Testing Strategies for IoT Systems: A Systematic Literature Review

Goian Sergiu

20.01.2026

ARTICLE INFO

Keywords:
Internet of Things (IoT)
Embedded Systems
Systematic Review
Testing
Quality

ABSTRACT
As Internet of Things (IoT) ecosystems become increasingly pervasive, the complexity of ensuring their reliability grows. Testing these systems presents unique challenges due to their heterogeneous nature, resource constraints and the integration of hardware and software together. This study aims to evaluate and synthesize current methodologies and frameworks for IoT systems testing, focusing on quality attributes such as reliability, security and performance. This paper is a systematic literature review (SLR), hence it contains a structured search strategy. From an initial pool of 30 relevant articles, a rigorous selection process and snowballing procedure were applied to filter the selection down to 10 studies for detailed qualitative analysis and data extraction. The review highlights a shift toward automated testing frameworks and identifies key metrics for evaluating IoT quality.

1 Introduction

The rapid growth of the Internet of Things (IoT) has transformed the modern computing, integrating physical devices with digital networks across a variety of domains such as healthcare, automotive and industrial automation. However, the complexity of these systems presents challenges for quality assurance and reliability.

This study performs a review with a focus on IoT systems' testing, analyzing a range of methodologies to ensure system reliability. The necessity of this review stems from the fragmentation of existing testing frameworks. While foundational surveys provide a broad overview of enabling technologies and protocols, there is a growing need to synthesize specialized testing techniques. Specifically, this review evaluates the transition from traditional software testing to IoT specific approaches, such as fuzzing techniques for vulnerability detection and Machine Learning (ML) intrusion detection for security validation [9, 10].

Furthermore, the research emphasizes the critical role of Quality Attributes. By analyzing architectural styles and their associated quality requirements, this paper categorizes current challenges, including testbed simulation and the integration of digital twins, while highlighting future perspectives in automated testing environments [1, 3]. This systematic approach aims to provide a detailed analysis about the evolving complexities of IoT quality engineering.

2 Methodology

This review follows the guidelines proposed by Kitchenham for Systematic Literature Reviews (SLR) in Software Engineering.

3 Research questions definition

The goal of this SLR is defined with the 2 following questions:

- **RQ1:** What are the most prevalent testing methodologies currently used to validate quality attributes (security, reliability and performance) in IoT systems?
- **RQ2:** What are the main challenges and issues that prevent the full automation of testing processes in integrated IoT environments?

3.1 Search Strategy

An initial search was conducted across digital libraries (ResearchGate, MDPI, IEEE) using keywords: "IoT Testing" and "Quality Attributes". This yielded 30 potential articles.

Table 1: Primary Studies Selected for the SLR: IoT Systems Testing

Reference	Title	Authors	Year
[1]	Current Challenges and Future Perspectives in Testing IoT Systems: A Comprehensive Review	B. Lima et al.	2022
[2]	Systematic Review of Fuzzing in IoT: Evaluating Techniques, Vulnerabilities, and Research Gaps	A. Touqi et al.	2024
[3]	PatrIoT: IoT Automated Interoperability and Integration Testing Framework	M. Bures et al.	2021
[4]	A Systematic Mapping Study on Internet of Things	A. Lepekhin et al.	2019
[5]	Aspects of Quality in Internet of Things (IoT): A Systematic Review	A. Ahmed, M. Bures	2019
[6]	Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications	A. Al-Fuqaha et al.	2015
[7]	Security Assessment of an Internet of Things Device: Penetration Testing Methodology	D. A. Cimpean et al.	2024
[8]	A Multi-Method Study of Internet of Things Systems Testing in Industry	F. Sabir et al.	2023
[9]	Advancements in Machine Learning-Based Intrusion Detection in IoT: Research Trends and Challenges	M. B. Bankó et al.	2025
[10]	A Survey on IoT Security Challenges and Their Solutions Using Machine Learning and Blockchain Technology	N. Sharma, P. Dhiman	2025

3.2 Selection Criteria

Papers were selected based on their relevance to IoT testing frameworks and quality metrics. Through a snowballing procedure and abstract and conclusion screening, the pool was reduced to the 10 most impactful studies.

