begins by constructing a recommendation network based on video suggestion patterns and then identifying focal structures. After that, we assess the impact of these structures by examining how their removal from the recommendation network affects its overall connectivity and resilience.

Through topic modeling and divergence metrics, we explore the topical consistency within these focal structures. Our study identifies content traps characterized by dominant, repetitive themes. Moreover, we examine user engagement trends to determine how interactions contribute to the persistence of these traps. Our study highlights how crucial it is to address algorithmic design in order to encourage content diversity and lessen the effects of content loops, which will ultimately help create a more equitable online interaction. This research highlights the broader implications of recommendation systems on user behavior and content consumption in online platforms.

Although algorithmic biases are becoming more widely recognized [5], to the best of our knowledge, there is currently no systematic method for identifying and assessing these content traps within YouTube's dynamic recommendation system. Our study aims to understand the presence of content traps within YouTube's recommendation network by leveraging the focal structure analysis approach. With this, we seek to answer the following research questions:

- **RQ1**. How can content traps or attractor contents be detected within focal structures from YouTube's recommendation network?
- **RQ2.** How can topical consistency within focal structures of YouTube's recommendation network be assessed to identify the presence of content traps or attractor contents?
- **RQ3.** Can engagement statistics be used to validate the findings of the content trap discovery model?

We examine how focal structures—defined as key sets of attractor nodes within YouTube's recommendation network—contribute to the emergence and persistence of content traps. By analyzing how these structures shape user experiences and exposure pathways, we reveal how recommendation algorithms can narrow informational diversity and reinforce homogeneous content loops. This study advances the understanding of algorithmic content traps by integrating structural, semantic, and behavioral perspectives. We show that Focal Structure Analysis systematically identifies cohesive subgraphs that act as attractor contents, that these structures display strong thematic uniformity as measured through BERTopic and divergence metrics, and that user engagement disproportionately amplifies the most uniform topics, empirically validating trap-like behavior. Together, these insights contribute to the broader discourse on algorithmic content control and offer a generalizable analytical approach for diagnosing and mitigating entrapment dynamics within large-scale recommendation platforms.

The rest of this paper is organized as follows: The Literature Review section reviews existing studies for identifying influential sets of nodes from a social network along with the identification and implications of content traps in social media. While the Methodology section outlines the approaches used for the overall study, the Experimental Results and Analysis section presents the findings of our research. Lastly, the final Conclusion and Future Work section summarizes the study and suggests directions for future research.

II. Literature Review

This section is divided into two parts. We first discuss the relevant literature related to identifying influential sets of nodes within a social network. After that, we delve deeper into the related literature around content traps and topical uniformity [6], exploring their formation and impact on social networks as well as the implications for user experience and societal impact from these attractor contents in social media.

A. Identifying Sets of Key Social Network Entities

Identifying key nodes or sets of nodes that are best connected and most influential in a social network is crucial for extracting actionable knowledge. Consequently, various methods have been proposed to identify these key nodes. While HITS determines hubs and authorities [7], PageRank assigns a numerical weight for each node in the network [8]. Both of these approaches can be used to identify influential nodes from social networks. In contrast, identifying the communities and clusters from a social network perspective has also been widely studied. Generally, in a community, similar nodes are more clustered together than nodes that don't share commonalities [9]. Previous researchers have also worked on a more sophisticated approach where their focus shifted from identifying the influential nodes or communities to detecting smaller key sets of players who maximized information diffusion. Authors in [10] devised a methodology where they identified focal patterns leveraging the Louvain method, which gave them more relevant information about the network than obtained from just the influential nodes [11]. When applying this method to large biological networks, they found more prominent, smaller, and relevant structures in protein-protein interaction networks [12]. Since this method couldn't extract structures with lower connection density, researchers extended their approach by combining highly connected candidate focal structures based on similarity values. This allowed the identification of both cliquish and small sparse yet connected structures [13]. An advanced version of this approach was proposed by [4], where they combined user-level centrality and group-level modularity methods to create a bi-level maximization network model that overcame the shortcomings of the previously described focal structures analysis methods.

While identifying these nodes is crucial for network analysis since they greatly influence the network's architecture, their impact assessment on the overall network is also imperative. As a consequence several of such assessment metrics have been developed [14] to quantify the network's resilience to disruptions, providing insights into its stability and connectivity.

B. Content Homogeneity in Social Media

Content traps occur when recommendation systems such as those used by social media platforms limit users' exposure to diversified content, which in turn can result in the reinforcement of existing beliefs and preferences. Pariser et al. (2011) first introduced the concept of filter bubbles, highlighting how algorithms can create information ecosystems that reinforce users' existing views and interests. More recent research by Bechmann et al. (2018) examined filter bubbles in Facebook news feeds by analyzing information similarity in link sources and content semantics. The role of recommender systems in increasing content homogeneity, particularly during significant political events such as the 2018 Brazilian presidential election, has also been studied. While the study of content bubbles has been conducted, solutions to mitigate these bubbles have also been researched. Diversification algorithms have been explored as a potential solution to mitigate these election polarization effects [17]. Studies on link prediction in social networks show that fairness-aware algorithms, which utilize network modularity measures, can decrease the segregation caused by homophily and mitigate filter bubbles [18]. On a behavioral level, awareness of filter bubbles influences how users respond to them, though personality traits have been found to play a less significant role in motivating corrective actions, particularly on platforms like Facebook [19]. Furthermore, an integrated tool synthesizing previous research has been proposed to provide users with a more comprehensive way to avoid filter bubbles across social networks, emphasizing the need for holistic approaches to diversify user experience and counteract content loops [20]. Together, these studies emphasize the growing importance of addressing filter bubbles through technological, algorithmic, and user-behavior-focused interventions.

Despite substantial attention to filter bubbles and algorithmic bias, there remains limited empirical work that systematically identifies and quantifies content traps within YouTube's recommendation network. Existing studies often address influential

nodes, content homogeneity, or user-level effects in isolation, but they do not provide an integrated framework that connects structural attractors, semantic uniformity, and engagement-driven reinforcement. This fragmentation leaves a gap in understanding how recommendation algorithms produce and sustain narrow content pathways. In this paper, we address this gap by identifying focal structures within YouTube's recommendation graph, assessing their structural influence through network fragmentation analysis, and evaluating their thematic consistency using topic modeling and divergence metrics. By linking these structural and semantic findings with engagement patterns, our study offers a unified approach for understanding how content traps emerge and persist within algorithmically curated environments.

III. Methodology

This section outlines our systematic approach to analyzing the YouTube recommendation network, detecting focal structures, and evaluating the presence of content traps. We start by summarizing our approach in collecting data, dataset background, and building YouTube recommendation networks. After that, we present the network resiliency approach taken to rank key focal structures. In addition to that, we lay out the foundation for the analysis of the topics using the BERTopic model. Lastly, this section explores several metrics to investigate the topical consistency across different topics within the focal structures.

A. Data Collection

The data collection process in this study was designed to systematically capture YouTube's algorithmic behavior through its 'watch-next' recommendations. In this study. we analyzed two contexts, the China–Uyghur conflict and the Cheng Ho propaganda on YouTube. Below we provide background details for these two contexts and the motivation for studying them.

1) Cheng Ho Propaganda: In recent years, the Chinese Communist Party (CCP) has adapted the story of the 15th-century admiral Zheng He, also known as Cheng Ho, to support its current political messages. Once known for his peaceful sea voyages, Zheng He is now depicted as a symbol of religious tolerance and diplomacy [21]. This shift aligns with China's efforts to address criticism of its treatment of Uyghur Muslims and to promote its Maritime Silk Road initiative. By rebranding this historical figure, the CCP aims to strengthen its soft power, particularly in Southeast Asia.

2) China Uyghur Conflict: The conflict in Xinjiang focuses on the struggles faced by the Uyghur Muslim minority in China, which includes issues like cultural suppression, ethnic tensions, and government policies[22]. Researchers have studied this situation through various lenses, such as identity politics, language policies, the dynamics between majority and minority groups, and the desire for self-governance [23]. The international response between 2018 and 2022 has drawn more attention to human rights issues, highlighting the severity of the conflict.

We selected the China–Uyghur conflict and Cheng Ho propaganda datasets for their geopolitical and ideological relevance in examining algorithmic content amplifications and recommendation dynamics within the recommendation network.

3) Keyword Generation and Crawling: We initiated the process by conducting workshops with subject matter experts to compile a targeted list of keywords related to the China–Uyghur conflict and the Cheng Ho propaganda. These keywords served as search queries on YouTube's search engine, generating an initial set of seed videos. It's important to note that the unbalanced count of keywords as shown in Table I and Table II between different datasets does not compromise the validity of our study.

Table I: Keywords related to China-Uyghur conflict

Keywords
Penindasan/oppression + Uighur/Uyghur, Kejam/cruel + Uighur/Uyghur, Saudara muslim/muslim brother + Uighur/Uyghur, Kalifah/capliph + Uighur/Uyghur, Khilafah/caliphate + Uighur/Uyghur, "China is Terrorist", "Stop Genocide", "Save Muslim Uyghur", "Get Out China", "I Love Muslim Uyghur", "peduli Uyghur" / "Care Uyghur", "bebaskan muslim uyghur dari penindasan china" / "free uyghur muslims from china's oppression", "do'a kan saudaramu" / Pray for Muslim Uyghur, Hizbul Tahrir / HTI + Uighur/Uyghur, Front Pembela Islam/FPI/Islamic Defenders Front + Uighur/Uyghur, Nahdlatul Ulama + Uighur/Uyghur, Muhammadiyah + Uighur/Uyghur, Hebibulla Tohti + Indonesia, Mohammed Salih Hajim + Indonesia, Yusuf Martak + Uighur/Uyghur, Slamet Ma'arif + Uighur/Uyghur, Xiao Qian + Uighur/Uyghur, Pendidikan/education + Uighur/Uyghur

Table II: Keywords related to Cheng Ho propaganda

Keywords
Cheng Ho, Zheng He, Sam Po Kong, Sam Poo Kong, Daerah Otonom Uighur Singkiang, Singkiang, Hatta + 1957, Novi Basuki, Sam Po Bo, Cheng Ho / "Zheng He" + laksamana + damai, Sam Po Kong + Islam + Indonesia, 1421 Saat China Menemukan Dunia + "Gavin Menzies", Gavin Menzies, Cheng Ho / "Zheng He" "Columbus"

Using custom-built crawlers, video recommendations were extracted recursively across multiple depths. Starting with the seed videos, the first depth of recommendations was captured, and each recommended video was subsequently treated as a parent node for further investigation [24]. Datasets with distinct videos were produced due to this iterative procedure, which proceeded until four recommendation depths were reached. This created the recommendation network with 5 levels. The decision to limit the collection of data for 5 levels was a strategic choice based on essential computational constraints. As we explored deeper recommendation levels, the data volume increased exponentially. Restricting the analysis to a maximum of 5 levels for a more manageable evaluation of the data without compromising the integrity of the findings. A similar strategy was adopted by the authors in [25]. We then gathered titles, descriptions, and engagement metrics such as likes, comments, and views for each video except for the transcript using YouTube Data API v3 [26]. However, for transcript generation, the method from [27] was used.

