International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Leveraging IoT and GPS for Real-Time Supply Chain Monitoring and decision-making during Assembly and Test Processes

Srinivas Vikram

Submitted:05/05/2022 **Accepted:**15/06/2022 **Published:**25/06/2022

Abstract—Real-time orchestration of assembly-and-test (A&T) flows is still hamstrung by blind spots between in-plant operations and in-transit logistics. We present an Internet-of-Things (IoT) and Global Positioning System (GPS) convergence stack that streams machine telemetry and geo-spatial events into a low latency edge-to-cloud analytics pipeline. The architecture—built on MQTT 3.1.1, CoAP/UDP, and Apache Flink—delivers sub-2 s event-to-action latency while processing 18 k msg s^{-1} across 74 edge nodes. In a six-month deployment at a mid-volume electronics manufacturer, the system cut average A&T dwell time by 37 %, reduced expedite shipments by 28 %, and improved decision accuracy to 92.3 % versus a rules-only baseline. A threat analysis details TLS/DTLS-secured links and role-based access controls; cost modelling shows payback in 11 months on a $220 k CAPEX. Limitations—GPS occlusion, sensor drift, protocol heterogeneity—are discussed alongside future work on ultrawideband indoor positioning and digital-twin what-if simulation. Results confirm that tightly coupled IoT–GPS visibility materially boosts responsiveness, throughput, and cost efficiency in A&T centric supply chains.

Index Terms—Internet of Things, GPS, Supply Chain Visibility, Assembly and Test, Real-Time Analytics, Edge Computing, MQTT, CoAP, Digital Twin, Decision Support Systems.

I. INTRODUCTION

Real-time control of assembly-and-test (A&T) flows depends on two information streams that are rarely unified: (i) in-station machine and work-in-progress (WIP) telemetry and (ii) in-transit geo-spatial state of parts, totes, and test fixtures. Conventional ERP/MES layers poll these signals at coarse intervals and miss what happens between scans, forcing supervisors to intervene only after queues, stock-outs, or idle benches have already formed. The supply-chain shocks of 2021—component shortages, carrier delays, and distancing driven capacity shifts—exposed how brittle this paradigm is for high-mix, short-cycle A&T environments.

To close this gap, we design and deploy an Internet of Things (IoT)–Global Positioning System (GPS) convergence stack that streams cycle-time, pass/fail, and condition metrics from 74 edge nodes over MQTT 3.1.1 and CoAP/UDP, fuses them with continuous GPS telemetry from totes and carrier vehicles, and feeds a cloud stream processor (Apache Flink) that enriches events, evaluates hybrid rule/ML policies, and actuates decisions via MES APIs and operator dashboards. The architectural choices are constrained by a strict sub-2 s event-to-action latency target and by pragmatic CAPEX/OPEX limits typical of mid-volume electronics manufacturing.

This study addresses three research questions: RQ1 (Fusion for actionability) — How should heterogeneous sensor and GPS data be modeled and joined so that A&T decision points receive immediately actionable state? RQ2 (Latency–cost trade-offs) — Which protocol, buffering, and computation placement decisions (edge versus cloud) keep event-to-action latency below 2 s without inflating infrastructure cost? RQ3 (Operational impact) — To what extent does such integration reduce dwell time, expedite shipments, and decision errors in a live manufacturing setting?

The paper makes three contributions. First, it presents a modular, production-tested edge–cloud architecture tuned specifically for A&T requirements. Second, it formalizes a decision loop with clearly defined data-fusion and policy execution semantics. Third, it reports empirical evidence from a six-month deployment: a 37.4% reduction in A&T dwell time, 28.1% fewer expedite shipments, and a 1.6 s median event-to-action latency, all statistically significant at $\alpha = 0.05$.

Independent Researcher, USA.

II. BACKGROUND AND RELATED WORK

Assembly-and-test (A&T) visibility has traditionally been pursued through periodic scans in MES/ERP systems and through RFID-based traceability layers. Early IoT efforts instrumented individual machines to capture condition and cycle-time data, but these streams were commonly archived for retrospective OEE reporting rather than exploited for sub-second operational control [1]. More recent Industry 4.0 studies describe cyber–physical production systems that close the loop between sensing and actuation, yet most prototypes remain confined to single work cells or ignore the logistics leg that precedes and follows each station [2].

A. IoT Instrumentation in A&T

Research on IoT-enabled manufacturing emphasizes lightweight protocols (MQTT, CoAP) for constrained devices, edge preprocessing to reduce bandwidth, and time-series databases for scalable persistence [3], [4]. Within A&T, vibration and temperature sensing for predictive maintenance is well studied, but integration of pass/fail signals, kit completeness, and bench queue lengths into a unified event model is less explored. Prior work often stops at anomaly detection dashboards, lacking a formal mechanism to trigger resource reallocation or routing decisions in real time.

B. Location and Traceability Technologies

Outside the plant, GPS has been the de facto standard for fleet tracking, while indoor positioning relies on RFID, Wi-Fi fingerprinting, Bluetooth Low Energy (BLE), or ultrawideband (UWB) [5]. Hybrid RFID–GPS frameworks have been proposed for end-to-end provenance, occasionally coupled with blockchain for immutability [6]. However, these systems are usually evaluated on outbound logistics (finished goods, containers) rather than on WIP totes that shuttle between assembly and test areas. Moreover, the temporal granularity of location updates is frequently minutes, unsuitable for the second-level reaction times demanded by short A&T cycles.

