Analytical evaluation of the performance of a high-loaded 5G-IoT ecosystem functioning

Viacheslav Kovtun[1,*,†], Oksana Kovtun[2,†], Maria Yukhimchuk[3,†]

[1] *Institute of Theoretical and Applied Informatics Polish Academy of Sciences, Bałtycka Str., 5, Gliwice, 44-100, Poland*
[2] *Vasyl' Stus Donetsk National University, 600-richchya Str., 21, Vinnytsia, 21000, Ukraine*
[3] *Vinnytsia National Technical University, Khmelnytske shose, 95, Vinnytsia, 21021, Ukraine*

Abstract

The article proposes a novel mathematical framework for the analytical evaluation of the performance of a high-loaded 5G-IoT ecosystem operating in an environment with multiple sensor networks. The relevance of the study is driven by the rapid development of critical-use Internet of Things (IoT) systems and the necessity to ensure guaranteed Quality of Service (QoS) under conditions of high traffic intensity. The constructed model accounts for the specific features of 5G technologies, including NR Scheduling, HARQ, DRX, as well as packet fragmentation mechanisms and adaptive frequency hopping (AFH) in NR-U mode. The developed approach integrates methods of mathematical modelling, queueing theory, and stochastic analysis, providing reliable predictive capabilities for key performance parameters, in particular, average response time, blocking probability, and system idleness. A comparative validation of the model was conducted using an open dataset of operational indicators from a real 5G-NB-IoT network, confirming its advantages over classical models. The findings of the study encompass models and methods for the analysis of modern information and communication technologies, intelligent traffic management, resource optimisation, and the evaluation of dependability in dynamic environments.

Keywords

Complex Systems & Network Data Analysis; analytical modelling; information technology; 5G-IoT ecosystem; sensor network; adaptive resource scheduling; energy efficiency; collision reduction.

1. Introduction

Due to the intensive deployment of sensor networks in critical sectors such as industry (in the context of the Industrial Internet of Things), energy, transport, telemetry, and security, the issue of efficiently aggregating information flows in 5G-IoT environments with a high number of nodes has become particularly pressing [1, 2]. This challenge is further exacerbated when the communication ecosystem operates under heavy load conditions, with an increased frequency of simultaneous request transmissions and stringent real-time Quality of Service (QoS) requirements [3]. International standards [4], including 3GPP TS 23.501, 3GPP TS 38.300, and 3GPP TS 22.104, clearly outline the requirements for supporting Ultra-Reliable Low-Latency Communications (URLLC) and Massive Machine-Type Communications (mMTC), which are key components of 5G-IoT. Furthermore, ITU-T standards Y.3101 and Y.3113 emphasise the need for analytical evaluation of resource availability, latency, and data loss probability in critical services [5]. However, neither these standards nor traditional approaches to modelling loaded networks fully capture the specific nature of multisensor systems, which dynamically alter their topology, traffic types, and processing requirements. Consequently, there is a growing need to develop advanced analytical models capable of reliably forecasting the performance of the target information and communication system, taking into account the interaction between multiple sensor networks and 5G base stations under real-world load and constrained resource conditions. These challenges, aligned with the provisions of the aforementioned international standards, define the relevance of the present study.

MoMLeT-2025: 7th International Workshop on Modern Machine Learning Technologies, June, 14, 2025, Lviv-Shatsk, Ukraine
* Corresponding author.
† These authors contributed equally.
vkovtun@iitis.pl (V. Kovtun); o.kovtun@donnu.edu.ua (O. Kovtun); umc1987@vntu.edu.ua (M. Yukhimchuk)
0000-0002-7624-7072 (V. Kovtun); 0000-0002-9139-8987 (O. Kovtun); 0000-0002-8131-9739 (M. Yukhimchuk)

Existing approaches to the analytical modelling of 5G-IoT ecosystem performance demonstrate a considerable diversity of methods – from classical queueing theory models to sophisticated hybrid frameworks that combine stochastic geometry, radio channel parameterisation, and service mechanisms [6–9]. Fundamental methods [10–12], such as the M/M/1, M/M/c, and MAP/PH/1 models and their modifications, are effective for a first approximation in modelling individual network components. However, they exhibit significant limitations in the context of 5G ecosystems due to their lack of adaptability to variable traffic loads, traffic prioritisation, NR Scheduling, and DRX/HARQ mechanisms, which are critically important for real-world operational scenarios [13].

Advanced approaches based on stochastic geometry, widely applied in LTE systems and early versions of 5G, are capable of accounting for the spatial distribution of users, signal blockage by obstacles, and environmental characteristics [14–16]. However, as noted in [17], these models insufficiently reflect the dynamics of service provision at the base station side, particularly in cases of adaptive multi-service traffic (eMBB, URLLC, mMTC). In 5G-IoT ecosystems, where sensor devices generate quasi-periodic or spontaneous traffic and transmission delay is of critical importance, there arises a need for multi-component analytical models that simultaneously cover both the physical transmission layer and service processes at the gNB level.

The study [18] proposes a two-level framework structure for modelling systems based on mmWave/THz communication, which can be adapted for 5G-IoT. The first level involves radio interface parameterisation (geometry, blockage, micro-mobility), while the second comprises a queueing service model. This approach enables the separate modelling of physical environment effects and traffic processing mechanisms at base stations. The application of such a framework facilitates the simulation of scenarios involving resource reservation, multi-connectivity, diverse traffic types, and prioritisation schemes.

Of particular interest are the models presented in [19], which are tailored to the specific requirements of URLLC and mMTC—key service categories within the 5G-IoT paradigm. Such models must account for both ultra-low latency and massive device connectivity, implying a necessary synergy between adaptive radio resource scheduling (NR Scheduling) and intelligent traffic optimisation algorithms. As stated in [20], traditional methods are incapable of accurately reflecting the effects of packet fragmentation, DRX mode switching, inter-flow prioritisation across varying levels of criticality, as well as system behaviour under excessive load conditions.

Another important trend is the advancement of analytical modelling methods for 5G networks operating in the millimetre-wave and terahertz frequency bands. The study [21] describes a modular framework that accounts for antenna directivity, atmospheric attenuation, blockage effects, and micro-mobility—factors characteristic of next-generation sensor networks. Such models can be integrated into broader 5G-IoT performance models to capture not only radio aspects but also traffic behaviour and request processing at the base station level, which is particularly relevant in scenarios characterised by high traffic density and critical latency requirements.

In addition, contemporary approaches [22], such as those involving Network Function Virtualization (NFV) and Network Slicing, introduce a new modelling paradigm in which each "slice" (logical subnetwork) is optimised for a specific traffic type. Within the context of analytical modelling, these approaches necessitate the construction of systems composed of multiple interacting service models, each with its own scheduling policy and acceptable QoS level.

In summary, existing approaches demonstrate considerable flexibility but fail to comprehensively capture all operational characteristics of real-world loaded 5G-IoT ecosystems. This underscores the need to develop new mathematical frameworks that integrate mechanisms for adaptive resource management, packet processing with fragmentation, dynamic frequency planning, and multi-channel routing services, all aimed at ensuring guaranteed Quality of Service under multisensor traffic conditions.

The object of this study is the operation of a 5G information and communication ecosystem involving the interaction of multiple sensor networks under high-load conditions.

The subject of the research is analytical models for evaluating the performance of 5G-IoT ecosystems, taking into account dynamic radio resource allocation, adaptive frequency planning, and other technologies inherent to the 5G platform relevant for the deployment of sensor networks.

The aim of the study is to develop an analytical framework for the reliable evaluation of the performance of a high-loaded 5G-IoT ecosystem, considering the specific features of radio resource scheduling, traffic typology, and the architectural characteristics of sensor networks.

To achieve this aim, the following tasks were accomplished:

1. The specific features of 5G-IoT ecosystem operation under high-load conditions were analysed.
2. A model of interaction between multiple sensor networks and a gNB base station was formulated, incorporating NR Scheduling, HARQ, and DRX.
3. Analytical expressions were formalised to evaluate blocking probabilities, system idleness, response time, and throughput.
4. The model was modified to account for packet fragmentation and its impact on communication efficiency.
5. The model was adapted to the conditions of dynamic frequency planning (AFH) in NR-U mode.
6. The proposed mathematical framework was validated using a real dataset and compared with classical models.
7. The representativeness of key performance metrics was investigated for characterising typical operational scenarios of 5G-IoT ecosystems.