B. Network Construction

The China–Uyghur conflict and Cheng Ho propaganda datasets consisted of 9,748 videos and 14,307 interconnections and 8,489 videos and 13,384 interconnections, respectively, that represented YouTube's recommendation paths across several hops, and they were used to start the network construction process. The initial networks had an average clustering coefficient of 0.067 and 0.062, respectively, suggesting comparatively sparse and loosely connected network structures.

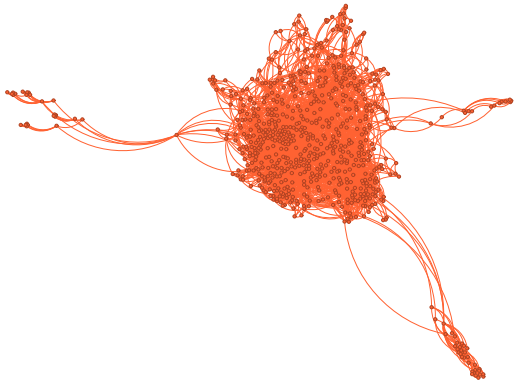


Figure 1: This graph illustrates the Cheng Ho propaganda recommendation network on YouTube after pruning, showing a denser, more interconnected structure with an average clustering coefficient of 0.362.

To make our study more efficient and less computationally constrained, we then refined the recommendation networks using a filtering criterion to concentrate on significant connections by retaining only videos that had a degree of 3 or more. This simple yet effective graph-pruning approach reduced network sparsity, enhanced structural coherence, and ensured that the analysis concentrated on nodes that were more influential, aligning with prior research that had demonstrated the effectiveness of graph pruning in improving computational feasibility and analytical precision [28], [29]. Thus, the filtering process reduced the networks, resulting in a denser graph with an average clustering coefficient of 0.38 and 0.362, respectively, for the China–Uyghur conflict and Cheng Ho propaganda recommendation networks, as depicted in Figures 1 and 2. The increase in the clustering coefficient indicates a more tightly-knit YouTube recommendation network, ensuring that our analysis captured well-connected focal structures rather than weakly linked, peripheral nodes. Due to this, we identified areas more likely to contribute to content traps by focusing on these denser network connections. This allowed us to assess the impact of focal structures on the topical consistency of recommended content and their influence on user engagement.

C. Focal Structure

Focal structures (FS) are key sets of individuals in a social network who may be responsible for coordinating events, protests, or leading citizen engagement efforts. In the context

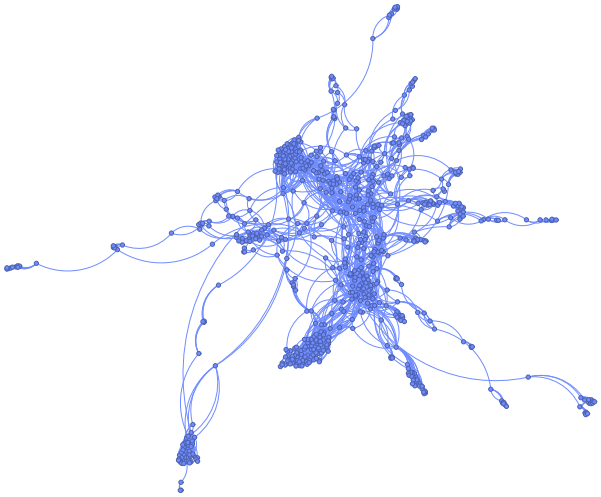


Figure 2: This graph illustrates the China–Uyghur conflict recommendation network on YouTube after pruning, showing a denser, more interconnected structure with an average clustering coefficient of 0.38.

of content traps, focal structures can be defined as a set of attractor content(s), e.g., a set of videos. By manipulating content recommendations, these structures limit exposure to diverse perspectives, reinforcing particular viewpoints, and ultimately drawing users deeper into a narrow content spectrum in the YouTube video recommendation network. Understanding these dynamics is crucial for assessing the societal impacts of content dissemination in digital environments.

In our study, we identified focal structures in YouTube video recommendation networks, represented as a social network $G = (V, E)$, where V is the set of vertices (videos) and E is the set of edges (recommendation links between videos). Focal structures are defined as subgraphs within this network. Formally, focal structures are defined by $F = \{G'\}$, where $G' = (V', E')$, such that $V' \subseteq V$ and $E' \subseteq E$. For all i and j, $i \neq j$, $G_i \in F$ and $G_j \in F$, such that no two focal structures can subsume each other, or $G_i \nsubseteq G_j$ and $G_j \nsubseteq G_i$. [4]. This ensures that the focal structures identified in the network are unique and non-overlapping, each representing a key distinct set of nodes (videos) and edges (recommendations).

D. Network Resiliency Assessment

To identify the key focal structures in our study, we first removed each focal structure from the pruned graph and analyzed the resulting cluster formation to assess network fragmentation. It can be posited that a higher number of clusters indicates greater fragmentation, which suggests that the removed focal structure played a more critical role in maintaining the overall cohesion of the network. In other words, these focal structures act as critical connectors within the overall network. When these key structures are removed, the network's cohesiveness decreases, demonstrating their influence to the overall network [30]. This method emphasizes the structural

significance of focal structures and provides insight into how they support content flow throughout the recommendation system and preserve network integrity.

E. Topic Modeling with BERTopic

In order to analyze the content and thematic focus within the focal structures, we applied the BERTopic [31] model, which is a powerful tool for topic modeling that can generate interpretable topics from large text datasets. In our analysis, we opted for BERTopic over conventional models such as Latent Dirichlet Allocation (LDA) [32] or Non-negative Matrix Factorization (NMF) due to its enhanced capability to capture semantic nuances and contextual information. This superiority enables BERTopic to generate more refined and granulated topics, thereby providing deeper insights into the underlying data structure. The primary goal of our study was to determine whether high topic uniformity within a focal structure, where one topic dominates beyond a specified threshold, indicated the presence of a content trap or attractor content, leading to more homogeneous content consumption.

Since YouTube video transcripts are often lengthy, we encountered a limitation due to BERTopic's processing constraint. The maximum length limit of tokens is 512 for the BERTopic model. Hence, we split the video transcripts into multiple chunks, each containing fewer than 512 words, ensuring that each chunk remained coherent by splitting at sentence boundaries. As a result, this method helped maintain the consistency and integrity of the information while ensuring that the chunks adhered to the token limit set by BERTopic. Furthermore, it helped reduce noise, leading to more accurate topic representations. Next, we mapped the identified topics to video IDs to analyze the distribution of videos by topic. This approach enabled a comprehensive understanding of thematic coverage within each focal structure, capturing the diversity or concentration of topics.

After that, we defined a threshold to detect content traps. Specifically, a content trap was identified when a particular topic appeared in more than 50% of the videos within a focal structure. This threshold was chosen based on the idea that a concentration of similar topics across a significant portion of the videos within a focal structure might limit the diversity of content, thereby trapping users in a narrow content trap. In other words, by limiting exposure to varied content, these tightly connected groups of videos increase the likelihood of users being repeatedly nudged toward the same themes, thereby acting as attractor content. The threshold can be formally represented as

$$T = \frac{n_{\text{topic}}}{n_{\text{total}}} > 0.5 \tag{1}$$

where

- T is the threshold for identifying a content trap within a focal structure,
- n_{topic} is the number of videos in the focal structure that share a specific topic, and
- n_{total} is the total number of videos within the focal structure.

We classified the focal structure with a content trap if the proportion T exceeded 0.5. This threshold helped identify clusters where the recommendation algorithm disproportionately favored a single topic that could result in a lack of content diversity. Consequently, this could lead to users continuously engaging with similar or attractor videos over extended periods. Identifying such focal structures can highlight specific areas within YouTube's recommendation system that may contribute to content traps, allowing us to further investigate the implications for user experience and content engagement patterns. Therefore, this methodology provided a systematic way to analyze the topic uniformity in a focal structure within the recommendation system, offering insights into how the role of YouTube's recommendation algorithm may influence content diversity and contribute to content traps.

F. Divergence Metrics

We utilized two key divergence metrics to quantify the similarity between the distributions of topics within focal structures, namely, Kullback-Leibler (KL) Divergence [33] and Jensen-Shannon (JS) Divergence [34]. These metrics provide insights into the diversity and uniformity of topics, which are critical in identifying content traps.

1) Kullback-Leibler (KL) Divergence: The Kullback-Leibler (KL) Divergence measures the difference between two probability distributions P and Q. It is defined as

$$D_{KL}(P||Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)} \tag{2}$$

where

- $P(i)$ represents the true topic distribution and
- $Q(i)$ represents the approximated topic distribution.

A low KL Divergence value indicates that the two distributions are similar, suggesting a limited diversity in the topics present within the focal structure that may signal a content trap.

2) Jensen-Shannon (JS) Divergence: The Jensen-Shannon (JS) Divergence is a symmetrized and smoothed version of KL Divergence. It is computed as

$$D_{JS}(P||Q) = \frac{1}{2}\left(D_{KL}(P||M) + D_{KL}(Q||M)\right) \tag{3}$$

where M is the average distribution

$$M = \frac{1}{2}(P + Q) \tag{4}$$

The JS Divergence is bounded between 0 and 1, making it more stable for comparing distributions. A lower JS Divergence value indicates a lack of topic diversity and the existence of content traps.

In our study, KL Divergence and JS Divergence were employed in order to assess the extent of topic uniformity within focal structures. High values of KL Divergence or JS Divergence suggested a diverse range of topics, indicating the absence of a content trap. In contrast, low KL and JS Divergence values pointed to significant topic uniformity, signaling a

potential content trap and providing a quantitative approach to evaluating content traps. Additionally, these approaches offered valuable insights into YouTube's recommendation algorithm dynamics and its role in shaping content diversity.

IV. Experimental Results and Analysis

This section focuses on key factors influencing content traps within YouTube's recommendation network. We start with Focal Structure Analysis and network fragmentation, which examine focal structures and their effects on network resiliency. Next, we explore topical uniformness and content trap identification, focusing on how topic dominance can lead to content homogeneity. The role of divergence metrics is then assessed in identifying content traps through KL and JS Divergence. Lastly, we investigate user engagement's role in reinforcing content homogeneity, highlighting how interactions influence content consumption patterns.

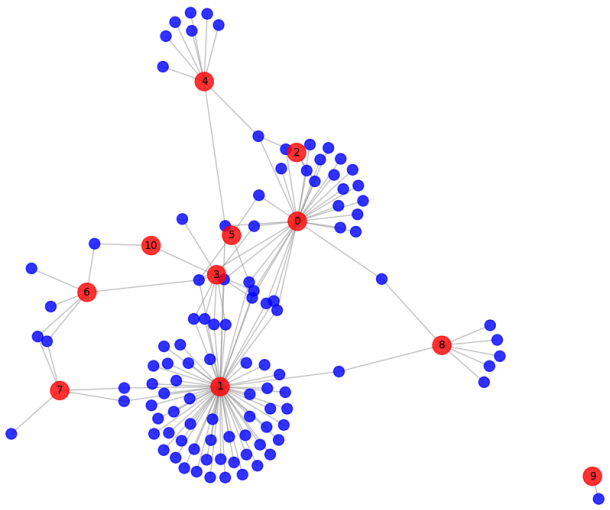


Figure 3: Network visualization of topics and associated video IDs, where red nodes represent topics, blue nodes represent video IDs, and edges denote the association between a video and its topic within the focal structure number 3 of China-Uyghur conflict network data.

