

Load Balancing Strategies for Terrestrial and Non-Terrestrial Network Coexistence

Francisco Cassimiro de Sousa Albuquerque, Francisco Rodrigo Porto Cavalcanti, Victor Farias Monteiro, and Lucas dos Santos Fonseca

Abstract—Non-Terrestrial Networks (NTNs) are commonly envisioned as a means of extending coverage beyond the reach of terrestrial infrastructure. However, they can also be used to complement Terrestrial Networks (TNs) in supporting traffic demand. This paper evaluates load balancing strategies for TN-NTN coexistence through bidirectional user offloading and resource allocation. Five offloading strategies are compared. The results show that bidirectional offloading reduces outage by redistributing traffic between both networks.

Keywords—Non-Terrestrial Networks, LEO Satellites, Load Balancing, Offloading.

I. INTRODUCTION

Non-Terrestrial Networks (NTNs) have emerged as a key component of future wireless networks, complementing terrestrial infrastructure through wide-area coverage and ubiquitous connectivity [1]. Beyond coverage extension, the integration of NTNs and Terrestrial Networks (TNs) also creates new opportunities for traffic management, allowing User Equipments (UEs) to be served by different access technologies according to network conditions and resource availability.

In integrated TN-NTN deployments, UE association plays a central role in determining how traffic is distributed across the terrestrial and satellite segments. When both networks are simultaneously available, assigning UEs exclusively according to radio conditions may lead to an uneven utilization of network resources. As a consequence, one segment may experience resource scarcity while capacity remains available in the other. Load balancing mechanisms address this issue by enabling UE migration between networks and redistributing traffic according to resource availability rather than solely channel quality [2].

While load balancing is widely studied in terrestrial networks, TN-NTN coexistence faces severe capacity asymmetry due to large satellite footprints and limited spectrum. Addressing this gap, this paper moves beyond conventional unidirectional offloading to propose a proactive bidirectional UE migration framework. Five distinct offloading heuristics are evaluated via Monte Carlo simulations, providing critical insights into how different selection criteria balance outage reduction, network load, and fairness trade-offs.

The authors are with the Wireless Telecom Research Group (GTEL), Federal University of Ceará (UFC), Fortaleza-CE, e-mail: cassimiro@alu.ufc.br. This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). V. F. Monteiro and R. P. Cavalcanti were supported by CNPq under the Grants DT-308267/2022-2 and DT-303625/2022-8, respectively. L. S. Fonseca was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Finance Code 001.

II. SYSTEM MODEL

We consider a coexistence scenario composed of a terrestrial cellular network and a Low Earth Orbit (LEO) satellite-based NTN operating over distinct frequency bands. The terrestrial segment consists of a single TN cell of radius R_{TN} , while the satellite segment is represented by a single NTN beam of radius R_{NTN} , with $R_{NTN} > R_{TN}$. The terrestrial base station is located at the center of the satellite footprint. A set K of UEs is distributed over the coverage area, and each UE may potentially access either network depending on channel conditions and resource availability.

The use of orthogonal frequency bands eliminates co-channel interference between the terrestrial and satellite segments. Consequently, the radio resources of each network can be managed independently. Let $N = \{\text{TN}, \text{NTN}\}$ denote the set of available access networks and let R_n represent the set of resource blocks (RBs) available in network $n \in N$. The objective of the subsequent load balancing procedure is to determine the network association of each UE.

The propagation conditions experienced by terrestrial and satellite links are modeled according to standardized 3rd Generation Partnership Project (3GPP) assumptions. For the terrestrial segment, the downlink channel follows the Rural Macro (RMa) non-line-of-sight (NLOS) propagation model specified in 3GPP TR 38.901 [3]. The corresponding path loss is computed according to the analytical expressions defined in the specification, while large-scale channel variations are represented through log-normal shadowing with standard deviation $\sigma_s = 8$ dB.

For the satellite segment, propagation follows the assumptions adopted for LEO NTNs in 3GPP TR 38.811 [1]. The channel model accounts for free-space path loss associated with the satellite slant range, beam gain variations due to the satellite radiation pattern, atmospheric attenuation, and polarization mismatch losses. Small-scale channel fluctuations are modeled through Rician-Shadowed fading, capturing both the dominant line-of-sight component and the effects of shadowing and multipath propagation [4].