2. Models and methods

2.1. Research Statement

In a 5G-IoT ecosystem, data transmission between the 5G base station (gNB) and sensor networks is implemented through dynamic radio resource allocation governed by NR Scheduling technology. All active sensor networks can initiate transmission via the Random Access Channel (RACH), after which the base station allocates uplink or downlink resources based on traffic demand and priority. The functioning of the 5G-IoT ecosystem during the aggregation of information flows from sensor networks can be conditionally divided into two operational modes:

- Time Division Mode (TDM) – an active traffic period during which sensor networks transmit data, and the gNB performs adaptive scheduling, allocating resources according to the current load.
- Traffic Control Mode (TCM) – a resource optimisation mode in which, under low-load conditions, the 5G base station may redistribute frequencies, enable Discontinuous Reception (DRX) for energy efficiency, or switch part of the radio resources to a low-activity state.

If, at any given moment, no sensor network has data to transmit, the system does not await the completion of TCM but can immediately activate DRX or adjust resource allocation. Consequently, TDM and TCM do not alternate according to a fixed schedule but are dynamically regulated based on current traffic conditions. Data transmission from a sensor network occurs in dedicated time slots (Transmission Time Intervals – TTI), which the gNB allocates based on real-time analysis of traffic requests. If a sensor network has data to send, it signals this through a Buffer Status Report (BSR), after which the gNB assigns the appropriate resources. Transmission success is monitored via the Hybrid Automatic Repeat Request (HARQ) mechanism, which enables retransmission with minimal delay in the event of packet loss. This approach enables flexible load management and prevents inefficient use of the radio channel. The present study considers a scenario in which the 5G base station operates under high-load conditions, where the probability of radio resource idleness due to traffic absence is negligible, and adaptive allocation mechanisms ensure a continuous flow of data from multiple sensor networks.

We describe a typical model of a 5G-IoT ecosystem designed to collect data from multiple sensor networks $S=\{1,\ldots,S\}$. Each sensor network $s \in S$ comprises a set of N_s end devices (sensors, IoT devices) and one or more edge servers that perform preliminary data processing prior to transmission via the 5G network. In this study, we consider the most common scenarios of information interaction:

- Requests directed to the edge server within the sensor network (local internal interaction within the sensor network).
- Requests directed to an external server (in the global network) via the gNB base station.
- Response transmission: either the local or external server processes the request and generates a response consisting of multiple packets, which are transmitted via the 5G connection.

Data transmission is carried out through dynamic resource allocation by the base station using NR Scheduling, while communication between sensor networks and the gNB is regulated via RACH and BSR mechanisms. If required, the HARQ mechanism ensures error correction and packet retransmission with minimal delay. This minimalist set of scenarios reliably reflects the actual data exchange mechanisms in real-world 5G-IoT ecosystems, optimising performance and minimising delays during the processing of information requests.

Since the aim of the study is to analyse the throughput of a 5G base station during data collection from multiple sensor networks, the interaction between individual sensor nodes within a local network $s \in S$ and beyond is modelled in a simplified manner. A request from a sensor terminal within the network s is generated at an average interval $1/\lambda_s$ following the reception of the last response packet from the gNB. Requests may be directed either to a local server within the sensor network or to external servers via the 5G infrastructure. External requests arrive at the base station as a Poisson flow with intensity Λ_0 and are forwarded to the corresponding server or sensor device capable of performing the required processing. The server's response is generated in the form of a set of packets or an aggregated Transport Block (TB) and transmitted back via the 5G connection. Within the scope of this study, it is assumed that the 5G network operates under high-load conditions, where the probability of traffic absence across all sensor networks is negligible. It is also assumed that the TCM interval significantly exceeds TDM, allowing periods of low activity to be disregarded when evaluating throughput.

After passing through the 5G base station (gNB), a request is either processed by the corresponding server within the sensor network or forwarded to external computing resources via cloud infrastructure. If the request is directed to a local server within sensor network r, $r > 0$, it is processed internally within that network. Conversely, if the request is intended for an external system, i.e., $r = 0$, the transmission is carried out through the 5G infrastructure in the form of an aggregated TB or a set of packets. The routing path of the request is determined by a routing matrix $P_{s,r}$, where $s = 0$ denotes external traffic originating from the global network. After being processed by the server r, the average response delay is $1/\mu_{s,r}$, and the resulting response stream consists of $F_{s,r}$ packets or aggregated data blocks. The response from the server is transmitted back via the 5G connection, undergoing buffering within the sensor network before being forwarded through the gNB base station. If the response originates from an external server $r = 0$, it is transmitted directly through the gNB to the sensor device that initiated the request. Upon receiving the final response packet, the sensor device may generate a new request to continue the data exchange process. The generation of requests and their processing by servers follow an exponential distribution, which is a typical approach for modelling high-loaded networks in 5G environments.

Based on the stated assumptions, a simulation was conducted to model the operation of a 5G base station (gNB) responsible for collecting and processing data from multiple sensor networks. Due to the high number of sensor nodes N_s and the substantial volume of responses from servers $F_{s,r}$, the simulation process proved to be resource-intensive, even for a relatively small number of sensor devices and their corresponding requests. The most effective approach for analysing the operation of a 5G-IoT ecosystem is analytical modelling, which enables the evaluation of system performance under varying load conditions. To this end, data transmission through sensor networks and the gNB is represented using a mixed exponential model, where requests from sensor devices circulate within the system as clustered data flows with dynamic routing.

In this context, each sensor network $s \in S$ consists of a set of Single-Channel Sensor Devices (SCs) operating in continuous data collection and transmission mode, and Buffering Controllers (BCs) that manage request queues before data is sent to the 5G base station. An SC generates a stream of requests with an average intensity Λ_s and transmits them via the RACH to the BC, which temporarily buffers the packets if no resources are available at the gNB. Once resources are allocated, the BC forwards the packet to the gNB, which assigns processing resources through NR Scheduling. To model the interaction between sensor networks and the base station, the following variables are introduced:

- $T_{resp}(s) = \dfrac{N_s}{\Lambda_s} - \dfrac{1}{\lambda_s}$ is the average response time to a request from a sensor device within the network s;
- $\Lambda = \sum_{s=1}^{S} \Lambda_s$ is the total throughput of the 5G-IoT ecosystem;
- $T_{proc}(s)$ is the average processing time of requests at the sensor network server or in the cloud environment;
- $\rho_s = 1 - T_{proc}(s) M_s$ is the load coefficient of the BC in sensor network s prior to transmitting data through the 5G infrastructure, where M_s represents the average number of requests waiting for processing in network s;
- $P_{vis}(s,r)$ is the visit ratio of server r by sensor network s, indicating the probability of routing a request to that server;
- $K_S = \sum_{r=0}^{R} P_{vis}(s,r)$ is a non-zero visit ratio of sensor network $s \in S$, indicating the frequency with which this network forwards requests to server $r \in R$, including both local and global routes.

Requests from sensor devices are received by the gNB, which assigns resources to them in accordance with NR Scheduling. The routing of requests is determined by the visit ratio matrix $P_{vis}(s,r)$, which specifies the target server for each request. In the case of a request directed to a local server of the sensor network $r \in R$, the average processing time is $T_{proc}(s)$, and the response is transmitted in the form of F_{rs} individual packets or an aggregated Transport Block (TB). If the request requires external processing on a Cloud Server, the response arrives from the global network and is routed through the gNB to the initiating sensor device. Upon receiving the final response packet, the sensor device may generate a new request, thus maintaining a continuous flow of data. The visit ratio coefficients K_S determine the frequency of traffic routing between sensor networks and servers. Fig. 1 visualises the interaction model between sensor networks and the gNB base station within the 5G-IoT ecosystem.

