Smart Cyber Security Intrusion Detection & Prevention System Using AI (Securityhub)

Proff. Jaypal Gedam[1], Mr. Sujal Ramteke[2], Mr. Atharva Sable[3], Mr. Swayam Wankhede[4], Mr. Aman Hirole[5]
[1, 2, 3, 4, 5] Dept of Electronics and Communication Engineering
[1, 2, 3, 4, 5] Jhulelal Institute of Technology

Abstract- *This project examines the combination of network-level security and physical biometric verification within a unified, hardware-accelerated edge environment. Conventional security setups frequently struggle with operational slowdowns, primarily caused by high-latency cloud dependencies and the large bandwidth needs of remote data handling. To fix these issues, we present a localized security hub designed for the Raspberry Pi 5 and the Hailo-8L AI accelerator. This setup allows for the simultaneous running of deep packet inspection through Suricata and real-time facial recognition using the ArcFace model.*

By integrating Keycloak as a centralized identity provider, the system ensures a unified security environment where network access rules are assigned based on biometrically verified roles. This research shows that enterprise-level security—capable of managing both digital threats and physical breaches—is possible on a low-power, single-board device. Our results show that using the dedicated PCIe Gen 3 interface on raspberry pi 5 for AI tasks significantly improves speed while protecting data privacy. Ultimately, this method offers a sturdy, small scale budget-friendly alternative to traditional isolated and expensive security tools, creating a reliable Zero-Trust environment at the network edge.

Keywords: Edge Computing, Raspberry Pi 5, Hailo-8L, Suricata, Keycloak, Face Recognition, Intrusion Detection System (IDS), IoT Security, ArcFace.

I. INTRODUCTION

The modern security sector is currently witnessing a significant shift as the boundaries between physical perimeters and digital networks continue to vanish. Historically, these two domains were managed through isolated infrastructures, where physical access control and network intrusion detection were operated as separate entities [1], [2]. However, the rise of the Internet of Things (IoT) and the increasing complexity of "Zero-Day" attacks have made these isolated models oudated. A physical breach in a restricted area, such as a server room, can now lead to immediate digital compromise, just as a network-based attack can be used to disable physical surveillance systems [19], [20].

Traditional security architectures have long depended on cloud-centric processing to handle the heavy computational requirements of deep packet inspection (DPI) and high-accuracy biometric recognition [3], [4]. While being powerful, this reliance introduces critical drawbacks, including high latency in emergency alert systems, excessive bandwidth consumption, and severe privacy risks associated with transmitting sensitive biometric data over public networks [7]. These challenges have pushed the transition toward "Edge Intelligence," where data is processed as close to the source as possible to ensure both speed and data sovereignty are maintained [8], [21].

This research explores the implementation of a unified security hub localized on a single-board computer (SBC). At the core of the proposed system is the Raspberry Pi 5, which provides the general-purpose compute power required to run the Suricata intrusion detection engine [10]. To overcome the hardware limitations of standard SBCs, the system utilizes the official raspberry pi's AI kit which includes Hailo-8L AI accelerator to offload the mathematical burden of the ArcFace biometric model [12], [14]. By leveraging the dedicated PCIe Gen 3 interface of the Pi 5, the hub achieves high-throughput inference that was previously only possible on desktop-class GPUs [13].

Furthermore, the integration of Keycloak as a centralized identity backbone allows the hub to link the gap between physical and digital security [11]. By mapping network-level permissions to biometrically verified user roles, the system implements a localized Zero-Trust Architecture (ZTA) [21]. This paper reviews the synergy between these hardware and software layers, demonstrating that a robust, cost-effective,small scale and privacy-focused security environment is now achievable at the network edge.

II. LITERATURE SURVEY

The development of localized security hubs is supported by decades of research in network defense,

computer vision, and hardware optimization. The following points represent the key research that informs this project:

2.1 Evolution of Network Defense:

Early intrusion detection systems relied on signature-matching, which struggled with high-speed traffic and new attack types [1], [2]. Paxson [3] and Roesch [4] introduced stateful protocol analysis and lightweight detection, providing the foundation for modern multi-threaded engines like Suricata [10].

2.2 Deep Learning for Traffic Analysis:

Recent research has pushed for the use of deep neural networks to identify zero-day threats. Shone et al. [6] proved that deep learning achieves higher accuracy than traditional algorithms, though Anderson and McGrew [5] noted the difficulty of inspecting encrypted malware traffic without high computational overhead.

2.3 The Rise of Edge Intelligence:

To overcome the latency and privacy issues of cloud-based AI, Xu et al. [7] and McMahan et al. [8] proposed moving intelligence to the network edge. This "Edge Intelligence" paradigm ensures that sensitive data stays on-site, which is crucial for maintaining data sovereignty [21].

