

DOI: https://doi.org/10.48009/3_iis_2026_110

AI-driven smart delivery lockbox with multi-modal authentication

Riley Nham, *Middle Georgia State University*, rileynham@gmail.com

Sahaj Patel, *Middle Georgia State University*, sahajmpatel08@gmail.com

Jonathan Jenkins, *Middle Georgia State University*, jonathan.jenkins2@mga.edu

Joobum Kim, *Middle Georgia State University*, joobum.kim@mga.edu

Package theft and missed deliveries have become increasingly prevalent challenges in residential settings, creating a demand for intelligent, autonomous security solutions. This paper presents an AI-driven smart delivery lockbox system that addresses the carrier-specific, cloud-dependent, and homeowner-reliant limitations of existing commercial solutions through a fully local, confidence-based multi-modal authentication framework. The proposed system integrates computer vision, voice recognition, and physical sensing to authenticate delivery personnel and secure package storage without proprietary hardware or external cloud services. Using a combination of Home Assistant, Frigate, Node-RED, ESPHome, and MQTT, the system continuously monitors its environment through five operational states: Scanning, Verification, Open, Security, and Theft. Person and package detection are performed using Frigate's AI-based object detection, while OpenAI Whisper handles speech-to-text transcription for company name and order number verification. A matrix keypad provides a fallback authentication method, and a load cell weight sensor confirms package placement upon delivery. A confidence scoring model drives all state transitions, dynamically escalating verification requirements based on accumulated evidence from multiple input modalities. Structured feasibility testing and quantitative trials across ten test cases, including unauthorized access attempts and degraded environmental conditions, demonstrate the viability of multi-modal, confidence-driven authentication for autonomous package reception in a residential setting.

Keywords: smart home security, multi-modal AI, IoT, authentication, speech recognition, porch piracy

Introduction

The rapid growth of e-commerce has fundamentally changed the way consumers receive goods, with residential package deliveries becoming an everyday occurrence for millions of households. However, this convenience has come at a cost. Package theft, colloquially known as "porch piracy," has emerged as a significant and growing problem across the United States, with surveys indicating that tens of millions of Americans have had at least one package stolen from their doorstep (Security.org, 2025). Beyond theft, missed deliveries create logistical burdens for both consumers and carriers, often requiring multiple re-delivery attempts or trips to carrier facilities.

Existing commercial solutions fall short in three critical ways: most are carrier-specific, requiring proprietary hardware or tight integration with a single delivery ecosystem; they depend on cloud

infrastructure, transmitting sensitive audio, video, and authentication data to external servers outside the homeowner's control; and they lack the ability to autonomously authenticate delivery personnel without active homeowner involvement. The proliferation of affordable IoT hardware, open-source software, and locally deployable AI models presents an opportunity to address all three limitations within a single unified system.

This paper presents an AI-driven smart delivery lockbox whose primary contribution is a confidence-based, multi-modal authentication framework that is carrier-agnostic, fully local, and fully autonomous. The system integrates computer vision, speech-to-text processing, physical sensing, and fallback keypad authentication, all coordinated locally over MQTT (Message Queuing Telemetry Transport), with no dependency on external cloud services or carrier-specific infrastructure. Because the order number used for authentication is printed directly on the package label and included in retailer confirmation emails, any delivery driver from any carrier can complete authentication without prior enrollment. The system is designed to defeat opportunistic and impersonation-based theft at the point of delivery; it does not defend against an attacker with access to the homeowner's email account or one capable of breaching the enclosure through brute force, both addressed later as boundaries of the threat model. Energy efficiency is achieved through smart plug integration and email parsing, ensuring the system is only active when deliveries are expected. The result is a fully local, privacy-preserving system that any motivated homeowner can build using commercially available, open-source components.

Background and Literature Review

Package Theft and the Porch Piracy Problem

Residential package theft in the United States has grown substantially alongside the expansion of e-commerce. Nearly half of Americans report having a package stolen, up from 34% in 2023 to 45% in 2024, with 29% of victims reporting multiple stolen parcels (Lombardo Homes, 2024). Porch piracy cost Americans more than \$37 billion in 2024 alone, concentrated in dense suburban and metropolitan areas ((Edwards, 2025), and is widely regarded as difficult to prosecute, with only 3.2% of reported incidents resulting in arrest (Red Stag Fulfillment, 2026).

Existing Solutions: Commercial Systems and Smart Lockers

Existing responses span passive physical solutions to active commercial integrations, each with meaningful limitations. Simple lockboxes require sharing credentials with every expected driver in advance, creating a static vulnerability with no autonomous verification. Amazon Key, the most sophisticated commercial alternative, grants one-time access through a proprietary app-based flow but is restricted to Amazon deliveries, requires compatible smart garage hardware, and depends on Amazon's cloud infrastructure. Ring Video Doorbell and similar systems provide passive monitoring but no physical package security, leaving packages exposed after delivery.

Smart lockers automate delivery by providing secure, unattended storage that holds packages until retrieval (Luis et al., 2022). Ooi et al. (2021) presented a smart modular parcel locker system using IoT to alleviate last-mile delivery pressure through concentrated, contactless stopping locations. These systems remain dependent on cloud services, carrier-specific integrations, or fixed-location infrastructure that limits individual residential deployment.

