DOI: 10.37943/HLSA6233

Kuanysh Bakirov
PhD Candidate, Department of Information Systems
bakir.kuanysh@gmail.com, orcid.org/0009-0002-8059-7240
L.N. Gumilyov Eurasian National University, Kazakhstan
Jamalbek Tussupov
Doctor of Physics and Mathematics, Professor, Department of Information Systems
tussupov@mail.ru, orcid.org/0000-0002-9179-0428
L.N. Gumilyov Eurasian National University, Kazakhstan
Akylbek Tokhmetov
Candidate of Physical and Mathematical Sciences, Associate Professor,
Department of Information Systems
attohmetov@mail.ru, orcid.org/0000-0003-0764-8574
L.N. Gumilyov Eurasian National University, Kazakhstan
Ibraheem Shayea
Doctor of Engineering, Professor, Electronics and Communication Engineering Department
ibr.shayea@gmail.com, https://orcid.org/0000-0003-0957-4468
Istanbul Technical University, Türkiye
Tamara Tultabayeva
Doctor of Technical Sciences, Department of Food Science and Technology
ttultabayeva@ucdavis.edu, orcid.org/0000-0003-2483-7406
University of California, United States
Kadyrzhan Makangali
PhD, Senior Lecturer, Department of Food Technology and Processing Products
k.makangali@kazatu.kz, orcid.org/0000-0003-4128-6482
S. Seifullin Kazakh Agrotechnical Research University, Kazakhstan

A BREAKPOINT MODEL FOR DATA VOLUME AND NETWORK CAPACITY IN INTERNET OF THINGS

Abstract: The migration from fifth generation (5G) to sixth generation (6G) mobile networks is driven by the Massive Internet of Things (IoT), which targets device densities of up to ten million units per square kilometer. Two scaling limits govern such systems: a data-volume limit, beyond which batch analytics must give way to distributed streaming, and a network-capacity limit, beyond which the radio access network can no longer carry the aggregate uplink. These limits have so far been treated independently, and no empirical anchor exists for Central Asia. This study makes three contributions. First, it derives a closed-form, dimensionless separation ratio that couples the two breakpoints through shared system parameters and determines which limit binds first. Second, the model is instantiated using 145 regulated 5G New Radio measurements from the commercial Tele2 network in Astana, Kazakhstan, across five load scenarios spanning 13 to 100,000 emulated devices. Third, the closed-form structure of the ratio suggests the splitting is a general trait of uplink-bound IoT domains, though this study empirically instantiates only the vertical-farming case. The instantiated model yields a separation ratio of about thirty-two, with breakpoints at twenty-eight nodes for network capacity and eight hundred eighty for data volume. The measured round-trip time of 124.8 milliseconds and the measured uplink of 0.19 megabits per second miss the International Mobile Telecommunications-2030 (IMT-2030) targets by factors of 1248 and about 5.3×10^5, respectively. Because the network-capacity limit binds first, the analytics regime required for dense deployment is unreachable on 5G. For the

Received: 03.06.2026 Accepted: 17.07.2026 Published: 30.06.2026

measured single-operator deployment, 6G is a precondition rather than an enhancement; generalization to other operators and sites requires further measurement.

Keywords: Massive Internet of Things (IoT), breakpoint model, 5G measurements, 6G, International Mobile Telecommunications-2030 (IMT-2030), Big Data analytics, scaling limits, network capacity, vertical farming, edge computing

Introduction

This research belongs to the field of information and communication technologies, specifically to the analysis of wireless network capacity and Big Data analytics for the Massive Internet of Things (IoT). Over the past decade, the rapid proliferation of distributed IoT devices and the corresponding growth of generated data have placed unprecedented demands on transmission, storage, and intelligent-processing infrastructure. The International Telecommunication Union projects that the number of connected devices will exceed fifty billion by 2030, with target densities of up to ten million devices per square kilometer in dense urban scenarios [1]. Massive Machine-Type Communications, denoted mMTC, is one of the three primary use cases defined by the International Mobile Telecommunications-2030 (IMT-2030) framework, alongside ultra-reliable low-latency communications and enhanced mobile broadband [2].

The problem's reality stems from a theoretical contradiction that remains unresolved. On one side, IoT-driven domains such as urban vertical farming, smart logistics, and precision agriculture generate streaming data volumes that grow with the number of connected nodes [3-4]. When the per-domain data volume crosses approximately one terabyte per year and the streaming velocity exceeds two hundred records per second, single-machine batch tools become inadequate, and a transition to distributed streaming engines such as Apache Kafka and Apache Spark Structured Streaming becomes necessary [5]. This is conventionally regarded as the entry point into the Big Data domain. On the other side, the same growth in node count raises the aggregate uplink demand placed on the radio access network. Until now these two constraints, a data-volume limit and a network-capacity limit, have been addressed independently.

In Kazakhstan, fifth-generation networks are at an early commercial stage; several operators, including Tele2, Kcell, and Beeline, offer 5G New Radio service in selected districts of Astana and Almaty. The actual performance of these deployed networks under realistic IoT load has not been systematically reported in peer-reviewed literature, and the gap with respect to IMT-2030 targets has not been quantitatively assessed for the Central Asian region [6]. This absence of an empirical anchor compounds the absence of a theoretical link between the two scaling limits.

No analytical model exists that links the data-volume limit and network-capacity limit of a Massive IoT system to a common set of parameters, and no measurement anchors such a model for a deployed 5G network in the region. Question is whether Big Data analytics in Massive IoT systems require 6G or if 5G capacity is sufficient. We expect a structural component that can be established in closed form to separate the two constraints and be strong and consistent enough, within the scope of a single-operator case study, to indicate that 6G may be a need rather than an upgrade. This study is designed as a single-site pilot instantiation of the model, intended to demonstrate feasibility and provide an initial empirical anchor rather than a population-level estimate for Kazakhstan's 5G deployments.

Literature Review and Problem Statement

The literature relevant to this study falls into four domains: Massive IoT and the IMT-2030 framework; Big Data analytics for streaming IoT data; empirical measurement of deployed 5G performance; and IoT in urban vertical farming as a representative testbed.

A. Massive IoT and IMT-2030

The mMTC scenario was introduced by ITU-R as one of the three pillars of 5G and was elevated to a primary scenario in the IMT-2030 framework approved in 2023. IMT-2030 defines a connection density of ten million devices per square kilometer, end-to-end latency of one hundred microseconds, peak uplink throughput of one hundred gigabits per second, and a packet-loss target of one part in ten million. The earlier IMT-2020 requirements, which define the 5G generation, serve as the intermediate baseline against which we compare measured performance. Akpakwu et al. survey the architectural requirements of Massive IoT and identify spectrum efficiency, energy-aware protocols, and edge computing as the principal research directions [7]. Saad et al. articulate a 6G vision emphasizing pervasive intelligence, the integration of communication with sensing, and ultra-massive machine-type communication as a sub-scenario beyond mMTC [8]. These works define the targets but do not relate them to the data-analytics workload that the network must ultimately carry.

