PIE

PATENT INFORMATION EXTRACTION FORM

TOPIC : PenBox-DMAS: A Distributed Multi-Agent Security Appliance for Autonomous Vulnerability Assessment and Remediation at the Edge

SCHOOL: Chitkara University, Chandigarh-Patiala National Highway (NH-64), Village Jansla, Tehsil Rajpura, Distt. Patiala, Punjab 140401

INVENTOR'S DETAIL WHO UPLOADED THE PATENT ON CHALKPAD

NAME:Vedant Sareen

ROLL NUMBER: 2310991318 MOBILE NO: +91 7087603933

MAIL ID (UNIVERSITY) – vedant1318.be23@chitkara.edu.in

MAIL ID (PERSONAL) - **securecybernetics@gmail.com**

OTHER INVENTORS DETAIL

NAME	EMP CODE/ROLL	ADDRESS WITH EMAIL AND MOBILE NO.
Dr. Himanshi Babbar	**CET1001727**	**Chitkara University, Punjab, India; Himanshi.babbar@chitkara.edu.in; 8557008265**

Answer the following questions in brief: (use extra sheets if needed, submit pics/videos etc. which can help us to understand the invention better)

1. **What was the problem?**

   The present IoT and edge network security systems experience three fundamental architectural problems. First, standalone scanning appliances operate with fixed vulnerability databases and cannot reason about previously uncharacterised threat patterns once deployed. The device becomes completely blind to novel exploitation chains which emerge outside the scanner's signature set. The security assessment system operates in a single processing unit because it needs to complete its host discovery and exploitation modelling and remediation process sequentially which causes the reconnaissance engine to wait for the reasoning engine to finish its result analysis. The assessment process takes three times longer to complete because of the serial bottleneck which prevents parallel processing. The existing edge-deployed tools create a system failure risk because all components operate as single points of failure: if the board loses power, overheats, or encounters a kernel panic during a multi-hour scan, the entire session state is lost with no recovery mechanism. The cloud-hosted solutions which include AWS Security Agent and CHATIOT remove hardware limitations, but they create unacceptable network latency and they expose raw packet captures to third-party servers and they require constant internet access which all three become unacceptable for air-gapped industrial and military and critical infrastructure environments.

   Additionally, existing tools lack a hardware-native EDR function— basically none of the current physical setups combine autonomous endpoint detection with full-stack AI-driven pentesting, plus that human-expert review loop with a Designated Review Server, all inside a single deployable node. PenBox-DMAS tries to address this gap as a Hardware-Integrated EDR appliance.

2. **How did you solve it (Inventive Step)?**

   The solution of PenBox-DMAS converts the complete security technology into multiple distinct processing units which work together as a unified system. The central system uses a Mini-ITX motherboard with high computing power which includes an x86 processor dedicated graphics processing unit and 64 gigabytes of random access memory to function as its primary planner and memory component. The system operates three to four Raspberry Pi 5 single-board computers which function as dedicated worker agents that continuously perform specific security duties which include reconnaissance and network mapping and exploitation validation and privilege escalation modelling and defensive hardening and patch deployment and adversarial reinforcement learning for predicting new multi-stage attack methods. The system communicates with a Designated Review Server (DRS), which is a controlled internal server where a senior certified penetration tester reviews AI generated findings, corrects false positives or negatives and approves remediation steps before they are carried out. Once that is done, the validated labels get sent

back to retrain the RL agent, so you have ever so slightly improving human–AI assessment loop.

The creative achievement manifests through a unified system that combines five essential components which no previous system has achieved. The system uses a two-tiered x86-plus-ARM hardware structure which enables separate processor operation for control and execution through three different messaging protocols that include MQTT for telemetry purposes and Redis for task queuing and gRPC for control signal transmission. The system uses a mathematical task-distribution function which automatically assigns work to different agents by monitoring real-time latency and compute expenses and queue fluctuations while allowing idle agents to take on extra work from their busy counterparts. The system implements a token-aware routing mechanism which processes standard analysis through the local quantised model until local token capacity reaches its limit or the system analysis confidence falls below the established calibration point. The device functions through complete offline capability while it provides access to cloud-level reasoning whenever users require it. The system uses a specialised reinforcement-learning agent which learns from previous vulnerability records to forecast unique exploitation patterns that the signature database has not documented. The system includes a dual-mode operator interface which enables operators to change hardware operation from offensive penetration testing to defensive SIEM-style hardening through one command.