Let $g_{k,n,r}$ denote the channel gain between network $n \in N$ and UE $k \in K$ on RB $r \in R_n$. The channel gain incorporates all propagation effects associated with the corresponding access technology, including path loss, shadowing, fading, antenna gains, and additional propagation losses. Assuming a fixed transmit power spectral density within each network, the transmit power allocated to a given RB is denoted by $P_{tx,n}$. The resulting Signal-to-Noise Ratio (SNR) experienced by UE

k on RB r is therefore given by

$$\text{SNR}_{k,n,r} = \frac{P_{tx,n} g_{k,n,r}}{\sigma^2}, \quad (1)$$

where σ^2 denotes the thermal noise power including receiver noise figure.

Assuming ideal link adaptation and neglecting coding and implementation losses, the achievable rate associated with RB r is approximated through Shannon's capacity expression

$$C_{k,n,r} = B_{RB} \log_2 (1 + \text{SNR}_{k,n,r}), \quad (2)$$

where B_{RB} denotes the bandwidth of a single RB. The total data rate achieved by a UE corresponds to the sum of the capacities of all RBs assigned to that UE. These rate values are subsequently used by the load balancing and resource allocation procedures described in the following sections.

III. DESIGN OBJECTIVES AND CONSTRAINTS

The considered TN-NTN coexistence scenario requires determining both the serving network of each UE and the allocation of the available radio resources within each network. Since the terrestrial and satellite segments operate over distinct frequency bands, their resource pools are independent and can be managed separately. Nevertheless, UE association decisions directly affect the resource availability in each network and, consequently, the achievable UE rates.

Let $x_{k,n}$ denote the association variable associated with UE k and network n , defined as

$$x_{k,n} = \begin{cases} 1, & \text{if UE } k \text{ is associated with network } n, \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

Since each UE can be served by at most one access network at a given time, the following association constraint must hold:

$$\sum_{n \in N} x_{k,n} \leq 1, \quad \forall k \in K. \quad (4)$$

Radio resource allocation is represented through the binary variable $y_{k,n,r}$, defined as

$$y_{k,n,r} = \begin{cases} 1, & \text{if RB } r \text{ of network } n \text{ is assigned to UE } k, \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

Because orthogonal multiple access is assumed, a given RB can be assigned to at most one UE within the same network. This requirement is expressed by

$$\sum_{k \in K} y_{k,n,r} \leq 1, \quad \forall n \in N, \forall r \in R_n. \quad (6)$$

Furthermore, radio resources can only be allocated by the network serving the corresponding UE. Therefore, resource allocation must remain consistent with the association decisions according to

$$y_{k,n,r} \leq x_{k,n}, \quad \forall k \in K, \forall n \in N, \forall r \in R_n. \quad (7)$$

Given the achievable rate $C_{k,n,r}$ associated with RB r in network n , the total rate experienced by UE k can be expressed as

$$R_k = \sum_{n \in N} \sum_{r \in R_n} y_{k,n,r} C_{k,n,r}. \quad (8)$$

In practical systems, UEs are typically associated with minimum service requirements. Accordingly, a UE is considered to be in outage whenever its achieved throughput falls below a predefined threshold $R_{\min}$. To facilitate the performance evaluation, we define the binary satisfaction indicator

$$z_k = \begin{cases} 1, & R_k \geq R_{\min}, \\ 0, & R_k < R_{\min}. \end{cases} \quad (9)$$

The total number of satisfied UEs can then be written as

$$N_{\text{sat}} = \sum_{k \in K} z_k, \quad (10)$$

while the outage probability is obtained from the fraction of UEs that fail to satisfy the minimum rate requirement.

The objective of the load balancing process is to increase the number of satisfied UEs while efficiently utilizing the radio resources available in both networks. Different association and offloading decisions may lead to distinct resource distributions between the terrestrial and satellite segments, affecting outage probability, UE throughput, fairness, and network utilization.

IV. LOAD BALANCING STRATEGIES

This work evaluates a set of load balancing strategies based on bidirectional UE offloading between the TN and NTN. The overall procedure consists of two sequential stages: UE association and resource allocation. The first stage determines the serving network of each UE, while the second stage distributes the available radio resources among the associated UEs.

Initially, UEs whose received SNR is below -5 dB in both networks are considered to be in physical outage and are excluded from the association process. For the remaining UEs, an initial association is established according to the highest instantaneous SNR. A UE is considered eligible for migration whenever the minimum SNR requirement is simultaneously satisfied in both networks. Such UEs constitute the candidate set from which offloading decisions are made.