Computer Vision for Person and Object Detection

Real-time object detection has become foundational in applied computer vision, with the YOLO architecture dominant across surveillance, robotics, and autonomous systems. Terven and Cordova-Esparza

(2023) document YOLO's evolution through YOLOv8, improving the balance between detection speed and accuracy for edge deployment, while Hussain (2024) shows that architectural innovations across YOLOv5, YOLOv8, and YOLOv10 have made these models increasingly viable for resource-constrained hardware. The proposed system applies these capabilities through Frigate, an open-source, locally hosted network video recorder that performs real-time object detection using TensorFlow without transmitting data externally, publishing detection events over MQTT (Frigate, n.d.). Frigate's edge-based inference ensures video data never leaves the homeowner's local network, directly addressing the privacy limitations of commercial alternatives.

Automatic Speech Recognition and Voice-Based Authentication

Automatic speech recognition has advanced rapidly with large-scale transformer-based models. Radford et al. (2023) introduced Whisper, trained through weak supervision on 680,000 hours of multilingual data, generalizing well without dataset-specific fine-tuning and running entirely on local hardware. Research on multimodal authentication has found that users place less trust in voice-based inputs than traditional modalities (Waked et al., 2023), motivating the decision to treat voice as one contributing factor within a broader confidence model rather than a standalone authentication method, with keypad fallback available when transcription fails.

Multi-Modal Authentication in IoT Systems

Multi-factor and multi-modal authentication have been widely studied for improving access control security, with research showing that combining independent signals raises the difficulty of unauthorized access beyond any single factor. Memos et al., as cited in Saleem et al. (2025), proposed a device authentication model that uses contextual IoT information and a Random Forest algorithm integrated with Belief-Desire-Intention reasoning agents to improve accuracy. The proposed system's confidence model reflects this work, dynamically accumulating evidence from multiple sources rather than relying on any single modality for a binary decision, with stricter authentication enforced precisely when risk is highest.

Gap in the Literature

While prior work has addressed individual components, including smart lockers, edge-based computer vision, voice authentication, and multi-modal access control, no existing system integrates all of these modalities within a unified, open-source, locally hosted, confidence-driven framework for residential autonomous package reception. Commercial offerings approaching this level of integration are proprietary, carrier-specific, and unavailable for independent replication. The system proposed here addresses this gap across three dimensions: it is carrier-agnostic, operating with any carrier whose confirmation emails include an order number; it is privacy-preserving, with all AI inference and authentication data remaining on local hardware; and it is fully autonomous, requiring no homeowner involvement at any delivery stage.

Research Questions

1. Can a confidence-based multi-modal authentication framework reliably distinguish authorized delivery personnel from unauthorized individuals in a residential setting?
2. How effectively can a state machine driven by AI-integrated vision, voice, and physical sensing manage autonomous package reception without human intervention?
3. Can IoT integration with email parsing and smart power management improve the energy efficiency and anticipatory capabilities of a smart delivery lockbox system?
4. What are the limits of package protection at both the authentication and physical layers within this threat model using open-source and accessible components?

System Design and Methodology

System Overview

The AI-driven smart delivery lockbox is designed as a fully local, event-driven system that autonomously manages the entire package delivery lifecycle, from anticipating an incoming delivery to securing the package. The system architecture integrates a combination of hardware components, local open-source software platforms, and AI-based decision making, managed through a central confidence-based state machine implemented in Home Assistant (Home Assistant, n.d.). All communication between platforms occurs over MQTT (OASIS, 2014), ensuring low latency and eliminating any dependency on external cloud services. Figure 1 illustrates the overall system architecture and component interactions.

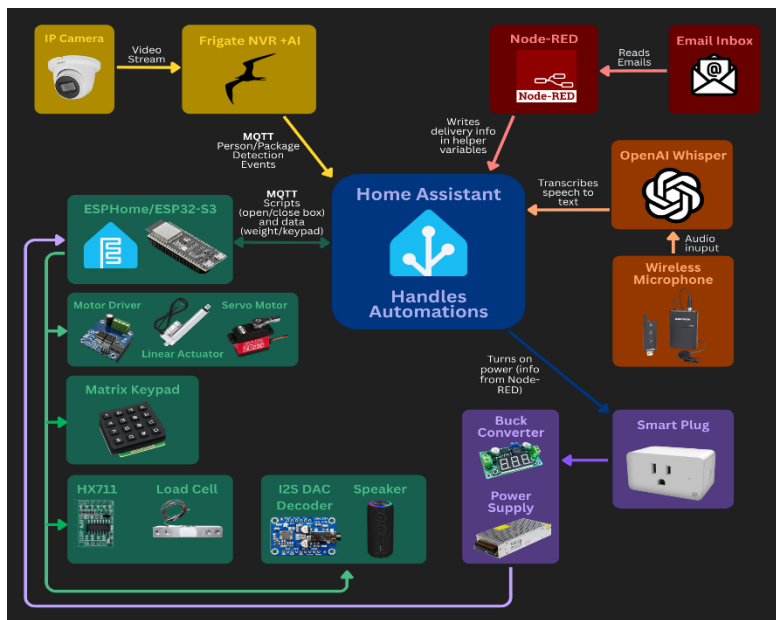


Figure 1. System Architecture

The delivery lifecycle begins before the delivery occurs. Node-RED (n.d.) continuously monitors the homeowner's email inbox for incoming delivery confirmation emails, parsing each message to extract the delivery carrier name, expected arrival date, and the last four digits of the order number, which serves as a unique per-delivery authentication code. These values are stored in a dual queue system maintained in Home Assistant helper entities, enabling the system to manage multiple pending deliveries across one or more days without homeowner intervention. Upon detecting an upcoming delivery, Node-RED triggers a smart plug to power on the lockbox hardware only

on days when deliveries are expected, ensuring energy is only consumed when needed. Once powered, the system continuously monitors its environment through five operational states: Scanning, Verification, Open, Security, and Theft. Each state is driven by inputs from the system's hardware components and evaluated against a confidence-based decision model that dynamically adjusts authentication requirements based on accumulated evidence from multiple inputs.