A. Focal Structure Analysis and Network Fragmentation

Focal Structure Analysis (FSA) is a social network analysis method that aims to find key sets of individuals rather than a set of key individuals within a social network. The goal of this analysis is to identify structures that can exert influence over the maximum number of entities across various parts and communities within social networks. As part of our research, we first applied the focal structure analysis method to our overall pruned recommendation networks. This analysis identified 105 focal structures within the China–Uyghur conflict dataset and 141 within the Cheng Ho propaganda dataset. After that, we analyzed the network fragmentation resulting from removing each focal structure to identify the key groups in the social network. By assessing the number of clusters formed after the removal, we extrapolated the role of each focal structure in maintaining the network's cohesion. When a focal structure is removed, a greater number of resulting clusters indicates a more significant disruption to the network. This suggests that the removed structure played a crucial role in maintaining connectivity and content flow. Therefore, we were able to identify the most important focal structures for our study. The top 5 of such focal structures for each dataset are listed in Table III.

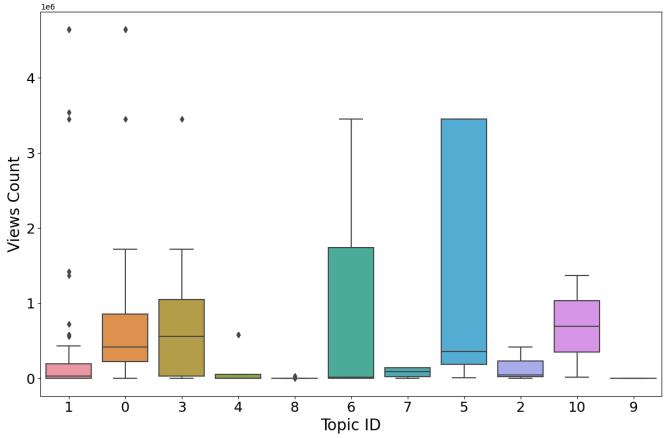


Figure 4: View count distribution across sorted topics based on topic uniformity for China-Uyghur data for focal structure 3. Topic 1 has the highest uniformity but low median views. Significant outliers indicate certain videos received disproportionately high attention, suggesting algorithmic bias and reinforcing the content trap.

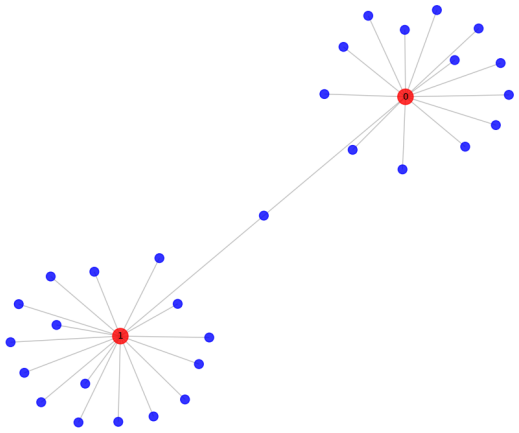


Figure 5: Network visualization of topics and associated video IDs, where red nodes represent topics, blue nodes represent video IDs, and edges denote the association between a video and its topic within the focal structure number 9 of China-Uyghur conflict network data.

Table III: Key metrics for focal structures in the recommendation networks, including size, dominant topic, topic uniformity, and divergence scores. Structures with uniformity above 50% are flagged as potential content traps.

Datasets	Focal Structure (FS)	No. Videos in FS	No. of Videos in Dominant Topic	No. of Clusters	Topic Uniformity	Content Trap	KL Divergence	JS Divergence
China-Uyghur Dataset	3	105	64	185	61%	YES	0.680	0.158
	9	30	17	44	57%	YES	0.004	0.001
	1	25	15	41	60%	YES	0.012	0.003
	102	13	7	31	54%	YES	0.067	0.234
	101	13	5	28	38%	NO	0.154	0.043
Cheng Ho Propaganda Dataset	13	68	45	134	66%	YES	0.700	0.166
	1	52	21	96	40%	NO	0.770	0.174
	6	28	22	81	79%	YES	0.446	0.120
	22	29	16	74	55%	YES	0.094	0.025
	3	28	18	74	64%	YES	0.277	0.069

B. Topic Uniformity and Content Trap Identification in Focal Structures

In this section, we delve deeper into analyzing the thematic focus within the identified key focal structures by measuring the topic uniformity. After identifying the top key focal structures from our networks, we applied the BERTopic model to extract interpretable topics from the video transcripts. Next, for each focal structure, we extracted the number of videos associated with each topic. Topic uniformity was then quantified by calculating the dominance of a particular topic within a set of videos, such as those within a focal structure. This measure reflects the proportion of content assigned to the dominant topic, which indicates the extent to which a focal structure is centered around a single or attractor topic. By setting up a threshold for topic uniformity, we identified content traps. A focal structure was classified as having a content trap if a single or attractor topic appeared in more than 50% of the videos within that structure. Table III presents the total number of videos in each focal structure, the number of videos in the dominant topic associated with each focal structure, topic uniformity, and whether the structure was identified as a content trap. This methodology provides valuable insight into how homogeneous topic concentration within focal structures can lead to a lack of content diversity, potentially trapping users in a narrow content trap and addressing our research question **RQ1**. In addition, to support these findings, Figures 3 and 5 illustrate the visualization of topics alongside their corresponding video IDs for two such focal groups within the China Uyghur conflict dataset which are also the two topmost focal structures that identified as content trap. Similarly, the top two focal structures that are identified as content traps from Table III are shown in Figures 7 and 9, which present comparable visualizations for the Cheng Ho propaganda network, where all the figures highlight a clear pattern of topics clustering together, suggesting the presence of attractor content. As a result, this phenomenon highlights a significant reduction in thematic diversity, where discourse becomes increasingly concentrated around a limited set of ideas or contents.

C. Divergence Metrics and Their Role in Identifying Content Traps

When evaluating the topical consistency within focal structures for our study, the Jensen-Shannon (JS) and Kullback-Leibler (KL) divergence metrics played a significant role.

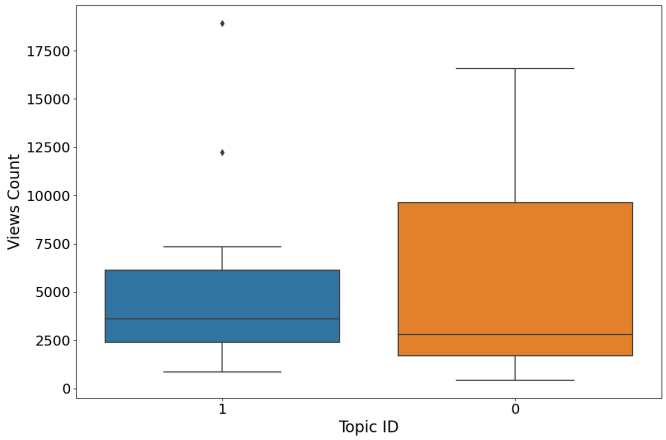


Figure 6: View count distribution across sorted topics based on topic uniformity for China-Uyghur data for focal structure 9. Topic 1 has the highest uniformity and higher median view count. Significant outliers indicate certain videos received disproportionately high attention, suggesting algorithmic bias and reinforcing the content trap.

These metrics helped us identify potential content traps within YouTube's recommendation system by measuring the distribution of topics within a focal structure. When a focal structure exhibits low divergence values, it suggests that the recommendation algorithm primarily favors a limited set of topical themes, reinforcing homogeneous content consumption. In contrast, higher divergence values indicate a broader distribution of topics with less topical uniformity, suggesting a more diverse range of content and thereby reducing the likelihood of a content trap. By analyzing these divergence metrics as shown in Table III, we identified content traps within focal structures. For instance, Focal Structure 3 for the China–Uyghur conflict dataset stands out with a relatively higher KL Divergence value compared to other focal structures, but it also exhibits a much lower JS Divergence. This combination suggests that although there may be some degree of variability in topic distribution as indicated by the KL Divergence, the overall diversity of topics within this structure remained limited, as evidenced by the low JS Divergence value. This leads to a content trap scenario, where a single or attractor topic dominates the recommendations, restricting the variety of content available

to users.

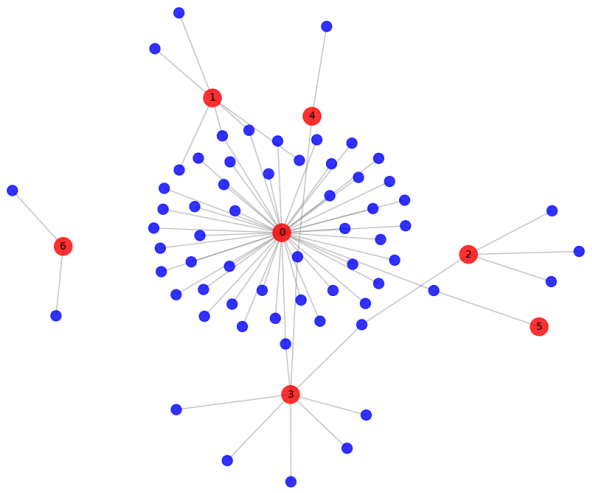


Figure 7: Network visualization of topics and associated video IDs, where red nodes represent topics, blue nodes represent video IDs, and edges denote the association between a video and its topic within the focal structure number 13 of Chengo Ho propaganda.

However, not all focal structures with high divergence scores resulted in content traps. For example, Focal Structure 1 in the Chengo Ho propaganda dataset exhibited both high KL and JS Divergence values, indicating that it still maintained internal topical diversity. This suggests that users exposed to this focal structure encountered a broader range of content rather than being constrained to a single dominant theme. As a result, this type of structure does not reinforce homogeneous content consumption and thus reduces the likelihood of content entrapment. By incorporating both cases, our analysis provides a nuanced understanding of how topical consistency influences content diversity within focal structures, thereby addressing our research question **RQ2**.

D. Role of User Engagement in Reinforcing Homogeneous Content

In our analysis, topics with higher uniformity exhibited notable outliers in user engagement metrics such as view count. In other words, while topics with high topic uniformity reflected a concentration around a single theme, the engagement distribution unveiled a more complex pattern, with some videos receiving disproportionately high attention. User engagement from the highly uniform topics showed disproportionately high engagement as well as surpassing the median engagement level in other topics for certain focal structures. However, the disproportionate high engagement persisted in most of the focal structures' high uniformity topics. This highlights how high-engagement videos reinforce homogeneous content consumption by amplifying specific content through the recommendation algorithm, further promoting uniformity within

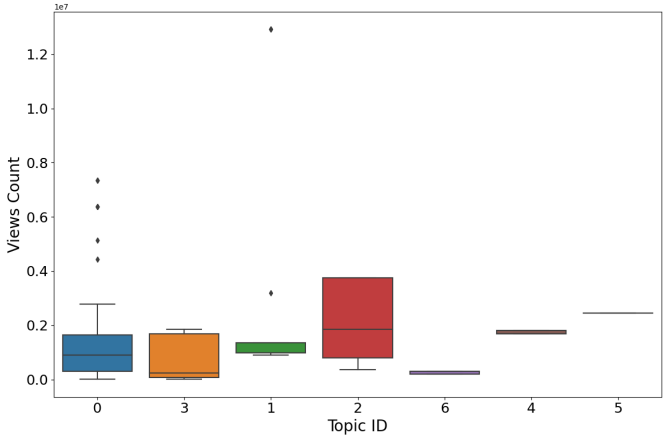


Figure 8: View count distribution across sorted topics based on topic uniformity for Cheng Ho data for focal structure 13. Topic 0 has the highest uniformity but low median views. Significant outliers indicate certain videos received disproportionately high attention, suggesting algorithmic bias and reinforcing the content trap.