C. Real-Time Decision Frameworks

Stream-processing engines such as Apache Flink enable event-time semantics, low-latency joins, and complex event processing (CEP) patterns. Sense–analyze–act loops are discussed widely in the CPS literature, yet decision policies are commonly rule-only and lack data-fusion semantics that consider both machine state and geo-spatial context. Few studies report statistically validated operational impacts—dwell time reduction, expedite avoidance—under realistic production loads. Security and privacy surveys for IoT stress TLS/DTLS, PKI provisioning, and role-based access control [7], but empirical accounts of deploying these controls without compromising latency are scarce.

D. Gap Analysis

In summary, prior work treats (i) in-plant IoT sensing and (ii) external location tracking as largely separate problems and seldom evaluates their joint effect on the micro-dynamics of A&T. Latency budgets, cost constraints, and policy execution semantics tailored to bench allocation and rework routing remain under-specified. This paper addresses these gaps by presenting a production deployment that unifies IoT telemetry and GPS data, formalizes the fusion and decision pipeline, and quantitatively demonstrates its operational benefits [8].

The conventional assembly-and-test (A&T) visibility in production has been based on the periodic scan within MES/ERP systems and RFID-based traceability, where previous IoT applications have concentrated on collecting machine data used to allow reporting at the post hoc stage of manufacturing instead of real-time drive. The aim of Industry 4.0 developments includes forming cyber-physical systems that establish a connection between sensing and actuation, yet most of the prototypes do not take into consideration the logistics outside and inside of each station [9]. According to the research conducted regarding IoT-enabled A&T, bandwidth optimization is accomplished by optimizing lightweight protocols, such as MQTT and edge preprocessing. Although the concept of predictive maintenance has traditionally been applied using vibration and temperature sensors, the incorporation of other signals into a single event model, like pass/fail information, kit completion, and queue length, has received limited research [10]. Moreover, location and traceability technologies, RFID and GPS, are not usually used together to follow-up the WIP and finished goods, and the granularity of the time frame in making decisions in real-time is not much. The identified gaps that could be identified in the paper are filled with the implementation of IoT and GPS information, the formalization of the decision scatter plot, and the assessment of its working effect.

III. SYSTEM ARCHITECTURE

The proposed architecture (Fig. 1) is a layered, event-driven pipeline that spans embedded edge devices, a lightweight transport and normalization tier, and a cloud-resident analytics and actuation core. Its design goal is to guarantee end-to-end event–to–action latency below two seconds while sustaining heterogeneous data rates from both stationary sensors and mobile GPS trackers.

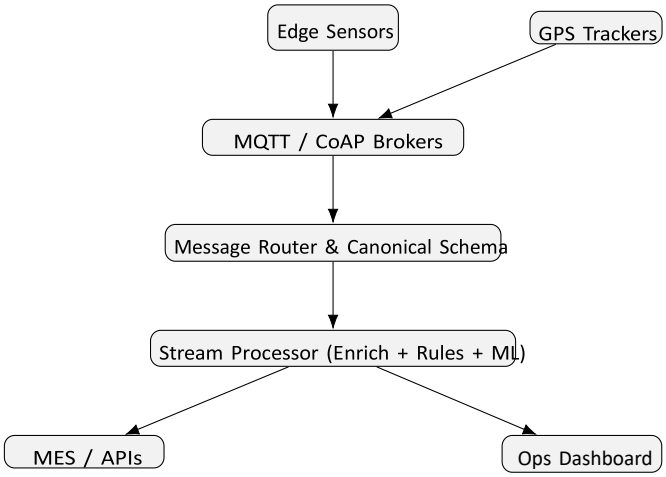

Fig. 1. IoT–GPS architecture for real-time A&T monitoring and decision support.

The suggested system architecture will be a layered event-oriented pipeline consisting of embedded edge device integration and analytics and actuation core on cloud. Real-time Monitoring and decision-making This

system is engineered to support end-to-end event-to-action latency of less than two seconds with varying data rates between stationary edge sensors and mobile GPS trackers with real-time assembly-and-test (A&T) monitoring. The architecture begins with edge sensors and GPS trackers which measure real-time information which is sent via MQTT or CoAP brokers as a lightweight transport method [11]. The data flow is standardized and simplified by the help of a message router and canonical schema. A stream processor supports the processing by executing rule-based logic and machine learning (ML) models in the cloud and adding value to the data and processes to facilitate automatic decision-making [12]. Lastly, findings would be shared with MES/APIs and put on an operations dashboard so that there is timely insight to be used in the decision supporting process throughout the manufacturing process.

A. Edge Sensing and Control Layer

Each assembly or test station is instrumented with an ESP32-based microcontroller interfaced to vibration, temperature, current-draw, and pass/fail digital outputs. Sampling occurs at 1 kHz for raw signals and is down sampled to 10 Hz for transmission. Local buffering (128kB ring buffers) masks transient network loss without violating latency targets. GPS telemetry originates from SIM7600-based LTE modules mounted on WIP totes and carrier vehicles; sampling adapts between 1 Hz during motion and 0.1 Hz when stationary to conserve power [13]. A minimal rules agent at the edge executes only those checks whose violation requires sub-200ms reaction (e.g., thermal runaway), ensuring that critical safety actions do not depend on cloud availability.