Hardware Components

The physical infrastructure of the smart delivery lockbox integrates computing, sensing, actuation, and audio components. A local computer hosts all software platforms and AI models. An ESP32-S3 microcontroller programmed via ESPHome bridges the software stack and physical components, managing the matrix keypad, motor driver, and HX711 load cell amplifier. A linear actuator and servo motor handle lid actuation and locking, an IP camera provides continuous video input, and a wireless microphone and speaker enable voice interaction with delivery personnel. A smart plug controls power to the entire system, and a load cell weight sensor confirms package placement inside the box. Figure 2 provides a labeled overview of the assembled hardware.

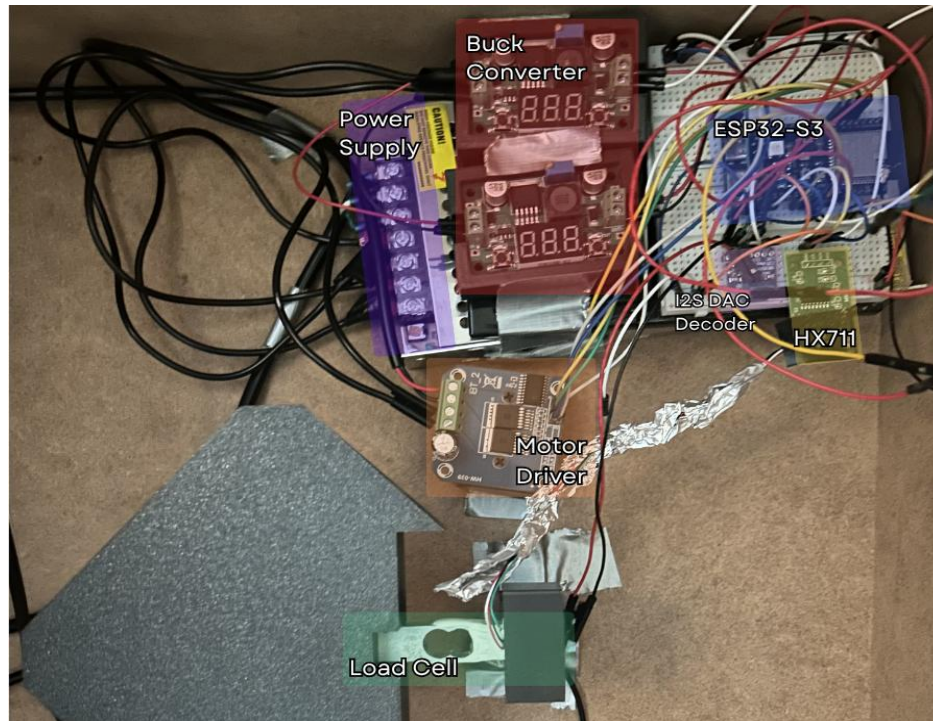


Figure 2. Assembled Hardware Components

Software and Platform Stack

The software stack is built entirely on open-source, locally hosted platforms communicating over MQTT, as illustrated in Figure 1. Home Assistant coordinates all state transitions and authentication logic; Frigate performs real-time AI-based object detection on the IP camera stream without transmitting data externally; OpenAI Whisper provides local speech-to-text transcription for voice authentication; and Node-RED handles email parsing and queue population. Because every authentication input except keypad entry is processed through AI-based perception, the keypad serves exclusively as a fallback for when AI processing fails rather than a primary authentication channel.

State Machine Design

The system's behavior is governed by a confidence-based state machine consisting of five operational states, each representing a distinct phase of the delivery interaction. Figure 3 illustrates the complete state machine, including the confidence point values and thresholds driving each transition.