The AI components carry out full stack penetration testing across the OSI span, like L1 physical all the way to L7 applications. The token-aware routing system takes those harder queries and forwards them to the DRS linked to AI pipeline, and the raw findings stay on-device. Any cloud connection is only for communication between the chosen pentesting server; no third-party cloud LLM APIs get called.

3. **What were the other possible solutions and why they could not be done?**

Alternative 1 — The assessment needs to operate completely through cloud systems which include AWS Security Agent and L2M-AID. The platforms deliver exceptional reasoning capabilities through advanced language models, yet they depend on users to maintain internet access, which is necessary for their operation. The system sends unprocessed vulnerability information to remote servers, which results in security problems that affect protected and controlled environments. The security system loses all its functions during an upstream outage, while latency issues from cloud round-trips make it impossible to respond instantly.

Alternative 2 — Software-only multi-agent frameworks (e.g., PENTEST-AI, PentestGPT, BreachSeek, PentAGI) divide work into specific tasks which their dedicated software agents handle yet run all agents on one central computer system. The system possesses all operational restrictions which arise from its single hardware structure because it cannot withstand errors and it cannot perform simultaneous

computations on multiple processors and it lacks any method for operating on devices with limited processing capabilities.

Alternative 3 — The security prototypes based on single-board design (for example ESA PenBox) demonstrate that Raspberry Pi can perform fundamental security scans but its single-board architecture limits processing capacity to what one ARM chip can achieve. The system has three critical limitations which include its inability to process large subnets, its failure to detect new security patterns, and its hardware malfunction which causes permanent session loss.

4. **What are the advantages of the solution proposed by you?**

- The system achieves faster assessment processing times through its hardware-based parallelism capabilities which use multiple ARM worker nodes instead of using a single board for sequential processing.
- The system achieves precise vulnerability detection through its high aggregate classification accuracy, which supports its ability to identify actual vulnerabilities throughout all testing environments better than L2M-AID and CHATIOT.
- The dedicated reinforcement-learning agent demonstrates strong detection capability for simulated unseen exploitation chains because it operates without needing any cloud connections.
- The mechanism of token-aware routing significantly lowers overhead costs arising from cloud API systems, providing cloud escalation only to queries that exceed the capabilities of the local model.

- .Cloud communication is routed exclusively to the Designated Review Server for expert human validation, not to commercial LLM APIs. This preserves operational security, ensures regulatory compliance, and eliminates third-party data exposure.
- The system exhibits hardware fault tolerance because the Main Brain system can identify node failures through heartbeat monitoring and it will reassign tasks from any failed agent node to operational nodes within a few seconds. The system allows dual-mode operation because it provides all necessary tools for both red-team and blue-team operations from a single hardware platform.
- The system provides complete offline functionality as its standard feature, which permits users to use encrypted cloud escalation that only sends hashed metadata while protecting all raw data and authentication information. The system achieves its permanent security state through faster risk score evaluation, which reaches stability before executing additional evaluation processes. .The system communicates via an encrypted channel with the Designated Review Server, enabling senior pentester review of all AI findings. Additionally, the system incorporates: (a) Blast Radius Business Impact Model (BRBIM) — computes a five-dimension BRS score per vulnerability measuring Confidentiality, Integrity, Availability, Lateral Reach, and Propagation impact, producing a Risk-Adjusted Priority Score (RAPS = $\alpha \cdot P(v) + \beta \cdot BRS(v)$) for remediation ranking; and (b) Mitigation, Remediation, and Response Model (MRRM) — three-tier response: Tier 1 fault isolation (network ACL + service restriction), Tier 2 Bug Fix Suggestion Engine (code/config/patch/architecture fixes), and Tier 3 IOC Quarantine Engine (active/dormant/residual classification, cryptographic containment, SHA-256 forensic logging).

5. Explain the stepwise working of the innovation explaining all the components used in the invention and the specific function they are performing.