4 Data Extraction

To ensure a consistent comparison across the selected studies, a standardized data extraction form was developed. For each of the 10 primary articles, the following information was recorded:

- **Testing Methodology:** The specific approach (Fuzzing, MBT, ML-based).
- **Quality Focus:** The primary quality attribute addressed (Security, Performance, Reliability).
- **Tools/Datasets:** Tools / Data used for validation.
- **Outcomes:** Significant findings and results.

5 Selected Articles Overview

6 Data Synthesis

The review by Lima and Pinto [1] evaluates the current state of IoT testing, focusing on the disconnect between research and industry practice. Instead of just looking at automation, the authors identify five core challenges: heterogeneity, scale, mobility, resource constraints and security. Their method involved a systematic analysis of existing literature to see how well current tools handle the things layer versus the cloud layer.

The paper emphasizes that while many frameworks exist, most struggle with the physical distribution of devices and real-time testing. The findings show that unit testing is well-addressed, but system-level interoperability remains a major gap. The authors conclude that future testing must move toward "test-as-a-service" models to handle the massive scale of industrial IoT deployments effectively.

The next work [2] comes with a deep technical investigation into the efficacy of fuzzing as a security testing methodology for the IoT ecosystem. Recognizing the fragmentation in existing security assessments, the authors utilize a systematic review process to evaluate fuzzing frameworks across three distinct layers: the perception layer, the network layer and the application layer. The approach focuses on identifying critical vulnerabilities such as memory corruption, buffer overflows and DoS flaws.

The used method involves a comparative analysis of different fuzzing strategies, specifically emphasizing the transition from traditional black-box mutation to more advanced grey-box and white-box techniques. This methodological diversity allows the researchers to benchmark how different tools handle the proprietary protocols and heterogeneous

hardware architectures typical of IoT devices.

In the evaluation phase, the research highlights the use of diverse datasets, including firmware from smart home devices and industrial controllers. The study benchmarks industry-standard tools like AFL, Peach, and Boofuzz, evaluating their performance in discovering "unique crashes" and "vulnerability types." The obtained results demonstrate that while grey-box fuzzers excel at path coverage in application-level software, they struggle with the deep state machines of IoT communication protocols. The review quantifies these findings by showing a significant research gap in "state-aware" fuzzing, which is necessary to improve vulnerability detection depth in non-standardized implementations. The research concludes that a hybrid approach, integrating machine learning-based seed generation with hardware-in-the-loop (HiL) testing, is essential to overcome the limitations of current automated security tools.

Paper [3] introduces PatrIoT, a specialized framework designed for automated interoperability and integration testing within complex IoT ecosystems. Addressing the challenge of testing heterogeneous devices, the authors implement a modular architecture that enables the validation of communication flows between IoT nodes, gateways and cloud services. The approach focuses on automating the test lifecycle, specifically targeting the integration points where data exchange frequently fails due to protocol mismatches.

The used method relies on Model-Based Testing (MBT) to describe system behavior through formal models. By utilizing these models, the framework automatically generates test cases that cover various integration scenarios, including edge to cloud and device to device interactions. This methodology allows for the systematic verification of "interoperability profiles" ensuring that diverse devices can communicate reliably regardless of their underlying hardware or software stack.

In the experimental evaluation, the research team deployed the PatrIoT framework in a controlled environment consisting of multiple smart home sensors and industrial actuators. The testing focused on critical performance metrics such as test execution time, fault detection effectiveness, and the reduction of manual effort. The obtained results demonstrate that the automated generation of test cases via MBT significantly decreases the time required for integration testing compared to traditional manual scripting. Furthermore, the study quantifies the framework's efficiency by showing high coverage of protocol level interactions with minimal computational overhead. The research concludes that PatrIoT provides a scalable solution for continuous integration (CI) environments, though it notes that the initial effort of model creation remains a challenge for rapid development cycles.

