International Journal of Advanced Research in Computer Science & Technology (IJARCST)

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Issue 3, May-June 2025||

DOI:10.15662/IJARCST.2025.0803002

AI-Enhanced Intrusion Detection and Prevention Systems

Rangeya Raghav

LNCT, Bhopal, India

ABSTRACT: With cyber threats growing increasingly sophisticated, traditional signature-based Intrusion Detection and Prevention Systems (IDPS) struggle to detect novel, polymorphic attacks. **AI-enhanced IDPS** offer powerful alternatives—leveraging machine learning (ML) and deep learning (DL) to model patterns, anomalies, and behaviors in network and host systems. These solutions enhance detection accuracy, reduce false positives, and provide dynamic adaptability to evolving threat landscapes.

This paper reviews foundational AI techniques in IDPS, including supervised classifiers (e.g., SVM, decision trees), ensemble methods, deep neural networks (like CNNs and RNNs), hybrid models, and reinforcement learning (RL). It outlines training methodologies, performance evaluation, system deployment architectures, and introduces explainability and federated learning for privacy and transparency.

Key findings suggest AI-IDPS can achieve accuracy exceeding 97%, significantly lower false positives, and real-time responsiveness surpassing traditional systems IJISEscienceacadpress.com. RL-based models like QL-IDS have reached near-perfect detection rates on benchmarking datasets arXiv. Federated learning enables decentralized, privacy-preserving deployment suitable for distributed infrastructures arXiv. Incorporation of explainable AI (XAI) techniques—such as LIME and SHAP—enhances analyst trust and transparency MDPIarXiv.

However, challenges persist: acquiring large, representative datasets; ensuring model interpretability; mitigating adversarial attacks; managing computational costs; and safeguarding privacy and fairness scienceacadpress.comAZoAiAICompetence.orgarXiv.

We present a workflow that includes stages such as data gathering, feature extraction, model training/tuning, deployment, monitoring, and feedback loops, all enhanced via XAI and federated learning modules to balance accuracy with explainability and privacy. The paper highlights key advantages (e.g., adaptability, accuracy, threat generalization) and disadvantages (e.g., complexity, data/resource demands, transparency concerns), discusses results and implications, and outlines future work such as adversarial robustness, continual learning, edge deployment, and standardization of XAI evaluation metrics.

KEYWORDS: Machine Learning, Deep Learning, Intrusion Detection Systems, Intrusion Prevention Systems, Anomaly Detection, Reinforcement Learning, Federated Learning, Explainable AI, Cybersecurity Automation

I. INTRODUCTION

Conventional Intrusion Detection and Prevention Systems rely heavily on predefined attack signatures and heuristic rules, limiting their effectiveness against zero-day exploits, polymorphic malware, and sophisticated multi-stage threats. To combat this, **AI-enhanced IDPS** have emerged—integrating ML and DL techniques to learn from vast datasets, discern nuanced patterns, and detect previously unseen intrusions. These systems support both network-based (NIDS) and host-based (HIDS) configurations, offering proactive, adaptive defense mechanisms.

AI-based IDPS utilize **supervised learning** for known attack classification, **unsupervised anomaly detection** to flag deviations from normal behavior, and **deep learning** architectures—like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs)—to extract complex hierarchical and temporal patterns in traffic and system logs IJISEMDPI. Hybrid systems, such as ensemble models and reinforcement learning agents (e.g., QL-IDS), offer further performance improvements, achieving near-perfect detection rates on benchmark datasets arXiv.

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Issue 3, May-June 2025||

DOI:10.15662/IJARCST.2025.0803002

Privacy-preserving training paradigms such as **federated learning** enable decentralized model development, crucial for distributed environments and sensitive data scenarios arXiv. Moreover, to address interpretability concerns inherent in black-box AI models, **Explainable AI (XAI)** tools like LIME, SHAP, and contrastive explanations are applied to clarify decision logic, improving operator trust and facilitating accurate incident response MDPIarXiv.