Phase 1: System Initialisation

The operator connects PenBox-DMAS to the target network and defines the engagement scope through the command interface which includes target IP range and permitted techniques and credential sets and access constraints. The Main Brain (Mini-ITX, x86 with GPU) establishes the security state vector which it uses to calculate the initial attack surface metric based on the ratio of exposed services and open network pathways to total enumerated attack paths. The Redis task queue is populated with the first batch of reconnaissance tasks. The task scheduler uses the dynamic distribution function to allocate tasks according to which agent will have the fastest response time with the lowest cost. During Phase 1, the BRBIM performs Business Asset Criticality Labelling (BACL) — classifying discovered nodes into Mission-Critical (Tier 1), Business-Important (Tier 2), and Supporting (Tier 3) tiers. This classification scales each asset's contribution to the Blast Radius Score computed in Phase 4.

Phase 2: Vulnerability Discovery (Recon Agent — Agent 1, Raspberry Pi 5)

The agent retrieves reconnaissance assignments from the Redis queue to conduct Nmap, masscan, passive discovery service, fingerprinting, Active Directory enumeration and banner grabbing tests on the designated subnet. The system transmits all operational results which include open port lists, OS fingerprints and service version information together with network topology graphs to the Main Brain through MQTT telemetry that operates in real-time. The Main Brain aggregates these results and applies the vulnerability discovery function which combines signature-based scan hits with a modified Dijkstra network traversal to produce the raw detection set.

Phase 3: Vulnerability Validation

The main Brain system processes the raw detection set through its validation filter which authenticates results according to its established false-positive and false-negative detection capabilities. The process generates the cleaned validated vulnerability set. The Main Brain uses its token-aware routing function to check two conditions during this phase. The analysis process advances to frontier agentic AI cloud models when either condition occurs. All reasoning processes take place on the local GPU.Cloud escalation refers exclusively to the Designated Review Server — the validated finding set is transmitted to the DRS where a senior pentester reviews confidence-flagged results. All raw packet data and credentials remain on the Main Brain. High-confidence findings proceed autonomously; low-confidence or complex findings are DRS-escalated.

Phase 4: TTP Mapping and Prioritisation (Main Brain)

The system uses the binary mapping matrix to connect every confirmed vulnerability with its corresponding MITRE ATT&CK techniques. The Main Brain computes a priority score for each vulnerability by multiplying technique prevalence weights by an inverse privilege-requirement factor which shows the vulnerabilities that an unauthenticated attacker could reach most easily. The top-K highest-priority vulnerabilities are dispatched to Agent 2 (Exploit Pi) via the Redis task queue for exploitation modelling. [Following TTP mapping, the BRBIM computes $BRS(v) = w_1 \cdot C(v) + w_2 \cdot I(v) + w_3 \cdot A(v) + w_4 \cdot R(v) + w_5 \cdot P(v)$ for each vulnerability, where C = confidentiality impact, I = integrity impact, A = availability impact, R = lateral reach via Dijkstra traversal, and P = propagation probability. A Blast Radius Heatmap and Business Impact Report are generated, with RAPS scores used to rank the exploitation queue dispatched to Agent 2.

Phase 5: Exploitation Modelling (Exploit Agent — Agent 2 + RL Agent — Agent 4, Raspberry Pi 5)

The system calculates single-stage exploitation probability for each candidate vulnerability using three factors which include exploit complexity (a normalised difficulty scale) and existing defence coverage (a normalised coverage scale) and the empirically measured LLM exploitation skill factor (a calibrated skill factor). The system

calculates multi-stage chain probabilities which exist between vulnerabilities by multiplying individual stage probabilities with escalation probabilities. Agent 4 (RL Pi) operates its Q-learning model which uses the Bellman optimality equation to forecast new multi-stage exploitation sequences by combining known vulnerability primitives into chains that remain unrecognized in the signature database. The RL agent detects simulated unseen exploit-chain scenarios with high accuracy without needing cloud connectivity.The AI agents perform full-stack exploitation modelling across all seven OSI layers — from hardware fingerprinting (L1) through application-layer web/API exploitation (L7). All results are structured into a validated finding set and transmitted to the DRS for senior pentester false-positive/negative review before proceeding to Phase 6.