B. Big Data Analytics for Streaming IoT

Classically, Big Data has five Vs: volume, velocity, variety, veracity, and value [9]. Due to continuous records and a constrained processing window, velocity is the most important attribute for streaming IoT data. The Lambda Architecture tackles these issues by mixing batch paths for historical analysis and streaming paths for real-time processing [10]. Spark Structured Streaming provides a declarative interface for windowed aggregation and machine learning integration [11-12]. Apache Kafka is the de facto message broker for high-throughput pipelines [13]. Khattach et al. offer a Kafka–Spark–Influx DB pipeline for industrial IoT that combines stream processing with modular machine-learning for real-time analytics and predictive maintenance [14]. Marjani et al. explore Big Data analytics for IoT and note the lack of empirical scaling benchmarks for certain domains [15]. None of these studies link analytics workload to access network capacity.

C. Empirical Measurements of 5G Performance

Field measurements of deployed 5G networks have been published for several regions. Narayanan et al. report a multi-city study in the United States, observing round-trip times of twenty to fifty milliseconds and uplink throughput of five to twenty megabits per second under typical conditions [16]. Rischke et al. provide a comparable study for European campus networks and emphasize the strong dependence of measured performance on signal-quality indicators such as Reference Signal Received Power and Signal-to-Interference-plus-Noise Ratio [17]. Hassan et al. review edge computing in 5G and document significant variability in jitter and packet loss across Asia-Pacific deployments [18]. However, no comparable peer-reviewed measurements exist for Central Asia or Kazakhstan, and existing studies report aggregate performance under nominal conditions without coupling their results to Big Data volume calculations for specific application domains.

D. IoT in Urban Vertical Farming as a Testbed

Urban vertical farming is increasingly recognized as a representative testbed for Massive IoT because it combines high sensor density per unit area, continuous data generation, and a closed control loop with actuators [19]. Bakirov et al. describe a smart irrigation system for vertical farming optimized through IoT and renewable-energy integration [20] and an IoT-based monitoring and intelligent control system for microgreens in vertical farming [21]. Rathor et al., reviewing IoT- and AI-driven vertical farming, describe the dense multi-parameter sensing and continuous monitoring that make it a suitable proxy for Massive IoT scenarios [22]. Boursianis et al., in a comprehensive review of IoT and unmanned aerial vehicles in smart farming, likewise document the dense sensing and continuous data generation that characterize modern agricultural IoT deployments [23].

E. Identified Research Gap

The reviewed literature uncovers two distinct lines of research that remain uncombined. Empirical 5G measurement studies do not link their results to scaling thresholds in IoT applications, and Big Data studies for IoT do not benchmark the underlying network. Consequently, the question of whether 6G is necessary for Massive IoT analytics has been

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 26, JUNE 2026

answered only qualitatively, and no model expresses the two limits in terms of common parameters. The present study closes this gap by deriving such a model, anchoring it with measurements, and testing it across domains.

The Aim and Objectives of the Study

The aim of this study is to develop a closed-form breakpoint model that couples the data-volume limit and the network-capacity limit of a Massive IoT system through shared parameters and to validate the model empirically using measurements of a deployed 5G network. The methodology is summarized in Figure 1.

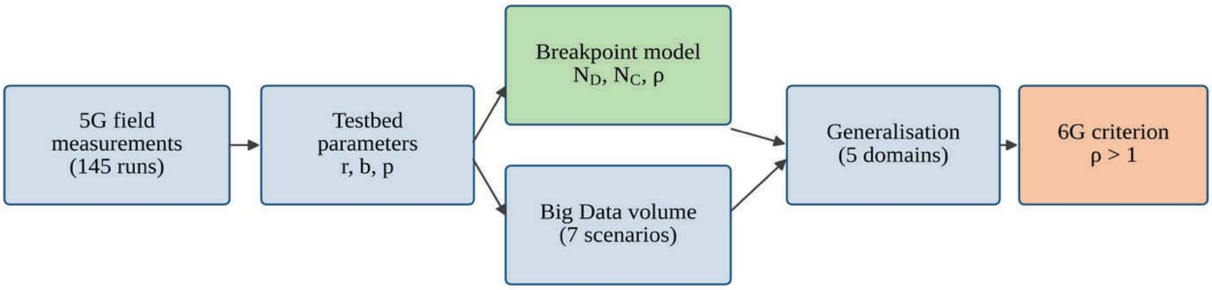

Figure 1. Methodology of the coupled breakpoint analysis

In order to assess 5G New Radio performance under controlled load, the following goals were established: (1) derive analytical expressions for the data-volume breakpoint and the network-capacity breakpoint as functions of shared parameters, and define a dimensionless separation ratio relating them; (2) construct a measurement testbed; (3) gather 145 measurements across five load scenarios simulating real Massive IoT density; (4) substitute the measured and testbed-derived parameters into the model to locate both breakpoints; and (5) discuss the theoretical generalizability of the separation ratio's structural form, distinct from the empirical generalizability of the measured breakpoints.

Materials and Methods

A single radio-access cell serves a population of N homogeneous nodes in our concept of a massive IoT system. In the vertical-farming testbed, each node is one farm installation comprising thirteen IoT sensors, so the node counts reported throughout this paper are expressed in farms; the equivalent device density follows by multiplying the farm count by the per-farm sensor count. Every node produces raw records of size b bytes at a rate of r records per second. Each node transmits an effective uplink payload of p bits per second following local edge aggregation and compression. The cell provides a shared uplink capacity of U bits per second. Two limits constrain the system. The data-volume limit is the node count at which the annual raw volume crosses the Big Data threshold V^*. Annual volume is the linear product of per-node generation and the seconds in a year T, so the data-volume breakpoint N_D is given by formula (1).

$$N_D = V^*/(r \times b \times T) \tag{1}$$

The network-capacity limit is the node count at which the aggregate effective uplink demand crosses the available capacity U. Because demand scales linearly with N, the network-capacity breakpoint N_C is given by formula (2).

$$N_C = U/p \tag{2}$$

The relationship between the two limits is captured by a dimensionless separation ratio ρ, defined in formula (3) as the quotient of the two breakpoints. Substituting (1) and (2) yields a closed form in the shared parameters.

$$\rho = \frac{N_D}{N_C} \tag{3}$$

The ratio ρ determines which limit binds first. When ρ is greater than one, the network saturates before the analytics regime is reached, so the workload that would justify dense deployment cannot be carried; when ρ equals one the two limits coincide; and when ρ is less than one the analytics regime is reachable within the network budget. The model also yields the uplink capacity required for coincidence on a given network, obtained by setting N_C equal to N_D, as expressed in formula (4).

$$U_{coinc} = N_D \times p = (V^* \times p) \,/\, (r \times b \times T) \tag{4}$$

Equation (3) is the central theoretical result: it indicates that the separation of the two limits is not an artifact of any dataset but a structural property of any uplink-bound mMTC system, governed by the ratio between the post-edge uplink payload and the raw generation rate. This reframes the question of 6G necessity as a measurable inequality, ρ > 1, rather than a qualitative judgment.