In this work [4], the authors provide a systematic mapping study to categorize the vast landscape of IoT research. The approach focuses on identifying the primary domains of application, such as smart cities and healthcare, while evaluating the maturity of existing software engineering practices. By utilizing a structured classification scheme, the authors analyze the distribution of research efforts across various IoT architecture layers.

This methodology allows the researchers to benchmark which areas, such as "connectivity" and "data processing" have reached technical saturation and which areas, like "standardized testing" and "maintainability" remain underdeveloped.

In the evaluation phase, the research identifies a significant imbalance: while architectural designs are heavily documented, empirical evidence regarding the long-term reliability of these systems is lacking. The obtained results highlight that only a small percentage of studies provide reproducible industrial benchmarks. The research concludes that the IoT field needs a shift from theoretical modeling toward standardized, practical validation frameworks to support large-scale industrial adoption.

This work [5] also addresses the technical characterization of quality attributes for resource limited IoT nodes. The approach is based on the decomposition of the ISO/IEC 25010 quality model, adapting standard sub-characteristics to the limitations of embedded hardware. The methodology implies a comparative analysis of quality measurement elements (QMEs), where the researchers extracted technical data regarding how reliability, security and performance are measured in heterogeneous environments. This process allowed for the identification of specific "Quality Profiles" for different IoT application domains (for instance healthcare vs. industrial).

The evaluation involved a dense dataset of 80+ technical studies to determine which metrics are most frequently validated in the field. The results provide a ranking of quality priorities, revealing that there is a significant lack of standardized metrics for recoverability and fault tolerance at the device level, while energy efficiency and interoperability are highly prioritized in the network layer. The study quantifies that only 18% of the reviewed methodologies provided a mathematical model for calculating Mean Time to Failure (MTTF) in a sensor network.

For the sixth paper [6], the focus is on the technical stack and communication protocols. This summary emphasizes the layers of testing required for the physical and network protocols that define IoT connectivity.

This survey provides a deep architectural break-

down of the IoT stack to identify the specific protocols and technologies that require rigorous validation. The approach is an exhaustive technical mapping of the five-layer IoT architecture, focusing on the intersection of the Sensing, Network and Application layers.

A performance comparison of constrained application protocols is performed, more specifically by benchmarking the overhead and throughput of MQTT, CoAP and AMQP. By extracting technical data from various protocol-specific datasets, the researchers evaluate the trade-offs between reliability and resource consumption (battery/CPU) across different wireless standards such as ZigBee, BLE and 6LoWPAN.

The obtained results provide technical benchmarks for latency and packet loss under high-congestion scenarios. The data reveals that CoAP exhibits lower overhead in low-power environments but lacks the robust state-management found in MQTT's Quality of Service (QoS) levels. The study quantifies that the selection of a communication protocol directly impacts the "Testability" of the system, with asynchronous protocols requiring more complex temporal testing models. The research concludes that testing frameworks must move beyond standard TCP/IP validation to account for the lossy, low-power nature of the Physical and MAC layers in heterogeneous IoT deployments.

Study [7] proposes a structured penetration testing methodology specifically tailored for the security assessment of IoT hardware and software. Recognizing the unique attack surface of connected devices, the approach focuses on a four steps process: reconnaissance, vulnerability scanning, exploitation and post exploitation analysis. The study aims to provide a repeatable security audit framework that addresses both network level and device level entry points.

The used method employs a combination of automated scanning tools and manual exploitation techniques. This includes analyzing the communication protocols between the IoT gateway and the end devices to identify weaknesses in encryption and authentication mechanisms. By simulating real world cyberattacks, the researchers benchmark the resilience of the target device against common threats such as credential brute-forcing and man-in-the-middle (MitM) attacks.