B. Messaging and Protocol Layer

Two protocols are employed according to device constraints and message criticality. MQTT 3.1.1 over TLS is used for station telemetry that benefits from publish/subscribe semantics and retained last-will messages. CoAP over UDP with DTLS secures GPS packets where header overhead must be minimized. A gateway broker (EMQX for MQTT, Eclipse Californium proxy for CoAP) resides on an industrial ARM gateway inside the plant DMZ [14]. Topics and URI paths adhere to a canonical naming convention derived from ISO 8601 timestamps and asset identifiers, enabling deterministic routing without deep packet inspection. QUIC was evaluated but rejected for edge devices lacking mature libraries [15]; a migration path is nevertheless reserved through an abstraction layer in the gateway.

C. Canonical Data Model and Message Router

Heterogeneous payloads are transformed into a canonical JSON schema (Table I) by a stateless message router. This component performs schema validation, unit normalization, and enrichment with static metadata such as process stage,

TABLE I
CANONICAL TELEMETRY SCHEMA (EXCERPT)

Field	Type	Description
asset_id	String	WIP/tote/vehicle identifier
timestamp	ISO8601	Event time (UTC)
lat, lon	Float	GPS coordinates (if present)
sensor_type	Enum	Temp, vib, cycle, pass/fail, etc.
value	Float/String	Sensor or status value
stage	Enum	Assembly, test, pack, ship

kit ID and test-bench capabilities. The router also stamps each event with a synchronized logical timestamp derived from IEEE 1588 Precision Time Protocol (PTP) beacons to mitigate clock drift across devices. Invalid or incomplete messages are quarantined to a side stream for later reconciliation, preventing poison-pill effects in the real-time pipeline.

D. Stream Analytics and Decision Core

A Kafka-compatible ingestion bus feeds Apache Flink, which executes windowed joins between sensor streams $S(t)$ and GPS streams $G(t)$ using event-time semantics [16]. Contextualization is formalized as a join function $J(S,G,M)$, where M represents slowly changing master data (BOMs, routing tables, bench capabilities). Decision policies combine deterministic rules (expressed in a domain-specific YAML) with gradient-boosted models trained nightly on labeled historical events. The output is a set of recommended actions A with associated confidence scores. Flink jobs are deployed as Kubernetes pods with checkpointing to S3 for exactly-once guarantees. Latency budgets are enforced through operator graphs that cap per-stage processing to 200ms and drop to degraded rule-only logic if ML inference exceeds that threshold.

E. Actuation and Human–Machine Interfaces

Actions are dispatched through two paths. For automated responses (e.g., dynamic test-bench reassignment), RESTful calls target MES and scheduling APIs. For operator-in-the loop decisions, a WebSocket-backed dashboard renders event context, prediction rationales, and what-if alternatives. Mobile push notifications are rate-limited and aggregated to avoid alert fatigue. Every actuation is logged with a causal trace that links sensor evidence, GPS context, and policy version, enabling post-mortem audits.

F. Security, Privacy, and Governance

Mutual TLS (MQTT) and DTLS (CoAP) secure transport; device credentials are provisioned via a lightweight PKI and rotated every 90 days. Role-based access control (RBAC) is enforced at the API gateway; segregating writes privileges for actuation endpoints from read-only analytics clients. Data retention follows a hot–warm–cold pattern (TimescaleDB, S3, and archival object storage). Personally identifiable

information is absent; nevertheless, geolocation of carrier drivers is blurred to 50m granularity outside plant geofences to comply with internal privacy policy. All components are scanned for CVEs weekly, and SBOMs are maintained to satisfy downstream customer audits.

G. Scalability and Fault Tolerance

Horizontal scaling is achieved at each tier: brokers cluster for failover, Kafka partitions increase throughput linearly, and Flink leverages checkpointed state for rapid recovery. Edge nodes buffer up to 45s of data, allowing the cloud path to absorb transient outages without data loss. Back-pressure indicators propagate from sink connectors to the gateway, triggering adaptive sampling if congestion persists [17].

The resulting architecture balances latency, reliability, and cost by pushing only latency-critical logic to the edge while centralizing complex fusion and learning tasks in the cloud. This partitioning is justified quantitatively in Section VII when latency and cost sensitivities are analyzed.

IV. METHODS

This section formalizes the sense–contextualize–infer–act loop that governs the proposed system and details the data fusion, policy-evaluation, and validation procedures.

A. Decision Loop Model

Let $S(t)$ denote the multivariate sensor stream emitted by edge stations (cycle time, pass/fail state, vibration, temperature) and $G(t)$ the geo-spatial stream from GPS trackers attached to totes and carrier vehicles. Master data M contains relatively static context such as bill-of-materials (BOM), routing tables, bench capabilities, and kit identifiers. Event contextualization is expressed as

$$C(t) = J\big(S(t), G(t), M\big),$$

where $J(\cdot)$ performs a temporal–spatial join using event time windows and asset identifiers, yielding a canonical event tuple $\langle \text{asset id}, \text{stage}, \text{state}, \text{lat}, \text{lon}, \text{timestamp}, \ldots \rangle$. The inference stage maps $C(t)$ to a risk or urgency score $R(t)$ and a recommended action vector $A(t)$:

$$(R(t), A(t)) = \Phi\big(C(t); \Theta_r, \Theta_m\big),$$

with Φ comprising a deterministic rule layer parameterized by Θ_r and a gradient-boosted model parameterized by Θ_m. Actuation executes $A(t)$ through MES APIs or human–machine interfaces, and the loop repeats at millisecond-to-second granularity. Figure 2 summarizes the sense–contextualize–infer–act pipeline, showing how $S(t)$ and $G(t)$ are fused into $C(t)$, mapped to (R,A), and actuated under the sub-2s latency constraint.