Phase 6: Remediation, State Evolution, and Convergence (Defence Agent — Agent 3, Raspberry Pi 5)

Vulnerabilities whose exploitation probability exceeds a configurable threshold are forwarded to Agent 3 (Defence Pi), which executes hardening playbooks, deploys patches, tests configuration rollbacks, and verifies that each remediation action is effective. The Main Brain updates the vulnerability state vector through the state evolution equation which decreases active vulnerabilities by the processed solution while new vulnerabilities are tracked. The system calculates the composite risk score again while it predicts how long it will take for the swarm to reach its destination. If the change in vulnerability count between iterations falls below the threshold a convergence threshold, the system has reached its equilibrium floor and the assessment concludes with a generated report. The cycle returns to Phase 1 for another full iteration across all six phases.Agent 3 (Defence Pi) additionally executes the three-tier MRRM: Tier 1 — automated isolation of exploitable paths via network ACL updates, service binding restrictions, and credential revocation flags; Tier 2 — Bug Fix Suggestion Engine generates code-layer, configuration, patch, and architecture remediation steps cross-referenced to the detected OS/framework version, ranked by RAPS, and DRS-approved before execution; Tier 3 — IOC Quarantine Engine classifies artefacts as Active (immediate quarantine), Dormant (monitored), or Residual (forensic hash preserved), moves Active IOCs to a cryptographically sealed container on Main Brain storage, and verifies quarantine success before updating the state vector S(t).

6. **Attach drawing hand-made/ computer made showing all the components of the Invention.**

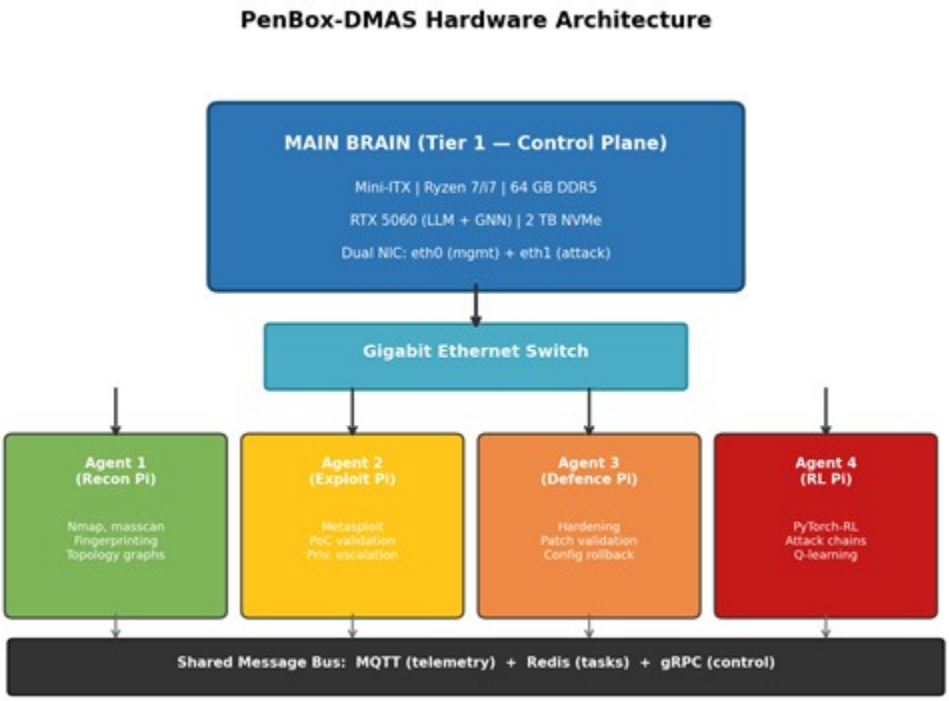

Figure 1: PenBox-DMAS Hardware Architecture — Tiered x86 Main Brain with Raspberry Pi 5 agent cluster connected via Gigabit Ethernet and shared message bus.

