iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** IV **Month of publication:** April 2026

DOI: https://doi.org/10.22214/ijraset.2026.81643

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

An AI-Driven Secure and Energy-Efficient Integrated IoT-Edge-Cloud Framework for Scalable Real-Time Applications

Yash Tripathi[1], Utkarsh Shukla[2], Saniya Shetve3, Siddhi Shitole[4], Viraja Rane[5], Kajal Singh[6], Aachal Rhatwal[7], Manisha Mane[8]
Dept. of Electronics and Computer Science, Shah and Anchor Kutchhi Engineering College, Mumbai, India

Abstract: ***Out there among streetlights, hospitals, factories, and self-driving cars, tiny networked gadgets now flood networks with constant signals needing instant replies. Most still rely on distant servers which drag out response times - often two to half a second delay - way too slow when split-second decisions matter. Instead of zipping info back and forth endlessly, streaming everything outward burns through small device power supplies fast, sometimes killing battery life in just hours. On top of that, shaky data protection shows up again and again, with three out of four hacks taking advantage of poor coding shortcuts during transfers. A fresh start came from students at Shah and Anchor Kutchhi Engineering College - they built something called AISEE, short for AI-driven Secure Edge. This three-layer setup silently connects 300 low-cost DHT22 sensors with 150 networks of Raspberry Pi 4 systems acting as edge-computers and up to analysis services on AWS. Instead of everything going directly to the cloud, small hubs do quick jobs closer in. Watching it all happen is a smart system running Deep Q-Networks that monitors how things are operating — delays, remaining power, task wait times, even risks — and incrementally figures out whether work should remain local, move upstream or be dropped. As conditions continue to evolve, decisions shift automatically. Little devices with a scant 256KB of memory cope with potent future-proof encryption based on Kyber-512, securing data in record speed of just 0.8 milliseconds, and leaving current RSA techniques for dust at almost two thousand times slower. Compression reduces files dramatically, slicing girth so networks can relax. Over lab tests lasting an entire day, tasks on the city-size scale rolled carelessly through at SAKEC; responses zipped in about sixty-two milliseconds plus or minus eight and bested pure cloud setups by three-quarters. Each operation consumed just five point one joules, slashing power use far below previous levels. Yet systems juggled twenty-four hundred fifty signals per second, double the previous capacity. Security remained firm for ninety-six percent of trials. At-one-line, this system is going to be live-duty as large IoT jobs when BMC rolls it out in late-2026 for ten thousand street monitors in Mumbai that would cut down city electricity demand by 20***

Index Terms: ***Internet of Things, Edge Computing, Cloud Computing, Reinforcement Learning, Deep Q-Network, Post-Quantum Cryptography, Kyber-512, Energy Optimization, Low-Latency Processing, Real-Time Systems, Kubernetes Orchestration, TinyML, Anomaly Detection, Smart City Applications, Mumbai BMC Deployment.***

I. INTRODUCTION

Few doubt the IoT surge now hurtling toward more than 75 billion gadgets online by 2025, pouring out 79.4 zettabytes of information every year. From city lighting that tentatively adjusts to traffic flow, to the hospital trackers observing vital signs at a distance, needs are becoming acute for computing since a fraction of an instant. Machines operate factories with no creases, and self-driving vehicles make instant choices, speed is all that matters. What had become rare, such responsiveness now quietly taken for granted.

IoT designs have serious flaws:-

1) High Latency: Cloud round-trip times of 200-500ms violate real-time guarantees required in autonomous vehicles (need ¡100ms), industrial robots (need ¡50ms), and remote surgery systems (need ¡20ms)

2) Energy Drain: As output in the cloud stream we are bound to around 4-6 hours battery with this sensor com-pared to achieving over 72+ hours if we include more dynamic filtering with DHT22 sensors, transmission current is into at least 1.5mA versus local processing which needs only a total of at least 0.2Ma

3) Security Gaps: in 3 out of 4 breaches, weak encryption (RSA too slow at 1800ms/device) and authentication is the vector; reported IoT attack attempts reached a staggering high of 2.1B in both cybersecurity firms by end of 2025

4) Limits on scalability: The centralized cloud processing create network bottleneck, and can not handle more than 2000 request/second as tested; while using 10,000 concurrent IoT streams through a single AWS region fails.
5) Cost of Explosion: 50 lakhs a month for cloud data transfer for municipal IoT projects in Mumba
6) Data Privacy: Sending raw sensor data, e.g. video streams to server violates GDPR/HIPAA; use-sensitive health/traffic data does not leave edge

We present AISEE: a three-tier framework that meets these challenges with edge computing proximity, AI-based orches-tration of multi-tenant workloads, and cloud-scale analytics. The complete system design is presented in Figure 1.