2.4 Advances in Facial Recognition:

Biometric verification has advanced from simple image matching to deep feature extraction. The MTCNN framework [15] standardized multi-task face detection, while the ArcFace model [14] introduced angular margin loss to significantly improve the precision of identity verification.

2.5 Hardware-Software Co-Optimization:

For resource-limited devices like the Raspberry Pi, research has focused on efficiency. Sandler et al. [16] developed MobileNet architectures for mobile processors, while Wang et al. [9] explored how specialized NPUs can offload mathematical tasks from the CPU to improve real-time performance [12].

2.6 High-Speed Interconnects on SBCs:

The feasibility of these systems depends on data throughput. Casneuf et al. [13] evaluated PCIe Gen 3 performance on single-board computers, proving that modern interfaces can handle the simultaneous data streams required for both AI and network monitoring.

2.7 Wireless Security and Standards:

Gast [17] and broader surveys [18] have highlighted the vulnerabilities in 802.11 standards. This research emphasizes the need for an integrated system that monitors both physical presence and wireless traffic to achieve a complete Zero-Trust environment [21].

III. METHODOLOGY

The methodology describes the practical implementation of a dual-layered security hub, designed to operate at the network edge with minimal latency. The approach is structured to validate the feasibility of running high-throughput packet inspection alongside real-time AI inference on a singular hardware footprint.

3.1 Hardware Requirements:

The hardware selection was determined by the need for a balance between low power consumption and high computational density.

(a) Primary Computing Unit: The Raspberry Pi 5 (8GB RAM) serves as the core processor, chosen for its improved ARM Cortex-A76 architecture and dedicated PCIe 2.0/3.0 support [13].

(b) AI Acceleration: A Hailo-8L M.2 module (13 TOPS) or official raspberry pi's AI kit HAT is integrated via a PCIe HAT to offload the mathematical burden of the ArcFace model from the CPU [12].

(c) Network Interface: A dual-interface configuration is utilized, involving the onboard Gigabit Ethernet for LAN traffic and a high-gain antenna with dual bandwidth support (2.4GHz/5GHz) for wireless monitoring [17].

(d) Thermal Management: Official raspberry pi's active PWM-controlled cooler is required to maintain stability during sustained AI and IDS workloads.

3.2 Software and Tools:

The software stack is built on open-source frameworks optimized for the ARM64 architecture:

(a) Operating System: A lightweight Debian-based distribution (Raspberry Pi OS Lite 64-bit) to maximize available system resources.

(b) Intrusion Detection: Suricata 7.0+ is deployed in AF_PACKET mode for efficient, multi-threaded traffic analysis [10].
(c) Identity Management: Keycloak is utilized as the OIDC-compliant identity provider to manage role-based access control [11].

(d) AI Framework: The Hailo Software Suite (HailoRT) is employed to run the ArcFace model in an optimized HEF (Hailo Executable Format) [14].

(e) Data Layer: A Redis in-memory database is used for the sub-millisecond storage and retrieval of facial embeddings.

3.3 System Workflow (How the Project Works):

The operational logic of the hub follows a three-stage sequential pipeline:

(a) Traffic & Vision Capture: The system simultaneously intercepts network packets via the Ethernet bridge and receives an RTSP video stream from a localized IP camera or CCTV camera through rstp link.

(b) Edge Processing: Suricata inspects packets for digital threats [1], while the Hailo-8L processes the video frames to identify individuals using the ArcFace algorithm.

(c)Identity Correlation: The "Decision Engine" compares the detected face against the Keycloak database. If an unauthorized individual is detected, the system [triggers] an alert and instructs Suricata to drop all network traffic from the intruder's associated IP address, enforcing a localized Zero-Trust policy [21].

IV. IMPLEMENTATION

The implementation phase focuses on the seamless integration of three distinct operational pipelines: the Network Telemetry Pipeline, the Biometric Inference Pipeline, and the Identity-Aware Alerting Pipeline. The primary engineering challenge was managing these services on a single-board footprint without causing resource contention or thermal instability.

4.1 Network Intelligence Integration:

The digital defense layer was [deployed] by configuring Suricata as a bridge-mode IDS/IPS [10]. To achieve "Identity-Aware" filtering, we utilized a custom Lua scripting interface within Suricata. All incoming packets are intercepted via the AF_PACKET interface, allowing for high-speed inspection [3]. The Lua script performs a real-time lookup in the Redis database to check if the source IP address corresponds to an authenticated user role from Keycloak [11]. If a mismatch is detected—such as a "Guest" attempting to access restricted server ports—the system dynamically blocks the traffic at the kernel level.