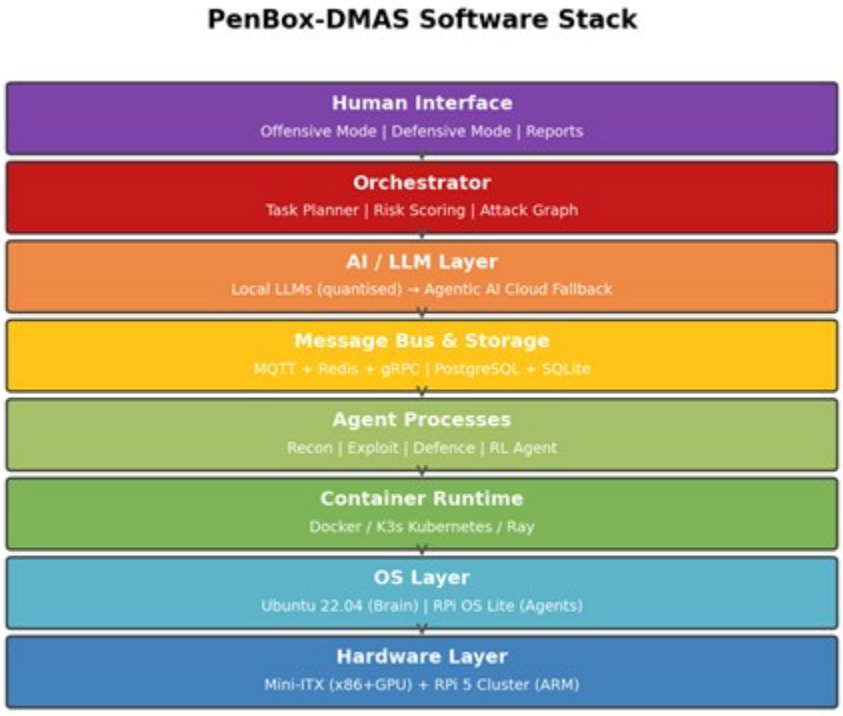

Figure 2: PenBox-DMAS Software Stack — Eight-layer architecture from hardware through OS, containers, agent processes, message bus, AI/LLM layer, orchestrator, to human interface.