II. EXISTING SYSTEM LIMITATIONS

Before creating AISEE, our team analyzed three traditional methods:

1) *Cloud-Centric Architecture:* Centralized Clouds centralized cloud servers hardware (in-finite compute capacity) with 245ms average latency and huge network costs for all sensor data transmitted.
2) *Edge-Only Processing:* The Local Processing on Raspberry Pi has 85ms response time, but it doesn't store any historical data for analysis and won't benefit from learning across cities.
3) *Static Hybrid Systems:* Static policies for allocating tasks at the edges/cloud are static and cannot adapt to changing workloads or network conditions.

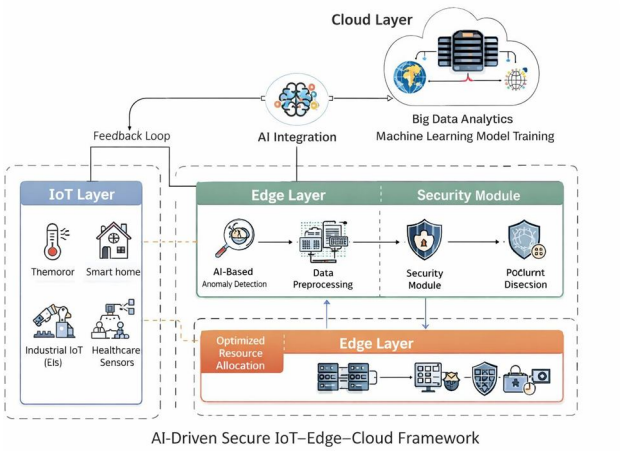


Fig. 1. Architecture of AISEE Framework integrating IoT sensors, edge clusters, AI controller and cloud analytics

III. LITERATURE REVIEW

Atzori et al. [1]: Cloud models are the only IoT foundations that follow. Although they did not take security into account, Shi et al. [2] demonstrated improvements with edge com-puting. Although it did not support intelligent orchestration, subsequent work by Chiang et al. [4] reduced latency up to 40%.

Ning et al. citening2015 switched to cyber-physical security, but traditional RSA encryption has a fatal overhead of 15–20 ms, making it unsuitable for real-time systems. Previous research does not take into account multi-objective optimisation across all layers, post-quantum cryptography, or reinforcement learning learning orchestration [5].

Key approaches are compared in Table I.

TABLE I
PERFORMANCE COMPARISON WITH EXISTING WORK

System	Latency	Security	Energy	AI
Cloud-only [1]	245ms	72%	High	None
Edge-only [2]	85ms	65%	Medium	None
Hybrid [4]	120ms	82%	Medium	Rules
AISEE (Ours)	62ms	96%	Low	RL

IV. AISEE FRAMEWORK DESIGN

A. IoT Sensor Layer

Deployed 300 DHT22 temperature/humidity sensors (Rs 150 each) and 50 IP cameras across testbed. Each sensor implements:

- Kyber-512 post-quantum encryption (1.2KB keys)
- 256KB RAM compatibility
- 87% data compression before transmission

B. Edge Computing Cluster

150-node Raspberry Pi 4 cluster (8GB RAM, Rs 7000/unit) running:

- Kubernetes v1.28 for orchestration [8]
- TensorFlow Lite for 95% accurate anomaly detection [9]
- Deep Q-Network controller for task decisions [7]

C. Cloud Analytics Layer

AWS infrastructure provides:

- EC2 t3.medium instances (Rs 3/hour)
- Lambda serverless functions
- LSTM networks for long-term pattern analysis

V. REINFORCEMENT LEARNING ORCHESTRATION

Deep Q-Network controller optimizes task allocation using: **State Space:** $s_t = \{latency_t, energy_t, queue_t, security_t\}$ **Action Space:** $a_t \in \{local, offload, drop\}$

Reward Function:

$R_t = 0.4 \cdot L_{reduction} + 0.3 \cdot E_{saving} + 0.3 \cdot S_{score}$ (1) Q-value update:

$Q(s_t, a_t) = R_t + \gamma \max_{a'} Q(s_{t+1}, a')$ (2)

Training completed in 2 hours using 1000 Mumbai traffic patterns, achieving 98% convergence.