In the experimental evaluation, the methodology was applied to a commercial IoT device to validate its effectiveness. The testing regimen identified critical security flaws, including unencrypted data transmission and the presence of hidden administrative interfaces. The obtained results demonstrate that a specialized IoT-centric methodology uncovers 40% more vulnerabilities than generic network security audits. The research concludes that for robust IoT protection, manufacturers must integrate periodic penetration testing into the device lifecycle to mitigate risks associated with evolving threat landscapes.

The authors in [8] present a multi-method study aimed at capturing the industrial reality of IoT testing. Moving beyond theoretical models, the approach focuses on identifying the specific hurdles practitioners face when validating large-scale, heterogeneous systems. To ensure a comprehensive analysis, the authors employed a triangulation methodology consisting of qualitative interviews with industrial experts followed by a broad quantitative survey of software practitioners.

The used method analyzes the current testing landscape across various domains, including smart homes and industrial automation. By categorizing feedback from these two phases, the researchers were able to benchmark the discrepancy between academic testing trends and the actual tools used in the field. This methodological approach allows for a deep dive into the "hardware-software" divide, highlighting how physical device constraints dictate the testing strategies used in commercial environments.

In the evaluation phase, the research highlights that while unit testing is highly automated, system-level integration remains largely manual due to the lack of specialized IoT testing tools. The testing focused on critical bottlenecks such as environment simulation and the high cost of physical testbeds. The obtained results demonstrate that a majority of industrial projects rely on emulators and simulators, which often fail to accurately replicate real-world network conditions like latency and packet loss. The research concludes by identifying a critical need for standardized, cost-effective testbeds and "test-as-a-service" models to support the rapid growth of industrial IoT deployments.

In the ninth work [9], the efficiency of ML algorithms is evaluated in detecting network-layer intrusions within IoT infrastructures. The approach treats the Intrusion Detection System (IDS) as a testable security component, benchmarking various supervised and unsupervised models for their robustness against adversarial inputs. It contains a comparison of deep learning architectures (specifically Recurrent Neural Networks - RNN and Long Short-Term Memory - LSTM) against traditional classifiers like Random Forest or Clustering. The researchers focused on technical metrics such as the "False Positive Rate" (FPR) and "Detection Accuracy" to determine which models best handle the high-velocity, low-payload data streams typical of IoT traffic.

The experimental phase utilized the UNSW-NB15 and BoT-IoT datasets, which contain a diverse range of simulated attack vectors including

Table 2: Technical Synthesis of Selected Studies: Methodology, Quality Focus and IoT Layer

Reference	IoT Layer	Quality Focus	Methodology	Key Technical Contribution
[1]	Things/Cloud	Interoperability	Systematic Review	Identified 5 core challenges and "Test-as-a-Service" model.
[2]	Cross-layer	Security	Hybrid Fuzzing	Benchmarked state-aware fuzzing and HiL testing gaps.
[3]	Edge/Cloud	Interoperability	MBT	PatrIoT framework for automated integration lifecycle.
[4]	Cross-layer	Reliability	Mapping Study	Classification of SE practices and industrial benchmark gaps.
[5]	Perception	Reliability	ISO 25010	Mapping of QMEs and domain-specific Quality Profiles.
[6]	Net/Physical	Testability	Protocol Benchmarking	Comparative overhead analysis of MQTT, CoAP, and AMQP.
[7]	Hardware/Net	Security	Penetration Testing	4-stage audit methodology uncovering 40% more flaws.
[8]	System-level	Practicality	Triangulation	Identification of environment simulation and testbed bottlenecks.
[9]	Network	Security	ML (LSTM/RNN)	98% accuracy in temporal intrusion detection patterns.
[10]	Network/App	Data Integrity	Blockchain/ML	Decentralized validation framework with hybrid detection.