This introduction establishes the evolution, capabilities, and challenges of AI-Enhanced IDPS—emphasizing their urgent relevance, technical innovations, and the necessary shift toward transparency and decentralized modeling.

II. LITERATURE REVIEW

Research on AI-enhanced IDPS is rich and varied:

- **Supervised and Anomaly-Based Learning**: Traditional ML approaches like decision trees, SVMs, random forests, and k-NN remain prevalent. Ensembles and hybrid methods—such as DT-SVM combinations—have demonstrated marked improvements, including up to 100% accuracy for specific attack classes MDPIAZoAi.
- **Deep Learning Approaches**: CNNs and RNNs enable powerful representation of complex patterns in network traffic, achieving accuracy exceeding 97% and outperforming classical systems with reduced latency and false positives IJISE.
- **Comparative Performance in Critical Infrastructures**: The ASCH-IDS, RBC-IDS, and RL-based QL-IDS models were evaluated using KDD'99 datasets; QL-IDS achieved a perfect detection rate, with SARSA and TD-based models close behind arXiv.
- **Federated Learning for IDS**: Agrawal et al. explore federated learning to train models without centralizing sensitive data. This method supports privacy while maintaining detection efficacy across distributed or IoT networks arXiv.
- **Explainability and Human Trust**: Mane & Rao propose leveraging SHAP, LIME, and other XAI methods to elucidate deep learning-based IDS decisions—bridging the transparency gap essential for analyst oversight arXiv.
- **Implementation Challenges and Trends**: AI-IDPS systems excel but depend on high-quality labeled data, balanced datasets, interpretability, and resilience against adversarial manipulation. Issues such as overfitting, model drift, computational load, and trainer bias remain unresolved scienceacadpress.comAICompetence.orgInsights to Tech Info.

Overall, the literature confirms both substantial progress in AI-powered intrusion detection and the need to address transparency, privacy, and deployment barriers.

III. RESEARCH METHODOLOGY

This study employs a multi-tiered methodology:

1. **Literature Synthesis**
2. Integrates insights from pre-2022 research on AI techniques (ML, DL, RL, federated learning, XAI) applied to IDPS—for performance evaluation, deployment, and transparency.
3. **Architectural Framework Formulation**
4. Proposes a layered AI-IDPS architecture comprising:
   - **Data Ingestion Layer**: capturing network and host logs.
   - **Feature Engineering Module**: selecting relevant attributes via statistical analysis or PCA.
   - **Modeling Layer**: supporting multiple AI models—supervised classifiers, deep neural networks, RL agents, ensemble hybrids.
   - **Federated Learning Module**: coordinating decentralized model updates without sharing raw data.
   - **Explainability Module**: integrating XAI tools (LIME, SHAP) to interpret model outputs.
   - **Response Engine**: converting detection signals into real or simulated prevention actions.
5. **Workflow Definition**
6. Defines a structured pipeline: data collection → preprocessing → feature extraction → model training/tuning → deployment → real-time detection → explanation generation → feedback looping.
7. **Case Study Evaluation**
8. Although empirical implementation is not conducted here, comparative findings are drawn from published performance metrics (e.g., QL-IDS rates, deep learning accuracy), highlighting trade-offs.
9. **Gap and Challenge Analysis**

International Journal of Advanced Research in Computer Science & Technology (IJARCST)

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Issue 3, May-June 2025||

DOI:10.15662/IJARCST.2025.0803002

10. Identification of barriers based on literature: data scarcity, interpretability, adversarial robustness, compute costs, and deployment scalability.

This methodology yields a well-defined, literature-backed AI-IDPS blueprint, with analysis of strengths, limitations, and future directions.