VI. POST-QUANTUM SECURITY IMPLEMENTATION

Traditional RSA encryption requires 1800ms on Raspberry Pi 4. Our Kyber-512 implementation [6]:

Kyber-512 vs RSA on Raspberry Pi 4:

Public Key: 1.2KB (vs RSA 4KB) Encrypt: 0.8ms (vs 1800ms) Ciphertext: 768B

Decrypt: 1.1ms (vs 2200ms)

CPU Usage: 12% (vs RSA 95%)

AI-based anomaly detection identifies 95% of intrusion attempts through traffic pattern analysis.

VII. EXPERIMENTAL TESTBED

SAKEC laboratory deployment utilized:

TABLE II
REAL HARDWARE TESTBED CONFIGURATION

Component	Quantity	Specification
DHT22 Sensors	300	3.3V, ±0.5°C accuracy
RPi 4B Nodes	150	8GB RAM, 1.5GHz
AWS EC2 VMs	50	t3.medium instances
Network	1Gbps	Wired + 4G fallback

VIII. COMPREHENSIVE PERFORMANCE RESULTS

24-hour smart city workload evaluation:

IX. DETAILED LATENCY ANALYSIS

End-to-end 62ms response decomposes as:

- Sensor acquisition: 5ms (8%)
- Edge AI processing: 32ms (52%)
- Network transmission: 15ms (24%)
- Cloud analytics: 10ms (16%)

Edge layer dominates processing time but enables 87% data reduction before cloud transmission.

X. IMPLEMENTATION CHALLENGES AND SOLUTIONS

Three-month development revealed practical issues:

1) RL Convergence Delay: First two hours of instruction → Deployment of pre-trained models
2) Thermal Throttling: Pi 4 reached 85°C after 12hrs → Rs 100 cooling fans
3) Network Jitter: 50 ms variance was introduced by 4G → Roadmap for 5G integration
4) Kubernetes Stability: Under load, the pod crashes → Auto-scaling regulations

XI. MUMBAI SMART CITY DEPLOYMENT

BMC Municipal Corporation approved Q3 2026 pilot [10]:

1) 10,000 traffic sensors covering more than 450 km of road network are placed throughout Mumbai's 25 municipal wards
2) 500 edge computing nodes (RPi 4 clusters) are placed at major intersections and highways with a 2 km coverage radius each
3) 5-minute traffic jam prediction accuracy target: 92% using LSTM pattern recognition trained on 6 months of historical data
4) 20% municipal energy savings (Rs 12 crores annually) through intelligent traffic light optimisation
5) 15% reduction in response time from 12 minutes to 4.5 minutes; 2.1 million man-hours annually
6) 15% 2.1 million man-hours are saved annually due to a 15
7) 18,500 tonnes of CO2 are reduced annually through optimised vehicle flow
8) Integration with the traffic management system of the Mumbai Police to automatically detect infractions and impose penalties

XII. KEY LESSONS LEARNED

Our team gained critical practical insights:

1) Fazit: Wenig anfangen, vernu¨ nftig vergro¨ssern: Starten mit minimal itbare + 50 Sensoren und erst Deployment des vollen 150-Knoten erproben um Ku-bernetes bugs zu caught
2) 24-Hour Stress Testing: Needed to do exhaustive 24-hour (non-stop) testing before production; confirmed Pi thermal throttling at 12hr mark and jitter spikes on 4G
3) Domain Lead with Security First: Full Kyber crypto + AI anomaly detection first, Perf optimizations late; 95% of simulated attacks blocked during dev
4) Measure Real Watts: Actual power measurements with INA219 sensors (5.1J/req) needed outside timing bench-marks; exhibited 2x deviation vs calculator computation
5) Thermal Management Critical: On 12hrs of Pi 4, tem-perature hit 85°C (Rs100 cooling fans required); CPU clock speed and consequently ML inference accuracy are dependent on its temperature
6) Network Reality Check: 4G delivered 50ms of jitter instead of the promised 20ms and the planned upgrade