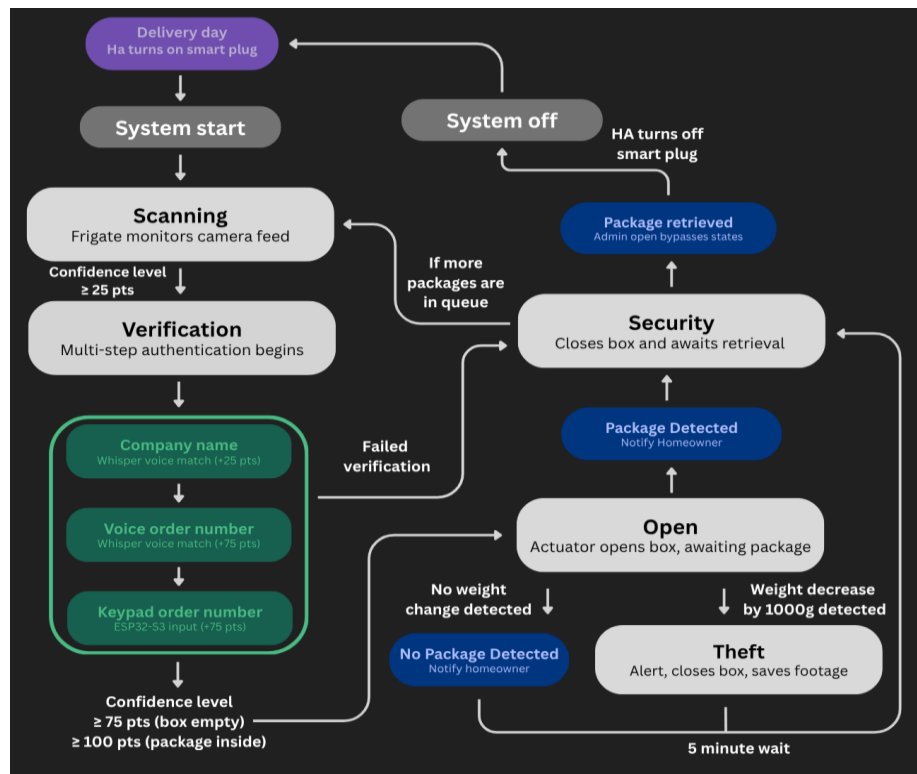


Figure 3. State Machine and Authentication Flow

The Scanning state is the system's default active state, entered automatically upon startup. Its purpose is to monitor the environment for evidence of an approaching delivery using Frigate's detection within a configured porch zone boundary, which excludes passersby outside the immediate delivery area. Upon person detection, the system transitions to Verification, the multi-step authentication phase that establishes the legitimacy of the detected individual before granting access to the box. The delivery person is guided through company name and order number verification by voice, with automatic keypad fallback if voice transcription fails. Once the cumulative confidence score reaches the applicable Open threshold, the system transitions to the Open state, where the ESP32-S3 releases the locking mechanism and the load cell monitors for package placement. Confirmed placement transitions the system to Security, in which the box remains locked, the delivered orders helper is updated with the received order number, and the system returns to Scanning if further deliveries are queued. The Theft state is a protective response triggered exclusively from Open when the load cell detects an unauthorized weight decrease, immediately securing the box, alerting the homeowner, and preserving timestamped video evidence before returning to Scanning or Security after a five-minute window.

Confidence-Based Decision Model

The confidence-based state machine is the core intelligence of the smart delivery lockbox, enabling adaptive, multi-modal authentication by accumulating evidence from multiple input sources and adjusting authentication requirements based on context. Rather than relying on a single authentication method, the system assigns confidence points to each verified input and compares the cumulative score against defined thresholds to drive state transitions. This approach mirrors how a human evaluates delivery legitimacy by weighing multiple pieces of evidence before making a decision.

Table 1 summarizes the confidence scoring model and state transition thresholds. All point values are assigned in increments of twenty-five to maintain a consistent and interpretable scoring framework. Person

and package detection each contribute twenty-five points, reflecting their role as foundational visual evidence of legitimate delivery intent. A person carrying a package strongly indicates a delivery is underway, as opportunistic theft typically involves taking an already-delivered package rather than arriving with a decoy. Company name verification also contributes twenty-five points. Although the company name is guessable, it provides meaningful corroborating evidence rather than serving as a standalone authentication factor. Order number verification, by voice or keypad, contributes seventy-five points because it provides significantly stronger authentication assurance. The order number is private information shared only between the retailer, homeowner, and delivery carrier, is printed on the package label, and cannot be reused once confirmed. A cumulative score of twenty-five points triggers the transition from Scanning to Verification.

The Open threshold adapts based on whether any packages are currently secured inside the lockbox, as tracked by the delivered orders helper variable. When the delivered orders helper contains no entries, indicating the box is empty, a threshold of seventy-five points is sufficient, permitting the combination of person detection, package detection, and company name verification to open the box without order number verification. This lower threshold is best understood as an availability and usability optimization rather than an authentication guarantee: because the vast majority of residential deliveries involve a single pending order, requiring full order-number verification on every delivery would impose unnecessary friction on delivery personnel for an asset, an empty box, with minimal value to protect. Since the box is empty, there is nothing to steal even if authentication is completed by an unauthorized individual, making the additional order number verification unnecessary in this context. The seventy-five-point threshold is therefore calibrated to open the box reliably for plausible delivery behavior, rather than to verify the deliverer's identity. When the delivered orders helper contains one or more entries, indicating a package is already secured inside the box, the threshold rises to one hundred points, making order number verification mandatory. This stricter requirement ensures that access to a box containing a secured package requires knowledge of a specific, valid, undelivered order number that is neither publicly available nor reusable, protecting previously delivered contents from unauthorized retrieval. Stronger authentication is therefore reserved for the state where it matters most: once a package is secured inside the box, genuine identity verification is enforced precisely when there is something of value to protect.

Table 1. Confidence Scoring Model

Input	Points	Condition
Person Detection	+25	Frigate AI Detection
Package Detection	+25	Frigate AI Detection
Company Name	+25	Whisper Voice Match
Voice Order number	+75	Whisper Voice Match
Keypad Order number	+75	ESP32-S3 Keypad Input
Verification Threshold	≥ 25	Enter Verification State
Open Threshold (box empty)	≥ 75	Enter Open State
Open Threshold (package inside)	≥ 100	Enter Open State

Email Parsing and Energy Management

Node-RED continuously monitors the homeowner's email inbox for incoming delivery confirmation emails. Upon detecting a qualifying message, Node-RED parses the email to extract the delivery carrier name, expected delivery date, and the last four digits of the order number, pushing these values into the delivery queue and delivery date queue maintained in Home Assistant helper variables. To ensure queue integrity across system restarts and new email arrivals, the Node-RED flow reads the live state of each queue directly from Home Assistant before every push operation, preventing previously stored entries from being

overwritten. Energy management is handled through a smart plug that keeps the system powered off when no deliveries are anticipated, triggering power-on only on days when a delivery is expected, ensuring the lockbox is ready while minimizing unnecessary energy consumption.