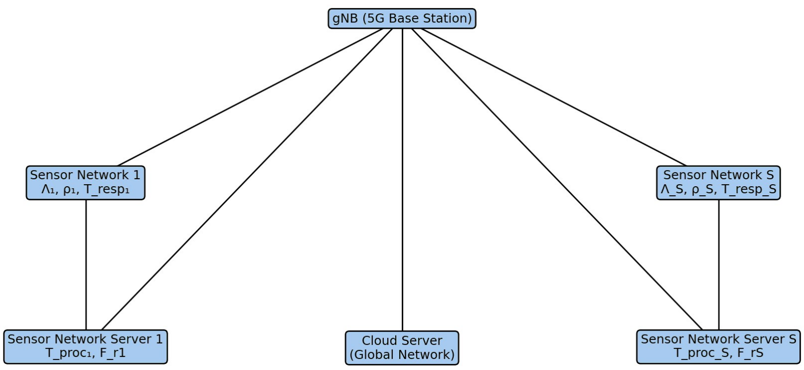

Figure 1: Structure of the 5G-IoT Ecosystem for Aggregating Information Flows from Sensor Networks.

Considering that sensor networks may utilise various communication mechanisms (NB-IoT, LTE-M, 5G NR), delays in data processing and transmission are modelled using an exponential distribution. The analytical characterisation of the performance of the 5G-IoT ecosystem presented in Fig. 1 is conducted in the following sections.

2.2. Mathematical Framework for Evaluating the Performance of the Target High-Loaded 5G-IoT Ecosystem

The model of the 5G-IoT ecosystem shown in Fig. 1 is multiplicative in nature (in accordance with [23–25]). Considering this, classical methods [23] can be applied to characterise each sensor network $s \in S$ in terms of performance metrics such as throughput Λ_s, average response time $T_{resp}(s)$, and probabilities $P_{idle}(s)$ and $P_{block}(s)$, which describe the absence of requests in the queue to the Buffering Controller (BC) of the sensor network and to the 5G base station (gNB), respectively. To carry out the evaluation, it is necessary to determine key characteristics of the target 5G-IoT infrastructure, such as the average request processing time on the sensor network server $T_{proc}(s)$ and the waiting time for resources at the gNB $T_{gNB}(s)$ for $\forall s \in S$.

To determine these characteristics, it is important to note that the operation of a 5G-IoT ecosystem under dynamic resource allocation resembles the structure of time-slotted networks [24]. Each sensor device within a sensor network may transmit data via the RACH, after which the gNB allocates resources in accordance with NR Scheduling. The duration of the time slot μ_s, assigned for data transmission from sensor network s, is a stochastic variable with an average value that depends on the dynamics of request flows within the system.

Depending on the resource scheduling mechanism at the gNB, sensor network s receives resource allocations m_s times during the observation period (in the case of uniform distribution – $m_s = 1$) for $\forall s \in S$. The average waiting time for a resource at the gNB can then be estimated as:

$$T_{gNB}(s) = \left(\frac{1}{H_{gNB}(s)} - \frac{q_{idle}(s)}{2} \right) \sum_{s=1}^{S} m_s \mu_s, \tag{1}$$

where $q_{idle}(s)$ is the probability that a request arrives at an empty queue of the BC of sensor network s, and $H_{gNB}(s)$ is the probability of successful resource assignment by the gNB for data transmission from sensor network s. The value $T_{proc}(s)$ is determined using a similar formula, but with consideration of the data processing time at the sensor network server.

We define the average time slot duration μ_s for sensor network s. Let us assume that the data transmission time of a packet, excluding the header, is denoted by t_d and is identical for all packets, while the average packet transmission time is t_p. The probabilities of frame corruption for frames containing and not containing payload data are denoted as z_d and z_0, respectively. Taking the introduced parameters into account, the average duration of a time slot is characterised by formula

$$\mu_s = Z_{poll}(s) + Z_{resp}(s) + T_{ack}(gNB), \tag{2}$$

where

$$Z_{poll}(s) = t_d + \left(1 - P_{block}(s)\right) t_p + \tau_s + \hat{z}_{block}(s)\left(\tau_{\max} - \tau_s\right), \tag{3}$$

$$Z_{resp}(s) = \left(1 - \hat{z}_{bloock}(s)\right)\left[t_d + \left(1 - P_{idle}(s)\right) t_p + \tau_s + 2z_0\right], \tag{4}$$

$$Z_{ack}(s)=\left[\left(1-\hat{z}_{block}(s-1)\right)\left(1P_{idle}(s-1)\right)+1-P_{block}(s)\right]\left(1-z_d\right), \quad (5)$$

and $T_{ack}(gNB)$ denotes the transmission time of the acknowledgement from the gNB, τ_s is the signal propagation time from the gNB to sensor network s, $\tau_{\max}$ is the maximum propagation delay in the system, $\hat{z}_{block}(s)=P_{block}(s)z_0+\left(1-P_{block}(s)\right)z_d$ is the effective transmission blocking probability, and the index $s-1$ in formula (5) is replaced by S if $s=1$ is used for cyclic resource allocation in the system.

The probabilities of successful data transmission in the 5G-IoT ecosystem are defined by formulas

$$H_{gNB}(s)=\left(1-\hat{z}_{block}(s)\right)\left(1-z_d\right)\left(1-\hat{z}_{block}(s+1)\right), \; H_{BC}(s)=\left(1-z_d\right)\left(1-\hat{z}_{idle}(s)\right), \quad (6)$$

where $\hat{z}_{idle}(s)$ is determined analogously to $\hat{z}_{block}(s)$, but in the context of probabilities $P_{idle}(s)$, and index $s+1$ is replaced by 1 in the case of $s=S$; $H_{gNB}(s)$ denotes the probability of successful resource assignment by the gNB for data transmission from sensor network s; $H_{BC}(s)$ represents the probability of successful packet passage through the BC of sensor network s.

The probability that a request arrives at an empty queue is defined analogously to [23, 24]:

$$q_{idle}(s)=P_{idle}(s)\left(\Lambda_s+\sum_{i=0,i\neq s}^{S}\Lambda_i p_{s,i}\right)\Big/M_{BC}(s),$$
$$q_{block}(s)=P_{block}(s)\left(\Lambda_s+\sum_{i=0,i\neq s}^{S}\Lambda_i p_{s,i}\right)\Big/M_{gNB}(s), \quad (7)$$

where $M_{BC}(s)=\Lambda_s+\sum_{i=0,i\neq s}^{S}\Lambda_i p_{s,i}F_{s,i}$, $M_{gNB}(s)=\Lambda_s\hat{z}+\sum_{i=0,i\neq s}^{S}\Lambda_i p_{s,i}$, and $P_{idle}(s)$ is the probability of no requests in the queue of the BC of sensor network s, $P_{block}(s)$ is the probability of transmission blocking through the gNB, Λ_s is the request generation intensity by sensor network s, $p_{s,i}$ is the probability of routing a request from network s to network i, and $F_{s,i}$ is the average number of packets in the response after request processing at the server of network s.

Thus, considering known values of $P_{idle}(s)$, $P_{block}(s)$, Λ_s for $\forall s\in S$, expressions (1)-(7) can be used to determine the average waiting time $T_{gNB}(s)$ and the average request processing time in the sensor network $T_{proc}(s)$. Clearly, the process of jointly determining all these characteristics is iterative in nature, which imposes constraints on computational resources when applying the corresponding equations. Therefore, to analyse a network with a large number of sensor nodes N_s, it is advisable to apply asymptotic methods [24], which yield the following approximations for $\forall s\in S$:

$$\Lambda_s=N_s/Q_s, \; T_{resp}(s)=N_s/\Lambda_s-1/\lambda_s, \; P_{block}(s)=1-T_{gNB}(s)M_{gNB}(s),$$
$$P_{idle}(s)=1-T_{proc}(s)M_{BC}(s), \quad (8)$$

where the value of Q_s is determined from the following system of equations:

$$Q_s=\frac{1}{\hat{\lambda}_s}+r_{proc}(s)+K_S r_{gNB}(s)+\sum_{i=1,i\neq s}^{S}p_{s,i}\left(r_{gNB}(i)+F_{s,i}r_{proc}(i)\right), \quad (9)$$

in which $\frac{1}{r_{proc}(s)}+\frac{N_s}{Q_s}+\sum_{i=1,i\neq s}^{S}\frac{N_i}{Q_i}p_{s,i}F_{s,i}=\mu_{proc}(s)$, $\frac{1}{r_{gNB}(s)}+\frac{N_s}{Q_s}K_S\sum_{i=1,i\neq s}^{S}\frac{N_i}{Q_i}p_{s,i}=\mu_{gNB}(s)$, $\frac{1}{\hat{\lambda}_s}=\frac{1}{\lambda_s}+\frac{p_{s,0}}{\mu_{s,0}}+\sum_{i=1,i\neq s}^{S}\frac{p_{s,i}}{\mu_{s,i}}$, $\mu_{proc}(s)=1/T_{proc}(s)-\Lambda_0 p_{s,0}F_{s,0}$, $\mu_{gNB}(s)=1/T_{gNB}(s)-\Lambda_0 p_{s,0}$, and $r_{proc}(s)$ is an auxiliary variable representing the average waiting time of a request prior to its processing within the sensor network, and $r_{gNB}(s)$ is an auxiliary variable representing the average waiting time for resource allocation by the gNB. The system of equations (8) should be solved under the condition $Q_s>0$, $r_{proc}(s)>0$, $r_{gNB}(s)>0$ for $\forall s>0$. Clearly, the intensity of the external request flow Λ_0 must satisfy the inequalities $1/p_{s,0}\Lambda_0>\max\{T_{gNB}(s),F_{s,0}T_{proc}(s)\}$ for $\forall s>0$.