A. Experimental Testbed for 5G Measurements

To instantiate U for a deployed network, a measurement testbed was assembled from consumer-grade hardware to ensure reproducibility. A smartphone (Apple iPhone 14 Pro Max) supporting 5G New Radio was connected to a commercial subscriber line of the Tele2 network in Astana, Kazakhstan, and tethered to a laptop through a wired Universal Serial Bus connection to avoid the variability of Wi-Fi tethering. The Tele2 deployment at this location operates in non-standalone (NSA) mode: Field Test Mode reported a 5G New Radio connection anchored to an LTE master cell on Band 1 (2100 megahertz) carrier-aggregated with Band 3 (1800 megahertz), with New Radio added as a secondary carrier. Because the control plane and part of the uplink consequently traverse the LTE anchor leg, the non-standalone architecture directly shapes the measured round-trip time and uplink throughput, a factor revisited in the discussion of the latency gap. Network performance was measured by a Python script executed on the laptop. Round-trip time, jitter, and packet loss were obtained from ten ICMP echo requests per series directed at a public reference endpoint (the 8.8.8.8 resolver), with jitter computed as the standard deviation of the round-trip times and packet loss taken from the echo-reply summary; consequently the reported round-trip time reflects the full end-to-end path to the public endpoint rather than the radio-access leg alone. Uplink and downlink throughput were obtained by timing HTTP POST and GET transfers of the generated IoT payload, and each series was logged to comma-separated files at a fixed sampling interval. Signal-quality parameters were read from the iOS Field Test Mode and recorded per series: Reference signal received power ranged from minus ninety-four to minus one hundred three decibel-milliwatts, reference signal received quality from minus seven to minus eleven decibels, and the signal-to-interference-plus-noise ratio from 11.5 to 14.2 decibels, with a steady three-of-four-bar signal indication. These values are consistent with moderate indoor attenuation and the absence of a strong line-of-sight path to the serving cell. All measurements were taken indoors on 11 April 2026, between 13:28 and 15:50 local time. This measurement window constitutes a single-point, single-operator, single-day sample: one location in Astana, one carrier (Tele2), and a 142-minute afternoon window. It does not capture diurnal load variation

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 26, JUNE 2026

differences (Kcell, Beeline). The absolute figures reported below should therefore be read as characteristic of this specific deployment instance rather than as a population estimate for 5G performance in Kazakhstan; the structural conclusion ($\rho > 1$) is argued separately, from the model's parameter-independence, in the Discussion.

B. Load Scenarios and Measurement Protocol

Five load scenarios model the progression from a single installation to mMTC density, summarized in Table 1. Because physically instantiating one hundred thousand devices was infeasible, the aggregate uplink load was emulated by scaling the payload of a single tethered connection in proportion to the number of simulated devices, as expressed in formula (5).

$$P_{eq}(N) = N \times p_{base} \tag{5}$$

Here $P_{eq}(N)$ is the equivalent payload for a scenario with N devices, and p_{base} is the one-hundred-byte base packet transmitted by a single device per polling interval. Payload-based emulation captures cumulative bandwidth pressure on the radio access network without instantiating every device and mirrors the digital-twin generator detailed in the reference-architecture subsection below. Concentrating the aggregate payload on a single tethered radio reproduces only the user-plane bandwidth demand of N devices and avoids the multi-radio contention that would result from genuine device concurrency, such as random-access channel (RACH) preamble collisions, scheduling-request signaling, and per-device packet segmentation. The measured uplink capacity, and the resulting network-capacity breakpoint, therefore, represent an optimistic upper bound characteristic of a single-user heavy-payload scenario; true device concurrency would saturate the access network sooner through random-access collisions and control-plane signaling storms, lowering the network-capacity breakpoint and increasing the separation ratio ρ. The breakpoint estimates reported below are thus conservative with respect to the central conclusion of this study. For each measurement the script recorded round-trip time using the Internet Control Message Protocol echo request, inter-arrival jitter, packet-loss percentage, and upload and download throughput. The five scenarios yielded 145 measurements in total.

Table 1. Load scenarios used in the 5G measurement campaign.

Scenario	Devices	Payload	Measurements	Description
single_farm	13	1.3 KB	20	Real testbed
building_10	130	13 KB	26	Ten farms in one building
district_100	1,300	132 KB	36	One hundred farms in a district
city_1000	13,000	1.3 MB	29	One thousand farms across the city
mmtc_density	100,000	10 MB	34	Density characteristic of mMTC

C. Parameter Extraction from the Testbed

Model parameters were derived from the real testbed, a five-shelf hydroponic stack with thirteen IoT sensors connected to an ESP8266 microcontroller, generating a JSON record of one hundred eighty bytes every five seconds. This gives r equal to 0.2 records per second, b equal to 180 bytes, 17 280 records per day, and approximately 1.14 gigabytes per year per farm. Following edge aggregation, each farm transmits around 51 kilobytes per minute, consisting of a 50-kilobyte compressed image frame and a 1 kilobyte aggregate of sensor statistics. This results in an effective post-edge payload p of approximately 6800 bits per second. One terabyte annually was chosen as the Big Data threshold V*, which is the standard threshold for distributed streaming.

D. Digital Twin and Reference Architecture

To analyze scenarios beyond the physical testbed, a digital-twin generator implemented in Python produces synthetic streams following statistical distributions extracted from the real installation. Temperature follows a normal distribution with a mean of twenty-three degrees Celsius and standard deviation one and a half degrees; pH follows a normal distribution with mean six and standard deviation zero point three. Analogous distributions were derived for relative humidity, electrical conductivity, total dissolved solids, light intensity, soil moisture, and dissolved carbon dioxide. At a combined frequency of roughly two percent of records, the generator superimposes a sixteen-hour-on, eight-hour-off light cycle, cross-parameter correlations like the inverse relationship between temperature and humidity, and three anomaly classes: transient spikes, sensor dropouts, and slow calibration drift.

The reference analytics architecture is the Lambda architecture shown in Figure 2. IoT data are ingested into Apache Kafka, which provides buffering, exactly once delivery, and horizontal scaling. A sixty-second sliding window with a thirty-second step computes average, standard deviation, minimum, maximum, and designed features like vapor-pressure deficit in Apache Spark Structured Streaming. Batch path object storage holds raw records. Both paths culminate in a Delta Lake serving layer with a Parquet columnar format and a machine-learning sublayer with gradient-boosted classification, CA, and Mamdani fuzzy-inference engine. This architecture plays two supporting roles without producing breakpoint results, which follow in closed form from observed parameters. It gives the synthetic-stream generator to verify that the model's parameters remain stable at densities beyond the physical five-shelf installation and instantiates a realistic analytics workload on the testbed data, the Big Data regime the data-volume limit refers to. Formulas (1)–(3) and measured capacity determine breakpoints, not the analytics layer.