4.2 Biometric Vision Pipeline:

The physical security layer uses the Hailo-8L NPU to handle the computational load of deep learning inference [12]. The implementation follows a "Capture-Extract-Match" logic:

(a) Frame Acquisition: The system pulls a high-definition 1080p stream from the IP or CCTV camera using the Real-Time Streaming Protocol (RTSP).

(b) Hardware Acceleration: Detected face crops are sent to the Hailo-8L via the PCIe Gen 3 interface. The NPU runs the ArcFace model to generate a 128-dimensional embedding [14].

(c) Vector Matching: This embedding is compared against authorized templates in Redis using Cosine Similarity. This localized processing ensures that biometric data never leaves the device, maintaining strict data sovereignty [21].

4.3 Identity and Alert Orchestration:

The final layer of implementation is the Centralized Alert API, which serves as the system's decision-maker. When the vision pipeline identifies an "Unknown" subject in a restricted zone, it sends a signed request to the API. The API verifies the request using a JSON Web Token (JWT) issued by Keycloak [11]. Once authenticated, the system pushes an encrypted intruder alert to a secure Telegram bot or a custom made android app with image stamps and simultaneously updates the live security dashboard via WebSockets.

V. RESULTS AND DISCUSSION

The performance of the integrated security hub was tested based on two primary metrics: the latency of the biometric inference pipeline and the packet-processing efficiency of the NIDS layer. The results demonstrate that hardware-accelerated edge intelligence significantly outperforms standard CPU-based processing.

Using the Hailo-8L AI accelerator, the ArcFace model achieved an average inference time of [12ms to 15ms] per frame. This is a substantial improvement over the standalone Raspberry Pi 5 CPU, which averaged over 180ms

per frame. The localized matching against the Redis database was found to be nearly instantaneous, ensuring that physical access alerts are triggered in under 0.5 seconds from the moment of detection.

During sustained testing, Suricata was assigned with monitoring a 1 Gbps link while simultaneously running the vision pipeline. The system maintained a zero-packet-loss rate at traffic loads up to 650 Mbps. While deep packet inspection (DPI) increased CPU utilization to 65%, the offloading of AI tasks to the NPU ensured that the kernel remained responsive. The detection accuracy for known CVE-based signatures remained consistent at 94.2%, proving that the integrated nature of the hub does not degrade digital security standards [10], [13].

The most notable result is the successful synchronization between physical identity and network access. When an unauthorized user was identified by the ArcFace pipeline, the system executed a firewall rule update via the Suricata API in less than 100ms. This real-time response confirms the effectiveness of the proposed Zero-Trust framework, proving that hardware-software co-optimization is a viable approch for decentralized security appliances [21].

VI. APPLICATIONS

1.Critical Infrastructure Protection: The hub can be deployed at the edge of SCADA networks in power plants or water treatment facilities. By requiring biometric presence to unlock specific network ports, it prevents remote attackers from manipulating physical machinery settings [19], [21].

2.Data Center Security: In high-density server environments, the system can detect "tailgating" incidents. If an unauthorized person follows a staff member into a "Hot Zone," the hub instantly alerts security and isolates the network traffic of that specific room to prevent data sniffing [7].

3.Secure Research Labs: For laboratories handling sensitive intellectual property, the system ensures that only verified researchers can access local high-speed file servers. This links physical identity directly to digital data access, fulfilling Zero-Trust requirements [11].

4.Air-Gapped Defense Units: Because the entire stack operates without cloud dependencies, it is ideal for mobile command centers or remote outposts. It provides a portable, low-power security solution that preserves data sovereignty in the field [8].

5.Smart Corporate Offices: The hub can manage smart office resources by dynamically adjusting network bandwidth and physical access based on the verified roles of the occupants, thereby enhancing both security and utility usage [18].

6.Automated Retail Surveillance: Beyond security, the hub can be used in high-end retail to identify VIP customers while simultaneously monitoring the store's Wi-Fi network for potential Point-of-Sale (POS) attacks [17].

7.Remote ATM Booths: By deploying the hub in standalone kiosks, banks can monitor for physical tampering while protecting the ATM's network connection from "black box" attacks using real-time packet inspection [5], [20].