We estimate from above the probability that the gNB base station of a 5G-IoT ecosystem transitions into an energy-saving state due to the absence of packets for transmission from all sensor networks during the TCM period. The probability that the last transmitted packet was the only one across all sensor networks can be indirectly evaluated through the probability P_0 of no pending requests in all BCs of the sensor networks:

$$P_{gNB\to idle}<P_0 TCM/T_{\min}, \tag{10}$$

where the minimum time $T_{\min}$ required for packet transmission, including acknowledgement, is defined as

$$T_{\min}=t_p+T_{ack}(gNB)+2\left(t_d+\tau_0+\min_s\{\tau_s\}\right), \tag{11}$$

with t_p denoting the packet transmission time, $T_{ack}(gNB)$ is the acknowledgement transmission time, τ_0 is the signal propagation time, and τ_s the maximum signal propagation time for sensor network s.

By analogy with [23], for a large number of sensor nodes N_s, the order of probability P_0 is estimated using formula

$$P_0 \sim \exp\left[\sum_{s=1}^{S}\left\lfloor \mu_{proc}(s)r_{proc}(s)+\mu_{gNB}(s)r_{gNB}(s)-N_s\ln\left(\hat{\lambda}_s Q_s\right)\right\rfloor\right]. \tag{12}$$

Expression (10) remains valid even when the number of polling cycle repetitions $n_0^{(IoT)}$ for the sensor networks is zero, indicating the absence of any traffic (an empty cycle). If $n_0^{(IoT)}>0$ holds, the right-hand side of expression (10) must be multiplied by the probability $P_0^{(IoT)}$ that no new requests are generated within the sensor networks during these empty cycles:

$$P_0^{(IoT)}=\exp\left[-2n_0^{(IoT)}\left[\sum_{s=1}^{S}\frac{N_s}{\hat{\lambda}_s}\right]\left(t_d+\tau_0+\sum_{s=1}^{S}\tau_s\right)\right]. \tag{13}$$

Thus, the developed analytical method can be applied if the upper bound $P_{gNB\to idle}$ computed using expression (10) is significantly less than one, and also if $TCM>>TDM$ holds.

To reduce the number of collisions in a 5G-IoT ecosystem, it is advisable to implement a packet fragmentation mechanism. This approach enables retransmission of only the affected fragments in case of channel errors, rather than the entire packet. Each packet, assuming a uniform average transmission time, is divided into n_f fragments, where the average transmission time of all fragments except the last is $t_m^{(f)}$ (category 1), and for the final fragment – $t_l^{(f)}$ (category 2). In this case, expression $1/H_{gNB}(s)$ in equation (6) is transformed into $\frac{n_f - 1}{H_s^{(f1)}} + \frac{1}{H_s^{(f2)}}$, where $H_s^{(f1)}$ and $H_s^{(f2)}$ denote the probabilities of successful transmission of category 1 and category 2 fragments, respectively, and parameter z_d is replaced with the fragment error probabilities z_m and z_l for categories 1 and 2.

In other equations of the mathematical framework, the parameters z_d and t_p are replaced with the averaged values $z_d^{(f)}$ and $t_{mean}^{(f)}$, which are determined using formulas

$$z_d^{(f)} = 1 - \frac{n_f}{(n_f - 1)(1 - z_m) + (1 - z_l)}, \tag{14}$$

$$t_{mean}^{(f)} = \frac{t_m^{(f)}(n_f - 1)(1 - z_m) + t_l^{(f)}(1 - z_l)}{(n_f - 1)(1 - z_f) + 1 - z_l}, \tag{15}$$

where $t_{mean}^{(f)}$ is the average transmission time of a fragment, taking into account that the majority of the packet consists of standard fragments of category 1 with length $t_m^{(f)}$, while only the final fragment belongs to category 2 and has a different length $t_l^{(f)}$. In condition (10), the minimum transmission time is updated to account for fragmentation:

$$T_{\min} = (n_f - 1)t_{mean}^{(f)} + t_l^{(f)} + n_f\left[T_{ack}(gNB) + 2\left(t_d + \tau_0 + \sum_{s=1}^{S}\tau_s\right)\right]. \tag{16}$$

Thus, the method can be readily modified to account for the distribution of packet lengths, which is particularly relevant for sensor networks where both short messages (e.g., control signals or telemetry data) and longer packets containing aggregated or multimedia data are present.

Another modification of the proposed mathematical framework involves adaptation to dynamic frequency planning in a 5G-IoT ecosystem based on Adaptive Frequency Hopping (AFH) technology in NR-U (Unlicensed NR). In this mode, all sensor devices operate on a single frequency for a fixed time-frequency interval T_{dewll}, after which they are switched to a different frequency (with a switching delay t_c). During frequency switching, terminals temporarily lose synchronisation with the network. Accordingly, a sensor device waiting for resource allocation to transmit a packet (or its fragment) postpones the transmission if the expected transmission time, calculated based on the maximum signal propagation time $\tau_{\max}$ (including acknowledgement), exceeds the remaining time in the current frequency interval. Such a scenario is referred to as delayed transmission.

The average waiting time between a delayed transmission and the end of the frequency interval (including the switching time t_c) can be estimated using formula

$$\Delta = t_c + T_{\min}/2, \tag{17}$$

where the minimum packet transmission time is defined according to equation (16), but with $\tau_{\max}$ replaced by $\min\{\tau_s\}$. In the case of fragmentation, instead of the average packet transmission time t_p, the averaged value from equation (15) is used.

As shown in [24], it can be assumed (and this is confirmed by simulation modelling) that the adaptive frequency hopping mode differs from the stationary regime only by silence periods with an average duration Δ at the end of each frequency interval T_{dewll}. Therefore, in formulas (8), which define the throughput Λ_s, the probabilities of no requests in the queue $P_{idle}(s)$, and transmission blocking $P_{block}(s)$ in the target 5G-IoT ecosystem, both the average waiting time at the 5G base station $T_{gNB}(s)$ and the average waiting time across all sensor networks $T_{proc}(s)$ must be divided by the coefficient $1-\Delta/(T_{dwell}+t_c)$. This adjustment accounts for the reduction in effective access time to radio resources in 5G-IoT systems caused by dynamic frequency hopping.

3. Results and Discussion

Let us begin this section by presenting the process and results of the verification of the mathematical framework proposed in the article. The verification involved a detailed evaluation of the adequacy of the analytical model presented, as well as empirical validation of its effectiveness and practical relevance. The objective of the verification was to empirically demonstrate that the developed model accurately describes the operation of a high-loaded 5G-IoT ecosystem, provides more reliable results compared to classical approaches, and can be effectively applied for the practical analysis of network performance.

To verify adequacy, we consistently carried out a series of actions structured into three key stages. The first stage involved the selection of classical mathematical models of network performance widely used in the field of information and communication systems. Particular attention was given to the ability of these conventional models to describe functional processes characteristic of high-loaded 5G-IoT ecosystems, with a specific focus on radio resource allocation, adaptive traffic scheduling, and dynamic request management. The second stage consisted in selecting a representative dataset containing sufficient information on the operational results of real 5G-IoT ecosystems in terms of the parameters used in the formalisation of the proposed mathematical framework. At the third stage, we visualised the comparison results between the proposed model and classical counterparts using the research-specific performance metrics, with the determination of confidence intervals and discussion of the findings. The analysis focused on key performance parameters such as latency, throughput, and request blocking probability, which are critical for the operation of a 5G-IoT ecosystem.