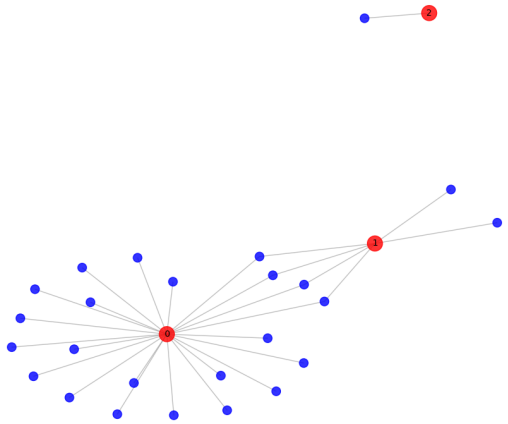


Figure 9: Network visualization of topics and associated video IDs, where red nodes represent topics, blue nodes represent video IDs, and edges denote the association between a video and its topic within the focal structure number 6 of Chengo Ho propaganda.

topics. Figures 4 and 6 illustrate that topics exhibiting the highest levels of topic uniformity also showed a greater number of outliers compared to other topics within the focal structure for the China–Uyghur conflict network, and Figures 8 and 10 portray a similar kind of pattern in the Cheng Ho propaganda network as well. This dynamic underscores how engagement metrics and topic uniformity can significantly boost content traps, where diversity is limited and users' exposure to new or varied content is constrained and thus addressing our research question **RQ3**.

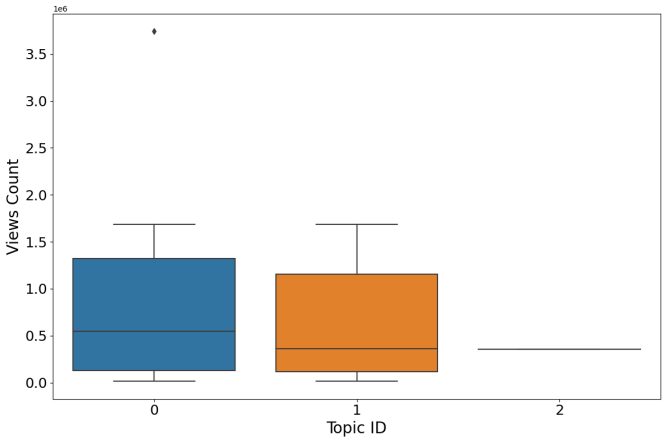


Figure 10: View count distribution across sorted topics based on topic uniformity for Cheng Ho data for focal structure 6. Topic 0 has the highest uniformity and higher median view count. Significant outliers indicate certain videos received disproportionately high attention, suggesting algorithmic bias and reinforcing the content trap.

V. Conclusions and Future Work

This study investigated how YouTube's recommendation network can give rise to content traps by jointly examining their structural, semantic, and behavioral properties. Our analysis shows that focal structures represent highly cohesive sub-networks that play a central role in maintaining the overall connectivity of the recommendation graph. Answering **RQ1**, we found that the removal of these structures causes substantial network fragmentation, indicating their importance as structural attractor groups. Addressing **RQ2**, topic modeling and divergence-based analyses demonstrated that several of these attractors exhibit pronounced topical uniformity, reflecting limited thematic diversity within the content they promote. Finally, in response to **RQ3**, engagement patterns revealed that videos within the most uniform topics receive disproportionately high user attention, suggesting that user behavior and algorithmic amplification jointly reinforce these content traps.

The lack of direct comparison with state-of-the-art community detection and influential node detection methods may seem to be a limitation of this study. However, prior research has shown that focal structure analysis outperforms traditional community and influential node discovery methods [4]. Given that, it can be posited that focal structure analysis offers better insights into content traps than traditional methods. Future work could expand this analysis across other online platforms which could further provide a deeper understanding of the dynamics shaping content flow and user engagement.

Acknowledgements

This research is funded in part by the U.S. National Science Foundation (OIA-1946391, OIA-1920920), U.S. Office of the Under Secretary of Defense for Research and Engineering (FA9550-22-1-0332), U.S. Army Research Office (W911NF-23-1-0011, W911NF-24-1-0078, W911NF-25-1-0147), U.S. Office of Naval Research (N00014-21-1-2121, N00014-21-1-2765, N00014-22-1-2318), U.S. Air Force Research Laboratory, U.S. Defense Advanced Research Projects Agency, the Australian Department of Defense Strategic Policy Grants Program, Arkansas Research Alliance, the Jerry L. Maulden/Entergy Endowment, and the Donaghey Foundation at the University of Arkansas at Little Rock. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding organizations. The researchers gratefully acknowledge the support.

References

[1] M. Bhuiyan and N. Agarwal, "Identification and characterization of content traps in youtube recommendation network", in *eKNOW 2025, The Seventeenth International Conference on Information, Process, and Knowledge Management*, IARIA, 2025, pp. 59–64.

[2] C. Agency, *Youtube statistics, https://www.charleagency.com/articles/youtube-statistics/*, Accessed: April 10, 2025, 2024.

[3] A. Gallagher, L. Cooper, R. Bhatnagar, and C. Gatewood, *Pulling back the curtain: An exploration of youtube's recommendation algorithm*, 2024.

[4] M. Alassad, M. N. Hussain, and N. Agarwal, "Comprehensive decomposition optimization method for locating key sets of commenters spreading conspiracy theory in complex social networks", *Central European Journal of Operations Research*, vol. 30, no. 1, pp. 367–394, 2022.

[5] M. A. Brown *et al.*, "Echo chambers, rabbit holes, and algorithmic bias: How youtube recommends content to real users", *Available at SSRN 4114905*, 2022.

[6] M. M. I. Bhuiyan and N. Agarwal, "Detecting algorithmic homophily in recommendation graphs via weighted topic distribution", *2025 IEEE 37th International Conference on Tools with Artificial Intelligence (ICTAI)*, 2025.

[7] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment", *Journal of the ACM (JACM)*, vol. 46, no. 5, pp. 604–632, 1999.

[8] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine", *Computer Networks and ISDN Systems*, vol. 30, no. 1-7, pp. 107–117, 1998.

[9] P. Bedi and C. Sharma, "Community detection in social networks", *Wiley interdisciplinary reviews: Data Mining and Knowledge Discovery*, vol. 6, no. 3, pp. 115–135, 2016.

[10] F. Sen, R. T. Wigand, N. Agarwal, D. Mahata, and H. Bisgin, "Identifying focal patterns in social networks", in *2012 Fourth International Conference on Computational Aspects of Social Networks (CASoN)*, IEEE, 2012, pp. 105–108.

[11] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks", *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, P10008, 2008.

[12] F. Şen, R. T. Wigand, N. Agarwal, M. Mete, and R. Kasprzyk, "Focal structure analysis in large biological networks", in *3rd International Conference on Environment, Energy and Biotechnology (ICEEB 2014)*, 2014.

[13] F. Şen, R. Wigand, N. Agarwal, S. Tokdemir, and R. Kasprzyk, "Focal structures analysis: Identifying influential sets of individuals in a social network", *Social Network Analysis and Mining*, vol. 6, pp. 1–22, 2016.

[14] M. J. Alenazi and J. P. Sterbenz, "Comprehensive comparison and accuracy of graph metrics in predicting network resilience", in *2015 11th International Conference on the Design of Reliable Communication Networks (DRCN)*, IEEE, 2015, pp. 157–164.

[15] E. Pariser, *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin, 2011.

[16] A. Bechmann and K. L. Nielbo, "Are we exposed to the same "news" in the news feed? an empirical analysis of filter bubbles as information similarity for danish facebook users", *Digital Journalism*, vol. 6, no. 8, pp. 990–1002, 2018.

[17] G. M. Lunardi, G. M. Machado, V. Maran, and J. P. M. de Oliveira, "A metric for filter bubble measurement in recommender algorithms considering the news domain", *Applied Soft Computing*, vol. 97, p. 106 771, 2020.

[18] F. Masrour, T. Wilson, H. Yan, P.-N. Tan, and A. Esfahanian, "Bursting the filter bubble: Fairness-aware network link prediction", in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, 2020, pp. 841–848.

[19] L. Burbach, P. Halbach, M. Ziefle, and A. Calero Valdez, "Bubble trouble: Strategies against filter bubbles in online social networks", in *Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management. Healthcare Applications: 10th International Conference, DHM 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26–31, 2019, Proceedings, Part II 21*, Springer, 2019, pp. 441–456.

[20] A. Amrollahi, "A conceptual tool to eliminate filter bubbles in social networks", *Australasian Journal of Information Systems*, vol. 25, pp. 1–16, 2021.

[21] R. Finlay, "The voyages of zheng he: Ideology, state power, and maritime trade in ming china", *Journal of the Historical Society*, vol. 8, no. 3, pp. 327–347, 2008.

[22] A. M. Dwyer, "The xinjiang conflict: Uyghur identity, language policy, and political discourse", *Policy Studies*, vol. 15, 2005.

[23] R. Hasmath, "What explains the rise of majority–minority tensions and conflict in xinjiang?", *Central Asian Survey*, vol. 38, no. 1, pp. 46–60, 2019.

[24] M. C. Cakmak, O. Okeke, U. Onyepunuka, B. Spann, and N. Agarwal, "Investigating bias in youtube recommendations: Emotion, morality, and network dynamics in china-uyghur content", in *International Conference on Complex Networks and Their Applications*, Springer, 2023, pp. 351–362.

[25] M. C. Cakmak, N. Agarwal, and R. Oni, "The bias beneath: Analyzing drift in youtube's algorithmic recommendations", *Social Network Analysis and Mining*, vol. 14, no. 1, p. 171, 2024.

[26] Google Developers, *YouTube Data API*, https://developers.google.com/youtube/v3, Accessed: April. 10, 2025, 2024.

[27] M. C. Cakmak and N. Agarwal, "High-speed transcript collection on multimedia platforms: Advancing social media research through parallel processing", in *2024 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*, 2024, pp. 857–860. DOI: 10.1109/IPDPSW63119.2024.00153.

[28] G. Zhang *et al.*, "Gder: Safeguarding efficiency, balancing, and robustness via prototypical graph pruning", *Advances in Neural Information Processing Systems*, vol. 37, pp. 50 285–50 312, 2024.

[29] J. Li *et al.*, "Less can be more: Unsupervised graph pruning for large-scale dynamic graphs", *arXiv preprint arXiv:2305.10673*, 2023.

[30] M. Bhuiyan, S. Shajari, and N. Agarwal, "Resilience and node impact assessment in youtube commenter networks leveraging focal structure analysis", *The Eleventh International Conference on Human and Social Analytics (HUSO 2025)*, 2025.

[31] M. Grootendorst, "Bertopic: Neural topic modeling with a class-based tf-idf procedure", *arXiv preprint arXiv:2203.05794*, 2022.