DDoS and protocol exploits. The obtained results demonstrate that LSTM-based models achieve an accuracy of 98% in detecting serialized attack patterns, significantly outperforming traditional methods in temporal sensitivity. However, the study identifies an important trade-off: the computational cost (especially the training time and inference latency) of these deep learning models, which increases the testing overhead. The research concludes that while ML provides superior detection capabilities for complex security testing, the deployment requires hardware-accelerated inference engines to maintain real-time performance in production IoT environments.

The last selected study [10] proposes a decentralized security framework to validate data integrity and mitigate centralized points of failure in IoT networks. The approach integrates Blockchain technology for immutable logging with Machine Learning for real-time anomaly detection. The used method involves a performance-security trade-off analysis, evaluating the overhead of smart contract execution on resource-constrained nodes.

By benchmarking the system against several datasets of IoT network traffic and transaction logs, the obtained results show that while the hybrid model achieves a 95% detection rate for data-tampering attacks, the blockchain layer introduces an up to 20% increase in communication latency, concluding that for high-security IoT testing, the decentralized validation of "proof-of-work" or "proof-of-stake" is necessary.

7 Similarities and Differences

7.1 Scope and Focus

Across the ten studies, a clear division emerges between survey-based research and solution-oriented or evaluations. Early works such as [6] and [4] (focus on architectural classification and research mapping, whereas more recent studies ([2], [7]) emphasize practical testing methodologies, security validation and applicability in industrial envirnomnets. This shift reflects the transition of IoT systems from conceptual architectures to deployable, testable infrastructures.

7.2 Methodological Diversity

A significant methodological contrast exists between formalized testing approaches (Model-Based Testing in PatrIoT) and exploratory or empirical methods (penetration testing, fuzzing, industrial surveys). While MBT-based approaches demonstrate higher automation and coverage, they require substantial upfront modeling effort. Industry-focused studies (Sabir's multi-method approach, Lima's implying "Test-as-a-Service") consistently highlight the lack of tooling capable of handling real world divergence and physical deployment constraints.

8 Results and Synthesis of Research Questions

In addressing the research questions posed in this SLR, the analysis of the 10 selected studies has yielded the following insights regarding the current state of IoT quality assurance.

Table 2 categorizes these studies by methodology and IoT layer, distinguishing between systematic literature mappings and applied framework research to provide a clear view of the IoT testing landscape.

8.1 Tools and Frameworks in IoT Systems Testing

The reviewed studies highlight a definitive trend toward specialized tools capable of handling IoT heterogeneity. Frameworks like PatrIoT were emphasized for managing automated integration between nodes and cloud services. For security validation, fuzzing tools were utilized for protocol-level vulnerability discovery [2].

The integration of penetration testing methodologies and automated scanners emerged as a critical theme in [7]. The use of simulators and emulators remains the primary industrial strategy for bypassing the high cost of physical testbeds. This tool diversity underscores a shift toward "test-as-a-service" models and hybrid environments designed to bridge the gap between virtual simulations and physical hardware constraints.

8.2 Methodologies for Validating Quality Attributes

The methodologies identified in the literature range from systematic mapping and ISO/IEC 25010-based quality modeling [4, 5] to dynamic validation methods.

MBT was a recurring theme, particularly for its ability to generate test cases for complex integration scenarios in PatrIoT [3]. Additionally, ML architectures, specifically LSTM and RNN, were noted for their 98% accuracy in identifying network intrusions [9]. Comparative protocol benchmarking also emerged as a vital methodology, allowing researchers to evaluate the trade-offs between CoAP's low overhead and MQTT's robust Quality of Service (QoS) levels [6]. These methodologies reflect a multi-layered strategy addressing the sensing, network, and application layers.

8.3 Synthesis of RQ1: Prevalent Methodologies

Our analysis reveals that methodologies are shifting from manual testing to automated MBT and decentralized security validation.