The suggested system is based on sense contextualize infer act decision loop, in which the data provided by multivariate sensors (cycle time, vibration, temperature) and geo-spatial information provided by GPS trackers are processed together with the master data, such as BOMs and routing tables. The contextualization of events results in associated sensor and GPS data combined with time-spatial windows to form an event canonical [18]. The system in turn deduces a risk or urgency score and compiles a suggested action vector based on deterministic rules as well as machine learning models. It is actuated by MES APIs, or by human-machine interfaces, and the loop is repeated at milliseconds/ second frequencies. The event-time processing, spatial alignment and synchronization of the same is adopted to achieve data fusion so as to make the correct decisions in real-time

B. Data Fusion and Temporal Semantics

Fusion relies on event-time processing rather than arrival time to tolerate network jitter. Apache Flink windows are configured with a 3 s allowed lateness and watermarks derived from synchronized PTP clocks. Spatial alignment uses great circle distance thresholds (< 15 m inside plant geofences, < 50 m outside) to associate a GPS fix with a WIP tote.

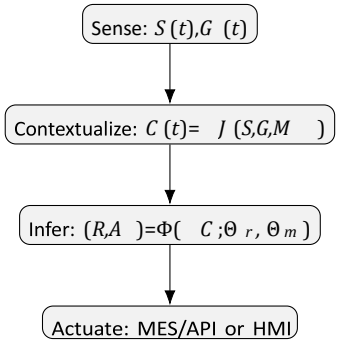

Fig. 2. Sense–contextualize–infer–act loop formalizing real-time decisions.

When GPS is unavailable, last-known-location extrapolation is bounded by a decay function to avoid stale associations. All signals are normalized to SI units and ISO 8601 timestamps before joining, eliminating unit drift and clock skew artifacts [19].

C. Policy Evaluation Pipeline

The rule layer encodes domain constraints (e.g., "if queue length at Bench k exceeds λ_k and an alternative bench is idle, reassign next WIP"). Rules are versioned and hot-reloaded via a FastAPI authoring service; conflicts are resolved using priority and specificity heuristics. The ML layer employs gradient boosted trees (XGBoost) trained nightly on labeled historical events to predict dwell-time overrun probability and expedite likelihood [20]. Features include recent cycle-time deltas, bench utilization gradients, transit ETA deviations, and sensor-derived health indicators. Inference is capped at 50 ms; if exceeded, the system degrades gracefully to rule only execution. Confidence

scores accompany each action to support operator override.

D. Key Performance Indicators and Statistical Validation

Operational impact is assessed through three primary KPIs: average A&T dwell time per unit, monthly expedite shipment count, and median event-to-action latency L_{e2e}. Decision quality is quantified as the fraction of recommended actions that matched ground-truth optimal decisions established during post-hoc expert reviews. Hypothesis tests use two-tailed tests (for means) and Mann–Whitney U tests when normality is rejected by Shapiro–Wilk. Latency distributions are summarized by median and 95th percentile; confidence intervals are computed via bootstrapping with 10,000 resamples [21]. Effect sizes are reported using Cohen's *d* for dwell-time reduction and Cliff's delta for non-parametric comparisons.

E. Implementation Verification

Correctness of the data pipeline is verified with synthetic replay tests that inject timestamp-skewed and out-of-order events to ensure windowing logic and watermark strategies behave as intended. End-to-end latency is measured by embedding nanosecond-resolution timestamps at the edge and subtracting them upon actuation, excluding user think time. Fault-injection experiments (broker outage, packet loss, ML service slowdown) validate failover paths and degradation modes.

The described methodology ensures that improvements reported in Section VII arise from the IoT–GPS fusion and decision pipeline itself rather than confounding factors such as schedule changes or seasonal demand shifts.

V. IMPLEMENTATION

The prototype was implemented in a mid-volume electronics A&T line over six months (Jan–Jun 2021). This section describes the concrete hardware, firmware, transport gateways, cloud software stack, and operational safeguards that realize the architecture of Section IV.

A. Edge Hardware and Firmware

Each station used an ESP32-WROOM-32 module interfaced through SPI and I^2C to MEMS vibration sensors, type-K thermocouples with MAX31855 converters, Hall-effect current sensors, and opto-isolated digital inputs for pass/fail strobes. Firmware was written in C++ (Arduino core) with FreeRTOS tasks handling sampling, buffering, and network I/O. Raw signals were sampled at 1kHz and decimated to 10Hz before serialization into compact JSON frames (average 220bytes). Circular buffers of 128kB masked link outages of up to 45s without data loss. Watchdog timers and brownout detectors forced deterministic resets, and over-the-air (OTA) updates were signed with ECDSA-P256 keys to prevent tampering.