[32] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation", *Journal of Machine Learning Research*, vol. 3, no. Jan, pp. 993–1022, 2003.

[33] S. Kullback and R. A. Leibler, "On information and sufficiency", *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[34] J. Lin, "Divergence measures based on the shannon entropy", *IEEE Transactions on Information Theory*, vol. 37, no. 1, pp. 145–151, 1991.

Millimeter-Wave Information-Centric Wireless-Sensor-Network Ecosystem: Evaluation under Non-Terrestrial and Long-Distance Environments in Actual City

Shintaro Mori
Department of Electronics Engineering and Computer Science
Fukuoka University
8-19-1 Nanakuma, Jonan-ku, Fukuoka 814-0180, Japan
E-mail: smori@fukuoka-u.ac.jp

Abstract — **This paper presents an evaluation of the previously presented information-centric wireless-sensor-network-based framework for smart-city applications, called mmICWSN. The framework uses millimeter-wave communications for future broadband wireless networks. To demonstrate the feasibility, the network performance, including application and network-layer throughput, was evaluated, and the video-streaming application was demonstrated in a non-terrestrial environment and in an actual city. In addition, the computer simulations are also performed. The experiments were conducted in the KOIL mobility field (Chiba, Japan) and Nogata City (Fukuoka, Japan). The demonstrations include the proof of concept for constructing a wireless network via multiple aerial nodes and establishing long-range wireless links in an actual city, connecting two 1-km-distant locations. The results indicate that our ecosystem can be verified to communicate with a point-to-point environment.**

Keywords-information-centric wireless sensor network; millimeter-wave communications; smart-city ecosystem.

I. INTRODUCTION

Internet of Things (IoT) and Wireless Sensor Network (WSN) technologies are essential and foundational for supporting smart-city applications to collect and distribute sensing data [2]. Traditional host-centric networking framework has limitations in managing sensing data due to addressing, inefficient mobility support, and unsophisticated in-network caching. Building IoT on top of Information-Centric Networking (ICN) [3][4] is believed to be a promising solution to tackle the above issues. ICN is an ideal candidate for future network architecture that shifts the focus from host locations to data. In short, ICN names data rather than addresses; hence, end-users can discover and obtain data on the basis of name, resulting in network abstraction. The data are copied and stored in cache memory for subsequent retrievals and can be self-certified and encrypted to improve security.

Underpinning future smart-city applications, wide-band and low-latency wireless communications will be necessity, thereby motivating to explore higher-frequency bands. In other words, it is necessary to consider using not only both Microwave bands (microWave) and Sub-gigahertz bands (sub-gigahertzWave), but also Millimeter wave bands (mmWave) [5]. Currently, mmWave-band radio is primarily used for radar and academic astronomy, while wireless communication is not so popular.

In the previous studies, a mmWave Information-Centric Wireless Sensor Network (mmICWSN) framework has been investigating and the framework integrates the three technologies, such as mmWave, ICN, and WSNs[6][7][8]. Among them, the mmICWSN test field was implemented and a preliminary evaluation was performed [6]. For practical operational demonstrations, long-term operational testing was conducted at the same field [7]. Through the experiments, serious weather effects in mmWaves were not observed, and the beam angle had a greater impact on network performance. In addition, a ground-to-air mmWave experiment using a single drone was conducted, assuming a Non-Terrestrial network (NTN) environment [8].

This paper addresses the following two matters that are insufficiently investigated in these studies, i.e., multi-hop air-to-air mmWave communications and experimental verification of long-distance mmWave communications. The baseline paper [1] presented some network-performance results; in this paper, additional evaluation results for an aerial multi-hop scenario and computer simulations are provided.

The remainder of this paper is organized as follows. Section II discusses related work. Section III provides a brief overview of the mmICWSN. Sections IV and V present the experimental evaluation results of the NTN environment and the long-distance demonstration, respectively. Sections VI presents the computer simulation result regarding IEEE 802.11 ad/ay-based communications to complement the experimental results. Finally, Section VII concludes this paper with a summary and mention of future work.

II. RELATED WORK

In the ecosystems of smart cities, IoT technology is an essential component due to its ability to utilize sustainable information and communication technologies [9]. As a wireless communication system underpinning smart-city application services, a comprehensive review and survey of the future evolution of next-generation technologies, including their principles, potential applications, current state-of-the-art research, and the related technical challenges, was surveyed [10][11][12]. The wireless backhaul has become a key enabler for future mobile communications systems, offering a cost-effective and scalable alternative to traditional fiber backhaul. Thanks to the availability of high bandwidths, data can be transmitted over fiber; however, the higher radio-

frequency bands suffer from propagation loss, distortion, and blockage. Ferreira et al. [13] presented an extensive measurement campaign and cross-layer analysis for outdoor mmWave environments. Doone et al. [14] used a commercial ray tracing engine to investigate two hypothetical use cases within future factories, such as inventory monitoring using an UAV and movement of stock using a forklift vehicle at microWaves and mmWaves.

Regarding the network performance of mmWaves, Zhang et al. [15] and Khorov et al. [16] surveyed the open challenges for transport- and network-layer protocols on the basis of a comprehensive simulation study. In particular, transport- and network-layer protocols are unsuitable for mmWaves because of their specific features compared with those of commonly used bands, due to high signal attenuation and blockage. Kumar et al. [17] experimentally found throughput degradations (collapses) in the 60-GHz band (one of the key frequency bands for mmWaves). Poorzare et al. [18] analyzed network performances related to this phenomenon under an urban deployment scenario.

Regarding ICN technology for smart-city applications, current implementation proposals for ICN-based IoT, as well as the caching and replacement policies presented, were surveyed [19]. Safitri et al. [20] proposed a mobile IoT optimization method for next-generation networks by evaluating a series of name-based techniques implemented in ICN. Gur et al. [21] found that consolidating ICN with MEC technology offers new opportunities to realize that vision and serve advanced use cases.

III. MMWAVE INFORMATION-CENTRIC WIRELESS SENSOR NETWORK FRAMEWORK

Figure 1 shows the network structure of the mmICWSN framework. In the framework, there are three types of network nodes: Private Base Station (PBS), Relay Node (RN), and Sensor Node (SN). PBS is a coordinator that manages a regional (local) network deployed in the field, and it acts as a gateway to external networks. RNs use mmWaves for backhaul and can be further classified into Ground RNs (GRNs) and Aerial RNs (ARNs) based on node location, such as NTNs. GRN provides high-performance relay functions under commercial power supply, while ARN mediates sensing data from ground-based SNs via Unmanned Aerial Vehicles (UAVs) with battery power.

On the other hand, RNs can be alternatively classified into mmRN and μRN, which utilize radio wavelengths at mmWaves and microWaves, respectively. The mmRN is a repeater node that enables high-speed wireless mesh networks to replace optical fiber networks in the field. If we construct a network using only mmRN, the construction costs will be expensive, and backward compatibility cannot be guaranteed. To mitigate this situation, μRN can promotes a low-cost deployment, enabling the usage of the 6-GHz band as well as the 5-GHz band. SNs are deployed in the field, and the sensing data are packaged and posted to the network. Sensing data includes not only text-based data, such as temperature, humidity, and illuminance, but also large-sized data, such as images, videos, and 3D sensing data. In addition, the data includes not only data that can tolerate delays, but also real-time and delay-sensitive data.

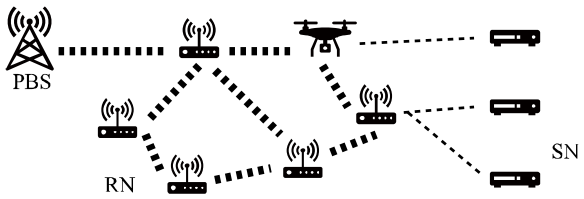


Figure 1. Network model of the mmICWSN framework

Regarding the selection of wireless communication systems for smart cities, particularly for outdoor environments such as smart agriculture, there are several options. Namely, there are cellular and satellite networks, Low-Power Wide-Area Networks (LPWANs), Personal Area Networks (PANs), Wireless Local Area Networks (WLANs), and, exceptionally, optical wired networks [22][23][24]. These network systems have both strengths and weaknesses; therefore, it is necessary to determine where they are suitable for deployment in terms of communication coverage, network communication (wireless) capacity, and the economic and technical costs of implementation, construction, and deployment.

Cellular and satellite telecommunications are the de facto systems, but their operation costs are high. LPWANs, such as LoRa and SigFox, have been widely adopted to provide wide-area coverage and low energy consumption. However, LPWANs are generally assigned 100-Hz bandwidth in the sub-gigahertz band; thus, the system can only transfer small amounts of data. PANs are used as traditional WSNs, but the deployment is limited to environments inside small-area networks. The wired network is the primary choice in areas where optical fiber lines have already been deployed; however, deploying new optical lines is unrealistic in rural areas due to economic reasons.

In contrast, the mmICWSN framework adopted the network structure of WLANs based on the IEEE 802.11 standard, also known as Wi-Fi (certified by its alliance). WLANs are widely recognized as another global communication system in comparison with cellular networks. In addition, the network system has several advantages, including low-cost wireless modules readily available, the ability to be constructed on IP networks, and the availability of unlicensed radio-frequency bands without regulations.

IV. EXPERIMENTAL RESULTS OF THE NTN ENVIRONMENT

In this section, we present a demonstration and an experiment on backhaul network construction using ARNs to answer the fundamental question: whether a high-speed wireless link can be established among aerial networks. The experiment was conducted in a simulated drone environment, using smart poles rather than actual drones, to demonstrate the feasibility of deploying multiple ARNs. In addition to evaluating fundamental network performance, we conducted video streaming trials to demonstrate that the mmICWSN framework can support delay-sensitive applications.

A. Experimental environment and setup

The experiment was conducted in the KOIL mobility field (Chiba, Japan), and its field view is shown in Figures 2 and 3. As shown in Figure 4, the field consists of three areas: a paved road for autonomous vehicle testing, a gravel (ballast) area for heavy vehicle testing, and a grass (weed) area for agricultural machinery testing. In the experiment, to deploy backhaul networks among mmRNs, we used Terragraph [25], which Meta (Facebook) offered as an mmWave mesh network to deploy an IEEE 802.11 ay-compliant network. Distribution Nodes (DNs) and Client Nodes (CNs) are used. With multiple DNs are interconnected to form a backhaul network that enables end-users to access the network via CNs. Note that Terragraph can communicate with DNs and CNs via multi-hop transmissions with a maximum of 15 hops, and the router node supports the Open/R routing protocol. In addition, Terragraph is compatible with the IEEE 802.11 ad/ay specification, but the device we used (BeMap MLTG-360 as a DN and MLTG-CN [26]) only supported single-carrier modulation at the physical layer.

Regarding the wireless communication devices, the maximum Effective Isotropic Radiated Powers (EIRP) are 43 dBm (DN) and 38 dBm (CN) at the transmitter, and the antenna gains are 28 dBi (DN) and 22 dBi (CN), respectively. In particular, their antenna consists of a phased array with 64 elements, and the steering angular ranges are [−45°, 45°] in the azimuth plane (φ) and [−25°, 25°] in the elevation plane (θ). Under Japan's Radio Act, Terragraph is assigned the unlicensed 60-GHz band (57–66 GHz) with four channels: 58.32, 60.48, 62.64, and 64.80 GHz (central) frequency bands, each with a 2.16-GHz bandwidth. For the ICN platform, we used Cefore [27], an open-source CCNx-based platform available on Linux (Ubuntu). Cefore consists of two daemon processes: cefnetd and csmgrd. Namely, cefnetd exchanges the data and forwards interest packets, and csmgrd provides an in-network caching scheme.