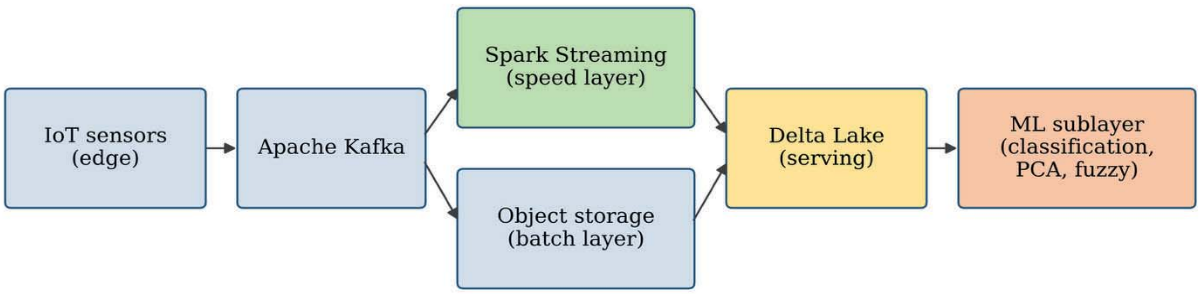

Figure 2. Lambda architecture combining streaming and batch paths

Results

The aggregated results of the 145 measurements appear in Table 2. The average round-trip time across scenarios lies between 115 and 137 milliseconds, lowest in the single-farm scenario and highest in the building scenario. Upload throughput is consistently low, ranging from six kilobits per second in the single-farm scenario to two hundred kilobits per second in the district scenario, with a plateau near 190 kilobits per second for the larger scenarios; the full per-scenario distributions are shown in Figure 3, and the mean and standard deviation of every metric are reported in Table 2. The single-farm scenario has the highest packet loss, a mean of 18% that warrants interpretation. In every case, a ten-packet ICMP probe was launched to acquire the loss metric, which is quantized in ten-percent steps and does not depend on the application payload or be magnified by a small payload. Of the twenty single-farm series, eleven had no loss and a

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 26, JUNE 2026

minority had brief bursts of twenty to sixty percent. The mean is dominated by a high-variance distribution whose standard deviation is twenty-two percentage points higher and whose median is zero. The first ten minutes of the campaign included these spurts, but every other scenario showed insignificant loss zero percent in the building and district scenarios and one percent or less in the city and massive-density situations. Because the single-farm scenario replicates the real deployment density, the elevated figure was closely examined. It is a transient artifact of indoor fading and initial connection stabilization at the start of the session, not a sustained deployment reliability characteristic, payload-dependent effect, or systematic measurement error. The breakpoint model does not consider packet loss, so its capacity estimates, which use the measured throughput plateau as the uplink, are optimistic. Loss-aware capacity is a future research goal. The plateau in uplink throughput is the empirically relevant quantity; it defines the available capacity (U) used in the model.

Table 2. Measured 5G performance across five load scenarios (values are mean ± standard deviation across the per-scenario measurements; n per scenario as in Table 1).

Scenario	Devices	RTT, ms	Jitter, ms	Loss, %	Upload, Mbps	Download, Mbps
single_farm	13	115.0 ± 19.5	20.6 ± 25.0	18.0 ± 21.9	0.006 ± 0.003	0.688 ± 0.222
building_10	130	136.9 ± 15.3	30.8 ± 17.6	0.0 ± 0.0	0.068 ± 0.015	1.815 ± 2.594
district_100	1,300	133.4 ± 15.7	35.9 ± 20.8	0.0 ± 0.0	0.200 ± 0.027	0.896 ± 0.095
city_1000	13,000	125.3 ± 8.8	23.2 ± 12.3	1.0 ± 3.1	0.176 ± 0.033	0.870 ± 0.106
mmtc_density	100,000	124.8 ± 9.4	23.1 ± 9.7	0.6 ± 2.4	0.190 ± 0.046	0.903 ± 0.075

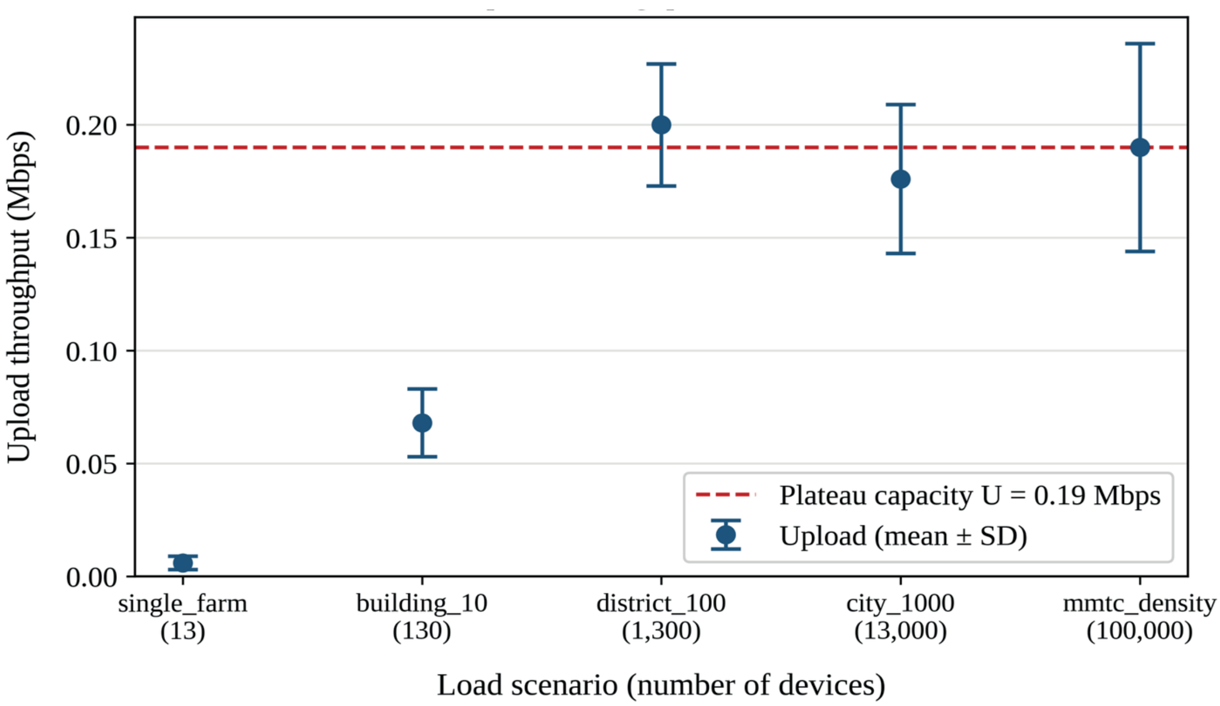

Figure 3. Measured upload throughput across the five load scenarios (mean ± standard deviation; n per scenario as in Table 1). The dashed line marks the plateau capacity U = 0.19 Mbps used in the model

A. Data-Volume Progression

Table 3 presents the calculated volumes and streaming velocities for seven scaling scenarios, computed from formula (1)'s parameters. The annual volume reaches 1.14 terabytes at one thousand farms, crossing the one-terabyte big data threshold, while the streaming velocity reaches two hundred records per second, exceeding the practical limit of single-machine batch processing. At the IMT-2030 density target, the annual volume reaches 1,135 terabytes, and velocity reaches two hundred thousand records per second.