IV. KEY FINDINGS

Key insights from existing research include:

- **High Detection Accuracy**: Deep learning models (e.g., CNN + RNN) have achieved >97% F1-scores, with latency under 100 ms and 35% fewer false positives than traditional systems IJISE.
- **RL-Based IDS Efficacy**: Models such as QL-IDS performed at 100% detection rates on KDD'99 datasets; SARSA and TD models closely followed (~99.5%) arXiv.
- **Privacy-Preserving Training**: Federated learning offers a viable route for distributed model training across decentralized devices or organizations, balancing detection performance with data confidentiality arXiv.
- **Improved Transparency via XAI**: Techniques like LIME and SHAP enhance human interpretability, explaining model predictions and feature influence—critical for analyst trust and actionable incident response MDPIarXiv.
- **Trade-offs and Constraints**: Despite their power, AI-IDPS are limited by the availability of labeled, representative datasets; lack of transparency in deep models; vulnerability to adversarial manipulation; high computational demands; and challenges in real-time, large-scale deployments scienceacadpress.comAICompetence.orgInsights to Tech InfoarXiv.

These findings affirm that AI-enhanced IDPS offer outstanding detection capabilities and adaptability, but their effectiveness hinges on addressing interpretability, privacy, and scalability.

V. WORKFLOW

A practical workflow for AI-Enhanced IDPS is envisioned as follows:

1. **Data Collection**
2. Gather network traffic (via packet captures, NetFlow, logs) and host-level metrics as raw inputs.
3. **Preprocessing & Feature Engineering**
4. Cleanse data, perform feature selection (e.g., PCA or statistical filters) to reduce dimensionality and enhance model performance.
5. **Model Training & Tuning**
   - Train supervised classifiers (e.g., SVM, decision trees) for known patterns.
   - Apply deep learning architectures (CNNs, RNNs) for complex pattern extraction.
   - Utilize reinforcement learning agents for sequential decision-making or threshold tuning.
   - Employ ensemble strategies to combine strengths of multiple models.
   - Leverage federated learning to train across decentralized nodes without sharing raw data.
6. **Explanation Generation**
7. Deploy XAI tools (LIME, SHAP) to interpret model outputs, identifying influential features and rationale behind detection.
8. **Deployment & Real-Time Detection**
9. Implement trained models in network or host environments for live threat detection, coupled with alerting or blocking mechanisms.
10. **Monitoring & Feedback Loop**
11. Aggregate detection performance metrics (accuracy, false positives, resource usage), feed into retraining and model refinement.
12. **Continuous Improvement**
13. Adapt models over time to concept drift, emerging attack vectors, and infrastructure changes, guided by XAI insights and federated updates.

This cyclical workflow unites accuracy, adaptability, transparency, and privacy—enabling robust AI-powered intrusion detection.

International Journal of Advanced Research in Computer Science & Technology (IJARCST)

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Issue 3, May-June 2025||

DOI:10.15662/IJARCST.2025.0803002

VI. ADVANTAGES & DISADVANTAGES

Advantages

- **Enhanced Detection**: Capable of identifying both known and novel threats via pattern generalization.
- **Reduced False Positives**: Advanced models and fine-tuning improve precision.
- **Adaptability**: Models can adapt over time through retraining and federated updates.
- **Interpretability**: XAI modules foster trust and actionable insights.
- **Privacy Preservation**: Federated learning mitigates centralized data sharing.

Disadvantages

- **Data Requirements**: High-quality, labeled data are essential and often scarce.
- **Model Opacity**: Deep models can be black boxes without XAI.
- **Computational Cost**: Training and deploying AI models—especially DL—are resource-intensive.
- **Adversarial Vulnerability**: ML models can be fooled via poisoning or evasion attacks.
- **Scalability & Maintenance**: Managing federated infrastructure and model drift adds operational complexity.

VII. RESULTS AND DISCUSSION

Although this work doesn't present new empirical testing, synthesized findings reflect substantial performance milestones:

- **Detection Excellence**: DL-based IDPS achieve high accuracy and low latency, outperforming classical IDS benchmarks IJISE.
- **Reinforcement Learning Gains**: QL-IDS and related RL approaches demonstrate near-perfect detection in controlled datasets arXiv.
- **Federated Learning Feasibility**: Decentralized training preserves privacy and supports scalable deployment—especially relevant in IoT and distributed environments—though research remains preliminary arXiv.
- **Explainability Impact**: XAI tools significantly improve usability and trust, enabling analysts to interpret and validate AI decisions MDPIarXiv.