System Evaluation Methodology

System evaluation was conducted through structured feasibility testing combined with quantitative trial-based measurement, designed to assess the reliability of the confidence-based authentication framework across a range of delivery scenarios and environmental conditions. Testing was performed on a local virtual machine running Home Assistant, Frigate, OpenAI Whisper, Node-RED, and ESPHome. Physical testing was conducted at a residential front porch to simulate real-world deployment conditions, including ambient noise, natural lighting variability, and adversarial access attempts.

Ten structured test cases were designed using an orthogonal test design to ensure every system feature was exercised by at least one scenario. The test cases included single-delivery authentication via voice (TC1), keypad fallback after failed voice attempts (TC2), multi-delivery disambiguation using order number verification (TC3), theft detection triggered by weight decrease (TC4), the no-package-detected notification path (TC5), unauthorized access attempts (TC6), authentication under degraded audio and ambient noise (TC7), person and package detection under low-light conditions (TC8), email parsing accuracy (TC9), and smart plug energy management (TC10). Table 2 maps each system feature to its corresponding test cases.

Each scenario was repeated across multiple trials, with trial counts, success rates, and timing data recorded. Testers included both system-familiar participants and individuals unfamiliar with the system to evaluate whether speaker-guided prompts alone enabled successful authentication. Evaluation metrics included Whisper transcription success under varying noise conditions, Frigate detection accuracy under different lighting conditions, weight-sensing reliability across package sizes, and overall authentication completion rates. The system was evaluated both as an integrated pipeline and, where applicable, through individual component performance to support the reliability claims in the Results section.

Table 2. Orthogonal Feature Coverage by Test Case

System Feature Category	TC1	TC2	TC3	TC4	TC5	TC6	TC7	TC8	TC9	TC10
Computer vision detection (person/package)	X	X	X	X	X	X	X	X		
Voice authentication (company/order number)	X	X	X		X	X	X			
Keypad fallback authentication		X				X				
Weight sensing (placement/theft)	X	X	X	X	X					
Confidence scoring and state transitions	X	X	X	X	X	X	X	X		
Security and access control			X			X				
Email parsing and delivery queue			X						X	X
Smart plug energy management										X
Notifications and alerts				X	X					

Results, Findings, and Discussion

Feasibility Testing Results

System evaluation combined structured feasibility testing with quantitative trial-based measurement across ten test cases, each targeting a distinct operational path within the system. All primary test cases are demonstrated in a recorded video (Patel, 2026). Table 3 summarizes trial counts, success rates, and relevant timing data across all test cases.

Table 3. Quantitative Evaluation Results

Test Case (TC)	Metric	Trials	Result
TC1: Single delivery, voice authentication	Completion rate	20	80%
TC1: Single delivery, voice authentication	Avg. authentication time	16	155 sec
TC2: Keypad fallback after voice failure	Fallback success rate	5	100%
TC3: Multi-delivery disambiguation	Correct order match rate	5	100%
TC4: Theft detection	Detection rate	8	75%
TC4: Homeowner retrieval (false positive check)	False trigger rate	8	0%
TC5: No package detected, notification accuracy	Notification success rate	5	100%
TC6: Unauthorized access attempt	Block rate	10	100%
TC7: Voice authentication under ambient noise	Transcription success rate	15	66%
TC8: Person/package detection under low light	Detection accuracy	15	100%
TC9: Email parsing accuracy	Parse success rate	20	95%
TC10: Smart plug activation on delivery day	Correct activation rate	3	100%

Single Delivery, Successful Voice Authentication

In the baseline delivery scenario with one package queued, the system successfully transitioned from Scanning to Verification upon detecting a person and package via Frigate, accumulating fifty confidence points. The delivery person followed the speaker prompts, stated the correct company name, and the system transitioned to Open state upon reaching the seventy-five point threshold. The linear actuator and servo motor opened the box, the load cell confirmed package placement via a weight increase exceeding one thousand grams, and the system secured itself and transitioned to the Security state. The delivered orders helper was updated with the corresponding order number upon successful completion. Across 20 trials conducted with both familiar and unfamiliar testers, the scenario completed successfully in 80% of attempts, with an average authentication time of 155 seconds across the 16 successful trials. Performance differed notably by familiarity: testers already acquainted with the system completed all 8 of their trials successfully, with authentication times ranging from 84 to 98 seconds. Testers unfamiliar with the system completed 8 of 12 trials successfully, with the majority of failures concentrated in each tester's earliest attempts. One unfamiliar tester failed their first three trials because they did not understand the spoken instructions and were unable to complete authentication, but succeeded on subsequent attempts once familiar with the prompts; a second unfamiliar tester failed their first attempt for similar reasons before succeeding on all later trials. Successful unfamiliar-tester trials averaged notably longer authentication times, ranging from 145 to 160 seconds, compared to 84 to 98 seconds for familiar testers. These results suggest that while the speaker-guided prompts are ultimately learnable, first-time users may require more than one attempt to understand the expected interaction, and a proposed solution of automatically restating the prompt if no input is detected within 10 seconds may reduce this initial confusion. The full delivery interaction required no homeowner involvement at any stage in either group.

Single Delivery, Voice Failure and Keypad Fallback

To test the keypad fallback mechanism under controlled conditions, the order number was intentionally misstated during voice authentication across two consecutive attempts, ensuring Whisper could not produce a usable transcription and triggering the fallback condition. The system automatically fell back to the matrix keypad, through which the correct digits were entered. Across 5 trials using this method, the keypad fallback succeeded in 100% of attempts, confirming that the fallback mechanism reliably activates and completes authentication when voice-based order number verification fails.