The offloading procedure is applied sequentially in both migration directions. It is first executed from the network exhibiting the highest outage probability towards the opposite network. After no further beneficial migrations can be identified, the same procedure is applied in the reverse direction. This bidirectional evaluation allows traffic redistribution across both networks and reduces the dependence of the final association on the initial UE distribution.

Candidate UEs are examined according to a predefined ordering criterion. UEs are migrated one at a time and a migration is accepted only if it does not increase the overall outage probability. Whenever the source network reaches zero outage, additional migrations are accepted only if they improve the overall median UE capacity. Consequently, the offloading process prioritizes outage reduction while avoiding unnecessary capacity degradation.

Five UE-selection strategies are investigated. The first strategy (M1) evaluates all migration candidates and prioritizes UEs with the lowest SNR in the source network. This greedy source-relief tactic assumes that offloading UEs with poor

channels removes the most resource-hungry connections from the origin, despite potential suboptimality at the destination.

The second strategy (M2) adopts the same ordering criterion as M1 but restricts the candidate set to UEs currently experiencing service outage. This policy focuses exclusively on UEs whose achieved rate is below the minimum service requirement.

The third strategy (M3) evaluates all migration candidates and prioritizes UEs with the highest SNR in the destination network. This proactive criterion ensures migrated UEs achieve high spectral efficiency, maximizing resource utility and minimizing bandwidth consumption after reassociation.

The fourth strategy (M4) combines the candidate filtering adopted in M2 with the destination-SNR ordering adopted in M3. Therefore, only UEs in outage are considered for migration and priority is given to those exhibiting the strongest channel conditions in the destination network.

Finally, the fifth strategy (M5) evaluates all migration candidates and prioritizes UEs according to their RB cost. The RB cost is defined as the number of resource blocks required for a UE to achieve the minimum target rate $R_{\min}$ in the source network. UEs demanding a larger amount of radio resources are therefore migrated first, potentially releasing additional resources for the remaining UEs.

After the UE association stage is completed, radio resources are allocated independently within each network. A Round Robin scheduler is employed to distribute the available RBs among the associated UEs. Since resource allocation is performed independently of channel quality, the resulting performance differences can be primarily attributed to the offloading decisions rather than to scheduler-specific throughput gains. This allows the impact of the evaluated load balancing strategies to be assessed in a consistent and comparable manner.

V. PERFORMANCE EVALUATION

A. Simulation Assumptions

The performance evaluation is conducted through snapshot-based Monte Carlo simulations. Each snapshot represents an independent realization of UE positions and channel conditions. The considered scenario consists of one terrestrial cell and one LEO satellite beam operating over distinct frequency bands, as described in the previous sections. The main simulation parameters are summarized in Table I.

For each simulation snapshot, UEs are spatially distributed over the considered coverage region according to a composite distribution. Half of the UEs are uniformly distributed inside the terrestrial cell, representing a local concentration of traffic within the TN coverage area. The remaining UEs are uniformly distributed over the satellite footprint. This spatial distribution creates a coexistence scenario in which part of the traffic demand is concentrated in the terrestrial cell, while another part is distributed over the wider NTN coverage area.

We consider different numbers of active UEs, i.e., $N_{\text{UE}} \in \{30, 40, 50, 60, 70, 80, 90, 100\}$ in order to evaluate the load balancing strategies under different traffic loads. For each value of N_{UE} , 1000 independent Monte Carlo snapshots are

TABLE I
PARAMETERS

Parameter	Terrestrial (TN)	Satellite (NTN)	Ref.
Layout	Open RMa	Scenario 9 (LEO-600)	[3], [5]
Carrier Frequency	700 MHz	2.0 GHz (S-band)	[1], [3]
System Bandwidth	20 MHz	30 MHz	[8]
Subcarrier Spacing	15 kHz	15 kHz	[8]
Num. of RBs	106	160	[8]
RB Bandwidth	180 kHz	180 kHz	[8]
Power / EIRP	46 dBm	78.8 dBm	[6], [5]
Eval. Radius	15 km	50 km	[3], [1]
Height	35 m	600 km	[1], [3]
Tx Gain	15 dBi	30 dBi	[3], [5]
Rx Gain	0 dBi	0 dBi	[3], [5]
Add. Losses (L)	2 dB	5.4 dB	[3], [5]
Prop. Model	RMa (NLOS)	FSPL + Beam roll-off	[3], [1]
Fading	Shadowing (8 dB)	Rician-Shadowed	[3], [7]
Chan. Params	-	$K = 15$ dB, $m = 15$	[7]
Noise Figure	9 dB	9 dB	[8]
Antenna Type	3GPP 3D	Spot beam	[1], [3]

generated. In each snapshot, UEs positions, channel gains, initial association, offloading decisions, and resource allocation are computed independently.