In implementing the first stage, we focused on selecting classical mathematical models for traffic flow and load evaluation in branched information and communication networks. Traditional Markov queueing models such as M/M/1 and M/M/c serve as fundamental analytical tools for systems with randomly arriving requests. The M/M/1 model assumes a single queue with exponentially distributed inter-arrival and service times, enabling analysis of individual network nodes. The M/M/c model extends this approach by incorporating multiple parallel service servers, offering a more adequate representation of modern multichannel 5G-IoT networks. However, these models have significant limitations, as they do not account for adaptive resource scheduling typical of 5G systems and fail to capture the impact of mechanisms such as NR Scheduling. Iterative analytical methods for evaluating network load allow for a more accurate analysis of radio resource status and allocation mechanisms. These methods are based on repeated computations of the steady state of the telecommunication system, considering dynamic load parameters and adaptive traffic control algorithms. This approach enables the estimation of request blocking probabilities and resource allocation efficiency in high-load scenarios. At the same time, classical analytical methods remain limited when dealing with the complex adaptive mechanisms of 5G-IoT, such as dynamic traffic prioritisation and the use of HARQ to minimise packet loss. In contrast, the proposed mathematical framework integrates elements of adaptive resource management and scheduling mechanisms tailored to high-loaded 5G-IoT ecosystems. This potentially offers a more realistic reflection of the functioning of such ecosystems compared to traditional approaches (M/M/1 and M/M/c).

For the verification of the proposed mathematical framework, we selected the open-access dataset “A Large-Scale Dataset of 4G, NB-IoT, and 5G Non-Standalone Network

Measurements" [26]. This dataset was developed by a research team from Simula Research Laboratory, Ericsson Research, Karlstad University, Sapienza University of Rome, and other institutions. It contains results from a seven-week campaign of experimental data collection on the real-world operation of commercial mobile networks in Rome, Italy. The dataset covers three major mobile communication technologies: LTE (4G), NB-IoT, and 5G NR (Non-Standalone), and includes measurements obtained across a variety of usage scenarios—ranging from stationary to dynamic (pedestrian and vehicular) contexts. Its representativeness for our study is explained by its scale, openness, and, most importantly, the breadth of captured characteristics directly related to the performance and reliability of 5G-IoT ecosystems. Specifically, the dataset includes parameters of radio coverage (RSRP, RSRQ, SINR), channel characteristics (TBS, Modulation and Coding Scheme), as well as indicators of data transmission quality such as round-trip time (RTT), packet delay variation (PDV), interface-level interactions, packet loss, and throughput. The RTT metric was measured under real-world conditions using Speedtest sessions and online gaming traffic, making the data highly relevant for modelling latency in practical 5G-IoT scenarios.

Within the scope of our study, the RTT parameter is employed as the primary verification metric. It is particularly critical for IoT applications with stringent response time requirements, such as telemetry, real-time control, and automated management systems. In the proposed mathematical framework, the response time (or the latency between request generation and receipt of the corresponding response) is modelled by the expression $T_{resp}(s)=1/(\lambda_s-\Lambda_s)$, where λ_s denotes the intensity of new request generation, and Λ_s represents the actual throughput of sensor network $s\in S$ within the 5G-IoT ecosystem. This expression is presented in formula (8) and is used as an analytical estimate of the average response time from the sensor device, which, in essence, corresponds to the RTT delay measured under real-world operating conditions. In addition, the related metric $T_{gNB}(s)$, as defined in formula (1), represents the average waiting time for resource allocation by the 5G base station. Combined with the average request processing time at the server $T_{proc}(s)$, the indicators $T_{resp}(s)$ and $T_{gNB}(s)$ enable the calculation of the overall estimate of the total request service time, which structurally corresponds to the empirical RTT.

As previously noted, the classical analogues selected for comparison with the proposed mathematical framework are the M/M/1 and M/M/c models. The M/M/1 model is used to describe a system with a single queue and a single server. In the context of this basic model, the average time a request spends in the system (including both queueing and actual service time) is given by $RTT_{M/M/1}=1(\mu-\lambda)$, where λ represents the intensity of the Poisson arrival process, and $\mu=12$ is the parameter of the exponential distribution corresponding to the average service rate of the single server. To avoid discontinuities in the corresponding function graph, a constraint $\lambda<\mu$ was introduced. The M/M/c model is a generalisation of M/M/1 and describes a system with c parallel servers (for example, in cases where the base station processes requests in parallel across multiple channels). The RTT in the M/M/c model is defined by expression $RTT_{M/M/c}=\frac{1}{\mu}+\frac{\rho^c}{c!(1-\rho)}\frac{1}{c\mu(1-\rho)}$, where $\rho=\lambda/c\mu$ is the system load metric. All parameters of the classical models were selected to ensure that the RTT estimates derived from them closely approximate the real-world data from the dataset.

Fig. 2 presents a comparison of RTT estimates in a high-loaded 5G-IoT ecosystem, calculated using the proposed model (red line), the M/M/1 model (blue line), and the M/M/c model (green line). The black line represents the empirical RTT values obtained from the dataset. The grey shaded area around this curve represents the confidence interval, reflecting the fluctuation of RTT under real-world operating conditions. The high data dispersion indicates significant variability in latency across different network load scenarios.

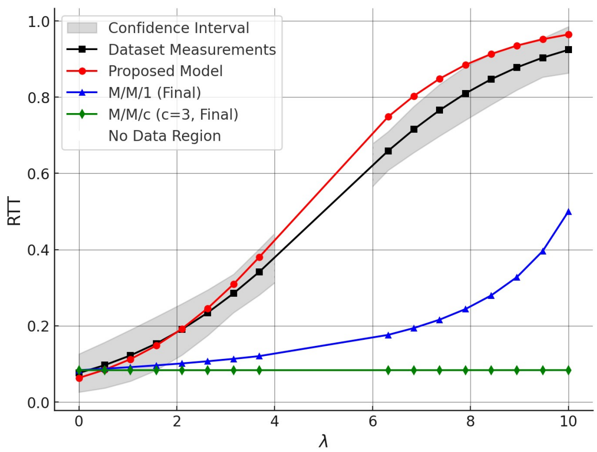

Figure 2: Verification Results of the Proposed Mathematical Framework.

An analysis of the real data dynamics in Fig. 2 reveals a gradual increase in RTT as the load grows, represented by the argument λ, which denotes the input flow intensity. In the low-load range $\lambda < 2$, RTT values remain low, as the system operates in a regime where the servicing capacity exceeds the volume of incoming requests, resulting in minimal delays. In the mid-load range $2 \le \lambda < 6$, a noticeable increase in RTT is observed, indicating a gradual saturation of network resources and the emergence of queue formation within the system. Under $\lambda \ge 6$, the latency continues to grow, though at a slower rate, as the system reaches a relatively stable operational region under high-load conditions. The confidence interval (grey area) reflects the variability of measured RTT values across different conditions, driven by the heterogeneity of the network environment and the influence of adaptive resource management in 5G. In the masked region $\lambda \in [4,6]$, empirical data were unavailable; however, interpolation of the confidence interval allows for an estimation of the likely RTT distribution within this range.

The red line representing the proposed model demonstrates a high degree of alignment with the empirical data, particularly in the mid- and high-load ranges λ. In the low-load region $\lambda < 2$, the model slightly overestimates RTT, which can be attributed to its design focus on providing more accurate predictions of system behaviour under high-load conditions. It is important to note that, beyond the masked region, the proposed model naturally continues the trend of the empirical data, reinforcing its predictive capability. In contrast, the classical M/M/1 model (blue line) exhibits substantially different behaviour. It predicts a sharp increase in RTT with increasing λ, which is a typical characteristic of systems with a single server and an unbounded queue. This can be observed in the graph, where for $\lambda \ge 6$, the RTT values predicted by the M/M/1 model significantly exceed the empirical results. The main reason for the discrepancy between M/M/1 and the real data lies in the model's inability to account for adaptive resource management, which is intrinsic to 5G-IoT systems, as well as its lack of support for parallel request servicing. The green line, representing the M/M/c model with $c = 3$, exhibits a more moderate increase in RTT compared to M/M/1, which can be attributed to the presence of multiple service channels. However, RTT values in this model remain substantially lower than the empirical measurements, particularly under high-load conditions λ, indicating its overly optimistic assumption of uniform request distribution across service channels. In reality, in 5G-IoT ecosystems, channel load is dynamically variable and influenced by numerous factors, including traffic type, adaptive radio resource scheduling, and retransmission mechanisms (such as HARQ and ARQ), which are not considered in the M/M/c model.