Multiple Deliveries, Same-Carrier Disambiguation

With two deliveries from the same carrier queued simultaneously, the system raises the Open threshold to one hundred confidence points, requiring the delivery person to provide the last four digits of the specific order number being delivered. Across 5 trials, the correct order number was matched against the delivery queue in 100% of attempts, with the corresponding entry removed and the remaining delivery staying in the queue unaffected. This test case validated the queue design and confirmed that the system correctly disambiguates concurrent deliveries from the same carrier without homeowner input.

Theft Detection

With a package confirmed inside the box via the load cell, a simulated theft was performed by removing the package while the system remained in the Open state. Across 8 trials, the load cell correctly triggered the Theft state within a 30-second detection window in 75% of attempts. The remaining trials were classified as undetected because the weight decrease was not registered within this window, though the exact delay in these cases was not recorded; further investigation is needed to determine whether the delay stemmed from sensor sensitivity, package placement, or removal speed. In all successful detections, the ESP32-S3 immediately closed and locked the box, the speaker announced a security alert three times, Frigate saved a timestamped video clip surrounding the event, and Home Assistant sent a push notification to the homeowner. After five minutes, the system returned to Scanning as additional deliveries remained in the queue. The homeowner retrieval toggle was separately verified across 8 trials to bypass the state machine correctly, with a 0% false trigger rate, confirming that normal package retrieval reliably does not trigger the Theft state.

No Package Detected After Open

In this scenario, authentication was completed successfully in single-delivery mode without order number verification, and the box opened. No package was placed inside within the Open state window. The load cell did not detect a sufficient weight increase, and Home Assistant sent a notification to the homeowner alerting them that the box had been opened but no package was detected. Across 5 trials, this notification fired correctly in 100% of cases, highlighting the relevance of order number verification as an additional safeguard in ambiguous delivery interactions.

Unauthorized Access Attempt

To assess the system's resistance to unauthorized access, 10 trials were conducted in which an individual without knowledge of the correct company name or order number attempted to authenticate. The system blocked unauthorized access in 100% of attempts, with no trial resulting in an unauthorized transition to the Open state when a package was present inside the box.

Authentication Under Ambient Noise Conditions

To evaluate the robustness of voice-based authentication outside controlled conditions, 15 trials were conducted under ambient outdoor noise, including wind, construction, and background conversation. Whisper-based authentication succeeded in 66% of attempts under these conditions, compared to the 80% baseline completion rate reported above. Failures clustered around a single, specific mechanism: under

sufficiently loud ambient noise, the microphone failed to register any input at all from the speaker, preventing the wake word and subsequent response from being captured, rather than producing an inaccurate transcription. This distinction is meaningful, as it points to a microphone sensitivity and input-detection issue rather than a limitation of Whisper's transcription accuracy itself. A proposed mitigation is to automatically fall back to the matrix keypad if no audio input is detected within 20 seconds, extending the existing keypad fallback mechanism to cover input-detection failures in addition to transcription failures.

Detection Accuracy Under Low-Light Conditions

To evaluate Frigate's reliability outside daylight conditions, 15 trials were conducted across evening and night conditions. Person and package detection succeeded in 100% of trials under both lighting conditions. This result is attributable to the IP camera's built-in night vision capability, which maintained reliable detection performance without requiring additional porch lighting or camera setting adjustments.

Authentication Accuracy and System Behavior

Authentication Framework Reliability (RQ1)

Across all test cases, the confidence-based authentication framework reliably differentiated authorized delivery interactions from unauthorized or incomplete ones. As detailed in the Unauthorized Access Attempt and theft detection results above, no trial resulted in an unauthorized transition to the Open state when a package was already secured inside the box, confirming that the adaptive threshold mechanism enforces stricter authentication precisely when the risk of unauthorized retrieval is highest. While voice-based authentication alone showed reduced reliability under degraded conditions, as discussed above, the keypad fallback path preserved the system's overall security guarantee in every tested scenario, indicating that the three-stage authentication sequence provides sufficient redundancy without compromising security.

Autonomous State Machine Performance (RQ2)

The five-state machine managed the full delivery lifecycle autonomously across all tested scenarios, including normal deliveries, authentication failures, theft events, and edge cases such as no package detection. State transitions were driven exclusively by sensor inputs and confidence score thresholds, with no homeowner intervention required during any delivery interaction. The homeowner's only point of engagement was receiving push notifications and, in the theft scenario, reviewing Frigate footage, both passive interactions. As discussed in the TC1 results above, familiar testers completed authentication reliably on every attempt, while unfamiliar testers required repeated exposure before achieving similar reliability, suggesting that the underlying state machine performs autonomously as designed, though the guided interaction layer may benefit from additional onboarding support to improve first-time success rates in real-world deployment.

IoT Integration and Energy Efficiency (RQ3)

The Node-RED email parsing pipeline extracted carrier names, expected delivery dates, and order numbers from incoming confirmation emails with a parse success rate of 95% across 20 test emails, populating the dual queue system in Home Assistant without manual homeowner input. The smart plug correctly activated the lockbox system on days when deliveries were anticipated in 100% of 3 trials, and kept the system powered off otherwise, ensuring the hardware consumed no standby energy on non-delivery days. The dual queue design correctly maintained queue integrity across multiple deliveries, system restarts, and concurrent same-carrier orders, confirming that anticipatory email parsing combined with smart power management can meaningfully improve the energy efficiency and autonomous preparedness of a residential delivery security system.