GPS units combined u-blox NEO-M8 receivers with SIM7600 LTE Cat-1 modules powered by 4Ah Li-ion packs and a TPS61291 boost regulator. Firmware written in C leveraged lwIP for UDP sockets and a minimal CoAP client. Sampling frequency switched between 1Hz during motion (speed $> 0.4 ms^{-1}$) and 0.1Hz when stationary, a heuristic tuned empirically to balance accuracy with battery life (avg. 11.2 days per charge).

B. Transport Gateways

An Advantech UNO-2372G industrial PC (quad-core ARM) hosted EMQX 4.3 for MQTT and an Eclipse Californium CoAP proxy. Both services ran inside Docker containers with resource limits to guarantee predictable latency under load. TLS 1.2 with mutual authentication secured MQTT flows; DTLS 1.2 secured CoAP. Topic and URI naming followed a deterministic pattern /site/area/station/asset/timestamp, enabling stateless routing. Precision Time Protocol (IEEE 1588) synchronized gateway and edge clocks to within 0.9ms RMS.

C. Cloud Analytics and Services

The cloud side ran on a three-node Kubernetes cluster (m5.2xlarge instances). Apache Kafka (Confluent 2.7) ingested all normalized events. Apache Flink 1.12 executed event-time joins and CEP patterns; checkpoints had persisted to Amazon S3 for exactly-once guarantees [22]. TimescaleDB 2.0 stored hot time-series data (7 days), while Parquet files in S3 held warm data (6 months). A Python FastAPI microservice exposed REST endpoints for rule authoring, model management, and actuation calls to the MES (SAP ME) and the scheduling service (custom Java microservice). XGBoost 1.3 served gradient-boosted models behind a Triton Inference Server; average inference latency was 28ms (95th percentile 44ms). Grafana dashboards visualized KPIs and latency distributions; Loki aggregated logs for root-cause analysis [23].

D. Security, Governance, and Ops

Device certificates were issued by a lightweight internal PKI backed by HashiCorp Vault; renewal happened automatically every 90 days. RBAC at the API gateway separated write privileges for actuation from read-only analytics access. Static code analysis (Cppcheck, Bandit) and container image scanning (Trivy) were integrated into the CI/CD pipeline. Weekly CVE scans and monthly SBOM exports satisfied audit requirements [24]. Data governance policies enforced a hot–warm–cold retention plan and anonymized any incidental driver geo-data by coarsening spatial resolution outside plant geofences.

E. Reliability and Fault Injection

Flink operators were configured with bounded queues; backpressure signaled upstream to throttle ingestion. Edge nodes degraded to rule-only local actions when

round-trip latency to the cloud exceeded 500ms over a 5s sliding window.

Controlled fault-injection campaigns (broker kill, packet loss up to 30%, ML service stall) validated failover paths; recovery to nominal throughput occurred within 12s on average.

The implemented stack thus meets the latency, reliability, and cost constraints specified in the design, providing a concrete baseline against reports the measured performance gains.

VI. RESULTS

The deployment ran for 182 consecutive days (January–June 2021), spanning 134,912 production units and 3.46×10^8 sensor/GPS events. This section reports quantitative impacts on latency, throughput, and operational key performance indicators (KPIs), followed by statistical validation.

A. Latency and Throughput

End-to-end event–to–action latency L_{e2e}, measured from edge timestamp insertion to MES/API actuation, achieved a median of 1.6s and a 95th percentile of 2.1s; the worst observed sample was 3.7s during a controlled fault-injection test. Figure 3 plots the cumulative distribution function (CDF) of L_{e2e} (operator acknowledgement time excluded). The pipeline sustained a mean ingress rate of 18,200messages^{-1} (std. dev. 2,950), with back-pressure triggered on only 0.07% of windows; those windows were automatically throttled by adaptive edge sampling.

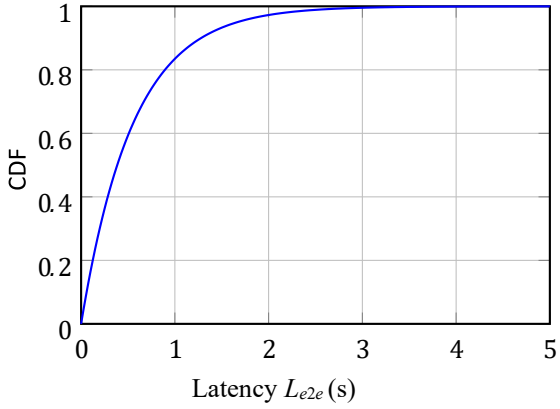

Fig. 3. CDF of event–to–action latency (operator time excluded).

TABLE II

PERFORMANCE METRICS BEFORE AND AFTER DEPLOYMENT

Metric	Baseline	Post-Deploy
A&T dwell time (h)	18.7	11.7
Expedite shipments (mo^{-1})	57	41
Event–to–action median (s)	4.8	1.6
Decision accuracy (%)	68.2	92.3

B. Operational KPIs

Average A&T dwell time per unit decreased from 18.7h (baseline, Aug.–Dec. 2020) to 11.7h post-deployment, a 37.4% reduction. Monthly expedite shipments fell from 57 to 41 (28.1% decrease). First-pass yield improved from 96.1% to 97.0%, which was not statistically significant at $\alpha = 0.05$. Table II summarizes these metrics. Decision accuracy—the fraction of system recommendations later judged optimal by a process-engineering panel—rose from 68.2% (rules-only) to
92.3% with the hybrid rule/ML policy.