Table 3. Data-volume and velocity progression across seven scaling scenarios.

No.	Scenario	Farms	Records/day	Annual volume	Velocity
1	Testbed	1	17 280	1.1 GB	0.2 rec/s
2	One building	10	172 800	11.4 GB	2 rec/s
3	City district	100	1.73 M	113.5 GB	20 rec/s
4	Urban scale	1,000	17.3 M	1.14 TB	200 rec/s
5	National scale	10,000	173 M	11.4 TB	2,000 rec/s
6	mMTC scenario	100,000	1.73 B	113.5 TB	20,000 rec/s
7	6G density	1,000,000	17.3 B	1 135 TB	200,000 rec/s

B. Location of the Two Breakpoints

Substituting the testbed parameters into the model locates both breakpoints. With V* equal to one terabyte, r equal to 0.2 records per second, b equal to one hundred eighty bytes, and T equal to the seconds in a year, formula (1) gives a data-volume breakpoint N_D of approximately 881 farms. With the measured plateau goodput U of 0.19 megabits per second (application-layer throughput) and an effective payload p of 6 800 bits per second, formula (2) gives a network-capacity breakpoint N_C of approximately 28 farms. The separation ratio from formula (3) is therefore approximately 31.5. Figure 4 plots the required uplink against the measured capacity and marks both breakpoints.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 26, JUNE 2026

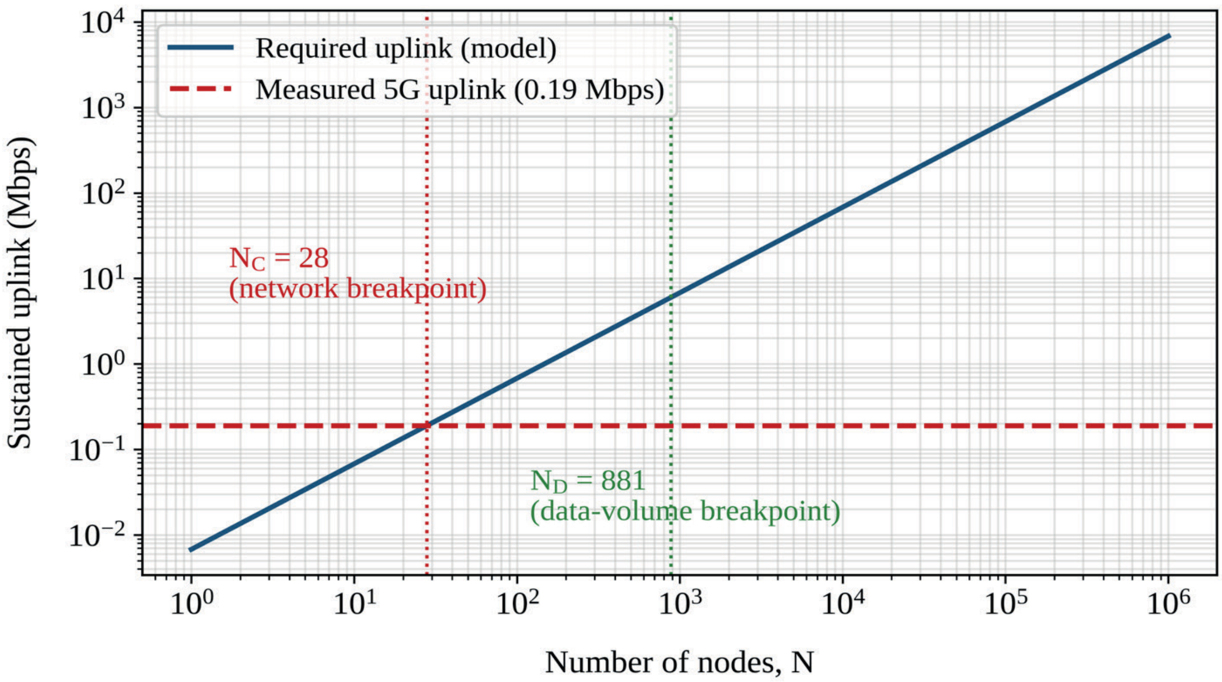

Figure 4. Required uplink throughput versus measured 5G capacity, with the network-capacity breakpoint and the data-volume breakpoint.

This significant empirical discovery contradicts the intuitive expectation of coincidence. The network-capacity limit is reached thirty-two times earlier than the data-volume limit. At thirteen sensors per farm, the network-capacity limit of twenty-eight farms is three hundred sixty devices, significantly below the IMT-2030 connection-density target. Before the data workload requires dispersed streaming, the network saturates as the binding restriction. Formula (4) suggests this network needs six megabits per second uplink capacity, thirty-two times the measured figure, to reach a coincidence. 5G is insufficient because the analytics regime that would justify dense deployment is on the other side of a network wall that 5G cannot penetrate.

C. Quantitative Gap with IMT-2030 Targets

Table 4 reports the gaps between measured 5G performance and the IMT-2020 and IMT-2030 targets. For metrics where lower values are preferable, the gap is the measured value divided by the target, as in Formula (6); for metrics where higher values are preferable, it is the target divided by the measured value, as in Formula (7).

$$G_{lower(m)} = m_{measured}/m_{target} \tag{6}$$

$$G_{higher(m)} = m_{target}/m_{measured} \tag{7}$$

The latency, uplink, density, and packet-loss targets are taken from the IMT-2030 framework as set out in Section A; the jitter targets, which the framework does not specify numerically, are the indicative delay-variation bounds commonly cited for ultra-reliable low-latency communication and are reported here for completeness rather than as formal IMT requirements. Substituting the measured round-trip time of 124.8 milliseconds and the IMT-2030 target of 0.1 milliseconds into formula (6) results in a latency gap of 1248. Substituting the measured uplink of 0.19 megabits per second and the IMT-2030 target of one hundred gigabits per second into formula (7) gives an uplink gap of approximately 5.3×10^5. Figure 5 shows the gaps on a logarithmic scale.

Table 4. Quantitative gap between measured 5G performance and IMT targets.