The baseline scenario associates each UE with the network providing the highest instantaneous SNR. No subsequent load balancing is performed in this case. The baseline therefore represents a channel-driven association policy, against which the evaluated offloading strategies are compared.

The minimum service requirement is set to $R_{\min} = 1$ Mbps. A UE whose achieved rate is below this threshold is considered to be in outage. The performance of each strategy is evaluated using four metrics. The first metric is the outage probability, defined as the fraction of UEs that do not achieve $R_{\min}$. The second metric is the median UE capacity, which represents the typical rate experienced by admitted UEs and is less affected by extreme values than the average throughput. The third metric is Jain's fairness index, computed from the achieved UE rates. Finally, the network load is measured as the fraction of UEs associated with each access network, allowing the effect of offloading on traffic distribution to be assessed.

B. Simulation Results

The purpose of this evaluation is to determine how different UE migration policies affect the performance of the integrated TN-NTN system. The evaluated strategies differ only in the criterion used to select UEs for migration. Therefore, the observed differences can be attributed to the load balancing policy rather than to changes in the propagation model, traffic model, or resource allocation procedure.

Fig. 1 shows the capacity distribution obtained for the baseline association with 30 UEs. In this case, UEs are associated exclusively according to the highest instantaneous SNR. The figure shows that the terrestrial and satellite segments provide substantially different capacity distributions. The terrestrial segment can provide higher data rates to a portion of the UEs, whereas the satellite segment exhibits a more concentrated distribution at lower rates. This difference is expected due to the distinct propagation conditions, coverage footprints, and resource sharing characteristics of the two access technologies. The result also indicates that a purely SNR-based association

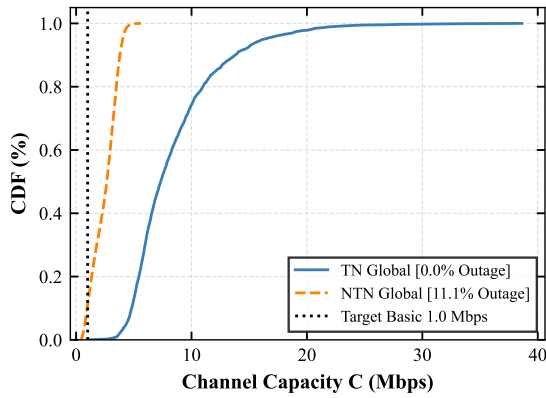


Fig. 1. Baseline scenario CDF evaluating 30 UEs.

may not fully exploit the resources available in both networks, motivating the use of load balancing mechanisms.

Fig. 2(a) presents the overall outage probability as the number of UEs increases. The baseline exhibits the highest outage over the evaluated load range, showing that maximum-SNR association alone is insufficient to balance the demand between the terrestrial and satellite segments. All offloading strategies reduce the outage relative to the baseline, confirming that UE migration can improve the utilization of the available radio resources.

The network-specific outage curves in Fig. 2(b) and Fig. 2(c) explain the origin of this improvement. For the considered scenario, most outage events occur in the satellite segment, while the terrestrial segment remains less congested under the baseline association. The offloading strategies reduce the NTN outage by transferring part of the traffic to the TN. This behavior demonstrates that, in the evaluated setup, terrestrial resources can be used not only for local coverage but also to relieve congestion in the satellite segment.

Among the evaluated strategies, M3 consistently provides the lowest overall outage. Prioritizing UEs with the highest destination SNR ensures they require very few RBs to achieve their target rate. This high spectral efficiency effectively prevents offloading from merely transferring congestion, thereby preserving capacity in the receiving network.

A second relevant trend is observed when comparing M1, M3, and M5 against M2 and M4. The former group evaluates all UEs eligible for migration, whereas the latter group restricts the candidate set to UEs already in outage. The results indicate that limiting the offloading process to outage UEs can be overly restrictive. UEs that are not currently in outage may still consume a significant amount of resources in the source network. Migrating such UEs can release resources for multiple remaining UEs and thereby reduce the overall outage more effectively than a purely reactive policy focused only on unsatisfied UEs.

The median capacity results in Fig. 2(d)–(f) show the trade-off associated with this redistribution. In the terrestrial segment, the median capacity decreases when load balancing is applied, as additional UEs are admitted by the TN. This reduction should be interpreted together with the outage

curves: the terrestrial network sacrifices part of its excess per-UE capacity in order to serve additional UEs and reduce the system outage. In contrast, the NTN median capacity tends to improve or remain more stable because fewer UEs compete for the satellite resources after offloading.