To visually demonstrate the effectiveness of the proposed mathematical framework, a block of three-dimensional visualisations was developed (see Figure 3), illustrating the dynamics of key performance parameters of the 5G-IoT ecosystem model for a selected dataset instance under varying operational conditions. The visualisation is based on the analytical expressions presented in Sections 2.2 and 2.3 of the article, particularly formulas (7)-(9), which describe the average response time T_{resp}, blocking probability P_{block}, and idleness probability P_{idle}; formula (1), which models resource waiting time at the base station T_{gNB}; and the expression for sensor network throughput $\Lambda_s = Q\lambda_s$, which determines the volume of processed requests.

As independent variables, we selected the input request intensity $\lambda \in [0.5, 1.0, \ldots, 10]$, which characterises the load arriving at the base station from multiple sensor devices, and the number of devices in the sensor network $N_s \in [1, 10]$, which influences the overall throughput of the subnet. The selected range for N_s reflects a typical volume of sensor nodes within a single communication domain in a distributed environment. The load coefficient was fixed at level $Q = 0.8$, simulating a near-saturation operating condition characteristic of critical applications. For the baseline performance visualisation, we considered a 5G-IoT ecosystem scenario with a single representative sensor network $s \in S$ under fixed routing conditions. This setup isolates the effects of request intensity and device count on performance metrics, avoiding interference from inter-network interaction effects, which will be the subject of subsequent experiments.

The three-dimensional visualisation $T_{resp}(\lambda_s, N_s)$ presented in the block Figs. 3 revealed a consistent increase in latency with rising input traffic intensity λ, particularly under conditions where the sensor network comprises a small number of devices. This result is based on formula (8), where latency is defined through the effective throughput Λ_s, which in turn depends on the load coefficient Q, the number of devices N_s, and the input intensity. In contrast to classical models—where such effects are either described in general terms or not accounted for at all—the proposed model enables the generation of a detailed latency curve, which is critically important for 5G scenarios with strict URLLC requirements. The plot of request blocking probability $T_{block}(\lambda_s, N_s)$ made it possible to identify high-risk zones for system overload. At low values of N_s and high λ, the model exhibits a sharp increase in blocking probability, as captured by formula (7), which accounts for radio resource limitations and the exponential nature of conflicts during simultaneous access. These findings are particularly significant for mMTC scenarios, where data loss can have a systemic impact and blocking at a single node may trigger a cascading failure effect. Of particular analytical value is the visualisation of resource idleness probability $T_{idle}(\lambda_s, N_s)$, which shows that under low traffic conditions, the system is underutilised and resources remain unused. This information enables assessment not only of the system's capacity limits but also of its efficiency under partial load conditions, which is critical for designing balanced and resilient 5G-IoT infrastructures. This is particularly important for distributed or energy-efficient architectures in 5G-IoT systems, where DRX or HARQ mechanisms introduce intermittency in data transmission and processing. The plot of resource allocation waiting time at the base station $T_{gNB}(\lambda_s, N_s)$ enables the assessment not of the final delay, but of the internal time losses associated with accessing local servers. In a real 5G-IoT ecosystem, this process is governed by complex NR Scheduling mechanisms, which take into account pre-established priorities and the dynamic switching between local servers and global cloud processing environments for sensor data. The proposed mathematical framework enables the modelling of these effects through variable service parameters, accounting not only for traffic intensity but also for the structural characteristics of the ecosystem itself.

To confirm the practical effectiveness of the proposed mathematical framework, three targeted experiments were conducted, the results of which are presented in Fig. 4. Each of these experiments provides informative representation of a corresponding key operational scenario within a 5G-IoT ecosystem, namely, the energy-saving mode of the base station (formula (12)), collision reduction through fragmentation (formulas (14), (15)), and adaptation to frequency dynamics under the AFH mode (formulas (16), (17)).

The objective of the first experiment was to evaluate the probability of the gNB transitioning into the DRX energy-saving state in the absence of requests during the control period TCM. The calculation was based on an exponential dependency (formula (12)), which aggregates the influence of traffic intensity, processing capacity, and the number of sensor networks. The modelling process employed both fixed and independent parameters denoted by $\lambda_s = 1$, $\mu_{proc} = 0.7$, $\mu_{gNB} = 0.65$, $Q = 0.85$, $S \in [5, 50]$. This experiment was designed to analyse the model's sensitivity to changes in topology and load at the gNB level, providing insights into its adaptability under varying 5G-IoT deployment conditions.

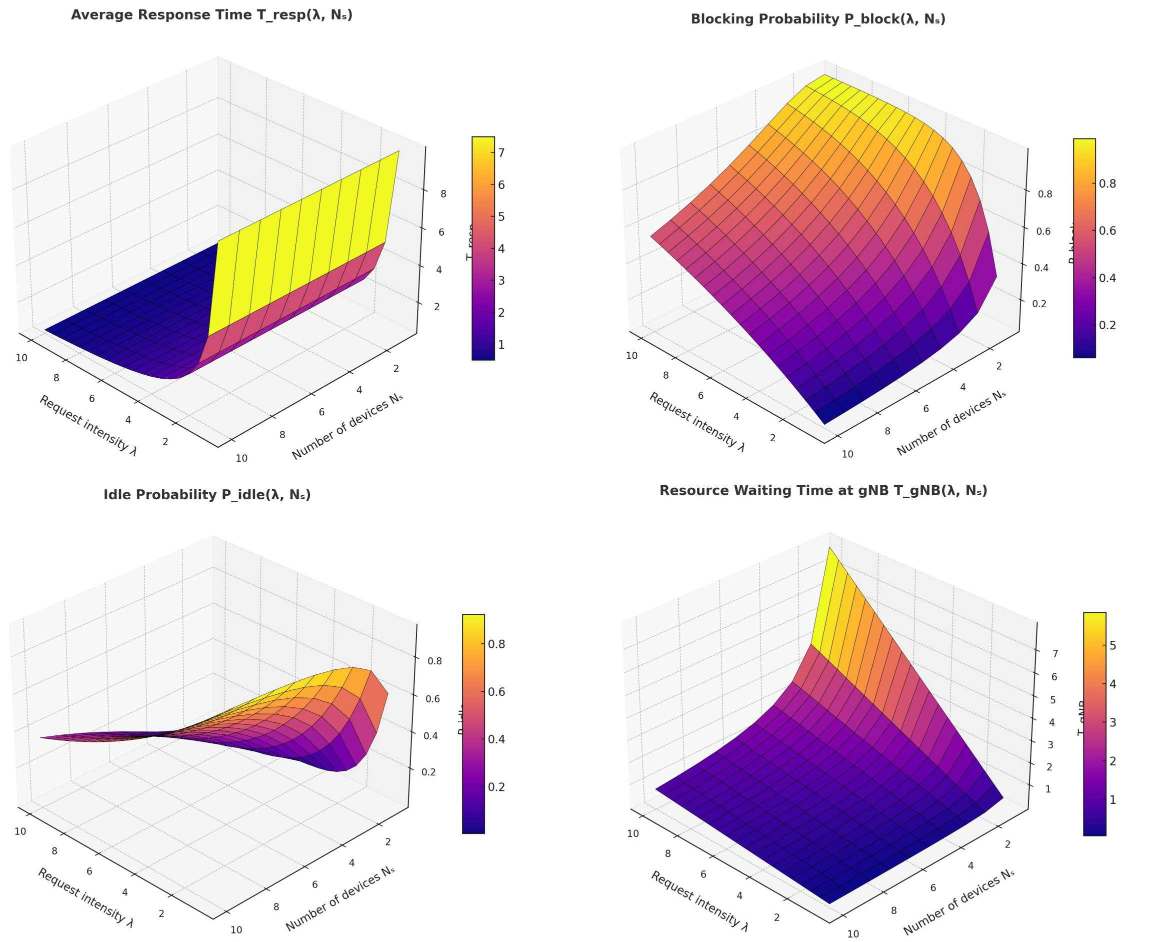

Figure 3: Performance Metrics of the 5G-IoT Ecosystem under Varying Traffic Intensity and Device Count.