To experiment in the NTN environment, we utilized smart poles [28]. This smart-pole-based airspace testbed validates the feasibility of drone networks and services in real-world settings, as illustrated in Figure 3. The smart-pole platform was utilized for several practical reasons, such as field restrictions and the high cost of drones. The smart pole has several propellers that can simulate drone movement, as shown in Figure 3(a). The smart pole requires two or more persons to lift it up, but it can stand on its own once the propellers are rotated. Namely, the propeller rotation can be automatically controlled to maintain stability by adjusting the angle of the pole, as shown in Figure 3(b).

The experimental devices were arranged in two layouts: all devices in a straight line (Figure 4(a)) and the smart poles at 90-degree angles to each other, forming a zigzag pattern (Figure 4(b)). Smart poles were placed on the gravel areas, and an SN was placed on the paved road. The distance between each wireless node was set to 11.7–29.4 m, which is summarized in Figure 4. For the three wireless sections, we have assigned a radio channel from four channels specified in IEEE 802.11 ay without any overlap, and we have suspended any other mmWave network in the field to prevent radio-wave

Figure 2. Field view of experimental site in the KOIL mobilty field

(a) (b)

Figure 3. Smart pole in which a city airspace testbed system

interference. The reasons for assuming these layouts are as follows. When ARN mediates SN in cooperation with GRNs that are placed in a grid (lattice) pattern around the field, the scenario involves UAVs being aviated along the vertical (or horizontal) lines. The scenario where all devices are placed in a straight line is reasonable for this situation. On the other hand, for the zigzag pattern, when the GRNs that are usually connected to the ARN cannot transfer data, another neighbor GRN will be switched.

Figure 5 shows the experimental network. As a PBS, Terragraph DN was placed on the fixed pole. We used a PC with Ubuntu installed on a Microsoft Surface as a user terminal, and the PC was connected to the PBS. The network links, including air-to-air and air-to-ground, were three-hop mmWaves. The pair of communication devices was utilized with two Terragraph CNs on the pole head, as shown in Figure 3(a). Since the weight limit for pole loadable, the mobile battery for power supply, and the control computer are located on the ground. The antenna was set to 4.2 m height, and its surface was oriented toward the opposite device. SN was used to implement a node with a web camera and CN.

B. Experimental results: Network performance

Figure 6 shows the experimental results, which include both (a) and (b), (c) and (d), and (e) and (f), respectively, for

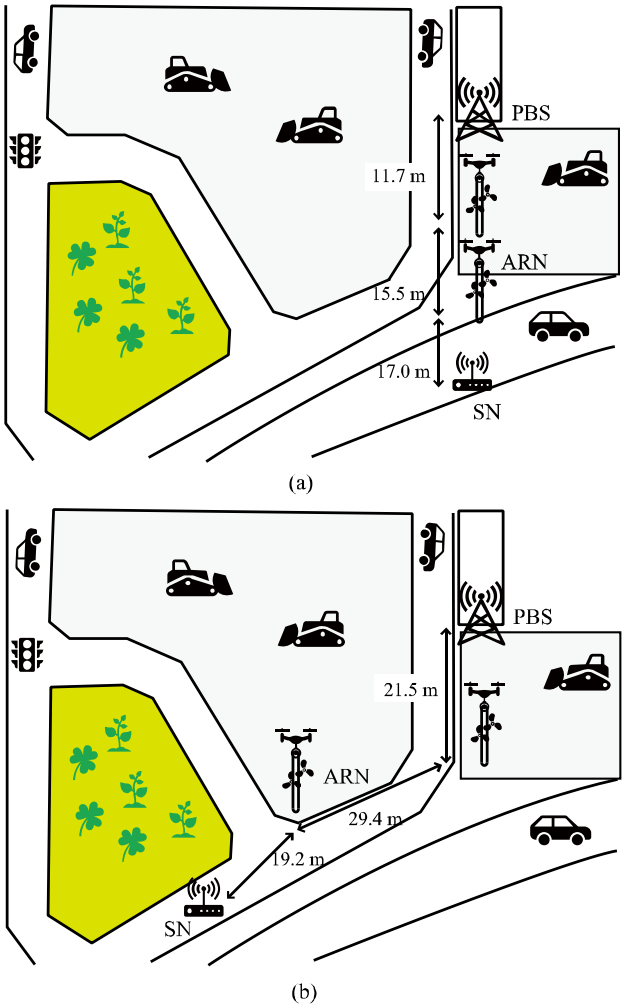


Figure 4. Field layout and node placement

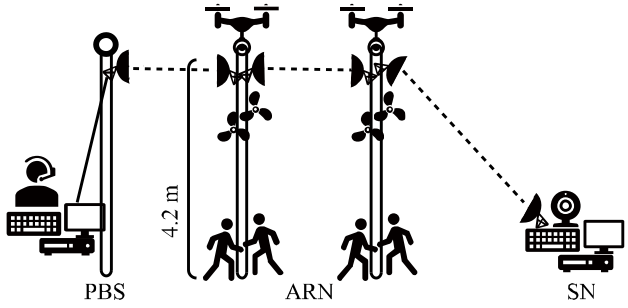


Figure 5. Experimental netowrk for mmWaves among ARNs

Transmission Control Protocol (TCP) with the CUBIC algorithm, User Datagram Protocol (UDP), and ICN performance. In Figures 7(a) to (d), iPerf3 [29], a well-known network-performance measurement tool, was used to measure TCP/UDP performance at 1 s interval for 90 s. Figures 6(e) and (f) show the results of retrieving the different data using Cefore.

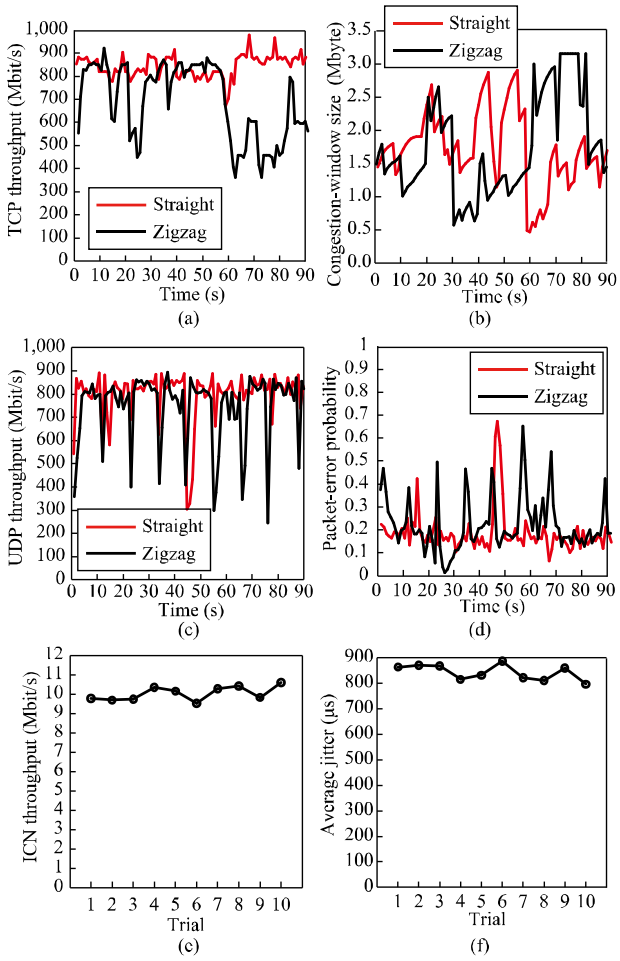


Figure 6. Experimental results regarding network performance in the NTN environment

As shown in Figure 6(a), the average TCP throughput was 841 and 693 Mbit/s for straight and zigzag placements, respectively. Compared to the TCP throughput of straight placement, that of zigzag was 17.7% lower, but it remained above several hundred megabits per second. For TCP congestion control, as shown in Figure 6(b), the average congestion-window size were 1.72 and 1.75 Mbytes; hence, there was no significant difference. As shown in Figure 6(c), the average UDP throughput was 805 and 752 Mbit/s for the two placements, respectively. Figure 6(d) shows the packet-error probability for UDP transfer; the averages were 0.193 and 0.225 for the two placements, respectively, in which there was no significant difference. As shown in Figure 6(e), the average ICN throughput was 10.0 Mbit/s for the straight placement. As shown in Figure 6(f), the average jitter was 843 μs for the straight placement. ICN throughput and jitter did not significantly differ between the two scenarios. In addition, the ICN throughput was significantly smaller than that of TCP or UDP because Cefore has a bottleneck. We also experimentally verified that we could obtain sufficient network performance for mmICWSN via multiple ARNs.

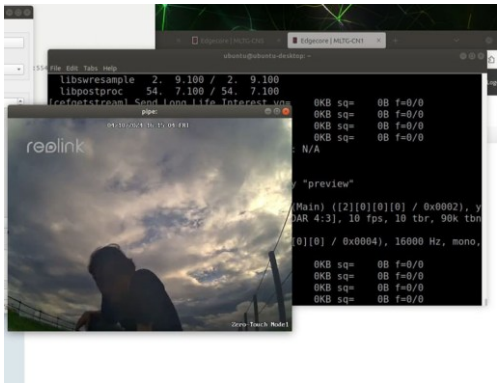


Figure 7. Experimental results regarding video-streaming demonstration in the NTN environment

Figure 8. Location map of transmitter- and receiver-side nodes

C. Experimental results: Video-streaming demonstration

In the experiments of this section, the camera video equipped with the SN device, shown in Figure 5, was retrieved as streaming data at the user terminal. Unlike the previous section's experiment, the cefgetstream and sefputstream commands in Cefore were used to transmit and retrieve real-time video streams. Note that, to integrate Cefore into the device, we can register and obtain the data from the application software using the following commands: cefputfile and cefgetfile for sending and receiving static data, respectively. Figure 7 shows the results of the video streaming experiment. The results show that the video can be delivered smoothly over the three-hop mmWave links in air-to-air environment. In addition, with the availability of video streaming applications, the mmICWSN framework can be effective for several applications requiring low-latency data flow.

V. Experimental Results of Long-Distance Communications in Actual City

As another proof-of-concept experiment in this paper, the feasibility of long-distance communication in an actual city deployment was evaluated. To develop ecosystems in actual cities, additional evaluations are necessary. To the best of our knowledge, there have been few experiments on mmWave long-distance data transmission; therefore, this paper's contribution is valuable.