However, key limitations persist: models face challenges in generalization across diverse network environments, adversarial attacks can undermine reliability, and infrastructure requirements may limit real-time, high-throughput adoption. Further, operationalizing federated systems requires coordination, security, and governance mechanisms—still emergent in academic exploration.

In sum, AI-enhanced IDPS hold transformative promise, but practical deployments must address interpretability, adversarial robustness, infrastructure readiness, and maintenance demands.

VIII. CONCLUSION

AI-enhanced Intrusion Detection and Prevention Systems represent a paradigm shift in cybersecurity—moving from static, signature-based defense to dynamic, learning-based detection capable of tackling evolving threats. Techniques ranging from supervised learning to deep neural networks, reinforcement learning, federated learning, and explainable AI have shown remarkable efficacy in benchmark evaluations.

The integrated architecture proposed herein balances detection performance with interpretability and privacy—via federated training and XAI modules—offering a practical blueprint for deployment.

Despite these advances, challenges remain in data access, model transparency, resource demands, and adversarial resilience. Real-world application will require attention to trust, continuous adaptation, operational scalability, and governance.

Ultimately, when responsibly designed and deployed, AI-enhanced IDPS offer powerful tools to safeguard modern networks—provided they are supported by robust data pipelines, interpretable models, and secure infrastructure.

International Journal of Advanced Research in Computer Science & Technology (IJARCST)

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Issue 3, May-June 2025||

DOI:10.15662/IJARCST.2025.0803002

IX. FUTURE WORK

Directions for future research and deployment include:

- **Adversarial Robustness**: Develop strategies (e.g., adversarial training, anomaly filters) to mitigate evasion and poisoning attacks.
- **Continual & Online Learning**: Implement models that adapt dynamically to concept drift and emerging threat vectors.
- **Edge & Resource-Constrained Deployment**: Explore lightweight and efficient models (e.g. TinyML) for deployment in IoT and remote environments.
- **Federated Learning Extensions**: Enhance federated techniques with secure aggregation, incentive mechanisms, and cross-domain collaboration.
- **Standardizing Explainability Metrics**: Establish benchmarks for XAI quality, stakeholder-tailored explanations, and usability evaluation.
- **Hybrid Architectures**: Combine rule-based, heuristic, and AI modules to leverage layered defense strategies.
- **Operational Tools & Governance**: Develop orchestration systems for managing federated learning, model updates, and auditability in enterprise contexts.
- **Dataset Diversity & Benchmarking**: Construct evolving, realistic datasets capturing modern threat environments, enabling valid comparison and evaluation.

REFERENCES

1. "AI-powered Intrusion Detection Systems: Real-World Performance Analysis." *Journal of AI-Assisted Scientific Discovery*
2. Yakub Reddy & ShankarLingam, "Artificial Intelligence in Intrusion Detection Systems: Trends, Frameworks, and Future Directions,"
3. Safa Otoum, Burak Kantarci, Hussein Mouftah, "A Comparative Study of AI-based Intrusion Detection Techniques in Critical Infrastructures,"
4. Shaashwat Agrawal et al., "Federated Learning for Intrusion Detection System: Concepts, Challenges and Future Directions,"
5. Shraddha Mane & Dattaraj Rao, "Explaining Network Intrusion Detection System Using Explainable AI Framework,"
6. Explainable Artificial Intelligence for Intrusion Detection System (X-IDS), *MDPI*, 2022
7. AI-based Intrusion Detection: Enhancing Cybersecurity Defence, *Insights2TechInfo*
8. The Rise of AI-Driven Network Intrusion Detection Systems: Innovations, Challenges, and Future Directions