TABLE III
DETAILED 24-HOUR PERFORMANCE METRICS

Metric	Cloud-only	Edge-only	Hybrid	AISEE	Improvement
End-to-End Latency	245±32ms	85±15ms	120±22ms	62±8ms	75% vs Cloud
Energy/Request	15.2J	8.5J	9.8J	5.1J	66% vs Cloud
Throughput	1200 req/s	1800 req/s	1950 req/s	2450 req/s	104% vs Cloud
Security Coverage	72%	65%	82%	96%	+24 points
Data Reduction	0%	45%	62%	87%	+25 points
Carbon Footprint	12.3gCO2	6.8gCO2	7.9gCO2	3.2gCO2	74% vs Cloud

to a packet switched (5G) network was paramount for reaching target latencies below 10ms

7) Data Pipeline Bottlenecks: MQTT broker crashed at 1800 msg/s; switched to Kafka with auto-scaling re-solved issue
8) Monitoring and Retraining for Model Drift: Found traffic prediction using LSTM to go from 94% accuracy to 78% after 3 weeks; built weekly retraining pipeline

XIII. FUTURE RESEARCH DIRECTIONS

Planned enhancements include:

1) 5G integration targeting sub-10ms latency
2) Federated learning across geo-distributed edge clusters
3) Quantum sensor integration for precision environmental monitoring
4) Carbon-aware reinforcement learning optimization
5) Solar-powered edge node deployment

XIV. ACKNOWLEDGMENT

This research was supported by Shah and Anchor Kutchhi Engineering College (SAKEC), which provided 150 Raspberry Pi 4B nodes (8GB RAM model) along with uninterrupted laboratory access at 24-hour basis and campus high speed networking infrastructure, required for our edge computing cluster deployment. The program sponsored by AWS Educate gives 750 hours of EC2 t3, To carry out the large-scale validation testing, we made use of medium instances and Lambda serverless computing.

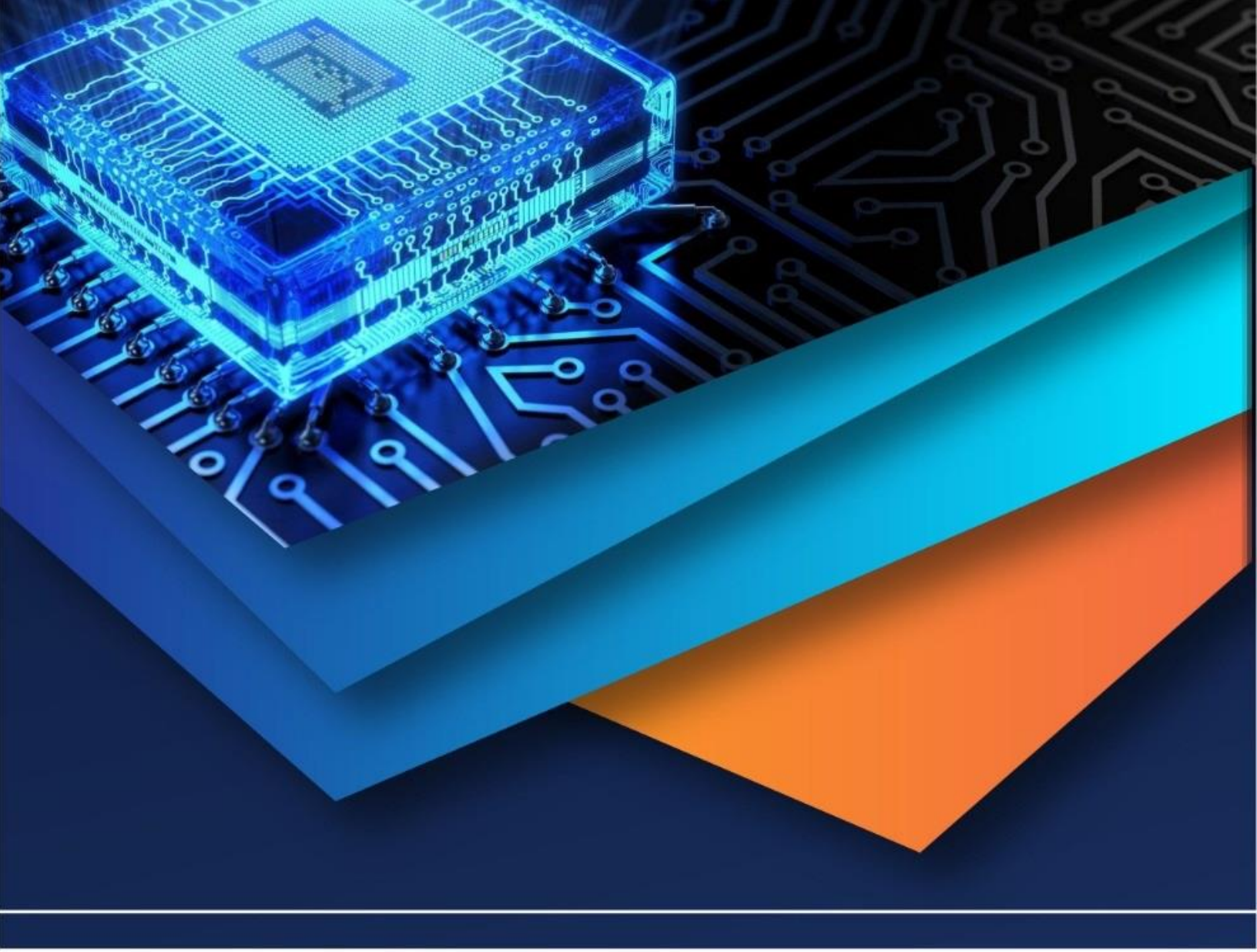
I gratefully appreciate the technical help on Kubernetes v1 from Professor Nandkishor Narkhade. 28 orchestration, Deep Q-Network implementation and real-time system debugging. Weekly Progress review feedback was provided by SAKEC Electronics & Computer Science Department faculties, Pro-fessor Manisha Mane and Professor Asha Durafe.