PenBox-DMAS Offensive Assessment Mode

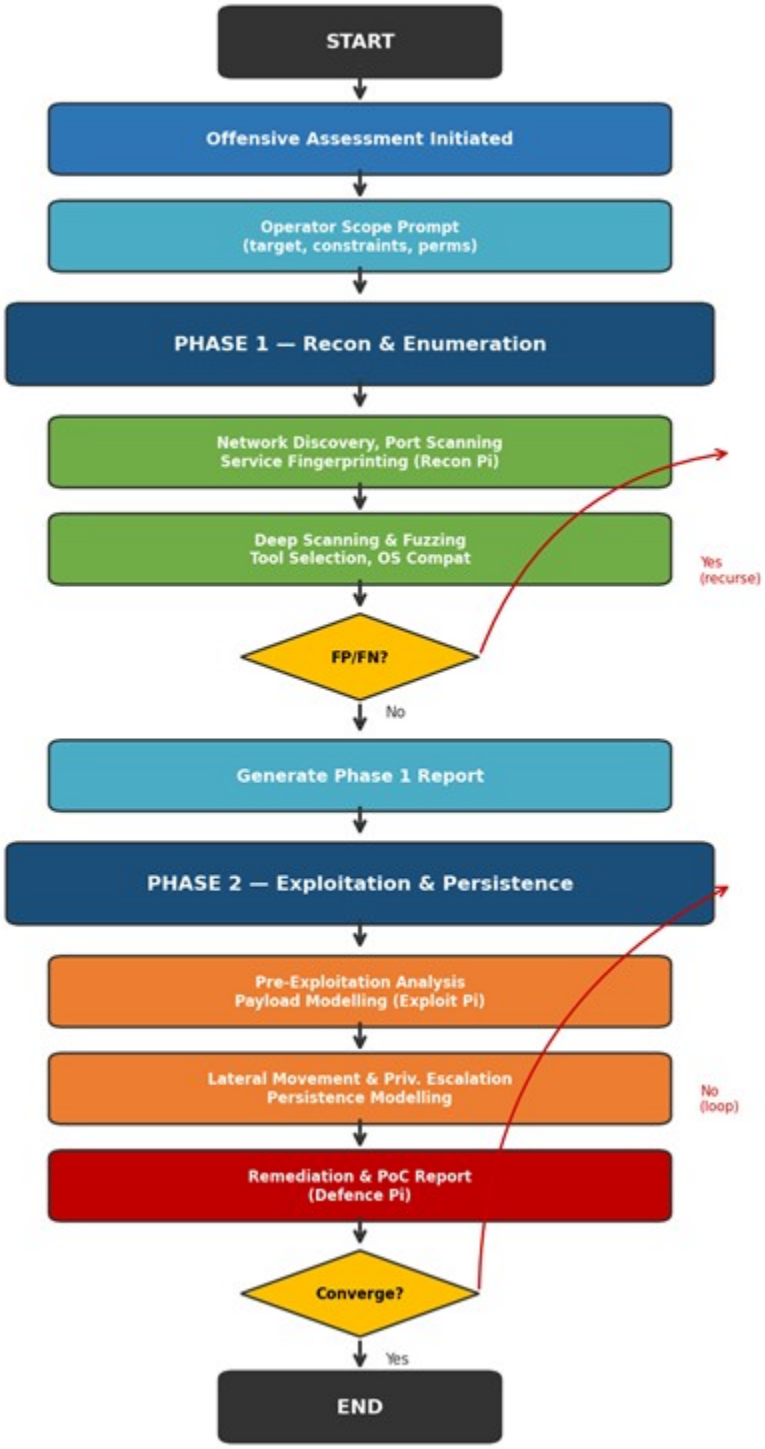

Figure 3: Offensive Assessment Mode Workflow — Two-phase assessment with recursive validation loops for false-positive/negative handling and convergence checking.

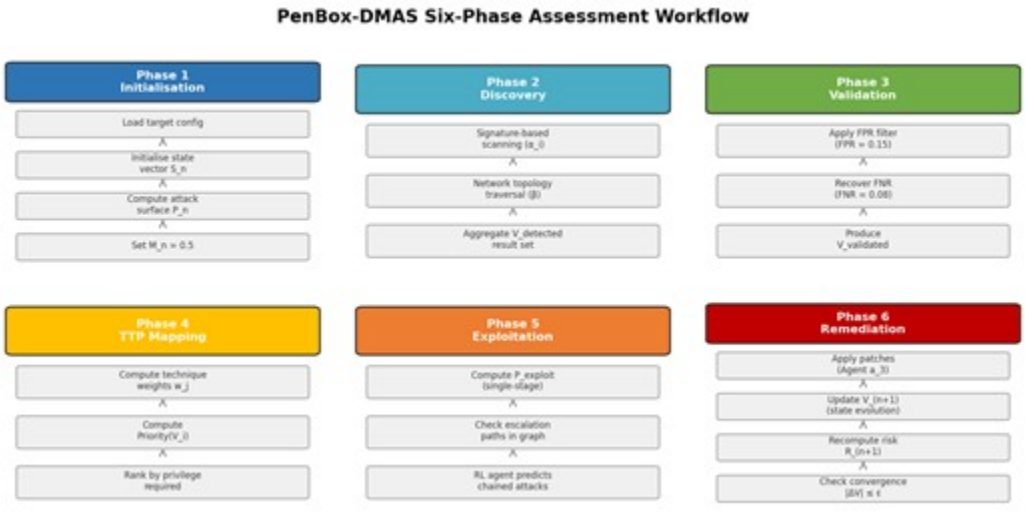

Figure 4: Six-Phase Assessment Workflow — Detailed step breakdown for each of the six phases from initialisation through remediation and state evolution.

Relevant Key phrases relating to your invention

S.NO. Key phrases

1. Distributed multi-agent security appliance
2. Tiered heterogeneous hardware architecture (x86 + ARM)
3. Autonomous vulnerability assessment and remediation
4. Edge-deployed penetration testing platform
5. Token-aware hybrid cloud-edge LLM routing

Confidence-threshold routing to Designated Review Server (DRS) for senior pentester validation — no third-party cloud LLM API usage

6. Adversarial reinforcement learning for exploitation chain prediction
7. Dynamic workload scheduling across heterogeneous agents
8. Swarm-based fault-tolerant security assessment
9. Dual-node offensive/defensive security appliance
10. MQTT-Redis-gRPC tri-protocol agent coordination bus
11. Recursive convergence-based vulnerability state evolution
12. Hardware-integrated AI security kiosk for IoT networks
13. Quantised local LLM with agentic AI cloud escalation
14. MITRE ATT&CK automated technique mapping engine

DRS-integrated full-stack pentesting AI with human expert review loop and RL agent retraining from validated labels

15. Real-time attack surface metric computation
16.

• Designated Review Server (DRS) with human-in-the-loop senior pentester validation

• Full-stack AI pentesting across OSI L1–L7

• Mitigation, Remediation, and Response Model (MRRM): isolation / bug-fix suggestion / IOC quarantine

• IOC Quarantine Engine with active/dormant/residual classification and cryptographic containment

• Blast Radius Business Impact Model (BRBIM): $BRS(v) = w_1C + w_2I + w_3A + w_4R + w_5P$

• Risk-Adjusted Priority Score $RAPS(v) = \alpha \cdot P(v) + \beta \cdot BRS(v)$