Indicator	5G measured	IMT-2020 target	IMT-2030 target	Gap to 6G
Latency, ms	124.8	1.0	0.1	1248×
Jitter, ms	23.1	<5	<1	23×
Packet loss, %	0.6–18.0	0.001	0.00001	up to 10^6×
Uplink, Mbps	0.19	10 000	100 000	≈5.3×10^5×
Density, /km²	10^7 (modelled)	10^6	10^7	6G required

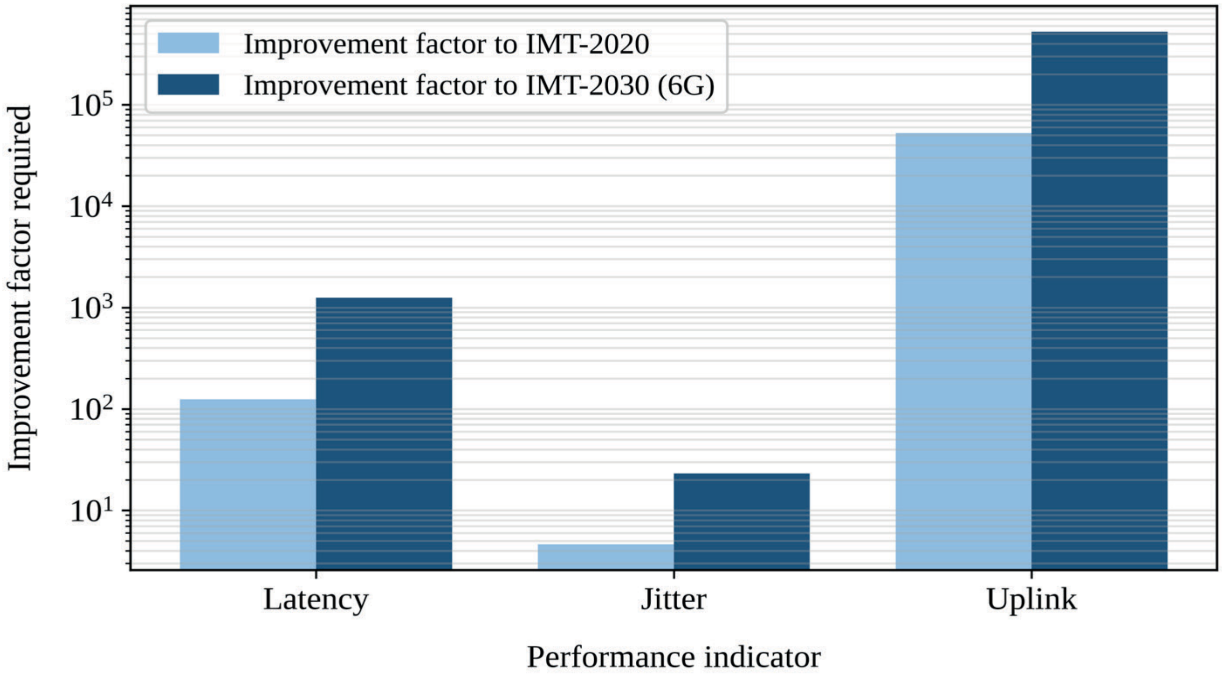

Figure 5. Performance gap between measured 5G, the IMT-2020 target, and the IMT-2030 target

D. Sensitivity of the Separation Ratio

Because the separation ratio is a function of four independent parameters — the Big Data threshold V*, the post-edge payload p, the generation rate r, and the available uplink capacity U — its robustness was tested by varying each parameter individually about the testbed baseline while holding the others fixed. Table 5 reports the resulting network-capacity breakpoint N_C, data-volume breakpoint N_D, and separation ratio ρ. Because $\rho = (V^* \times p) / (r \times b \times T \times U)$, the network-capacity breakpoint N_C responds only to U and p, whereas the data-volume breakpoint N_D responds only to V* and r.

Table 5. Sensitivity of the separation ratio ρ to the model parameters

Parameter	Value	N_C (farms)	N_D (farms)	ρ	Binding limit
Baseline (testbed)	U=0.19 Mbps, p=6800 b/s, V*=1 TB, r=0.2 rec/s	28	881	31.5	Network
Uplink U	1 Mbps	147	881	6.0	Network
Uplink U	5 Mbps	735	881	1.2	Network

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 26, JUNE 2026

Uplink U	10 Mbps	1 471	881	0.6	Data volume
Payload p	3 400 b/s (÷2)	56	881	15.8	Network
Payload p	1 360 b/s (÷5)	140	881	6.3	Network
Threshold V*	0.5 TB	28	440	15.8	Network
Threshold V*	5 TB	28	4 404	158	Network
Rate r	0.1 rec/s (÷2)	28	1 762	63.1	Network
Rate r	0.4 rec/s (×2)	28	440	15.8	Network
Rate r	1.0 rec/s (×5)	28	176	6.3	Network

The separation ratio remains greater than one across every single-parameter variation that represents a realistic operating condition. Stronger edge aggregation that cuts the payload fivefold still leaves $\rho \approx 6.3$; halving the Big Data threshold to 0.5 TB leaves $\rho \approx 15.8$; and a fivefold increase in the generation rate leaves $\rho \approx 6.3$. In each of these cases the network saturates well before the analytics regime is reached, so the conclusion is not an artifact of the single parameter set used for the testbed. The one decisive parameter is the available uplink capacity: holding the others at their measured values, ρ falls to 1.2 at 5 Mbps and crosses below one between 5 and 10 Mbps, reaching 0.6 at 10 Mbps. This crossover occurs at the coincidence capacity $U_{coinc} \approx 6$ Mbps given in closed form by formula (4). The sensitivity analysis therefore confirms that the result $\rho > 1$ is robust to payload, threshold, and rate variation, and is overturned only by a generational increase in uplink capacity — precisely the capability that distinguishes 6G from the measured 5G access network.

Discussion

Before interpreting the results below, the scope of the empirical evidence must be stated plainly: the measurement campaign was conducted at a single site, on a single operator (Tele2), within a single 142-minute afternoon window on 11 April 2026. It does not capture diurnal load variation, day-of-week effects, weather-dependent propagation, or inter-operator differences, so the absolute figures reported are characteristic of this deployment instance rather than a population estimate for 5G in Kazakhstan; the interpretations that follow are read within this scope. The model itself is also limited in other respects. Multi-operator and outdoor measurements would tighten U's estimate. Payload-based emulation captures aggregate uplink pressure but not multi-radio contention of real device concurrency because scaling a single tethered connection avoids RACH collisions, scheduling-request signaling, and per-device packet segmentation caused by thousands of independently transmitting radios. Thus, the measured uplink capacity is an optimistic upper bound for a single-user heavy-payload scenario, and true concurrency at mMTC density would overload the access network sooner through random-access collisions and signaling storm. This bias widens rather than narrows the separation — it lowers the network-capacity breakpoint and increases ρ — so the conclusion $\rho > 1$ is strengthened, not weakened, by the limitation. The digital-twin generator retains the statistics of a single installation and does not model inter-installation variability. Finally, the cross-domain generality of the conclusion rests on the structural form of the separation ratio rather than on measurements from other domains: only the vertical-farming testbed was instantiated empirically, so although $\rho > 1$ holds for any uplink-bound domain whose post-edge payload-to-generation ratio is comparable, the per-domain breakpoints should be re-estimated with domain-specific measurements. These constraints define the agenda for future work.