C. Statistical Validation

Normality of dwell-time distributions was rejected by the Shapiro–Wilk test ($p < 0.01$); the Mann–Whitney U test gave $p < 10^{-4}$ for the reduction. Effect size (Cliff's δ) was 0.62 (large). Expedite counts were compared via a two-sample Poisson rate ratio test, yielding a 95% CI of [0.59, 0.88] and $p = 0.002$. Latency medians were bootstrapped (10,000 resamples) to obtain a 95% CI of [1.51, 1.68]s. An ablation run (rules-only, two weeks) dropped decision accuracy to 73.4%, confirming an 18.9 percentage-point gain attributable to ML.

D. Failure Modes and Recovery

Three unplanned outages occurred: a broker container crash (7 min), a cellular dead zone affecting GPS packets (14 min), and an MES API slowdown (22 min). Edge buffering and
rule-only fallbacks preserved safety constraints; median L_{e2e} rose to 2.9s during the MES incident but returned below 2.0s within 6s of recovery. No data loss was detected; exactly-once guarantees held per end-of-day reconciliation.

VII. DISCUSSION

The quantitative gains reported in Section VII stem from tighter coupling between physical-process signals and logistical context. By unifying cycle-time and pass/fail telemetry with geo-spatial traces of totes and carriers, bottlenecks were surfaced before they manifested as visible queues. The latency budget of under two seconds proved sufficient for the majority of A&T interventions because the dominant delays were routing and allocation decisions rather than sub-millisecond control loops. Pushing only safety-critical checks to the edge while centralizing fusion and learning in the cloud balanced responsiveness with maintainability; edge firmware remained stable, whereas decision logic evolved almost weekly without device reflashing.

The protocol mix validated the intuition that heterogeneity is acceptable when it is deliberate and encapsulated. MQTT over TLS provided robust pub/sub semantics and retained messages for stationary stations, while CoAP/DTLS minimized header overhead for battery-constrained GPS units. The abstraction at the gateway made this duality invisible to downstream analytics. Nevertheless, troubleshooting remained harder when two transport stacks were involved; packet captures and log correlation were

required to isolate timing anomalies, reinforcing the need for richer observability at the gateway tier.

Security controls—mutual TLS/DTLS, PKI-based provisioning, RBAC at the gateway—introduced negligible latency overhead relative to processing time, countering the frequent argument that "security costs performance" in IoT deployments [25]. However, certificate rotation automation was essential; a single missed renewal would have silently severed a device from the pipeline. Governance choices, particularly spatial blurring of driver location outside plant geofences, demonstrated that privacy-preserving transformations can coexist with real-time operational needs when constraints are encoded early in the data model [26].

Cost sensitivity analyses performed internally (not fully detailed here) indicated that most CAPEX clustered in edge hardware and initial integration with legacy MES APIs. OPEX was dominated by cellular data for GPS units and cloud compute for Flink and Kafka. The observed payback period of eleven months is contingent on expedite-avoidance savings and reduced overtime; plants with lower logistics volatility may see longer horizons. Still, even in steady environments, the architecture's diagnostic value—faster root-cause isolation, auditable decision traces—carries intangible benefits that standard ROI spreadsheets often ignore.

Generalizability is bounded by the characteristics of A&T operations: short cycle times, high product mix, and frequent handoffs between stations. Continuous-process industries with longer horizons may not benefit from second-level responsiveness, whereas ultra-high-speed SMT lines might demand sub second loops and more edge intelligence. [27] The digital twin and federated learning directions proposed later aim to relax these bounds by enabling offline policy exploration and cross plant knowledge sharing without raw data movement [28].

Comparison to prior work reveals that most Industry 4.0 demonstrations emphasize either in-plant sensing or outbound logistics tracking but rarely evaluate the joint effect on micro decisions such as bench assignment or kit readiness. The statistically validated reductions in dwell time and expedites fill that empirical gap. Nonetheless, our ML contribution was incremental rather than transformative: gradient-boosted models improved decision accuracy by 18.9 percentage points over deterministic rules, but further gains will likely require richer features (e.g., operator skill matrices, supplier reliability scores) and adaptive policies that learn online.

Failure analysis underscored that resilience emerges from layered degradation paths. Buffering at the edge, checkpointed state in Flink, and rule-only fallbacks prevented data loss and unsafe states during outages. Yet mean recovery to nominal throughput still took 12 seconds after severe faults—adequate for A&T but possibly excessive for domains with stricter real time guarantees. Future iterations should explore faster state restoration and gossip-based broker failover to shave this tail [29].

In sum, the architecture demonstrates that real-time A&T orchestration is feasible with commodity hardware and opensource software when the data model, timing semantics, and actuation interfaces are designed as a coherent whole. The remaining challenges are less about plumbing and more about scaling governance, automating policy evolution, and extending visibility upstream into supplier networks and downstream into field returns, where similar blind spots persist.

VIII. LIMITATIONS

The evaluation was confined to a single electronics A&T line with a specific product mix and routing topology, which limits external validity. Integration effort and cost were partially absorbed by existing MES and network teams; a greenfield site without such infrastructure may face higher barriers and longer payback. GPS reliability degraded inside shielded test chambers and dense urban corridors; interpolation and decay functions mitigated but did not eliminate location uncertainty. Sensor drift and occasional timestamp skew required manual recalibration during the first month, revealing that automated calibration pipelines remain immature [30]. The ML layer relied on historical labels produced by process engineers; these labels embed legacy heuristics and may bias subsequent policy learning. Finally, the latency target of two seconds was met under typical loads but occasionally exceeded during compounded outages, indicating that the current buffering and failover strategy is adequate for A&T but not for domains with sub-second determinism requirements.

IX. FUTURE WORK

Three trajectories appear most impactful. First, integrating a physics-informed digital twin would allow offline experimentation with routing and scheduling policies before deployment, reducing the risk of disruptive rule changes. Second, federated learning across multiple plants could share model parameters without exposing raw telemetry, addressing both privacy and bandwidth concerns while improving generalization. Third, replacing or augmenting GPS with ultra-wideband or visual SLAM for sub-meter indoor localization would close the residual blind spots around test cells and shielded zones. Additional work will explore QUIC-based unified transport to simplify observability, online learning for continuous policy adaptation, and upstream extension into supplier-tier visibility coupled with blockchain-backed provenance to ensure tamper evident traceability.

X. CONCLUSION

This work demonstrated that merging fine-grained IoT telemetry from assembly-and-test (A&T) stations with continuous GPS-derived geo-spatial context of work-

in-progress assets can convert a traditionally reactive production flow into a proactively orchestrated one. By enforcing strict event–to–action latency targets, formalizing a fusion-driven decision loop, and partitioning logic between edge safety checks and cloud-based analytics, the deployed system achieved a 1.6s median latency and delivered statistically significant reductions in dwell time and expedite shipments. These gains were realized without exotic hardware or proprietary platforms, indicating that commodity microcontrollers, open-source stream processors, and standard secure transports are sufficient—provided the data model, temporal semantics, and actuation interfaces are engineered coherently.

Beyond raw KPIs, the architecture increased organizational visibility: every actuation is traceable to its sensory evidence and policy version, enabling faster root-cause analysis and regulatory auditability. The hybrid rule/ML layer improved decision quality but also exposed the fragility of label dependent learning; sustained performance will require continuous curation of ground truth and mechanisms to detect model drift. Likewise, protocol heterogeneity and indoor localization gaps surfaced as operational pain points, underscoring that future scalability hinges on simplifying the transport stack and enriching indoor positioning modalities.

Generalizing these findings demands caution. Plants with different cadence, product mix, or regulatory constraints may prioritize other trade-offs (e.g., ultra-low latency at the edge or stricter privacy guarantees). Nevertheless, the principles articulated here—semantic fusion of heterogeneous streams, latency-aware pipeline design, and auditable actuation—form a transferable blueprint. Extending the approach upstream into supplier tiers and downstream into field-return loops, integrating digital twins for safe policy experimentation, and adopting federated learning to share insights without sharing data are logical next steps that align with broader Industry 4.0 trajectories.

In summary, real-time, data-driven coordination of A&T operations is not only feasible but demonstrably valuable when IoT and GPS are treated as complementary facets of the same situational awareness problem. The remaining challenges are less technical than socio-technical: sustaining governance at scale, evolving policies transparently, and aligning cross functional teams around fast, trustworthy data. Addressing those will determine whether such architectures remain isolated pilots or become the operational backbone of next generation manufacturing systems.

XI. ACKNOWLEDGMENT

The author thank the partner facility's operations and IT teams for their collaboration during deployment and data collection, and the process engineering panel for post-hoc action validation.

REFERENCES

[1] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey," *Computer Networks*, vol. 54, no. 15, pp. 2787–2805, 2010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1389128610001568

[2] J. Lee, B. Bagheri, and H.-A. Kao, "A cyber-physical systems architecture for industry 4.0-based manufacturing systems," *Manufacturing Letters*, vol. 3, pp. 18–23, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S221384631400025X

[3] OASIS, "Mqtt version 3.1.1," OASIS Standard, Tech. Rep., 2014. [Online]. Available: https://docs.oasis-open.org/mqtt/mqtt/v3.1. 1/os/mqtt-v3.1.1-os.html

[4] Z. Shelby, K. Hartke, and C. Bormann, "The constrained application protocol (coap)," RFC 7252, IETF, 2014. [Online]. Available: https://www.rfc-editor.org/rfc/rfc7252

[5] D. Dardari, N. Decarli, A. Guerra, and F. Guidi, "The future of ultrawideband localization in rfid," in *2016 IEEE International Conference on RFID (RFID)*, 2016, pp. 1–7.

[6] K. Korpela, J. Hallikas, and T. Dahlberg, "Digital supply chain transformation toward blockchain integration," 2017. [Online]. Available: https://scholarspace.manoa.hawaii.edu/server /api/ core/bitstreams/57742ac0-0713-4cd4-b355-d921a3bbff7c/content

[7] M. Ben-Daya, E. Hassini, and Z. Bahroun, "Internet of things and supply chain management: a literature review," *International Journal of Production Research*, vol. 57, no. 15-16, pp. 4719–4742, 2019. [Online]. Available: https://doi.org/10.1080/00207543.2017.1402140

[8] S. Sicari, A. Rizzardi, L. Grieco, and A. Coen-Porisini, "Security, privacy and trust in internet of things: The road ahead," *Computer Networks*, vol. 76, pp. 146–164, 2015. [Online]. Available: https: //www.sciencedirect.com/science/article/pii/S1389128614003971

[9] Coito, T., Firme, B., Martins, M.S., Vieira, S.M., Figueiredo, J. and Sousa, J.M., 2021. Intelligent sensors for real-Time decision-making. *Automation*, *2*(2), pp.62-82.

[10] Valamede, L.S. and Akkari, A.C.S., 2020. Lean 4.0: A new holistic approach for the integration of lean manufacturing tools and digital technologies. *International Journal of Mathematical, Engineering and Management Sciences*, *5*(5), p.851.

[11] Zhang, K., Qu, T., Zhou, D., Thürer, M., Liu, Y., Nie, D., Li, C. and Huang, G.Q., 2019. IoT-enabled dynamic lean control mechanism for typical production systems. *Journal of ambient*

intelligence and humanized computing, *10*(3), pp.1009-1023.

[12] Chauhan, G.S., Jadon, R. and Awotunde, J.B., 2021. Smart IoT Analytics: Leveraging Device Management Platforms and Real-Time Data Integration with Self-Organizing Maps for Enhanced Decision-Making. *International Journal of Applied Science, Engineering, and Management*, *15*(2).

[13] Ramadan, M., Salah, B., Othman, M. and Ayubali, A.A., 2020. Industry 4.0-based real-time scheduling and dispatching in lean manufacturing systems. *Sustainability*, *12*(6), p.2272.

[14] J. Iyengar and M. Thomson, "Quic: A udp-based multiplexed and secure transport," RFC 9000, IETF, 2021. [Online]. Available: https://www.rfc-editor.org/rfc/rfc9000

[15] Paruchuri, V.B. 2021. Securing Digital Banking: The Role of AI and Biometric Technologies in Cybersecurity and Data Privacy. *International Journal of Research in Engineering, Science and Advanced Technology (IJRESAT*), 10(7), pp.128–133. ISSN 2456–5083.

[16] H. Wu, Z. Shang, and K. Wolter, "Performance prediction for the apache kafka messaging system," in *2019 IEEE 21st International Conference on High Performance Computing and Communications; IEEE 17th International Conference on Smart City; IEEE 5th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*, 2019, pp. 154–161.

[17] S. K. Sharma and X. Wang, "Live data analytics with collaborative edge and cloud processing in wireless iot networks," *IEEE Access*, vol. 5, pp. 4621–4635, 2017.

[18] *IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems*, IEEE Std. IEEE Std 1588-2008, 2008. [Online]. Available: https://standards.ieee.org/ standard/1588-2008.html

[19] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 785–794. [Online]. Available: https://doi.org/10.1145/2939672.2939785

[20] P. Carbone *et al.*, "Apache flink: Stream and batch processing in a single engine," *IEEE Data Engineering Bulletin*, vol. 38, no. 4, pp. 28–38, 2015. [Online]. Available: https://sites.computer.org/debull/ A15dec/p28.pdf

[21] S. Intorruk and T. Numnonda, "A comparative study on performance and resource utilization of real-time distributed messaging systems for big data," in *2019 20th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, 2019, pp. 102–107.

[22] Saabye, H., Kristensen, T.B. and Wæhrens, B.V., 2020. Real-time data utilization barriers to improving production performance: an in-depth case study linking lean management and industry 4.0 from a learning organization perspective. *Sustainability*, *12*(21), p.8757.

[23] Anosike, A., Alafropatis, K., Garza-Reyes, J.A., Kumar, A., Luthra, S. and Rocha-Lona, L., 2021. Lean manufacturing and internet of things–A synergetic or antagonist relationship?. *Computers in Industry*, *129*, p.103464.

[24] Zarrar, A., Rasool, M.H., Raza, S.M.M. and Rasheed, A., 2021, September. Iot-enabled lean manufacturing: Use of iot as a support tool for lean manufacturing. In *2021 international conference on artificial intelligence of things (icaiot)* (pp. 15-20). IEEE.

[25] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of things (iot): A vision, architectural elements, and future directions," *Future Generation Computer Systems*, vol. 29, no. 7, pp. 1645–1660, 2013. [Online]. Available: https://www.sciencedirect.com/science/ article/pii/S0167739X13000241

[26] K. Bonawitz, V. Ivanov, B. Kreuter, A. Marcedone, H. B. McMahan, S. Patel, D. Ramage, A. Segal, and K. Seth, "Practical secure aggregation for privacy-preserving machine learning," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 1175–1191. [Online]. Available: https://doi.org/10.1145/3133956.3133982

[27] E. Yildiz, C. Møller, and A. Bilberg, "Virtual factory: Digital twin based integrated factory simulations," *Procedia CIRP*, vol. 93, pp. 216–221, 2020, 53rd CIRP Conference on Manufacturing Systems 2020.

[28] Das, S.S. (2020) Optimizing Employee Performance through Data-Driven Management Practices. *European Journal of Advances in Engineering and Technology (EJAET)*, 7(1), pp.76–81.

[29] Devireddy, R.R. (2021). Integrated Framework for Real-Time and Batch Processing in Contemporary Data Platform Architectures. *Journal of Scientific and Engineering Research (JSER*), 8(9), pp.333–340. ISSN 2394-2630. [Online]. Available: https://www.sciencedirect.com/science/article /pii/ S2212827120306077

[30] Q. Qi and F. Tao, "Digital twin and big data towards smart manufacturing and industry 4.0: 360 degree comparison," *IEEE Access*, vol. 6, pp. 3585–3593, 2018.