A. Experimental environment and setup

The node devices were deployed at a community center and school in Nogata City (Fukuoka, Japan), as shown in Figure 8. The community center and school are three-story buildings, and the node devices were placed on their rooftops. In accordance with the three-dimensional map provided by the National Geographical Institute [30], their altitudes are respectively 7.5 and 16 m, and the straight-line distance between them is 1 km. Across the wireless link, there is a river, a road, a bridge, and a car park, as shown in Figure 9, which might affect radio propagation. The river is the Onga River and the riverside area is well maintained and covered with

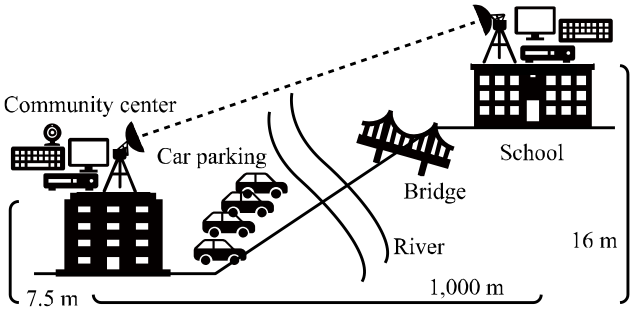


Figure 9. Outline of experimental environment regarding long-distance testing

grass and aquatic plants. During the experiment, the river surface was flat and calm, with no significant waves, i.e., factors affecting mmWaves propagation were not observed. The Kanroku Bridge spans the river and connects to the main national road. Nogata City is an inter- and suburban city between large cities (e.g., Fukuoka City and Kitakyushu City), but traffic is not dense. The riverside area in front of the community center is used as a parking lot, with several dozen cars parked there.

Figure 10 shows the field view of the experimental site. Figures 10(a) and (b) and Figures 10(c) and (d) show the field views of the rooftops of the community center and school, respectively. As shown in Figures 10(a) to (d), the mmICWSN node devices were connected to the Terragraph devices. In the experiment, two MLTG-CNLR devices communicated over distances up to 1 km, as specified in the catalog. In the MLTG-CNLR device, the EIRP is 56 dBm, and the antenna gain is 40 dBi. In addition, the MLTG-CNLR's antenna is a dish (parabola) type, and its scan range and beam width are 3° and 1°, respectively (that is, narrow compared with the former two devices). Throughout the experiment, the Modulation and Coding Scheme (MCS) index was automatically set as 9. Table I shows the parameter settings regarding adaptive rate control in the IEEE 802.11 ay. Note that IEEE 802.11-compliant Wi-Fi systems achieve effective (high-throughput) data transmission using the control modulation scheme, the error-correcting code rate, and a

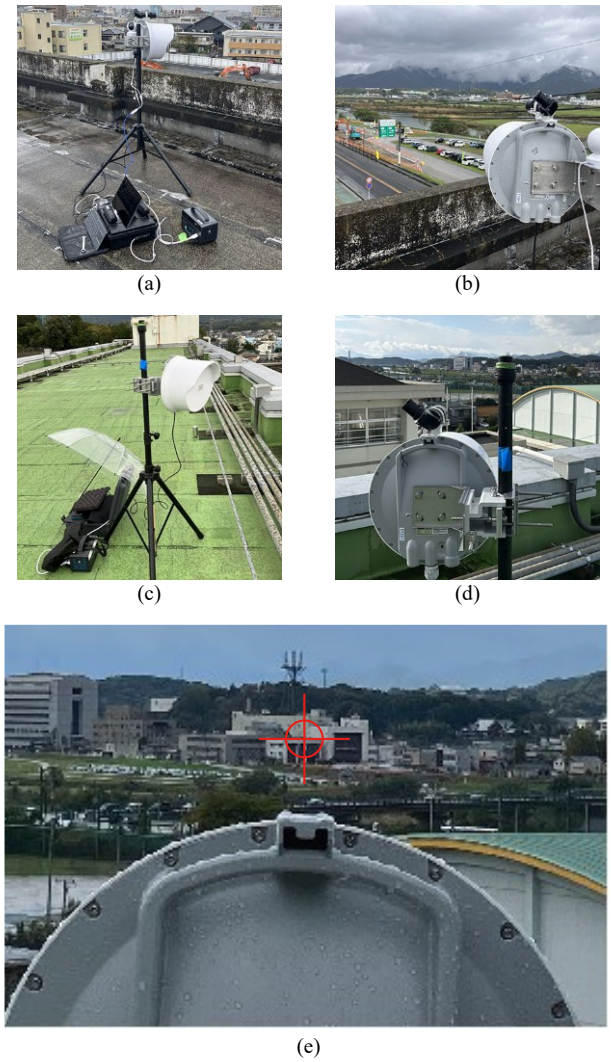


Figure 10. Field view of transmitter- and receiver-side nodes on rooftop of community center and school buildings

repetition code based on the wireless channel condition, and their combination is predefined as the MCS settings.

Figure 10(e) shows a photo taken behind the dish antenna on the school rooftop toward the community center. The community center is located at the red marking, where the opposite node was placed. As shown in Figure 10(e), the line of sight between the transmitter- and receiver-side nodes can be clearly maintained. The weather was cloudy during the experiment. Note that mmWaves have been used as an alternative for backhaul, short-range and high-capacity indoor communications, and radar. Compared with the radio-frequency bands currently widely used, additional attenuation in the mmWave link budget, such as rain, oxygen, and hydrophilic materials (e.g., trees, leaves, and humans) must be considered. Note that radio waves in the 60-GHz band are particularly affected by the rain and oxygen. Nevertheless, the weather did not significantly affect network performance, and

TABLE I. MCS SETTINGS IN SINGLE CARRIER PHYSICAL MODE

Index	*Modulation method*	*Code rate*	*Repetition*	*Data rate* (Mbit/s)
1	BPSK	1/2	2	385
2	BPSK	1/2	1	770
3	BPSK	5/8	1	963
4	BPSK	3/4	1	1,155
5	BPSK	13/16	1	1,251
6	QPSK	1/2	1	1,540
7	QPSK	5/8	1	1,925
8	QPSK	3/4	1	2,310
9	QPSK	13/16	1	2,503
10	16-QAM	1/2	1	3,080
11	16-QAM	5/8	1	3,850
12	16-QAM	3/4	1	4,620

TABLE II. PHYSICAL-LAYER INFORMATION IN EXPERIMENT

Terms	*Antennas are matched*	*Antennas are mis-matched*
Radio channel	Ch 2 (60.48 GHz with 2.16 GHz)	
RSSI	-64 dBm	-62–63 dBm
MCS settings	8–9	6–9
Beam index	30 / 30	30 / 5

the beam direction was severely affected, as indicated by the experimental results [8]. Namely, the experiment was conducted in two scenarios: one in which both elevation and azimuth angles were appropriately adjusted (the antennas were matched), and the other in which they were slightly offset (the antennas were mismatched).

B. Experimental results: Network performance

Figure 11 shows the experimental results, which include both (a) and (b), (c) and (d), and (e) and (f) are TCP with the CUBIC algorithm, UDP, and ICN performance, respectively. In Figures 11(a) to (d), iPerf3 [36] was used to measure TCP/UDP performance at every 1 s interval for 90 s. Figures 11(e) and (f) show the results of retrieving the different data using Cefore. The status information of the physical layer for these scenarios is summarized in Table II. Note that, in the CNLR device, the antenna's front space is divided into a grid pattern of elevation- and azimuth- angles, and each sub-region is assigned a beamforming index. The most central beam direction on the antenna surface is when the beamforming index is 30.

As shown in Figure 11(a), the average TCP throughput was 941 and 94.4 Mbit/s when the antennas were matched and mismatched, respectively. The TG antenna is a parabolic dish; thus, even a few degrees of angular misalignment can cause significant degradation in TCP throughput. For TCP congestion control, as shown in Figure 11(b), the average congestion-window sizes were 1.26 and 0.967 Mbytes; hence, there was a 39.3% difference. In the curve, when the antennas were matched, several attempts were made to increase the congestion-window size.

As shown in Figure 11(c), the average UDP throughput was 902 and 93.3 Mbit/s for the two scenarios, respectively. In the curve, when the antennas were matched, there were

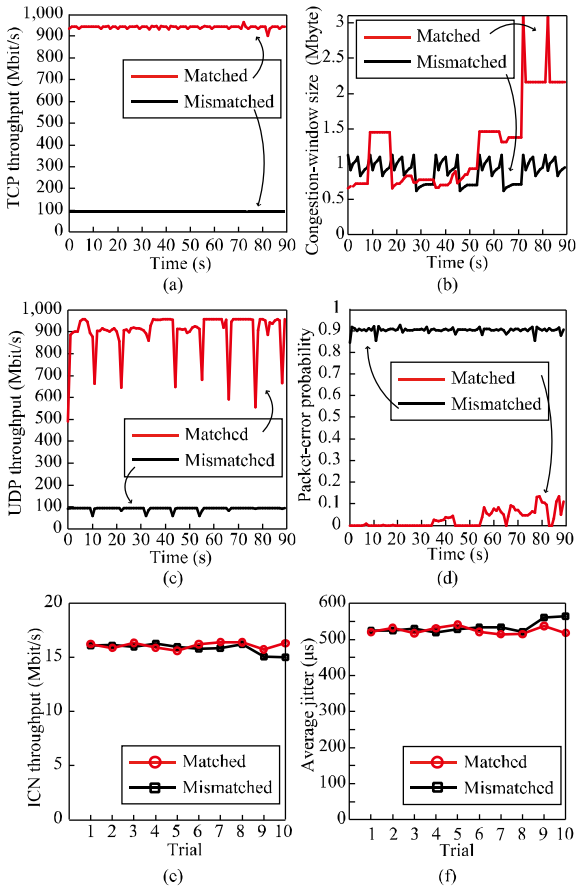


Figure 11. Experimental results in actual city

regions where UDP throughput temporarily decreased. The reason for this decrease is that automatic retransmission requests and forwarding-error-control mechanisms are omitted, resulting in these dramatic degradations. The results in Figure 11(a) indicate no degradation, as the TCP congestion control mechanism is available and effective. Figure 11(d) shows the packet-error probability for UDP transfer; the averages were 0.0294 and 0.903 for the matched and mismatched scenarios, respectively. When the antennas were mismatched, many packet losses occurred, affecting not only UDP throughput but also TCP throughput, as shown in Figures 11(a) and (c).

As shown in Figure 11(e), the average ICN throughput was 16.1 and 15.8 Mbit/s for the two scenarios, respectively. The ICN throughput was significantly smaller than that of TCP or UDP because Cefore has a bottleneck. In mmICWSN, the ICN layer was stacked on the TCP/UDP layers. Thus, due to the middleware implementation in Cefore, if the maximum data-transmission bandwidth is set to its maximum value, the failure probability of data registration, storage, and transfer worsens. As shown in Figure 11(f), the average jitter was 525 and 534 μs for the two scenarios, respectively. ICN throughput and jitter did not significantly differ between the two scenarios. In line with these results, we found that

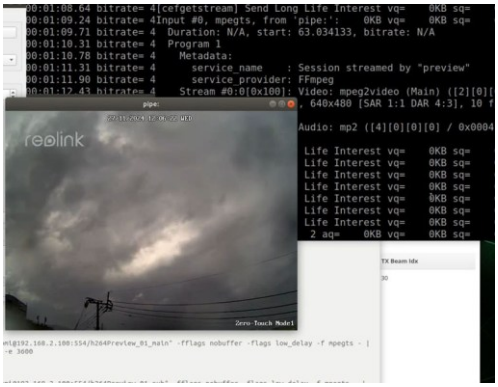


Figure 12. Experimental results regarding video-streaming demonstration in the long-distance environment

TCP/UDP/IP protocol stacks were not affected by the ICN-layer protocol. We also experimentally verified that we could obtain sufficient network performance for mmICWSN in an actual city.