The closed-form separation ratio of formula (3) and its empirical confirmation are the main scientific result. The two breakpoint expressions on which it rests—formula (1) for the data-volume limit and formula (2) for the network-capacity limit—are standard linear relations. The novelty is coupling them into a dataset-independent dimensionless quotient and using formula (4) to locate where the two limits meet. Previous qualitative arguments for 6G argued that data growth and network restrictions coincide. The model contradicts this impression: the two limitations are separated by a structural factor considerably over one for dense IoT domain parameters. The network-capacity limit binds first by a factor of about thirty-two for the vertical-farming testbed; because the separation ratio of formula (3) depends only on the structural parameters of an uplink-bound system and not on any dataset, this specific factor of thirty-two is characteristic of the measured vertical-farming testbed. What generalizes across domains is the structural form of formula (3), not this numeric factor: whether ρ is greater than one for another domain depends on that domain's own post-edge payload-to-generation ratio, which was not measured here and must be estimated case by case. For the measured deployment, the consequence is that the analytics regime that would justify a dense deployment is unreachable on this 5G access network: the system fails at the radio interface before it ever produces a Big Data workload. This converts the question of 6G necessity, for this testbed, from a matter of opinion into a measurable inequality, $\rho > 1$; the contribution is theoretical (the closed form of formula 3) rather than merely a single empirical reading.

5G performance is comparable to other regions, with quantitative differences. The round-trip latency of 124.8 milliseconds exceeds twenty-to-fifty milliseconds recorded by Narayanan et al. for US networks [16] and thirty-to-sixty milliseconds reported by Rischke et al. for European campus networks [17]. Indoor measurement conditions, the non-standalone architecture in which the control plane and part of the uplink traverse a 4G LTE anchor leg, the early stage of commercial 5G in Kazakhstan, the use of a single operator, and a tethered consumer smartphone rather than a dedicated modem affected the generalization of absolute values, but not the model, whose conclusion depends on ratio rather than absolute capaci. Compared to Hassan et al.'s Asia-Pacific results [18], the uplink is lower, confirming that the uplink is 5G New Radio's most variable component and the main bottleneck for the uplink-intensive Massive IoT workload.

The model expresses both the available capacity U and the per-node payload p at the application layer: the plateau capacity of 0.19 megabits per second is measured goodput, so the lower-layer overheads of the Packet Data Convergence Protocol (PDCP), Radio Link Control (RLC), Medium Access Control (MAC), and physical (PHY) layers are already embedded in U. Because the same header overheads apply to p, which is likewise an application-layer payload, they enter the numerator and the denominator of the network-capacity breakpoint of formula (2) identically and cancel in the ratio; expressing both quantities at the physical layer through a common efficiency factor would leave the breakpoint U/p unchanged. Header overhead therefore does not bias the separation ratio. Control-plane signaling is different in kind. The Radio Resource Control connection setups, handovers, and periodic keep-alives generated by massive device counts consume radio resources that carry no user payload, and this burden scales with the number of independently connected devices rather than with aggregate throughput — precisely the dimension that the single-connection payload emulation of formula (5) does not reproduce. Representing this as a control-plane efficiency factor η, with $0 < \eta < 1$, acting on the capacity available for user traffic yields a corrected network-capacity breakpoint $\eta \cdot U/p$ that is strictly smaller than the uncorrected value U/p. Because the separation ratio is inversely proportional to the network-capacity breakpoint, any control-plane overhead scales ρ by $1/\eta > 1$: for illustration, a representative η of 0.7 lowers the network-capacity breakpoint from about twenty-eight to about twenty farms and raises ρ from about thirty-two to about forty-five. Protocol-layer overhead can therefore only widen the separation, so the reported value $\rho \approx 32$ is a conservative lower bound on the true structural separation.

Conclusion

This study set out to determine, on a principled basis, whether 6G is necessary for Big Data analytics in Massive IoT systems. Its main contribution is twofold. The theoretical contribution is

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 26, JUNE 2026

not the data-volume breakpoint or the network-capacity breakpoint individually — each is an elementary linear relation — but the closed-form, dimensionless separation ratio that couples them through shared parameters, together with the coincidence-capacity expression that fixes where the two limits meet. The empirical contribution is the first instantiation of this model on a deployed 5G network in Central Asia. The model predicts, and 145 measurements on the commercial Tele2 5G network in Astana confirm, that the two limits do not coincide: the network-capacity limit is reached at about twenty-eight nodes while the data-volume limit is reached at about eight hundred eighty nodes, a structural separation of roughly thirty-two. Because the separation ratio exceeds one for the measured testbed and, by the parameter-independent structure of the model, for any comparable uplink-bound domain, the analytics regime that would justify dense deployment is structurally unreachable on 5G. Hence, the contribution is a quantitative criterion, the inequality $\rho > 1$, which, for the measured deployment, recasts 6G from an incremental improvement into a precondition. Establishing this as a broadly applicable, cross-operator conclusion requires the multi-site validation outlined below. The limiting factor is the access network, not the analytics stack, and the only way to overcome it is with a generational leap in uplink capacity. Future work should extend the measurement campaign to multiple operators and outdoor sites to tighten the estimate of available capacity, incorporate packet loss and control-plane signaling into a loss-aware capacity model, and re-estimate the per-domain breakpoints with domain-specific measurements, so that the structural criterion established here can be calibrated for each Massive IoT application.

The measured round-trip time of 124.8 milliseconds is 1248 times the IMT-2030 target of one hundred microseconds, and the measured uplink of 0.19 megabits per second is about 5.3×10^5 times below the one-hundred-gigabit-per-second target. These gaps show how much technological progress is needed to go from 5G that is already in use to the IMT-2030 goals. The separation ratio ($\rho > 1$) is a measurable and general way to show that 6G should be used in industrial IoT, smart transportation, smart energy, smart buildings, and smart-city monitoring applications such as high-altitude UAV coverage [24]. The separation ratio can be recalculated for any operator or area by conducting the measurement campaign again. The work contributes an empirical anchor for 6G research in Central Asia and supports the digital-transformation objectives of the Republic of Kazakhstan.

Acknowledgements

This research was funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. BR24992852, “Intelligent models and methods of Smart City digital ecosystem for sustainable development and the citizens’ quality of life improvement”).