Limits of the Defense Model (RQ4)

The system closes off the most common porch piracy attack vector, opportunistic theft of unattended packages, by securing the package inside a locked enclosure immediately upon confirmed delivery; the package remains physically unattended by a person, but access requires a specific, undelivered order number that is not publicly available and cannot be reused. This same requirement mitigates impersonation: an empty box has nothing to steal even if authentication succeeds, while a box already holding a package requires knowledge of that specific order number to access. The primary remaining authentication-layer vulnerability is email compromise, since a moderately motivated attacker with access to the homeowner's inbox could extract a pending order number; protecting the homeowner's email account falls outside this system's protection model. The obstruction-based attack, in which an attacker wedges an object inside the box to gain leverage for forced entry, is mitigated by the linear actuator's 1,000-Newton force rating, which makes leverage-based forced opening effectively infeasible, with Frigate's continuous recording providing video evidence of any attempt. The remaining physical vulnerabilities are the box material itself giving way under sufficient force, and the detection latency suggested by the 75% theft detection rate within the tested 30-second window, both of which fall outside the scope of force-based mitigations.

Implications

This work demonstrates that open-source, locally hosted components can be composed into a capable, sophisticated residential security system without proprietary hardware or cloud infrastructure. The confidence-based state machine introduced here represents a generalizable framework applicable to any residential access control scenario requiring adaptive, multi-modal authentication, and the increasing availability of affordable IoT hardware and locally deployable AI models will make systems of this kind accessible to everyday homeowners without specialized technical backgrounds. This accessibility extends to cost: the complete hardware bill of materials, including the microcontroller, sensors, actuator, camera, and audio components, totaled under \$300, well below comparable proprietary smart-lock ecosystems.

Because every authentication input except keypad entry is processed through AI-based perception (Frigate's object detection and Whisper's speech-to-text transcription) running entirely on local hardware, no audio, video, or authentication data is transmitted to external servers at any point, preserving homeowner privacy and eliminating exposure to third-party data handling practices associated with cloud-dependent alternatives. Ethically, the system saves video footage only in response to a detected theft event, limiting surveillance to security-relevant interactions rather than continuous monitoring of public space. This reflects a deliberate effort to balance security benefits against the privacy interests of delivery personnel, who are recorded and authenticated without the ongoing opt-in consent typically associated with biometric systems; future iterations should consider explicit disclosure mechanisms, such as posted signage or verbal notice during the Verification state, even though the data itself never leaves the homeowner's local infrastructure.

Limitations

This section addresses the current constraints of the smart delivery lockbox prototype, including queue capacity and failure handling, voice recognition and deployment considerations, and testing scope. As a prototype, the system is designed to demonstrate architectural viability and the feasibility of confidence-based, multi-modal authentication rather than serve as a finished consumer product. While the system successfully performs all described operations across its five operational states, it has not undergone hardened production deployment or extended real-world validation beyond structured feasibility and quantitative testing, and the following limitations should be interpreted accordingly.

Queue Capacity and Failure Handling

Home Assistant input text helper entities enforce a 255-character limit on stored values. Under the current dual queue implementation, this constrains the system to approximately five to six simultaneous pending deliveries before the queue reaches capacity. While this is sufficient for typical residential delivery volumes, this represents an architectural constraint tied to Home Assistant's native data model. The current implementation has not been tested for its behavior under two specific failure conditions: what occurs when an incoming delivery confirmation email arrives after the queue has reached capacity, and what occurs when an email fails to parse correctly due to an unrecognized carrier format. Neither condition produced a defined fallback behavior in the current prototype, and both represent meaningful gaps to address prior to production deployment, since silent failure in either case could result in a missed delivery being assigned no entry in the queue.

Voice Recognition and Deployment Considerations

OpenAI Whisper's transcription accuracy is dependent on audio input quality, which may degrade under challenging outdoor conditions such as wind noise, crowd noise, or heavy accents; the three-stage authentication sequence mitigates this risk through keypad fallback, but extended outdoor deployment may reveal edge cases not encountered during testing. Outdoor installation more broadly exposes the microphone, camera, and weight sensor to environmental conditions including wind, precipitation, and temperature variation, each of which may affect system performance, while carrier-specific email formats vary and may require Node-RED flow adjustments as retailers update their notification templates. Delivery personnel unfamiliar with voice-authenticated lockboxes may require brief onboarding, though feasibility testing suggests that the speaker-guided prompts are sufficient for first-time users to complete authentication without prior instruction.

Testing Scope

Feasibility and quantitative testing were conducted by the development team across a defined set of structured scenarios in a controlled residential setting. While testing confirmed correct system behavior across all primary operational paths and provided quantitative reliability metrics across ten test cases, the evaluation did not encompass extended longitudinal deployment, a statistically significant sample of delivery personnel, or the full range of environmental and hardware conditions that a production deployment would encounter. Results should therefore be interpreted as a demonstration of feasibility rather than a comprehensive performance benchmark.