The overall median capacity in Fig. 2(d) reflects the combined effect of resource redistribution between the two access networks. Although the terrestrial median capacity decreases when additional UEs are associated with the TN, the NTN median capacity tends to remain slightly higher for some offloading strategies. As a result, the overall median capacity varies only moderately across the evaluated methods. This indicates that median capacity and outage probability capture different aspects of the system behavior and should be interpreted jointly rather than as equivalent performance indicators.

Figs. 2(h) and Fig. 2(i) show the distribution of UEs between the two networks. Compared with the baseline, the offloading strategies increase the fraction of UEs associated with the TN and reduce the load assigned to the NTN. This behavior is consistent with the outage results and confirms that the reduction in outage is achieved through a redistribution of UEs across the two access networks.

Finally, Fig. 2(g) presents Jain's fairness index computed from the achieved UE rates [9]. The fairness values remain limited by the heterogeneous nature of the coexistence scenario, since terrestrial and satellite links exhibit different channel conditions and resource availability. Nevertheless, the offloading strategies do not produce a substantial fairness degradation relative to the baseline. This suggests that the outage reduction is mainly obtained through improved traffic distribution rather than through excessive concentration of resources on a small subset of UEs.

Overall, the results show that bidirectional offloading is effective in reducing outage in the considered TN–NTN coexistence scenario. The comparison among the strategies also provides an important design insight: considering all eligible UEs for migration tends to be more effective than considering only UEs already in outage. This indicates that load balancing should be viewed as a proactive resource redistribution mechanism rather than only as a corrective action for UEs that have already failed to meet the target rate.

VI. CONCLUSION AND FUTURE PERSPECTIVES

This paper evaluated load balancing strategies for the coexistence of terrestrial and non-terrestrial networks operating over distinct frequency bands. Five bidirectional UE offloading policies were investigated under different traffic loads and compared in terms of outage probability, median UE capacity, fairness, and network load distribution.

The results showed that different UE-selection criteria lead to distinct traffic distributions and performance trade-offs across the considered metrics.

A particularly relevant observation is that strategies evaluating the complete set of eligible migration candidates consistently outperform those restricted to UEs already experiencing outage. This behavior suggests that effective load balancing should not be viewed solely as a corrective mechanism for