The objective of the second experiment was to determine the impact of packet fragmentation on the probability of communication channel blocking. The calculation of the effective length of a fragmented packet was performed using formula (14), taking into account the average fragment length and the length of the final segment. To evaluate the blocking probability, expression (15) was applied, which models its dependency on traffic with exponential saturation characteristics. The computational procedure was implemented in two modes: with fragmentation and without it, using the following fixed and independent parameters: $n_f = 4$, $z_m = 0.045$, $z_l = 0.12$, $\lambda \in [0.5, 10]$. This experiment characterises the model's capability to account for resource holding time reduction effects resulting from traffic segmentation.

The third experiment tested the model's adaptability to frequency dynamics under NR-U conditions. Within this scenario, the processing and response times were adjusted according to formula (16), which accounts for time losses due to frequency switching. The total response delay was evaluated using formula (17), which incorporates the non-linear influence of parameter T_{dwell}, varied within the range $T_{dwell} \in [5, 30]$ ms. During modelling, the values of parameters T_{resp}, T_{proc}, Λ_s were balanced using appropriate correction coefficients, which were set proportionally to changes in the available frequency window. This experiment focused on validating the model's adequacy under adaptive frequency planning, which is typical of scenarios with high interference levels and spectrum usage constraints.

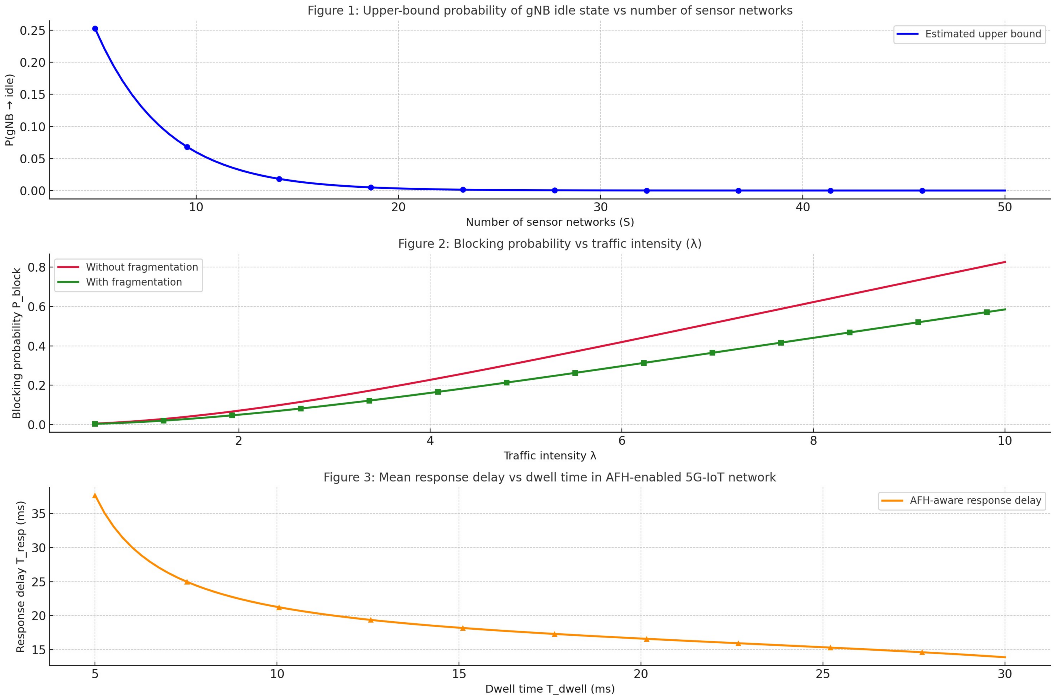

Figure 4: Analytical Validation of Performance Metrics in a High-Loaded 5G-IoT Ecosystem under Packet- and Frequency-Level Adaptation.

The results of the first experiment, presented in Fig. 4, demonstrate an exponential decrease in the probability of the gNB base station transitioning to the DRX energy-saving state as the number of sensor networks increases. This leads to a key practical conclusion: scaling the ecosystem by increasing the number of devices contributes to the stabilisation of its active state. As the number of nodes grows, the probability of a complete absence of global traffic drops, even when the load on each individual local sensor network remains moderate. From the second experiment, a clear distinction can be observed between the packet fragmentation and non-fragmentation modes. Fragmentation proves to be an effective strategy for reducing blocking probability, especially under moderate load conditions. This effect can be explained by the fact that dividing transmitted data reduces the resource holding time per device, which in turn lowers contention for access channels. Practically, this supports the application of fragmentation as a flexible access control mechanism in overloaded network segments, particularly at the MAC layer. The third experiment analysed response delay under AFH operation mode in the NR-U frequency band. The dependencies visualised in Fig. 4 show a sharp increase in latency as the parameter T_{dwell}, which defines the dwell time on a single frequency, decreases. The region below 10 ms is particularly critical, where time losses due to frequency switching begin to outweigh effective data transmission. From a practical standpoint, this highlights the need for careful tuning of frequency planning parameters in high-interference environments. Taken together, the results of the three experiments confirm the model's capability to accurately capture the complex interplay between architectural, protocol-level, and frequency-level components of a 5G-IoT ecosystem.

4. Conclusions

The relevance of this research topic is driven by the rapid rise in the importance of 5G-IoT ecosystems in mission-critical domains, where the efficient aggregation of traffic from multiple sensor networks with guaranteed real-time Quality of Service (QoS) is a fundamental requirement for system operation. Traditional analytical approaches fail to account for the dynamic nature of such networks, which manifests in topological variability, traffic fragmentation, the need for energy efficiency, and adaptation to frequency instability. Therefore, the development of a novel mathematical model capable of accurately representing the operation of high-loaded 5G-IoT ecosystems is both scientifically justified and practically essential.

The scientific novelty of this work lies in the formalisation of a multi-component mathematical framework for the analytical evaluation of the performance of 5G-IoT ecosystems, taking into account key factors such as NR Scheduling, packet fragmentation, DRX mode, HARQ mechanisms, and adaptive frequency hopping in NR-U mode. The proposed model is the first to integrate the interaction of multiple sensor networks with a gNB base station within a generalised system of equations, enabling the estimation of blocking and idleness probabilities, average response, processing, and waiting times, as well as the evaluation of boundary operating regimes with consideration for energy-saving mechanisms.

The conducted experiments confirmed the adequacy and high effectiveness of the proposed model. Verification was performed using the open-access dataset from Simula Research Lab, which contains seven weeks of data on the real-world operation of 5G, LTE, and NB-IoT networks. The analytical estimation of the T_{resp} showed a deviation of less than 12% compared to empirical RTT values, and within the medium load range $2 \le \lambda < 6$, this error was only 6–9%, significantly outperforming the accuracy of classical models such as M/M/1 and M/M/c, which exhibited errors of up to 40%. Within the experimental block, it was also demonstrated that the probability of the gNB transitioning to the energy-saving DRX mode does not exceed 2% when operating with more than 30 sensor networks at a request generation intensity $\lambda = 1$, confirming efficient system utilisation. Moreover, packet fragmentation with parameters $n_f = 4$, $z_m = 0.045$, $z_l = 0.12$ resulted in a 17% reduction in channel blocking probability compared to the non-fragmented scenario. Adaptation to frequency dynamics in AFH mode, with parameters $T_{dwell} = 10$ and $t_c = 2$, led to an 8–11% reduction in effective throughput, which was accounted for in the model through modified waiting time coefficients.

The practical value of the developed analytical framework lies in its ability to safely predict system behaviour under load without the need for resource-intensive simulations. This opens promising opportunities for applying the model in real-world 5G-IoT infrastructure design, overload risk assessment, optimisation of energy-saving modes, and adaptation to frequency instability. The proposed approach accounts for dynamics at both the physical channel level and the logical traffic routing level, making it suitable for multi-service scenarios in mission-critical applications.

Future research will focus on extending the model to incorporate heterogeneous QoS traffic classes (URLLC, mMTC, eMBB), developing integrated models for multi-gNB topologies, integrating Network Slicing functionality, and designing analytical mechanisms for monitoring performance recovery following peak load conditions.