Conclusion and Future Work

Conclusion

This paper presented the design, implementation, and evaluation of an AI-driven smart delivery lockbox that authenticates delivery personnel and secures packages through a confidence-based, multi-modal authentication framework. Unlike existing commercial solutions, which are typically carrier-specific, cloud-dependent, or reliant on homeowner involvement, the proposed system is carrier-agnostic, fully local, and fully autonomous, integrating computer vision, speech-to-text processing, weight sensing, and keypad fallback authentication, all coordinated locally over MQTT through Home Assistant. A dual queue system populated through Node-RED email parsing manages multiple concurrent deliveries without homeowner intervention, while smart plug integration ensures no standby energy is consumed on non-delivery days. Structured feasibility testing and quantitative trials across ten test cases confirmed correct behavior across all five operational states, including authentication, theft detection, unauthorized access resistance, and performance under degraded environmental conditions, demonstrating that open-source components can be

composed into a capable, privacy-preserving residential security system without proprietary hardware or cloud services.

Future Work

Several directions for future development were identified based on the prototype's current capabilities and limitations. The most significant planned addition is homeowner biometric authentication using face and voice recognition, completing the delivery lifecycle as a fully autonomous end-to-end system without requiring homeowner interaction with the Home Assistant interface. Expanded carrier and retailer support is also planned, updating the Node-RED email parsing flow to handle a wider range of confirmation formats and defining fallback behavior for queue capacity limits and parsing failures. Extended real-world testing with actual delivery personnel across varying conditions, carrier types, and delivery volumes represents the most important next validation step, alongside outdoor hardening of the physical enclosure to address weather and sustained exposure.

A longer-term direction is replacing the current rule-based logic with a purpose-built AI model trained specifically for this system, allowing sensor inputs, voice responses, and visual data to be weighed together more flexibly than fixed confidence thresholds allow. The existing feasibility and quantitative test data could serve as a foundation for building a labeled training dataset grounded in real system behavior, moving the system from hand-crafted logic toward a more adaptive, generalizable decision-making core and from a functional prototype toward a deployable residential delivery security solution.

Acknowledgements

The authors acknowledge the assistance of Claude (Anthropic) in drafting and refining this manuscript. All content was reviewed and approved by the authors, who take full responsibility for the work presented. The authors acknowledge Dr. Myungjae Kwak for funding support for some components and Brian Ye, Shirui Pan, and Daniel Dong for assistance in project tasks.

References

- Frew, P. & Petrino, G.. (2025). 2025 Package Theft Report and Statistics. Retrieved on 04/15/2026 from <https://www.security.org/package-theft/annual-report/>
- Frew, P. & Petrino, G.. (2025). 2025 Package Theft Report and Statistics. Retrieved on 04/15/2026 from <https://www.security.org/package-theft/annual-report/>
- Frigate. (n.d.). Introduction. Retrieved on 05/02/2026 from <https://docs.frigate.video/>
- Home Assistant. (n.d.). Home Assistant Documentation. Retrieved on 05/12/2026 from <https://www.home-assistant.io/docs/>
- Hussain, M. (2024). YOLOv5, YOLOv8 and YOLOv10: The Go-To Detectors for Real-Time Vision. arXiv preprint arXiv:2407.02988. Retrieved on 05/02/2026 from <https://arxiv.org/abs/2407.02988>
- Lombardo Homes. (2024). Package Theft Statistics 2024: Porch Pirates Study. Retrieved on 04/15/2026 from <https://lombardohomes.com/porch-pirates-and-stolen-packages/>

- Luís, A. F. S., Martins, G. M. C., Caldeira, J. M. L. P. & Soares, V. N. G. J. (2022). Smart Lockers: Approaches, Challenges and Opportunities. *International Journal of Engineering and Advanced Technology*, 11(3). Retrieved on 04/15/2026 from <https://www.ijeat.org/wp-content/uploads/papers/v11i3/C33740211322.pdf>
- Node-RED. (n.d.). Node-RED: Low-code programming for event-driven applications. Retrieved on 05/12/2026 from <https://nodered.org/docs/>
- OASIS. (2014). MQTT Version 3.1.1. OASIS Standard. Retrieved on 05/12/2026 from <https://docs.oasis-open.org/mqtt/mqtt/v3.1.1/os/mqtt-v3.1.1-os.html>
- Ooi, J. Z., Ling, Y. X., Ng, Y. H., Yap, J. S. & Yong, J. Y. (2021). Smart Modular Parcel Locker System Using Internet of Things (IoT). Retrieved on 04/15/2026 from <https://www.researchgate.net/publication/356668049>
- Patel, S. (2026). System Demonstration and Test Cases. Retrieved on 05/14/2026 from https://youtu.be/c_MYeUWRegM
- Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C. & Sutskever, I. (2023). Robust Speech Recognition via Large-Scale Weak Supervision. *Proceedings of the 40th International Conference on Machine Learning*, 202, 28492–28518. Retrieved on 04/15/2026 from <https://proceedings.mlr.press/v202/radford23a.html>
- Red Stag Fulfillment. (2026, January 20). How Many Packages Are Stolen Annually? (2026 Statistics). Retrieved on 05/02/2026 from <https://redstagfulfillment.com/how-many-packages-are-stolen-annually/>
- Saleem, J., Raza, U., Hammoudeh, M. & Holderbaum, W. (2025). Machine Learning-Enhanced Attribute-Based Authentication for Secure IoT Access Control. *Sensors*, 25(9), 2779. Retrieved on 04/15/2026 from <https://doi.org/10.3390/s25092779>
- Terven, J. R. & Cordova-Esparza, D. M. (2023). A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *arXiv preprint arXiv:2304.00501*. Retrieved on 05/02/2026 from <https://arxiv.org/abs/2304.00501>
- Waked, R., Mannan, M. & Youssef, A. (2023). Multimodal User Authentication in Smart Environments. *arXiv preprint arXiv:2305.03699*. Retrieved on 04/15/2026 from <https://arxiv.org/abs/2305.03699>