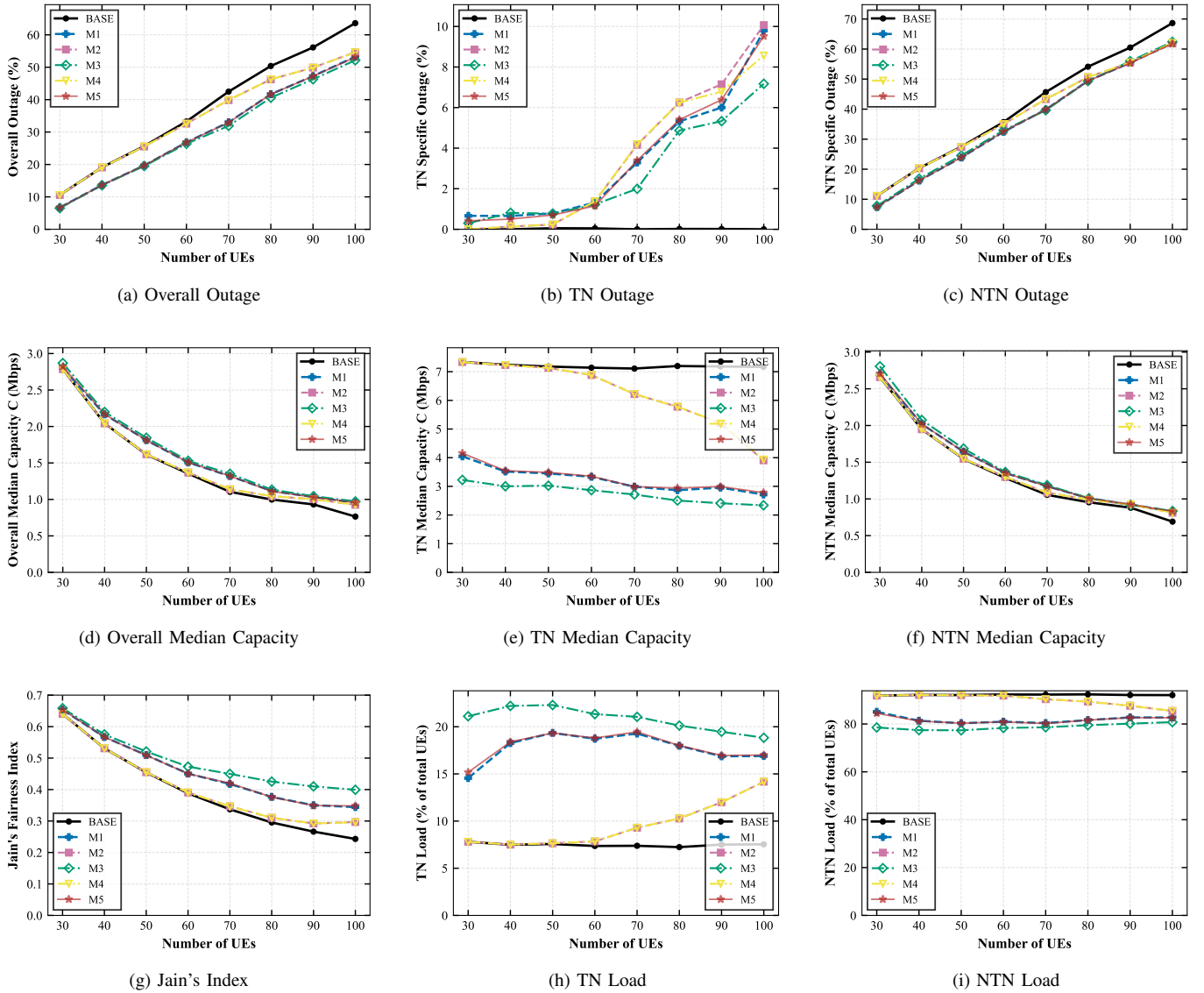


Fig. 2. Performance evaluation metrics: (a)-(c) Outage probability comparison; (d)-(f) Median capacity performance; (g)-(i) System fairness and network load distribution.

unsatisfied UEs. Instead, proactive traffic redistribution can release resources in congested segments and indirectly benefit a larger number of UEs.

The results also showed that the terrestrial infrastructure can play an important role in supporting traffic demand within integrated TN-NTN deployments. By redistributing UEs across the available access networks, the evaluated strategies modify both the load distribution and the resource utilization patterns observed in the system.

Future work includes multi-cell scenarios, UE mobility, QoS-aware schedulers, and statistical sensitivity analyses to further validate the heuristics' robustness.

REFERENCES

- [1] 3GPP, "Study on New Radio (NR) to support non-terrestrial networks," Technical Report (TR) 38.811, Rev. 15.4.0, 2020.
- [2] J. G. Andrews, et al., "An overview of load balancing in HetNets: Old myths and open problems," *IEEE Wireless Communications*, vol. 21, no. 3, 2014.
- [3] 3GPP, "Study on channel model for frequencies from 0.5 to 100 GHz," Technical Report (TR) 38.901, Rev. 17.0.0, 2022.
- [4] A. Vanelli-Coralli, N. Chuberre, G. Masini, A. Guidotti, and M. El Jaafari, *5G Non-Terrestrial Networks: Technologies, Standards, and System Design*. Hoboken, NJ, USA: Wiley-IEEE Press, 2024.
- [5] 3GPP, "Solutions for NR to support non-terrestrial networks (NTN)," Technical Report (TR) 38.821, Rev. 16.1.0, 2021.
- [6] 3GPP, "Non-terrestrial networks (NTN) related RF and co-existence aspects," Technical Report (TR) 38.863, Rev. 18.3.0, 2024.
- [7] A. Abdi et al., "A new statistical model for land mobile satellite channels," *IEEE Trans. Veh. Technol.*, vol. 52, no. 1, 2003.
- [8] 3GPP, "NR; Base Station (BS) radio transmission and reception," Technical Specification (TS) 38.104, Rev. 17.8.0, 2022.
- [9] R. Jain, D. Chiu, and W. Hawe, "A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Computer Systems," DEC Research Report TR-301, 1984.
- [10] Y. Fu and X. Wang, "Joint User Association and Resource Allocation for Tailored QoS Provisioning in 6G HetNets," *arXiv preprint arXiv:2403.03492*, 2024.
- [11] A. Jain, E. Lopez-Aguilera, and I. Demirkol, "User Association and Resource Allocation in 5G (AURA-5G): A Joint Optimization Framework," *Computer Networks*, vol. 178, p. 107439, 2020.