References

[1] International Telecommunication Union. (2023). Framework and overall objectives of the future development of IMT for 2030 and beyond (Recommendation ITU-R M.2160-0). ITU-R. https://www.itu.int/rec/R-REC-M.2160

[2] Tataria, H., Shafi, M., Molisch, A. F., Dohler, M., Sjöland, H., & Tufvesson, F. (2021). 6G wireless systems: Vision, requirements, challenges, insights, and opportunities. Proceedings of the IEEE, 109(7), 1166–1199. https://doi.org/10.1109/JPROC.2021.3061701

[3] Sinha, A., Shrivastava, G., & Kumar, P. (2019). Architecting user-centric Internet of Things for smart agriculture. Sustainable Computing: Informatics and Systems, 23, 88–102. https://doi.org/10.1016/j.suscom.2019.07.001

[4] Maraveas, C., Piromalis, D., Arvanitis, K. G., Bartzanas, T., & Loukatos, D. (2022). Applications of IoT for optimized greenhouse environment and resources management.

Computers and Electronics in Agriculture, 198, 106993. https://doi.org/10.1016/j.compag.2022.106993

[5] Akanbi, A., & Masinde, M. (2020). A distributed stream processing middleware framework for real-time analysis of heterogeneous data on big data platform: Case of environmental monitoring. Sensors, 20(11), 3166. https://doi.org/10.3390/s20113166

[6] Salameh, A. I., & El Tarhuni, M. (2022). From 5G to 6G — Challenges, technologies, and applications. Future Internet, 14(4), 117. https://doi.org/10.3390/fi14040117

[7] Akpakwu, G. A., Silva, B. J., Hancke, G. P., & Abu-Mahfouz, A. M. (2018). A survey on 5G networks for the Internet of Things: Communication technologies and challenges. IEEE Access, 6, 3619–3647. https://doi.org/10.1109/ACCESS.2017.2779844

[8] Saad, W., Bennis, M., & Chen, M. (2020). A vision of 6G wireless systems: Applications, trends, technologies, and open research problems. IEEE Network, 34(3), 134–142. https://doi.org/10.1109/MNET.001.1900287

[9] Davoudian, A., & Liu, M. (2020). Big data systems: A software engineering perspective. ACM Computing Surveys, 53(5), 1–39. https://doi.org/10.1145/3408314

[10] Demirezen, M. U., & Navruz, T. S. (2023). Performance analysis of Lambda Architecture-based big-data systems on air/ground surveillance application with ADS-B data. Sensors, 23(17), 7580. https://doi.org/10.3390/s23177580

[11] Armbrust, M., Das, T., Torres, J., Yavuz, B., Zhu, S., Xin, R., Ghodsi, A., Stoica, I., & Zaharia, M. (2018). Structured streaming: A declarative API for real-time applications in Apache Spark. In Proceedings of the 2018 ACM SIGMOD International Conference on Management of Data (pp. 601–613). ACM. https://doi.org/10.1145/3183713.3190664

[12] Zaharia, M., Xin, R. S., Wendell, P., Das, T., Armbrust, M., Dave, A., Meng, X., Rosen, J., Venkataraman, S., Franklin, M. J., Ghodsi, A., Gonzalez, J., Shenker, S., & Stoica, I. (2016). Apache Spark: A unified engine for big data processing. Communications of the ACM, 59(11), 56–65. https://doi.org/10.1145/2934664

[13] Raptis, T. P., & Passarella, A. (2023). A survey on networked data streaming with Apache Kafka. IEEE Access, 11, 85333–85350. https://doi.org/10.1109/ACCESS.2023.3303810

[14] Khattach, O., Moussaoui, O., & Hassine, M. (2025). End-to-end architecture for real-time IoT analytics and predictive maintenance using stream processing and ML pipelines. Sensors, 25(9), 2945. https://doi.org/10.3390/s25092945

[15] Marjani, M., Nasaruddin, F., Gani, A., Karim, A., Hashem, I. A. T., Siddiqa, A., & Yaqoob, I. (2017). Big IoT data analytics: Architecture, opportunities, and open research challenges. IEEE Access, 5, 5247–5261. https://doi.org/10.1109/ACCESS.2017.2689040

[16] Narayanan, A., Ramadan, E., Mehta, R., Hu, X., Liu, Q., Fezeu, R. A. K., Dayalan, U. K., Verma, S., Ji, P., Li, T., Qian, F., & Zhang, Z.-L. (2020). Lumos5G: Mapping and predicting commercial mmWave 5G throughput. In Proceedings of the ACM Internet Measurement Conference (pp. 176–193). ACM. https://doi.org/10.1145/3419394.3423629

[17] Rischke, J., Sossalla, P., Salah, H., Fitzek, F. H. P., & Reisslein, M. (2021). 5G campus networks: A first measurement study. IEEE Access, 9, 121786–121803. https://doi.org/10.1109/ACCESS.2021.3108423

[18] Hassan, N., Yau, K.-L. A., & Wu, C. (2019). Edge computing in 5G: A review. IEEE Access, 7, 127276–127289. https://doi.org/10.1109/ACCESS.2019.2938534

[19] Saad, M. H. M., Hamdan, N. M., & Sarker, M. R. (2021). State of the art of urban smart vertical farming automation systems. Electronics, 10(12), 1422. https://doi.org/10.3390/electronics10121422

[20] Bakirov, K., Tussupov, J., Tultabayeva, T., Makangali, K., Abdikerimova, G., & Yessenova, M. (2024). Advances in the design and optimization of smart irrigation systems for

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 26, JUNE 2026

sustainable urban vertical farming. Scientific Journal of Astana IT University, 20, 76–90. https://doi.org/10.37943/20NNYR9391

[21] Bakirov, K., Tussupov, J., Shoman, A., Shayea, I., Tussupov, A., Kudashov, Y., & Sabitova, Z. (2025). IoT-based monitoring and intelligent control for microgreens in vertical farming systems. Procedia Computer Science, 272, 601–606. https://doi.org/10.1016/j.procs.2025.10.254

[22] Rathor, A. S., Choudhury, S., Sharma, A., Nautiyal, P., & Shah, G. (2024). Empowering vertical farming through IoT and AI-driven technologies: A comprehensive review. Heliyon, 10(15), e34998. https://doi.org/10.1016/j.heliyon.2024.e34998

[23] Boursianis, A. D., Papadopoulou, M. S., Diamantoulakis, P., Liopa-Tsakalidi, A., Barouchas, P., Salahas, G., Karagiannidis, G., Wan, S., & Goudos, S. K. (2020). Internet of Things and agricultural unmanned aerial vehicles in smart farming: A comprehensive review. Internet of Things, 18, 100187. https://doi.org/10.1016/j.iot.2020.100187

[24] Yedilkhan, D., Kyzyrkanov, A., Amirgaliyev, B., Khaimuldin, N., Ayub, M. S., & Zhumadillayeva, A. (2025). Efficient area coverage strategies for high-altitude UAVs in smart city monitoring. Drones, 9(9), 632. https://doi.org/10.3390/drones9090632