C. *Experimental results: Video-streaming demonstration*

Similar to Section IV.C, the video-streaming experiment was conducted. As shown in Figure 9, the camera was located on the device at the community center, and the device received the video stream at the school. Figure 12 shows the experimental result in which the cefgetstream and sefputstream commands in Cefore were used to transmit and receive real-time video streams. Under matched and mismatched conditions, the video stream was delivered and received smoothly. Although there were differences in UDP network performance between the two conditions, the ICN performance did not differ. Therefore, we can conclude that the degradation in network performance did not affect the ICN layer for video-streaming applications.

VI. COMPUTER SIMULATION

In this section, the computer simulations regarding the frame-error-probability performance of mmWaves were performed. As a simulation environment, Mathworks Matlab (2025b) on a PC (Panasonic Let's Note FV1 (Core i7 1185 G7 (4-core, 3 GHz), 32-GB RAM, and Windows 11 Pro OS) was used. The parameter settings of the transceiver and receiver were determined based on the previous experiments. Namely, for the simulation scenario between ARNs, the height between the transmitter and receiver was set as 4.2 m, and their distance was set as 10, 20, and 30 m, respectively. In addition, as in the case of ARN and SN, the heights of the transmitter and receiver were set to 4.2 m and 1.2 m, respectively, and the distance between them was set to 20 m. The antenna has a 16 (= 4×4) element-based array layout, and the two antenna surfaces are placed face-to-face between the transmitters and receivers. In addition, we assume that the beamforming mechanism ideally works. In the radio-propagation environment, the transmission-side device is regarded as an access point (hotspot), located in an open space, and the receiver-side device is connected to it. For the wireless channel model, there was no terminal movement, including

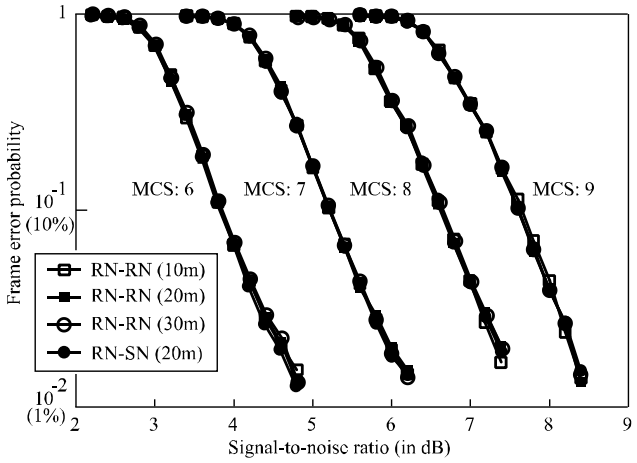


Figure 13. Computer-simulation result: frame-error probability versus signal-to-noise ratio in decibel

transmitter and receiver; thus, we assume it as an Additive White Gaussian Noise (AWGN) channel. In adaptive rate control, the MCS index was evaluated based on the actual observed values in Section IV B, specifically for the range of 6–9.

The computer simulation was based on exhaustive Monte Carlo simulations, with the number of iterations set to either 1,000 frame error detections or 10,000 frame transmissions. In addition, the frame length was set as 4,096 bytes (= 32,768 bits). Figure 13 shows the results of the frame-error probability versus the Signal-to-Noise Ratio (SNR) in decibels. As shown in Figure 13, the curves for the four cases were overlapped (similar); therefore, there is no difference between the node distances in these cases. The desired SNR at a frame error rate of 10% and 5% were 3.8, 5.2, 6.6, and 7.7 dB, and 4.2, 5.5, 7.0, and 8.0 dB for the case where the index of MCS was 6, 7, 8, and 9, respectively.

VII. Conclusion and Future Work

This paper evaluated and presented the feasibility of the network performance in the TCP, UDP, and ICN protocols with mmICWSN framework. The experimental results indicated that it was necessary to improve the ICN throughput by modifying the Cefore settings, and the antenna placement for mmWaves was sensitive to a few degrees of angle. Through the demonstration of the mmWaves experiment, the developed system could be applicable to multi-hop aerial nodes and long-distance wireless transmission in an actual city. For future work, we plan to deploy mmICWSN for practical smart-city applications, such as smart agriculture. In detail, we will develop a new ecosystem that supports an on-demand and real-time video and image forwarding platform for common demand in smart agriculture applications.

Acknowledgments

This work was partly supported by JSPS KAKENHI Grant Number JP25K15104 and NICT Japan, Grant Number JPJ012368C05601. We are grateful to Dr. Kenji Kanai and Dr. Taku Yamazaki for their helpful discussions, and to Nogata City and Advantech Japan for their help with the experiments.

References

[1] S. Mori, “Network-performance evaluation for millimeter-wave information-centric wireless-sensor-network ecosystem in actual city,” *Proc. IARIA ICN 2025*, pp. 7–11, Nice, France, May 2025.

[2] P. Mishra and G. Singh, “6G-IoT framework for sustainable smart city: Vision and challenges,” *IEEE Consumer Electronics Mag.*, pp. 1–8, Sept. 2023.

[3] B. Ahlgren, C. Dannewitz, C. Imbrenda, D. Kutscher, and B. Ohlman, “A survey of information-centric networking,” *IEEE Communications Mag.*, vol. 50, no. 7, pp. 26–36, July 2012.

[4] L. C. M. Hurali and A. P. Patil, “Application areas of information-centric networking: State-of-the-art and challenges,” *IEEE Access*, vol. 10, pp. 122431–122446, Nov. 2022.

[5] K. Aldubaikhy, W. Wu, N. Zhang, N. Cheng, and X. Shen, “MmWave IEEE 802.11ay for 5G fixed wireless access,” *IEEE Wireless Communications*, vol. 27, no. 2, pp. 88–95, Apr. 2020.

[6] S. Mori, “Test-field development for ICWSNs and preliminary evaluation for mmWave-band wireless communications,” *Proc. IEEE CCNC 2024*, Las Vegas, USA, Jan. 2024, pp. 1–2, doi: 10.1109/CCNC51664.2024.10454799.

[7] S. Mori, “Air-to-ground-integrated wireless information-centric networks with mmWave communications: Long-term operational testing and aerial node evaluation,” *Proc. IEEE APWCS 2024*, Singapore, Aug. 2024, pp. 1–5, doi: 10.1109/APWCS61586.2024.10679312.

[8] S. Mori, “Development of UAV-aided information-centric wireless sensor network platform in mmWaves for smart-city deployment,” *International J. Advances in Networks and Services*, vol. 17, no. 3&4, pp. 105–115, Dec. 2024.

[9] O. J. Adeleke, K. D. Jovanovich, S. Ogunbunmi, O. Samuel, and T. O. Kehinde, “Comprehensive exploration of smart cities: A systematic review of benefits, challenges, and future directions in telecommunications and urban development,” *IEEE Sensors Reviews*, vol. 2, pp. 228–245, May 2025.

[10] F. Tariq et al., “A speculative study on 6G,” I*EEE Wireless Communications*, vol. 27, no. 4, pp. 118–125, Aug. 2020.

[11] M. Alsabah et al., “6G wireless communications networks: A comprehensive survey,” *IEEE Access*, vol. 9, pp. 148191–148243, Nov. 2021.

[12] M. Vaezi et al., “Cellular, wide-Area, and non-terrestrial IoT: A survey on 5G advances and the road toward 6G,” *IEEE Communications Surveys & Tutorials*, vol. 24, no. 2, pp. 1117–1174, Feb. 2022.

[13] T. Ferreira et al., “Millimeter-wave feasibility in 5G backhaul: A cross-layer analysis of blockage impact,” *IEEE Access*, vol. 11, pp. 5178–5192, Jan. 2023.

[14] M. G. Doone et al., “Signal propagation characteristics at 6 GHz and 60 GHz for Wi-Fi 8 UHR in future factories,” *Proc. AP-S/INC-USNC-URSI 2024*, Firenze, Italy, July 2024, pp. 2189–2190, doi: 10.1109/AP-S/INC-USNC-URSI52054.2024.10687229.

[15] M. Zhang et al., “Will TCP work in mmWave 5G cellular networks?,” *IEEE Communications Mag.*, vol. 57, no. 1, pp. 65–71, Jan. 2019.

[16] E. Khorov, A. Krasilov, M. Susloparov, and L. Kong, “Boosting TCP & QUIC performance in mmWave, Terahertz, and Lightwave wireless networks: A survey,” *IEEE Communications Surveys & Tutorials*, vol. 25, no. 4, pp. 2862–2891, Aug. 2023.

[17] R. Kumar et al., “TCP BBR for ultra-low latency networking: Challenges, analysis, and solutions,” *Proc. IFIP Networking 2019*, Warsaw, Poland, May 2019, pp. 1–9, doi: 10.23919/IFIPNetworking.2019.8816856.

[18] R. Poorzare and A. C. Augé, “How sufficient is TCP when deployed in 5G mmWave networks over the urban deployment?,” *IEEE Access*, vol. 9, pp. 36342–36355, Mar. 2021.

[19] Z. Zhang et al., "In-network caching for ICN-based IoT (ICN-IoT): A comprehensive survey," *IEEE Internet of Things Journal*, vol. 10, no. 16, pp. 14595–14620, Aug 2023.

[20] C. Safitri, Q. N. Nguyen, M. Anugerah Ayu, and T. Mantoro, "Robust implementation of ICN-based mobile IoT for next-generation network," *Proc. IEEE ICCED 2022*, Sukabumi, Indonesia, July 2022, pp. 1–5, doi: 10.1109/ICCED56140.2022.10010561.

[21] G. Gür et al., "Integration of ICN and MEC in 5G and beyond networks: Mutual benefits, use cases, challenges, standardization, and future research," *IEEE Open Journal of the Communications Society*, vol. 3, pp. 1382–1412, Aug. 2022, doi: 10.1109/OJCOMS.2022.3195125.

[22] A. Pagano, D. Croce, I. Tinnirello, and G. Vitale, "A survey on LoRa for smart agriculture: Current trends and future perspectives," *IEEE Internet of Things Journal*, vol. 10, no. 4, pp. 3664–3679, Feb. 2023.

[23] M. N. Mowla, N. Mowla, A. F. M. S. Shah, K. M. Rabie, and T. Shongwe, "Internet of Things and wireless sensor networks for smart agriculture applications: A survey," *IEEE Access*, vol. 11, pp. 145813–145852, Dec. 2023.

[24] A. U. H. Hashmi et al., "Effects of IoT communication protocols for precision agriculture in outdoor environments," *IEEE Access*, vol. 12, pp. 46410–46421, Mar. 2024.

[25] Terragrah, https://terragraph.com/ (retrieved: Nov. 2025).

[26] BeMap, https://www.bemap.co.jp/ (retrieved: Nov. 2025).

[27] Cefore, https://cefore.net/ (retrieved: Nov. 2025).

[28] T. Yamazaki, S. Miyata, and T. Miyoshi, "A city airspace testbed for drone networks in future smart cities," *Advances in Engineering and Information Science Toward Smart City and Beyond*, pp. 141–162, May 2023.

[29] iPerf3, https://iperf.fr/ (retrieved: Nov. 2025).

[30] NGI, https://www.gsi.go.jp/ENGLISH/ (retrieved: Nov. 2025).