Acknowledgements

The authors are grateful to all colleagues and institutions that contributed to the research and made it possible to publish its results.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

Funding

This research is part of the project No. 2022/45/P/ST7/03450 co-funded by the National Science Centre and the European Union Framework Programme for Research and Innovation Horizon 2020 under the Marie Skłodowska-Curie grant agreement No. 945339.

References

[1] A. da Silva, A. J. Marques Cardoso, Industrial Internet of Things-Driven Coopetition Networks for SMEs, in: Procedia Computer Science, vol. 253, Elsevier, 2025, pp. 227–236. doi:10.1016/j.procs.2025.01.086.

[2] D. Sah, V. Kumar, M. Vahabi, H. Fotouhi, Federated learning at the edge in Industrial Internet of Things: A review, in: Sustainable Computing: Informatics and Systems, vol. 46, Elsevier, 2025, p. 101087. doi:10.1016/j.suscom.2025.101087.

[3] B. P. Wong, B. Kerkez, Real-time environmental sensor data: An application to water quality using web services, in: Environmental Modelling & Software, vol. 84, Elsevier, 2016, pp. 505–517. doi:10.1016/j.envsoft.2016.07.020.

[4] S. Rommer, C. Mulligan, P. Hedman, M. Olsson, L. Frid, S. Sultana, Architecture overview, in: The Core Network for 5G Advanced, Elsevier, 2025, pp. 17–81. doi:10.1016/b978-0-443-29188-3.00006-5.

[5] [M. Marcozzi, O. Gemikonakli, E. Gemikonakli, E. Ever, L. Mostarda, Availability evaluation of IoT systems with Byzantine fault-tolerance for mission-critical applications, in: Internet of Things, vol. 23, Elsevier, 2023, p. 100889. doi:10.1016/j.iot.2023.100889.

[6] D. Narciandi-Rodriguez, J. Aveleira-Mata, M. T. García-Ordás, J. Alfonso-Cendón, C. Benavides, H. Alaiz-Moretón, A cybersecurity review in IoT 5G networks, in: Internet of Things, vol. 30, Elsevier, 2025, p. 101478. doi:10.1016/j.iot.2024.101478.

[7] S. Popli, R. K. Jha, S. Jain, A comprehensive survey on Green ICT with 5G-NB-IoT: Towards sustainable planet, in: Computer Networks, vol. 199, Elsevier, 2021, p. 108433. doi:10.1016/j.comnet.2021.108433.

[8] L. Liu, X. Guo, C. Lee, Promoting smart cities into the 5G era with multi-field IoT applications powered with advanced mechanical energy harvesters, in: Nano Energy, vol. 88, Elsevier, 2021, p. 106304. doi:10.1016/j.nanoen.2021.106304.

[9] W. U. Rehman, M. A. Koondhar, S. K. Afridi, L. Albasha, I. H. Smaili, E. Touti, M. Aoudia, W. Zahrouni, I. Mahariq, M. M. R. Ahmed, The role of 5G network in revolutionizing agriculture for sustainable development: A comprehensive review, in: Energy Nexus, vol. 17, Elsevier, 2025, p. 100368. doi:10.1016/j.nexus.2025.100368.

[10] B. T. Jijo, S. R. M. Zeebaree, R. R. Zebari, M. A. M. Sadeeq, A. B. Sallow, S. Mohsin, Z. S. Ageed, A Comprehensive Survey of 5G mm-Wave Technology Design Challenges, in: Asian Journal of Research in Computer Science, Sciencedomain International, 2021, pp. 1–20. doi:10.9734/ajrcos/2021/v8i130190.

[11] D. A. Sehrai, M. Asif, J. Khan, M. Abdullah, W. A. Shah, S. Alotaibi, N. Ullah, A High-Gain and Wideband MIMO Antenna for 5G mm-Wave-Based IoT Communication Networks, in: Applied Sciences, vol. 12, no. 19, MDPI, 2022, p. 9530. doi:10.3390/app12199530.

[12] M. M. Ahamed, S. Faruque, 5G Network Coverage Planning and Analysis of the Deployment Challenges, in: Sensors, vol. 21, no. 19, MDPI, 2021, p. 6608. doi:10.3390/s21196608.

[13] I. Selinis, K. Katsaros, M. Allayioti, S. Vahid, R. Tafazolli, The Race to 5G Era; LTE and Wi-Fi, in: IEEE Access, vol. 6, IEEE, 2018, pp. 56598–56636. doi:10.1109/access.2018.2867729.
[14] M. A. Ouamri, Stochastic geometry modeling and analysis of downlink coverage and rate in small cell network, in: Telecommunication Systems, vol. 77, no. 4, Springer, 2021, pp. 767–779. doi:10.1007/s11235-021-00770-5.
[15] Y. Hmamouche, M. Benjillali, S. Saoudi, H. Yanikomeroglu, M. D. Renzo, New Trends in Stochastic Geometry for Wireless Networks: A Tutorial and Survey, in: Proceedings of the IEEE, vol. 109, no. 7, IEEE, 2021, pp. 1200–1252. doi:10.1109/jproc.2021.3061778. Y. O. Imam-Fulani et al., 5G Frequency Standardization, Technologies, Channel Models, and Network Deployment: Advances, Challenges, and Future Directions, in: Sustainability, vol. 15, no. 6, MDPI, 2023, p. 5173. doi:10.3390/su15065173.
[16] B. S. Khan, S. Jangsher, A. Ahmed, A. Al-Dweik, URLLC and eMBB in 5G Industrial IoT: A Survey, in: IEEE Open Journal of the Communications Society, vol. 3, IEEE, 2022, pp. 1134–1163. doi:10.1109/ojcoms.2022.3189013.
[17] X. Ye, L. Fu, J. M. Cioffi, Joint Codebook Selection and MCS Adaptation for MmWave eMBB Services Based on Deep Reinforcement Learning, in: IEEE Internet of Things Journal, vol. 11, no. 19, IEEE, 2024, pp. 31545–31560. doi:10.1109/jiot.2024.3419900.
[18] D. Feng, L. Lai, J. Luo, Y. Zhong, C. Zheng, K. Ying, Ultra-reliable and low-latency communications: applications, opportunities and challenges, in: Science China Information Sciences, vol. 64, no. 2, Springer, 2021. doi:10.1007/s11432-020-2852-1.
[19] B. S. Khan, S. Jangsher, A. Ahmed, A. Al-Dweik, URLLC and eMBB in 5G Industrial IoT: A Survey, in: IEEE Open Journal of the Communications Society, vol. 3, IEEE, 2022, pp. 1134–1163. doi:10.1109/ojcoms.2022.3189013.
[20] S. Tripathi, N. V. Sabu, A. K. Gupta, H. S. Dhillon, Millimeter-Wave and Terahertz Spectrum for 6G Wireless, in: Computer Communications and Networks, Springer, 2021, pp. 83–121. doi:10.1007/978-3-030-72777-2_6.
[21] S. Zahoor, I. Ahmad, M. T. B. Othman, A. Mamoon, A. U. Rehman, M. Shafiq, H. Hamam, Comprehensive Analysis of Network Slicing for the Developing Commercial Needs and Networking Challenges, in: Sensors, vol. 22, no. 17, MDPI, 2022, p. 6623. doi:10.3390/s22176623.
[22] V. Kovtun, I. Izonin, M. Gregus, Formalization of the metric of parameters for quality evaluation of the subject-system interaction session in the 5G-IoT ecosystem, in: Alexandria Engineering Journal, vol. 61, no. 10, Elsevier, 2022, pp. 7941–7952. doi:10.1016/j.aej.2022.01.054.
[23] V. Kovtun, I. Izonin, M. Gregus, Mathematical models of the information interaction process in 5G-IoT ecosystem: Different functional scenarios, in: ICT Express, vol. 9, no. 2, Elsevier, 2023, pp. 264–269. doi:10.1016/j.icte.2021.11.008.
[24] V. Kovtun, O. Kovtun, K. Grochla, O. Yasniy, The quality of service assessment of eMBB and mMTC traffic in a clustered 5G ecosystem of a smart factory, in: Egyptian Informatics Journal, vol. 29, Elsevier, 2025, p. 100598. doi:10.1016/j.eij.2024.100598.
[25] K. K. Kousias et al., A Large-Scale Dataset of 4G, NB-IoT, and 5G Non-Standalone Network Measurements [Dataset], in: IEEE DataPort, IEEE, 2022. doi:10.21227/7A8S-NT68.