• Business Asset Criticality Labelling (BACL) tiers: Mission-Critical / Business-Important / Supporting

• Blast Radius Heatmap for topological business risk visualisation

• Regulatory compliance correlation: GDPR / HIPAA / PCI-DSS / ISO 27001

PenBox-DMAS is a Hardware-Integrated EDR appliance. Unlike software EDR (CrowdStrike, SentinelOne), it is a physical node requiring no agent on monitored endpoints, operating in industrial, enterprise, and IoT environments. It uniquely combines hardware-tiered x86+ARM EDR substrate; full-stack AI pentesting; MRRM three-tier response; BRBIM blast radius scoring; and DRS human-expert validation — none of the cited references (CN121966949A, CN121051758A, US12184683B2, CN120342896A) disclose this combination.

Claim 16 (Independent — MRRM): A Mitigation, Remediation, and Response Model comprising: (a) Tier 1 isolation module executing network ACL updates, service binding restrictions, and credential revocation upon exploitation probability threshold breach; (b) Tier 2 Bug Fix Suggestion Engine generating layered code/config/patch/architecture fixes ranked by RAPS; and (c) Tier 3 IOC Quarantine Engine classifying IOCs into active/dormant/residual categories, executing cryptographic quarantine, and generating SHA-256 forensic logs — all DRS-approved before execution.

Claim 17 (Independent — BRBIM): A Blast Radius Business Impact Model computing $BRS(v) = w_1 \cdot C(v) + w_2 \cdot I(v) + w_3 \cdot A(v) + w_4 \cdot R(v) + w_5 \cdot P(v)$ and $RAPS(v) = \alpha \cdot P(v) + \beta \cdot BRS(v)$, with Business Asset Criticality Labelling, Blast Radius Heatmap generation, and regulatory compliance correlation (GDPR/HIPAA/PCI-DSS/ISO 27001).

Claim 18 (Independent — DRS): A Designated Review Server communication layer transmitting structured vulnerability findings to a senior certified penetration tester for false-positive/negative validation, returning expert-annotated corrections that retrain the on-device RL agent and inform MRRM execution.

§3(k) (Computer Programme Per Se): Rebutted — PenBox-DMAS is a physical hardware appliance (x86 Mini-ITX + ARM cluster + Gigabit LAN). The technical effects (fault-tolerant EDR, zero-day RL prediction, MRRM quarantine, BRBIM blast scoring) require the specific hardware substrate and cannot be reproduced on a general-purpose computer. Ref: IPO CRI Guidelines 2017, §4.4.2.

§3(e) (Mere Admixture): Rebutted — The combination produces synergistic effects not predictable from components individually: (i) RL agent + DRS expert loop creates continuously improving prediction accuracy; (ii) MRRM + BRBIM together prioritise both exploitability and business impact — neither alone achieves this; (iii) tri-protocol bus enables fault-tolerant sub-second task migration across heterogeneous CPUs. These are new technical effects, not a mere admixture.

§3(f) (Mere Arrangement): Rebutted — The physical arrangement produces: (i) 3x faster assessment via parallel ARM workload isolation; (ii) DRS-validated professional-grade accuracy unachievable by fully autonomous tools; (iii) MRRM three-tier response that no cited reference implements; (iv) RAPS-ranked remediation that integrates exploitation probability with business blast radius. These constitute new technical results.

CN121966949A (IoT security, CN): Software-only; no hardware EDR; no DRS; no MRRM; no BRBIM. PenBox-DMAS adds physical appliance, human-in-the-loop DRS validation, MRRM isolation and quarantine, blast radius scoring.

CN121051758A (Supply Chain Vuln, CN): Software supply chain domain; no physical deployment; no DRS; no hardware fault tolerance; no BRBIM. PenBox-DMAS: physical EDR appliance, full-stack network pentesting, DRS, MRRM, BRBIM.

US12184683B2 (Cyber resilience, US Army): Enterprise software; no physical appliance; no DRS; LSTM-only (no RL chain prediction); no MRRM; no BRBIM; no convergence termination. PenBox-DMAS: physical hardware, RL zero-day prediction, convergence-based termination, MRRM, BRBIM, DRS.

CN120342896A (Smart building O&M, CN): Building automation domain; no security function; no pentesting; no MRRM; no BRBIM. Overlap (edge gateway, LLM) is domain-specific and not applicable to security assessment.