- **Security:** Fuzzing (AFL, Boofuzz) [2], Penetration Testing [7] and LSTM / RNN-based detection [9] are the dominant approaches for finding memory corruption and network exploits.
- **Reliability:** MBT is the primary method used to handle interoperability and integration between heterogeneous devices [3].
- **Quality Measurement:** Decomposition of the ISO 25010 model is the standard for adapting quality metrics to resource-limited nodes [5].

8.4 Synthesis of RQ2: Challenges to Automation

The main challenges preventing full automation identified in the primary studies include:

- **Protocol Complexity:** Asynchronous communication and proprietary protocols limit the effectiveness of standard "state-unaware" fuzzers [2, 6].
- **Hardware Barriers:** The high cost of maintaining physical testbeds and the "reality gap" of simulators prevent accurate real-world validation [8].
- **Computational Overhead:** Deep learning models for security and blockchain layers for data integrity introduce latency increases of 20-30%, complicating real-time performance testing [9, 10].

9 Discussion and Future Work

The analysis reveals a significant disconnect between academic research and industrial needs. While 2022-2025 research focuses heavily on automation and ML, practitioners still rely on manual testing for system-level integration due to a lack of specialized tools [8, 1]. The shift toward hybrid security models, combining Blockchain for integrity and ML for detection, represents a promising but computationally expensive path [10].

Future research should focus on standardizing metrics for fault tolerance and recoverability at the device level, as these remain underdeveloped [5]. Additionally, there is a critical need for "state-aware" fuzzing tools and lightweight ML models that can operate within the resource constraints of the perception layer without incurring prohibitive latency.

10 Conclusion

This SLR synthesized 10 primary studies to map the evolving landscape of IoT testing. We identified a clear transition from manual validation to automated frameworks like PatrIoT and ML-driven security systems. While specialized methodologies uncover up to 40% more vulnerabilities than generic audits [7], significant gaps remain in environment fidelity and standardized industrial benchmarks. As IoT systems scale, the convergence of Model-Based Testing and decentralized security validation will be vital for ensuring the robustness of global digital infrastructures.

References

[1] B. Lima and R. Pinto, "Current Challenges and Future Perspectives in Testing IoT Systems: A Comprehensive Review," *IEEE Sensors Review*, 2022.

[2] A. Touqi et al., "Systematic Review of Fuzzing in IoT: Evaluating Techniques, Vulnerabilities, and Research Gaps," *Research Square*, 2024.

[3] M. Bures, B. S. Ahmed, V. Rechtberger, and M. Klima, "PatrIoT: IoT Automated Interoperability and Integration Testing Framework," in *14th IEEE Conference on Software Testing, Verification and Validation (ICST)*, 2021.

[4] A. Lepekhin, A. Borremans, I. Ilin, and S. Jantunen, "A Systematic Mapping Study on Internet of Things," in *IEEE/ACM 1st International Workshop on Software Engineering Research & Practices for the Internet of Things*, 2019.

[5] A. Ahmed and M. Bures, "Aspects of Quality in Internet of Things (IoT): A Systematic Review," *IEEE Access*, 2019.

[6] A. Al-Fuqaha et al., "Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 2347–2376, 2015.

[7] D. A. Cimpean, M. Vochin, R. Craciunescu, and A. C. Dragulinescu, "Security Assessment of an Internet of Things Device: Penetration Testing Methodology," *Good Practices and New Perspectives in Information Systems and Technologies*, pp. 284-294, 2024.

[8] F. Sabir, Y. Gueheneuc, and J. B. Minani, "A Multi-Method Study of Internet of Things Systems Testing in Industry," *IEEE Internet of Things Journal*, 2023.

[9] M. B. Bankó et al., "Advancements in Machine Learning-Based Intrusion Detection in IoT: Research Trends and Challenges," *Research Trends and Challenges*, 2025.

[10] N. Sharma and P. Dhiman, "A Survey on IoT Security Challenges and Their Solutions Using Machine Learning and Blockchain Technology," *Cluster Computing*, 28(5):3-40, 2025.