VI. LIMITATIONS

While the integration of the Raspberry Pi 5 and the Hailo-8L offers a high-performance solution for edge security, several technical constraints remain noticeable. One primary challenge is the system's reliance on ambient environmental conditions; the accuracy of the ArcFace biometric engine can decrease in low-light scenarios or when subjects utilize heavy facial occlusions, requiring the use of specialized infrared sensors for consistent performance [14]. Additionally, while the hub manages gigabit-level traffic efficiently, it is not currently scaled for backbone-level monitoring in large-scale data centers where throughput exceeds 10 Gbps, as the interrupt handling of a single-board computer becomes a hardware bottleneck [13]. Thermal stability also remains a challenge, as sustained high-load AI and network inspection tasks generate significant heat, making active cooling mandatory for long-term reliability. Furthermore, the decentralized nature of the hub lacks a global policy management engine, meaning that updates to biometric databases or Suricata rulesets must be managed on a per-device basis, which may hinder large-scale campus deployments without further orchestration software [10], [21].

VII. CONCLUSION

This research has proved the feasibility of merging enterprise-grade network security with real-time biometric authentication on a localized edge footprint. By utilizing the Raspberry Pi 5 and the Hailo-8L AI accelerator, we have developed a system that effectively overcomes the latency and privacy bottlenecks inherent in cloud-based security models [7], [12]. The successful integration of Suricata for deep packet inspection and ArcFace for high-accuracy vision proves that hardware-software co-optimization can handle complex security workloads without compromising

performance [13], [14]. Furthermore, the use of Keycloak to bridge physical and digital identity establishes a robust foundation for localized Zero-Trust Architectures [11], [21]

VIII. FUTURE WORK

Planning ahead, future development will focus on enhancing the hub's scalability and intelligence. One key area of exploration involves the adoption of Federated Learning, which would allow a fleet of hubs to collectively improve their threat detection models without sharing raw, sensitive data [8]. Additionally, we aim to expand the system's wireless defense capabilities by integrating automated countermeasures against deauthentication attacks and rogue access points [18]. Finally, moving the core logic to a custom-designed PCB could further improve thermal efficiency and allow for more compact deployments in industrial environments. Ultimately, this integrated approach represents a significant step towards a more resilient, privacy-centric, and autonomous security ecosystem at the network edge.

REFERENCES

[1] Scarfone, K., & Mell, P. (2007). Guide to Intrusion Detection and Prevention Systems (IDPS). NIST Special Publication 800-94.
[2] Northcutt, S., & Novak, J. (2002). Network Intrusion Detection. New Riders Publishing.
[3] Paxson, V. (1999). "Bro: A System for Detecting Network Intruders in Real-Time." Computer Networks.
[4] Roesch, M. (1999). "Snort - Lightweight Intrusion Detection for Networks." Proc 13th USENIX Conf on System administration (LISA).
[5] Anderson, B., & McGrew, D. (2022). "Identifying Encrypted Malware Traffic with Contextual Flow Data." IEEE Transactions on Network and Service Management.
[6] Shone, N., Ngoc, T. N., Phai, V. D., & Shi, Q. (2018). "A Deep Learning Approach to Network Intrusion Detection." IEEE Trans on Emerging Topics in Computational Intelligence.
[7] Xu, R., et al. (2020). "Edge Intelligence for IoT: A Survey." IEEE Internet of Things Journal.
[8] McMahan, B., et al. (2017). “Communication-Efficient Learning of Deep Networks from Decentralized Data.” AISTATS.
[9] Wang, Y., et al. (2023). "Hardware-Software Co-Optimization for Deep Learning on Edge Devices." ACM Computing Surveys.
[10] Suricata IDS (2025). Suricata User Manual v7.0.x. Open Information Security Foundation (OISF). Suricata.io.
[11] Keycloak (2026). Keycloak: Open Source Identity and Access Management. Keycloak.org.
[12] Hailo (2024). "Hailo-8L: Performance Benchmarks for Edge AI Applications." Technical Documentation. Hailo.ai.
[13] Casneuf, L., et al. (2024). "Evaluating PCIe Gen 3 Throughput for AI Accelerators on Single-Board Computers." Journal of Real-Time Image Processing.
[14] Deng, J., et al. (2019). "ArcFace: Additive Angular Margin Loss for Deep Face Recognition." IEEE/CVF CVPR.
[15] Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). “Joint Face Detection and Alignment using MTCNN.” IEEE Signal Processing Letters.
[16] Sandler, M., et al. (2018). "MobileNetV2: Inverted Residuals and Linear Bottlenecks." IEEE/CVF CVPR.
[17] Gast, M. S. (2012). 802.11 Wireless Networks: The Definitive Guide. 2nd ed. O'Reilly Media.
[18] "Wireless Intrusion Detection Systems: A Survey." (2019). IEEE Communications Surveys & Tutorials.
[19] Cybersecurity and Infrastructure Security Agency (CISA). Best Practices for Intrusion Detection and Response.
[20] ENISA (2025). Threat Landscape Report. Annual Publication.
[21] Rose, S., et al. (2020). "Zero Trust Architecture." NIST Special Publication 800-207.