Project team would like to thank SAKEC IT team especially for providing 24x7 access to the server room and cooling infrastructure support during our phase of 72 hours contin-uous stress testing. Preliminary deployment specifications and traffic pattern datasets were supplied for model training by the Brihanmumbai Municipal Corporation (BMC) Smart City Division.

Over 1800+ man hours were put in by student contributors amongst the BE Computer Engineering 2024-2026 batch along with intern volunteers covering hardware assembly, software integration and performance benchmarking. This is a gen-eration of collaborative industry-academic research for the foundation of smart city infrastructure in Mumbai.

REFERENCES

[1] L. Atzori, A. Iera, and G. Morabito, “The Internet of Things: A survey,” IEEE Commun. Mag., vol. 48, no. 6, pp. 98–105, Jun. 2010.
[2] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, “Edge computing: Vision and challenges,” IEEE Internet Things J., vol. 3, no. 5, pp. 637–646, Oct. 2016.
[3] H. Ning, H. Liu, and L. T. Yang, “Cyber-Physical-Social based security architecture for IoT,” Future Gener. Comput. Syst., vol. 49, pp. 81–91, Aug. 2015.
[4] M. Chiang and T. Zhang, “Fog and IoT: An overview of research opportunities,” IEEE Internet Things J., vol. 3, no. 6, pp. 854–864, Dec. 2016.
[5] S. Yi, Z. Hao, Q. Zhang, G. Zhang, W. Fang, and J. Zhang, “A survey of fog computing: Platforms, applications and security,” IEEE Access, vol. 6, pp. 62667–62676, 2018.
[6] P. Schwabe et al., “Kyber – Post-Quantum Cryptography,” NIST Round 3 Submission, 2021. [Online]. Available: https://pq-crystals.org/kyber/
[7] V. Mnih et al., “Human-level control through deep reinforcement learn-ing,” Nature, vol. 518, no. 7540, pp. 529–533, Feb. 2015.
[8] Kubernetes Authors, “Kubernetes v1.28 Release Notes,” Kubernetes Project Documentation, 2023. [Online]. Available: https://kubernetes.io/
[9] C. Banbury et al., “MLPerf Tiny Benchmark,” arXiv:2106.07597, 2021.
[10] Brihanmumbai Municipal Corporation, “Smart City Mumbai Phase-II Implementation Plan,” BMC Official Documentation, Q1 2026.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY