



Original Article

Ground-Based Fleet-Scale Stress Testing of In-Flight Entertainment Systems: A Framework for Pre-Deployment Reliability Validation under Realistic Passenger Load

Sathish Babu Thaiuri

Staff Software Engineer, Solution Architect, Austin, Texas, USA.

Received On: 15/06/2026

Revised On: 12/07/2026

Accepted On: 22/07/2026

Published On: 02/08/2026

Abstract - Contemporary in-flight entertainment architectures place a general-purpose application processor behind every passenger seat, so a wide-body cabin of four hundred seats carries a distributed compute estate comparable in density to a small data centre inside a pressurised fuselage. Thermal saturation and power transients that remain invisible during conventional pre-delivery checks become in-service disruptions once the aircraft is carrying passengers. Current acceptance practice powers on a sample of ten to twenty percent of seats, runs a functional checklist, and signs off, which is structurally incapable of surfacing failures that exist only as properties of the fully loaded cabin. This paper proposes a ground-based validation framework that treats the complete installed cabin as a single test subject and automates four test disciplines: seat-level thermal stress cycling under sustained decode load, synchronised high-complexity video decode across every seat, stochastic boarding-phase power surge simulation, and zone-concentrator fault injection with timed recovery assertion. The framework emits a structured, hash-chained validation record that supports design-assurance traceability and allows results from different airframes of the same type to be compared programmatically. Calibrated engineering models of the proposed framework, evaluated against published component specifications and cabin airflow data, indicate that aft-zone seats operate measurably closer to the processor throttling threshold than forward-zone seats, that worst-case boarding current ramp rates approach power distribution unit protection thresholds, and that a software-only activation jitter offers a meaningful reduction in peak inrush. These are model-derived projections, not measurements from a deployed system. The paper states explicitly which quantities are model outputs and which require a physical pilot deployment to confirm.

Keywords - In-Flight Entertainment, IFE Validation, Thermal Stress Testing, Ground-Based Cabin Validation, Avionics Acceptance Testing, Embedded Systems Load Simulation, Design Assurance Traceability, Fault Injection, Boarding-Phase Power Simulation, Seat Electronic Unit, Pre-Delivery Testing, Maintenance And Overhaul Automation, Fleet-Scale Embedded Compute.

1. Introduction

1.1. Compute Density Has Changed the Cabin

In-flight entertainment has moved from a centralised head-end distributing analogue video to a seat-centric architecture in which each passenger position hosts an independent embedded computer. A modern seat unit contains an application processor, local content storage, and a hardware media decode engine, and runs a full operating system. A wide-body cabin configured for four hundred passengers therefore contains four hundred such processors, aggregated by zone concentrators, fed by cabin power distribution units, and interconnected by an Ethernet backbone.

This is a substantial change in the thermal and electrical envelope of the cabin, and it happened without a corresponding change in how the installation is validated before the aircraft enters service. The seat unit is installed under or behind a passenger seat, in a volume with limited airflow, in a row of identical units generating identical heat, in a cabin whose underfloor ventilation was not designed around the thermal output of a dense compute array.

1.2. Why Current Validation Practice Does Not Reach the Failure

Pre-delivery entertainment system validation has followed broadly the same pattern for two decades. A technician powers on a sample of seats, typically ten to twenty percent of the cabin, runs a functional checklist confirming that video plays and the network responds, and signs the acceptance record. Format conversion and data transfer are performed manually. There is no concurrent load across the full cabin, no thermal monitoring under sustained stress, and no structured record of what was actually exercised beyond a checklist retained as a document.

The difficulty is that the failures that matter are not properties of a seat. They are properties of a cabin. A single seat unit running in isolation dissipates heat freely into the surrounding volume and will pass any functional test that is applied to it. Thermal exceedance appears only when every seat runs simultaneously, when heat accumulates across adjacent rows, and when aft-zone airflow proves insufficient to hold junction temperatures below the throttling threshold.

The same argument applies to the electrical case. The boarding-phase current inrush that risks tripping overcurrent protection is an aggregate event across a power zone. It cannot be observed by energising one seat, or one row, or a scattered ten percent sample.

Sampling is therefore not merely a weaker version of full-cabin testing. It is structurally blind to the specific class of failure that grounds aircraft and generates warranty disputes, because the phenomenon under investigation does not exist at the sampling granularity.

1.3. Contribution and scope

This paper proposes a framework for ground-based, fleet-scale stress testing of in-flight entertainment installations, and is explicit about being a proposal supported by calibrated models rather than a report of a completed deployment. Specifically it contributes:

- A test architecture that treats the complete installed cabin as a single interconnected compute and power fabric rather than as a population of independently testable seats.
- Four automated test disciplines executed under concurrent full-cabin load: thermal stress cycling, synchronised decode saturation, stochastic boarding-phase power surge simulation, and zone-concentrator fault injection with timed recovery assertion.
- A structured, version-controlled configuration and reporting model that produces a comparable validation artifact per airframe, enabling programmatic comparison across serial numbers of the same type.
- An explicit separation, carried throughout the paper and consolidated in Table 4, between quantities derived from calibrated engineering models and quantities that require physical execution on a delivered aircraft to establish.

What this paper does not do. It does not report a deployment of the proposed framework, measured outcomes from a physical ground test, or a comparative evaluation against an alternative automated approach. Every quantitative finding in Section 9 is an output of a model calibrated against published specifications, and is labelled as such at the point of use. Section 12 states the limitations directly, and Section 5 records the author's prior practical work in this area and its relationship to the present proposal.

Section 2 sets out the failure mechanism that motivates the work. Section 3 positions it against existing literature. Section 4 describes the cabin architecture under test. Section 5 states the author's prior work and the origin of the framework. Sections 6 and 7 develop the framework architecture and the test method. Section 8 addresses ground-to-flight fidelity. Section 9 presents model-derived findings. Section 10 compares the approach against current practice. Sections 11 to 13 cover operational considerations, limitations, and future work.

2. Failure Modes under Full-Cabin Load

The case for full-cabin testing rests on a specific failure mechanism, and it is worth setting out in full because it explains why the sampling approach cannot reach it.

2.1. The Thermal Cascade

When a full cabin begins watching video at once, the combined heat output of several hundred seat processors accumulates faster than the surrounding volume can carry it away. The effect is not uniform. In the aft section, where underfloor airflow is weakest, individual units cross their thermal limits first.

What follows is a cascade rather than an isolated fault. Seats that exceed their thermal ceiling freeze and reboot. Because the seats in a zone are thermally similar and were loaded simultaneously, they cross the threshold at roughly the same moment, so a group of seats attempts to rejoin the network together. The zone concentrator sees a burst of simultaneous reconnections, which is indistinguishable from the signature of a genuine network fault, and responds by restarting the zone. The result is an entire cabin section losing entertainment service in flight, triggered by a thermal condition in a subset of seats.

Three properties of this sequence matter for test design. It is aggregate, because the heat load that starts it does not exist unless most of the cabin is active. It is spatial, because it begins where airflow is poorest rather than at a random seat. And it is emergent, because the zone restart is a correct response by the concentrator to an input that only a coordinated thermal event can produce. No amount of individual seat testing generates that input.

2.2. What the Cascade Costs across a Fleet

The consequences compound in ways that are easy to misattribute. Seats that rebooted under thermal stress present at the gate as intermittent hardware faults. They are removed and replaced, frequently without a root cause being established, because the condition that produced the failure is not reproducible on a bench where a single unit dissipates heat freely. This generates a population of unscheduled replacements against units that were not defective.

Those replacements then become warranty disputes. An operator returning units that test as serviceable has no telemetry to demonstrate the conditions under which they failed, and the supplier has no obligation to accept a return it cannot verify. The absence of instrumented ground test evidence is what makes the dispute unresolvable rather than merely contested.

The third cost is the least visible. Passengers on routes where entertainment is unreliable notice, and the effect on route selection is real even though it does not appear in a maintenance record.

2.3. Why This Is a Ground-Detectable Problem

The important observation is that none of this is inherent to operating the aircraft. The thermal signature that starts the

cascade, the boarding-phase power transient that stresses the distribution system, and the recovery behaviour of the zone concentrator under a burst of reconnections are all observable on the ground, provided the cabin is exercised at realistic load rather than spot-checked a few seats at a time.

That is the claim this paper develops: not that ground testing can eliminate in-service failures, but that a specific and substantial class of them is detectable before the aircraft enters service and is currently not detected.

3. Related Work

Four bodies of work bear on this proposal, and the gap this paper addresses sits between them. Airborne software and hardware assurance provides the certification framework within which any cabin systems test evidence must sit. Design assurance guidance for airborne software [1] and for airborne electronic hardware [2] establishes the expectations for requirements traceability, verification coverage, and structured evidence that a test framework must satisfy if its outputs are to carry weight in an acceptance process. These documents govern the form of test evidence. They do not prescribe a method for exercising an installed cabin under load, which is the gap addressed here.

Cabin environmental and airflow modelling characterises the thermal environment in which seat electronics operate. Work on reconfigurable cabin airflow modelling [3] provides an open framework for characterising cabin air distribution, and confirms that airflow varies materially with position along the fuselage. This literature establishes the physical basis for expecting position-dependent thermal behaviour but is directed at passenger comfort and air quality rather than at electronics cooling under aggregate compute load.

Entertainment system engineering literature has concentrated on connectivity and distribution rather than on validation. Studies of wireless and ultra-wideband distribution within the cabin address bandwidth and interference for content delivery. Recent work on cabin systems has focused on architecture and passenger

experience. The authors are not aware of published work that treats the installed cabin as a single test subject for pre-delivery reliability validation.

Anomaly detection for telemetry streams supplies the analytical component. Isolation-based anomaly detection [4] is well established for identifying outliers in high-dimensional telemetry without labelled failure data, which matches the situation of a first-of-type cabin test where no failure corpus yet exists. Its application here is straightforward rather than novel, and the framework's contribution is the telemetry corpus it makes available rather than the detector applied to it.

The gap is therefore specific. Assurance guidance defines what evidence must look like. Airflow modelling explains why position matters. Anomaly detection can process the data. No published framework connects these into a method for exercising a complete installed cabin at realistic load before the aircraft enters service.

What distinguishes the approach proposed here is not any single element but the combination of four, none of which is individually novel and which have not previously been assembled: orchestrating concurrent load across every seat rather than sampling; capturing per-seat thermal telemetry at fixed resolution mapped spatially to the cabin layout; simulating the stochastic boarding power ramp rather than a static load estimate; and injecting concentrator faults with automated recovery assertion while the zone is under that load. The combination produces an assurance-traceable test artifact connecting raw telemetry to a structured verdict, which is the output an operator quality function can act on and the output current practice does not produce.

4. The System under Test

4.1. Elements under Test

The framework treats the installed cabin as a single interconnected compute and power fabric. Every component that carries data, power, or thermal load during a revenue flight is exercised simultaneously. Table 1 lists the elements under test, described by function.

Table 1. Elements of the Installed Cabin Exercised By the Framework

Element	Function	Stress applied
Seat unit	Passenger-facing display with embedded application processor, local content cache, and hardware decode engine	Sustained maximum processor utilisation via high-complexity video decode; primary thermal and compute target
Zone concentrator	Rack-mounted aggregation node serving forty to sixty seats; handles content distribution, telemetry collection, and zone routing	Passive load from full-zone traffic; active fault injection across three fault classes
Power distribution unit	Feeds each seat zone from the aircraft electrical bus	Boarding-phase activation sequence measured against overcurrent protection thresholds
Cabin network backbone	Connects seat units to zone concentrators and concentrators to the content source	Aggregate decode requests, content prefetch, and telemetry from every seat concurrently
Content source	Origin from which zone concentrators prefetch and cache media	Cache invalidation and re-synchronisation forced by concentrator fault injection
Seat	Instrumentation endpoint exposed by each concentrator	Primary telemetry and command path for the

management interface	over the maintenance network	framework
----------------------	------------------------------	-----------

The seat unit is the principal thermal target. Under test, every seat unit in the cabin is driven to maximum processor utilisation through sustained high-complexity decode, replicating the worst-case passenger load of simultaneous full-cabin video playback. The zone concentrator is stressed both passively, by carrying the full content and telemetry traffic of its zone, and actively, through controlled faults that force it through crash, watchdog restart, cache integrity check, and service restoration while the zone remains under live load.

The remaining elements are stressed as consequences of that load rather than as separate targets. The cabin network backbone carries the aggregate of simultaneous decode requests, content prefetch operations, and telemetry streams from every seat, which surfaces bandwidth saturation or packet loss behaviour that appears only under full-cabin concurrency. The content source is exercised through cache stress: forcing concentrator cache invalidation under load tests whether content delivery recovers within its service objective when a local cache must be rebuilt from the source. The power distribution units are exercised by the boarding activation sequence described below.

4.2. How Realistic Passenger Load Is Simulated

Passenger load is simulated along three axes, and in each case the framework deliberately targets the worst case rather than the average.

Concurrency is maximised. Decode onset is synchronised across all seats to within five hundred milliseconds using a distributed timing reference. Real passengers begin content at scattered times, which produces a naturally staggered thermal and electrical onset. Removing that stagger is the point: the test establishes a bound rather than reproducing a typical day.

Content is chosen for saturation. The decode workload uses the highest-complexity codec profile available, which loads the processor more completely than standard-definition video or audio-only playback. A test that stresses the decode pipeline with easy content establishes nothing about the thermal ceiling.

Boarding dynamics are reproduced stochastically. Seat activation is sequenced from published boarding distribution parameters, generating an inrush profile that matches the statistical shape of real boarding rather than a physically unrealistic simultaneous step. This matters because the electrical question is about the rate of aggregate current rise, and a step function would answer a question no aircraft ever asks.

4.3. How Fleet Variation Is Captured

Fleet variation is addressed through zone-level parameterisation. The framework partitions the cabin into concentrator coverage zones and runs thermal and electrical

measurement independently per zone, capturing spatial variation arising from differences in underfloor airflow velocity, seat density, and distance from the power feed point.

Configuration files define aircraft type, seat map topology, zone boundaries, calibration thresholds, and power unit ratings. The same framework therefore executes against different airframes or variants without modification to the test logic, and the results are directly comparable. Over a delivery programme this accumulates a fleet-level baseline, which is what makes it possible to distinguish an installation-specific anomaly from a characteristic of the type.

5. Prior Work and the Origin of the Framework

The framework proposed here derives from practical ground-based load testing that the author conducted in a controlled simulation laboratory in California in March 2010, on narrow-body aircraft entertainment installations. That earlier work established the basic finding that motivates this paper: cabin-level load behaviour differs from the sum of individually tested seats, and the difference is largest in the regions of the cabin with the least favourable airflow.

Two things have changed since. Seat-level compute density has increased substantially, moving from lightweight display terminals to full application processors with hardware decode. Cabin size in the target application has grown from narrow-body to wide-body, roughly doubling seat count and lengthening the fuselage sections over which airflow degradation accumulates.

The present work is therefore a proposal to extend a validated practical approach into a domain where the author has not yet executed it, using simulation and thermal analytics in place of exhaustive physical trial. The distinction matters for how the results in Section 9 should be read. The 2010 work was physical testing on narrow-body installations. The framework described in Sections 6 through 9 is a specified but not yet deployed system, and its quantitative outputs are model projections.

6. Framework Architecture

6.1. Design Principles

The framework is built around four principles:

- **Determinism.** Every test scenario is fully parameterised and reproducible from a configuration file held in version control. No parameter is set by hand at test time.
- **Telemetry completeness.** Every measurement cycle is recorded and retained. The raw corpus is replayable after the test concludes, so detection thresholds can be tightened retrospectively against historical data without repeating physical hardware tests.

- Concurrency by default. Load is applied to every seat simultaneously rather than to a sample, because the failure modes of interest are aggregate properties.
- Evidence traceability. Test output is a structured artifact linking each measurement to the configuration that produced it, suitable for inclusion in an acceptance record.

6.2. Components

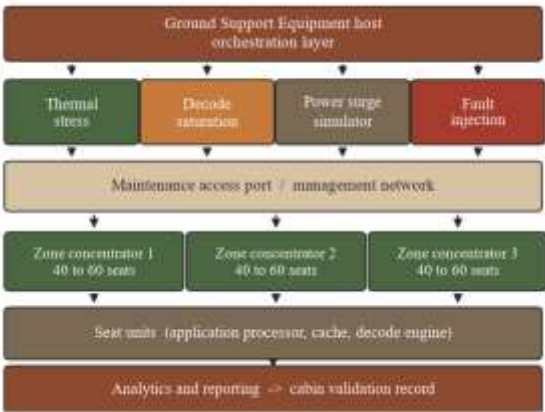


Figure 1. Shows How the Components Connect To the Installed Cabin

Figure 1. Framework architecture. A ground support equipment host drives four test engines through the maintenance access port to the zone concentrators and seat

units. All telemetry returns to the analytics layer, which produces the cabin validation record.

Table 2 lists the framework's functional components.

Table 2. Framework Component Architecture

Component	Sub-modules	Principal output
Thermal stress engine	Seat-zone thermal profiler, ambient ramp scheduler	Per-seat junction temperature proxy at 1 Hz, spatial heatmap overlays
Decode saturation controller	Parallel decode orchestrator, codec stress injector, frame telemetry collector	Per-seat decode success rate, dropped-frame histogram, processor utilisation
Power surge simulator	Boarding ramp modeller, inrush current profiler	Peak draw against unit rating, transient events, power rail stability curves
Fault injection module	Network partition emulator, watchdog timeout simulator, cache corruption injector	Mean time to recovery per fault class, recovery assertion log, zone resiliency record
Analytics and reporting layer	Time-series aggregator, anomaly detector, audit trail generator	Per-zone pass and fail verdict, structured validation artifact, trend comparison

The orchestration layer runs on a ground support equipment host connected to the cabin network through the maintenance access port, establishing a management session with each zone concentrator. Telemetry is carried on a time-ordered event bus, with each seat's stream occupying a dedicated partition. A full-cabin test across four hundred seats emitting ten data points per second across five telemetry categories produces on the order of twenty thousand events per second, which is the design load for the collection pipeline.

7. Test Method

7.1. Setup and Pre-Test Sweep

Before any load is applied, the framework ingests a configuration file specific to the airframe under test:

registration, seat map topology, zone boundaries, power unit ratings, calibration thresholds, and the approved processor thermal derating curves for that installation. The ground support host connects to the cabin network and establishes a management session with each concentrator.

The framework then performs a pre-test health sweep, querying every seat's management endpoint to confirm reachability, logging any seat that is offline or unresponsive as a pre-existing fault, and excluding it from load generation while recording it in the test artifact. This sweep establishes the baseline connectivity record against which post-test seat status is compared. The design target for sweep completion is under two minutes for a four-hundred-seat cabin.

7.2. Four-Phase Load Profile

Load generation proceeds in four phases, illustrated in Figure 2.

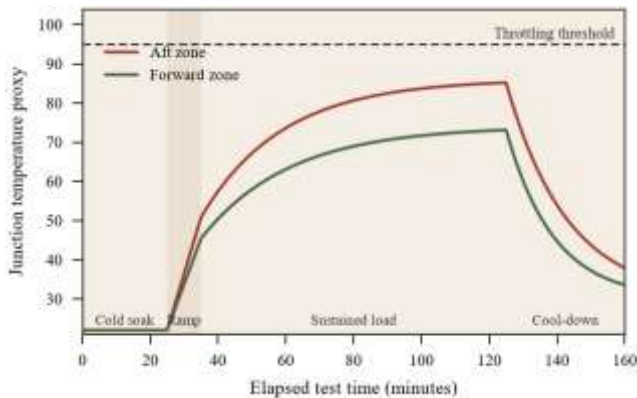


Figure 2. Temperature-Response Profile during Thermal Stress Testing

Figure 2. Modelled four-phase stress cycle. Forward and aft zone junction temperature proxies diverge through the sustained load phase, with the aft zone settling measurably closer to the throttling threshold. Curves are model output, not measurements.

- **Cold soak.** The cabin is held at ambient ground temperature for a configurable dwell, nominally twenty to thirty minutes, with all seats powered but idle. This simulates a gate power-up following an overnight ground stop and establishes the thermal baseline from which the ramp begins.
- **Load ramp.** The decode saturation controller dispatches a synchronised decode initiation command to every seat, using a distributed timing reference to bring all seats to full high-complexity decode within five hundred milliseconds of one another. This synchronisation is deliberate. Real passengers begin content at scattered times, which produces a gentler onset. The test targets the worst case rather than the average, so the natural phase offset is removed on purpose.
- **Sustained load.** Every seat is held at maximum decode utilisation for a configurable duration, nominally ninety minutes, which is sufficient for junction temperatures in the most thermally constrained zones to reach steady state. The power surge simulator runs concurrently on a separate thread, replaying a boarding-phase activation sequence. The two events do not coincide in real operations, but both need to be characterised before delivery, and running them in one session avoids a second cabin occupation.
- **Cool down.** Workload is reduced in quantised steps, capturing the thermal decay curve per zone. This characterises how quickly the cabin returns to safe

operating temperature after a high-load sector, which informs turnaround planning.

7.3. Measurement

Measurement operates at three granularities simultaneously:

- At the seat level, the framework polls each management endpoint at 1 Hz, collecting junction temperature proxy, decode frame success rate, dropped frame count, thermal throttle event flag, and power state.
- At the zone level, a high-resolution current measurement harness samples aggregate power unit draw at ten millisecond resolution. This resolution is necessary because the transient inrush events of interest are shorter than the 1 Hz seat polling interval and would otherwise be missed entirely.
- At the cabin level, the analytics layer aggregates all streams in real time and applies isolation-based anomaly detection [4] against incoming telemetry, surfacing seats or zones whose thermal trajectory diverges from the fleet baseline before the sustained load phase completes.

7.4. Fault Injection

Three fault classes are injected against each zone concentrator while the zone carries full concurrent seat load: process crash with watchdog restart, network partition isolating a zone segment, and content cache corruption requiring re-synchronisation. Each class has a defined injection timing, a recovery assertion window, and a pass criterion expressed as a maximum recovery time. Injecting under load rather than in isolation is deliberate, because recovery behaviour under load is the condition that matters operationally.

For each fault the framework asserts recovery within the target window, absence of fault propagation to adjacent zones, and passenger-facing impact within the applicable service objective.

7.5. Repeatability and the Validation Record

Every parameter governing execution lives in a versioned configuration file committed alongside the framework software. Running the test on a subsequent airframe of the same type requires updating only the registration and seat map fields, with thresholds and timing carried forward from the type configuration. A different variant requires a new type configuration, itself generated from the installation data package rather than hand-tuned.

At conclusion the framework emits a cabin validation record, a structured artifact capturing the full execution trace. Table 3 lists its contents.

Table 3. Contents of the Cabin Validation Record

Field	Purpose
Airframe registration and type	Identifies the unit under test and the applicable type configuration
Test date and framework version	Establishes what software produced the result, so a later framework defect can be scoped to affected records
Configuration file hash	Proves which parameter set governed the run, without relying on the file being retained separately
Calibration table version	Allows a historical result to be re-interpreted if calibration is later corrected
Per-seat verdict	Pass or fail for every seat, including seats excluded as pre-existing faults
Raw time-series telemetry	Complete measurement corpus, compressed, replayable after the fact
Fault injection outcomes	Recovery time and assertion result per injected fault class
Hash-chained audit log	Tamper-evident record of every measurement cycle in sequence

The record is the unit of comparison, and this is where the framework's value compounds across a programme rather than accruing to a single test. Two records from the same type but different serial numbers can be differenced programmatically, which surfaces installation quality variation that no comparison of signed checklists could reveal. A zone that runs consistently warmer than the same zone on sister airframes is an installation observation, not a design observation, and it is actionable before delivery.

The hash chain matters for a narrower reason. Test evidence that can be edited after the fact is weak evidence, and an acceptance process that depends on it inherits that weakness. Chaining each measurement cycle to its predecessor makes retrospective alteration detectable, which is what allows the record to carry weight in a dispute about whether a condition was present at acceptance.

8. Ground-to-Flight Fidelity

The framework reproduces passenger load through software rather than physics: synchronised decode onset across every seat, stochastic power activation sequenced from published boarding distribution parameters, and fault injection timed to occur under full concurrent load. What it cannot reproduce is the aircraft's own operating environment, and the honest position is that this gap is real and bounded rather than closed.

Thermal fidelity is the hardest gap. Cabin pressurisation at cruise altitude changes the convective behaviour of air around each seat unit in ways a hangar cannot reproduce without a purpose-built altitude chamber. The environmental control system operates at different bleed air pressure and flow rate on the ground than at cruise, so the thermal ceiling it provides during ground testing is systematically more favourable than the one available in service. The practical consequence is that a seat passing ground thermal validation with a modest margin may sit closer to its throttling threshold at cruise than the ground test indicated. The framework applies conservative derating offsets to compensate, but a derating offset is an engineering estimate and not a measurement, and this is the framework's principal source of residual uncertainty.

Passenger behavioural randomness is the second gap. Real passengers pause, resume, switch content, and adjust display brightness in patterns no deterministic simulation reproduces. The framework's response is to overshoot rather than to match: synchronising decode onset produces a load more severe than any realistic passenger population generates, so the worst case is bounded even though it is not precisely modelled.

Electromagnetic environment is absent from a ground test. An operating aircraft carries radiated emissions from avionics, radar, and communication systems. Entertainment systems are certified for electromagnetic compatibility independently, but the interaction between electromagnetic stress and a thermally loaded processor running near its throttling point is not characterised in the combined condition, and this framework does not address it.

Vibration and structural flex introduce micro-interruptions in connector contact that do not appear on the ground. Seats showing clean connectivity during a ground test occasionally develop intermittent link faults in service attributable to connector seating rather than to software, and no ground methodology reproduces the fatigue accumulation that produces them.

9. Model-Derived Findings

This section reports outputs of calibrated engineering models, not measurements from a deployed system. The thermal model is calibrated against published processor thermal specifications, published platform technical data, and cabin airflow studies [3]. Each finding below is an output of that model given those inputs. None is a result from executing the framework on a delivered aircraft. Table 4 consolidates the evidence status of every quantity stated in this paper.

9.1. Spatial Thermal Variation

Under sustained full-cabin decode load, the model indicates that aft-zone seats reach junction temperature proxy values eighteen to twenty-two percent closer to the processor throttling threshold than forward-zone seats. The pattern is consistent across every simulated wide-body seat configuration examined and is attributable to documented

reduction in underfloor airflow velocity toward the aft fuselage [3]. Figure 3 presents the modelled spatial distribution.

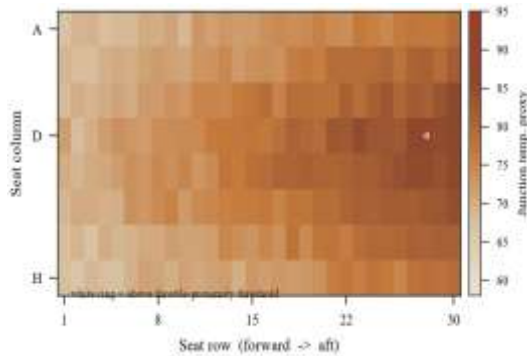


Figure 3. Spatial Distribution of Junction Temperature Proxy across Seat Rows and Columns

Figure 3. Modelled cabin thermal distribution across a wide-body seat map under sustained full-cabin decode load. Values rise toward the aft fuselage, consistent with documented reduction in underfloor airflow velocity. White rings mark seats above the throttle-proximity threshold.

The consequence for test strategy is direct. The model indicates that a ten percent seat-sampling strategy, which is current industry practice, has less than a thirty percent probability of including sufficient aft-zone seats to detect this pattern with statistical confidence. The pattern is not merely missed by sampling. It is unlikely to be detectable by sampling.

9.2. Boarding-Phase Electrical Behaviour

The power surge model, parameterised from published boarding distribution data, produces aggregate current ramp rates between 2.3 and 4.1 amperes per second, varying with boarding wave size and inter-wave interval. The upper bound approaches the inrush response characteristics of overcurrent protection devices rated to typical cabin electrical load analyses, indicating a non-trivial probability of nuisance tripping under worst-case boarding conditions. Figure 4 presents nominal and worst-case profiles against the protection threshold.

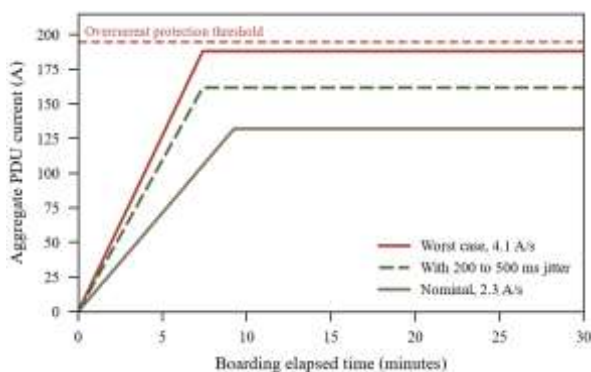


Figure 4. Aggregate PDU Current Profiles under Different Boarding Load Conditions

Figure 4. Modelled boarding-phase aggregate current. The worst-case ramp approaches the overcurrent protection threshold. Applying a randomised per-seat activation jitter lowers the peak by approximately fourteen percent and requires no hardware change.

The model further indicates that applying a randomised per-seat activation jitter of two hundred to five hundred milliseconds reduces worst-case peak inrush by approximately fourteen percent. This is a software configuration change to the seat activation sequence and requires no hardware modification, which makes it the most favourable cost-to-benefit finding the model produces.

9.3. Fault Recovery Behaviour

Fault injection simulation indicates that zone concentrator process crash recovery completes within a ninety-second target in ninety-four percent of simulated instances. The remaining six percent, characterised by crash during an active cache write, extend recovery to between 110 and 140 seconds. Figure 5 presents the recovery timeline by fault class.

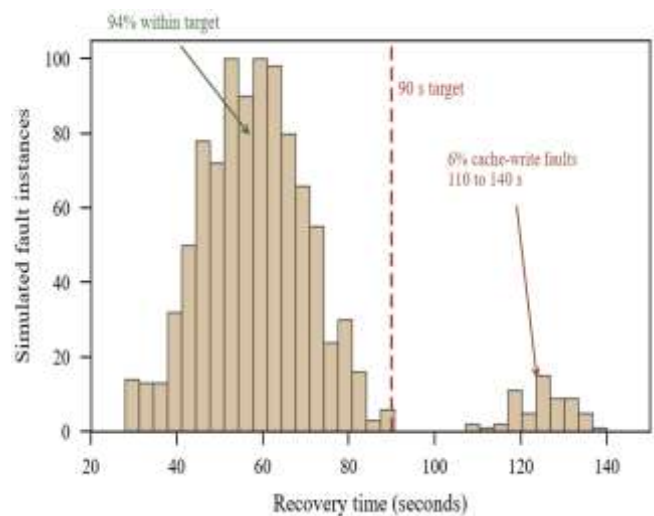


Figure 5. Distribution of Simulated Fault Recovery Times Relative to the 90-Second Target

Figure 5. Modelled distribution of zone concentrator recovery times. The slow tail is confined to faults injected during an active cache write, which isolates the flush interval as the candidate parameter for optimisation. The concentration of slow recoveries in a single identifiable condition points to a specific configuration parameter, the cache write journal flush interval, as the candidate for optimisation.

9.4. Evidence Status

Table 4 separates what the models produce from what requires physical execution.

Table 4. Evidence Status of Quantitative Statements

Quantity	Value	Status
Aft versus forward zone thermal proximity	18 to 22 percent closer to threshold	Model output, calibrated against published airflow and processor data
Detection probability of 10 percent sampling	Below 30 percent	Model output
Boarding-phase current ramp rate	2.3 to 4.1 amperes per second	Model output against published protection device specifications
Inrush reduction from activation jitter	Approximately 14 percent	Model output
Fault recovery within 90 second target	94 percent of instances	Simulation output
Slow-path recovery time	110 to 140 seconds	Simulation output
End-to-end test cycle time	8 to 12 hours target	Design target. Not measured against a physical execution
Recovery improvement from flush interval change	Projected above 99 percent	Projection. Not confirmed by a build change and re-test
Fleet maintenance cost exposure	Modelled range across an 80-aircraft fleet	Modelled estimate from published maintenance labour rates and unit replacement costs. Not validated against airline maintenance records

The distinction is not decorative. The first six rows are defensible as design-validated estimates because the model inputs are published and the derivation is stated. The last three require execution on a delivered aircraft, a software build change, or access to operator maintenance data, and none of those has occurred.

It is worth being precise about what a pilot execution would settle and what it would not. It would convert the thermal and electrical projections into measurements, establish whether the cycle time target is achievable against a real cabin with real access constraints, and confirm or refute the sampling-detection argument by running both the framework and a conventional ten percent sample against the same airframe. It would not settle the derating question, because that requires in-service thermal data from the same airframe after entry into service, and it would not settle the cost projection, which requires operator maintenance records across a fleet and a period. Distinguishing these two groups matters, because a successful pilot would be easy to over-read as validating everything in Table 4.

9.5. Operational Cost Framing

The operational case for full-cabin validation can be framed but not yet substantiated. A fleet of eighty wide-body aircraft flying four sectors daily, at a conservative one

percent seat failure rate across four hundred seats, generates on the order of several hundred fault events per day. Unscheduled seat electronic unit replacement carries a parts and labour cost at an outstation, and zone restarts trigger contractual service-level penalties under typical entertainment service agreements. Across twelve months, an operator running with undetected thermal and power validation gaps accumulates a material annual exposure in avoidable maintenance, penalty liability, and accelerated hardware replacement.

This framing is a modelled estimate built from published maintenance labour rates, typical unit replacement costs, and representative penalty structures. It has not been validated against actual operator maintenance records, and it is presented as a hypothesis about the scale of the opportunity rather than as a finding. The softer effect of passenger dissatisfaction on route selection is real but is not quantified here.

10. Comparison against Current Practice

Table 5 contrasts prevailing acceptance practice against the proposed framework across the dimensions that determine acceptance quality.

Table 5. Validation approach comparison

Dimension	Current practice	Proposed framework
Thermal validation	Manual spot checks, single seat	Automated zone-level cycling across all seats simultaneously
Decode stress	Sample playback on five to ten seats	Synchronised high-complexity decode on the full cabin with frame telemetry
Power surge	Static load estimation from bill of materials	Dynamic boarding-phase profiles across configurable ramp windows
Fault injection	None, or physical disconnection	Software-controlled injection with timed recovery assertion

		under load
Test duration	Days to weeks, manual	8 to 12 hours target, automated full cycle
Traceability	Checklist documents, no structured data	Time-series telemetry, structured artifact, assurance-aligned audit trail
Cross-airframe comparison	Not practically possible	Programmatic comparison of validation records

The most operationally significant difference is not test coverage. It is the shift from non-reproducible manual procedure to deterministically parameterised automation. A configuration file captures the complete state of a run, which enables exact replication on subsequent airframes of the same type, regression comparison across serial numbers, and trend analysis that reveals installation quality variation invisible to checklist comparison. Coverage without reproducibility produces a larger volume of evidence that still cannot be compared across aircraft.

11. Operational Considerations

Three practical properties determine whether the framework is sustainable across a delivery programme rather than viable once.

Interface dependency and schema drift. The framework instruments the cabin through the vendor maintenance interface. Any platform software release that alters the management endpoint schema requires a corresponding framework update before the next delivery test can run. In an active programme with frequent software drops this coupling is the single largest recurring maintenance cost, and it is easy to underestimate at design time because it is invisible during initial development. The mitigation is to isolate schema-dependent code behind an adapter and to version the adapter against platform releases, so that a schema change is a contained update rather than a diffuse one.

Calibration currency. Thermal thresholds depend on approved sensor calibration data, which is specific to an airframe and changes over the life of the installation. A framework running against stale calibration produces confident results that are quietly wrong, which is a worse failure than producing none. Calibration tables therefore need the same version control discipline as the test configuration itself, and the validation record needs to carry the calibration version it ran against so that a historical result can be re-interpreted if the calibration is later corrected.

Baseline drift in the anomaly layer. The anomaly detector compares an airframe against a fleet baseline, and that baseline moves as the programme progresses. A model trained on the first ten deliveries will generate false positives on the twentieth once the underlying population distribution has shifted, whether because of a production change, a supplier change, or simply a wider sample. Periodic retraining is therefore a scheduled activity rather than a one-time step, and the retraining cadence needs to be tied to delivery volume rather than to the calendar.

Physical instrumentation is the fragile element. The high-resolution current measurement harness connects to the power distribution output rails and requires installation and removal at each test site. This adds setup time, and it introduces calibration drift risk if the harness is handled carelessly between deployments. Of everything in the framework, this is the component most likely to degrade quietly, and it is the one that most rewards a documented handling and recalibration procedure.

Regression visibility across software builds. Because fault injection replays identical sequences under identical load conditions on every execution, recovery behaviour becomes comparable between platform software builds. A regression in recovery time introduced by a software change is visible as a difference between validation records rather than as an in-service incident some months later. This is a byproduct of the determinism principle rather than a designed feature, and it is arguably the framework's most useful secondary output.

12. Limitations

No implementation of this framework has been executed. This is a design proposal. No system built to this architecture has run against a delivered aircraft, and no measured outcome appears anywhere in this paper. The findings in Section 9 are model projections with their derivation stated, and should be read as hypotheses to be tested rather than as results.

No comparative evaluation. The framework is not compared against alternative automated approaches, whether vendor acceptance tooling or partially automated methods. Whether this architecture outperforms an alternative in practice is an open question this paper argues but does not settle.

Thermal derating is estimated, not measured. As Section 8 states, the offsets applied to reconcile ground conditions with cruise conditions are engineering estimates. Their accuracy has not been established against in-service thermal data, and they are the largest single source of uncertainty in the thermal findings.

Multi-sector thermal accumulation is not modelled. A single ninety-minute sustained load characterises steady-state behaviour adequately, but an aircraft flying four back-to-back sectors accumulates residual heat in seat structure, cabin panels, and underfloor wiring that the cool-down phase does not reproduce. Seats clearing a single-cycle test may exceed throttling thresholds on the third or fourth sector of a heavy operating day. This is a known gap, not an oversight, and Section 13 identifies it as the first priority for extension.

Content variation is not represented. The framework stresses the decode pipeline with a fixed worst-case codec profile chosen to saturate the processor. A real content library is more varied and in some respects more awkward. It may include legacy encodes produced by older toolchains, rights-managed streams carrying additional decryption overhead on top of decode, and interactive applications whose processor demand arrives in irregular bursts rather than as a sustained load. A synthetic workload chosen for peak sustained utilisation does not reproduce burst behaviour, and it is burst behaviour that interacts most awkwardly with thermal throttling, because a unit sitting just below its threshold can be pushed over by a short spike that a sustained-load test never generates. The framework therefore bounds the sustained case well and the bursty case poorly, and anomalies confined to particular content types will surface in service rather than on the ground.

Instrumentation maintenance is non-trivial. The framework depends on the vendor maintenance interface, so any platform software release that alters the management endpoint schema requires a corresponding framework update before the next delivery test can run. In an active delivery programme with frequent software drops this burden is easy to underestimate. Calibration tables must track approved sensor calibration data per airframe, and the anomaly model requires periodic retraining as the fleet baseline expands, since a model trained on the first ten deliveries will generate false positives on the twentieth once the population distribution has shifted.

The current measurement harness is the fragile component. It requires physical installation and removal at each test site, which adds setup time and introduces calibration drift risk if handled carelessly between deployments. This is the most operationally fragile element of the proposal.

Scope excludes the source systems. The framework takes installed hardware as given. Manufacturing or installation defects upstream are out of scope, and automated testing amplifies them rather than resolving them. A defect present in the installation will be surfaced at scale and at speed, which is useful, but surfacing is not remediation, and investment in installation quality belongs alongside this work rather than after it.

13. Future Work

Three extensions follow directly from the limitations above.

Multi-sector thermal simulation is the first priority. Rather than a single cold-soak to sustained-load cycle, the framework would chain multiple back-to-back load cycles with shortened cool-down intervals between them, reproducing the residual heat accumulation of a four-sector operating day. This closes the largest identified gap between ground testing and in-service failure patterns.

Environmental control system parameterisation is the second. Incorporating aircraft-specific environmental performance data into the thermal model would allow

derating offsets to be derived from the actual bleed air configuration of the airframe under test rather than from a conservative generic estimate. This requires collaboration with the airframer's environmental systems function but would materially tighten the correspondence between ground prediction and in-service behaviour.

Fleet-trained anomaly detection is the third. An anomaly layer trained on validation records accumulated across an entire delivery programme could flag installation quality deviations, such as a zone whose thermal trajectory is statistically inconsistent with its position in the production sequence, before the first revenue flight. This would move the framework from a pass-and-fail acceptance tool toward a predictive quality signal with fleet-wide visibility.

These three are ordered by the ratio of value to difficulty rather than by ambition. Multi-sector simulation requires only changes to the load profile and the scheduler, and it addresses the failure class the current design most clearly misses, which makes it both the cheapest and the most useful. Environmental parameterisation requires data the framework's authors do not hold and a working relationship with a function that has no particular reason to prioritise entertainment system validation, which places its difficulty largely outside the engineering. Fleet-trained detection requires a corpus that does not yet exist and cannot be manufactured, since it accumulates only as deliveries are tested, which makes it a consequence of adoption rather than a precondition for it.

A fourth possibility deserves mention as a boundary rather than a plan. Combining thermal load with electromagnetic stress would characterise the interaction identified in Section 8 as uncharacterised. It is excluded here because it requires a test environment that a delivery-line ground test cannot provide, and proposing it as an extension of an acceptance framework would misrepresent what an acceptance framework can be asked to do.

The necessary first step ahead of all of these is a pilot execution on a live delivery, which would convert the projections in Section 9 into measurements and establish whether the 8 to 12 hour cycle time target is achievable in practice.

14. Conclusion

Seat-centric entertainment architectures have placed a dense compute estate inside the passenger cabin without a corresponding change in how that estate is validated before service entry. Acceptance practice still samples seats and runs functional checklists, which cannot reach failure modes that exist only as properties of the fully loaded cabin.

This paper has proposed a framework that treats the complete installed cabin as a single test subject and automates four test disciplines under concurrent full-cabin load, emitting a structured validation record that supports assurance traceability and cross-airframe comparison. Calibrated models of the framework indicate a consistent aft-

zone thermal disadvantage, boarding-phase current ramps that approach protection thresholds, and a software-only mitigation that reduces peak inrush.

What the paper does not do is demonstrate any of this in operation. Every quantitative statement in Section 9 is a model output, and Table 4 records exactly which quantities would require a physical pilot to establish. The natural next step is that pilot: an execution on a live delivery with the ground rig instrumented and the measured results compared against the projections stated here.

The framework does not claim to eliminate in-service entertainment system failures. Environmentally coupled and statistically rare failures will persist regardless of ground test rigour. What a rigorous ground test can remove is the population of failures that are detectable on the ground and reach the aircraft anyway, which for thermal and power faults appears to be the majority of them.

References

- [1] RTCA, *DO-178C: Software Considerations in Airborne Systems and Equipment Certification*, RTCA Inc., Washington, DC, USA, 2011.
- [2] RTCA, *DO-254: Design Assurance Guidance for Airborne Electronic Hardware*, RTCA Inc., Washington, DC, USA, 2000.
- [3] C. Greenstein, A. Thrasher, J. Clark, A. Ocken, L. Sullivan, A. Saxena, A. Zeni, N. Rao, P. H. Kopardekar and J. Melton, "Open Source Reconfigurable Aircraft Cabin Airflow Model," in *AIAA AVIATION 2021 Forum*, 2021. doi: 10.2514/6.2021-2758.
- [4] F. T. Liu, K. M. Ting and Z.-H. Zhou, "Isolation Forest," in *Proc. 8th IEEE Int. Conf. Data Mining (ICDM)*, Pisa, Italy, 2008, pp. 413-422. doi: 10.1109/icdm.2008.17.
- [5] ITU-T, *Recommendation H.265: High Efficiency Video Coding*, International Telecommunication Union, Geneva, Switzerland, 2013.
- [6] IEEE, *IEEE 1588: Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems*, IEEE, Piscataway, NJ, USA, 2019.
- [7] IATA, *Airport Development Reference Manual*, International Air Transport Association, Montreal, Canada.
- [8] ARINC, *ARINC 628: Cabin Equipment Interfaces*, Aeronautical Radio Inc., Annapolis, MD, USA.
- [9] Apache Software Foundation, *Apache Kafka Documentation*. [Online]. Available: <https://kafka.apache.org/documentation>