

Article

An Autonomous AI-Driven Framework for Adaptive Cyber Deception with Real-Time Threat Detection and Behaviour-Based Attribution

Muhammad Shahzad , Muhsin Hassanu Saleh *  and Raja Ujjan

School of Computing, Engineering and Physical Sciences, University of the West of Scotland, London Campus, London E14 2BA, UK; b01798648@studentmail.uws.ac.uk (M.S.); raja.ujjan@uws.ac.uk (R.U.)

* Correspondence: muhsin.hassanu@uws.ac.uk

Abstract

Contemporary cyber threats increasingly employ multi-stage and behaviourally adaptive strategies that challenge static intrusion detection and non-adaptive deception mechanisms. Existing approaches typically treat threat detection, deception deployment, and adversarial attribution as separate functions, limiting timely response and underusing the behavioural evidence generated during attacker interaction. This study develops and evaluates a theory-informed computational and operational framework for autonomous cyber deception. The principal research artefact is a reusable closed-loop architecture rather than a single predictive model: it specifies the interacting components, interfaces, data and control flows, decision rules, and feedback mechanisms that connect detection, deception, telemetry, and attribution. Methodologically, the study follows an engineering design-and-evaluation approach comprising problem and requirement identification from the literature, architectural synthesis, component-level mathematical modelling, prototype implementation, and controlled cyber-range evaluation. In this context, modelling refers to the distinct computational models embedded within the framework: a hybrid detection model combining supervised classification, anomaly detection, and temporal sequence analysis; a Markov Decision Process and reinforcement-learning policy model for selecting and reconfiguring deception actions under engagement, intelligence-gain, resource, and containment objectives; and similarity-based and Bayesian attribution models for estimating MITRE ATT&CK techniques from incomplete behavioural evidence. The component models were developed offline using the NSL-KDD, CICIDS2017, UNSW-NB15, and ToN-IoT datasets, while the integrated prototype was evaluated separately in a controlled enterprise-like cyber range using reconnaissance, brute-force, exploitation, and multi-stage attack scenarios. The reported classification metrics were calculated from the labelled cyber-range evaluation events, not by pooling the four benchmark datasets. On this integrated cyber-range evaluation set, the system achieved 95.4% detection accuracy, 93.6% precision, 94.7% recall, and a 94.1% F1-score, with a mean detection latency of 85 ms. It also achieved 100% honeypot deployment reliability, 92% dynamic reconfiguration success, 88% fingerprinting resistance, and attacker engagement durations of up to 280 s. The attribution component demonstrated end-to-end generation of ATT&CK-aligned technique hypotheses from deception-derived telemetry; however, the present archived evaluation does not support per-technique or baseline-comparative performance claims. These findings show that specialised models and operational services can be coordinated within a unified adaptive defence process, while also identifying the additional class-level and ablation evidence required for rigorous attribution validation.



Academic Editor: Yang Yu

Received: 20 June 2026

Revised: 15 July 2026

Accepted: 17 July 2026

Published: 21 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: cyber deception; intrusion detection; machine learning; reinforcement learning; honeypots; behavioural attribution; autonomous cyber defence; MITRE ATT&CK

1. Introduction

The contemporary cyber threat landscape is characterised by a fundamental shift in adversarial capability, intent, and operational sophistication. Modern attackers increasingly operate through coordinated, multi-stage campaigns that combine reconnaissance, exploitation, lateral movement, and persistence while continuously adapting their behaviour in response to defensive controls [1–3]. This evolution has rendered many traditional security mechanisms inadequate, particularly those designed under assumptions of static attack patterns and clearly separable detection and response phases.

Intrusion detection systems (IDSs) have historically formed the backbone of network and host-based defence strategies. Early IDS research established foundational taxonomies and operational models for misuse-based and anomaly based detection [1,4]. While signature-based systems demonstrated effectiveness against known attack patterns, they inherently lacked resilience against novel or obfuscated threats. Anomaly-based approaches sought to address this limitation by modelling deviations from normal behaviour; however, decades of research have shown that such systems often suffer from elevated false-positive rates, sensitivity to environmental drift, and limited operational trust [5,6]. These challenges persist despite significant advances in sensing, logging, and computational capacity.

The introduction of machine learning into intrusion detection marked a significant step forward, enabling the modelling of complex, non-linear relationships within high-dimensional security data. Temporal sequence models, including LSTM architectures, further enable attack context to be retained across ordered observations and improve the detection of slow, stealthy, and multi-stage activity [2,7,8]. Other deep learning approaches have extended this capability to complex network- and host-level classification tasks [9,10]. Publicly available datasets such as UNSW-NB15 and CICIDS-derived traffic corpora have facilitated systematic benchmarking and comparative evaluation of learning-based IDS techniques [11–13]. Nevertheless, despite demonstrable improvements in classification accuracy, most AI-based IDS solutions remain architecturally isolated within security pipelines. Their outputs are typically reduced to alerts, leaving subsequent response and analysis to manual or loosely coupled downstream processes.

In parallel with advances in detection, cyber deception has gained renewed attention as a means of shifting defensive posture from passive observation to active adversarial engagement. Honeypots and decoy systems have long been recognised for their ability to attract malicious activity, delay attacker progress, and generate high-fidelity behavioural data [14,15]. Comprehensive surveys and taxonomies have highlighted the strategic value of deception in increasing attacker uncertainty and asymmetry in defender–attacker interactions [14,16,17]. However, empirical studies have consistently demonstrated that static or manually configured deception environments are vulnerable to fingerprinting and rapid abandonment, particularly by skilled adversaries employing reconnaissance and environment validation techniques [15,16].

To address these limitations, recent research has explored adaptive and decision-theoretic approaches to cyber deception. Game-theoretic formulations model deception as signalling under asymmetric information, capturing strategic interactions between attackers and defenders [16,18,19]. Reinforcement learning-based approaches further extend this paradigm by enabling defenders to optimise deception strategies through interaction

and feedback, balancing engagement benefits against operational cost and risk [20–22]. While these approaches demonstrate conceptual promise, they are frequently evaluated in constrained or abstract settings and are often decoupled from real-time intrusion detection mechanisms. As a result, detection, deception, and learning are commonly treated as separate research problems rather than as components of an integrated defensive system.

A further structural weakness in existing cyber defence architectures lies in the disconnect between low-level detection events and higher-level behavioural understanding. The MITRE ATT&CK framework has emerged as a de facto standard for representing adversarial tactics, techniques, and procedures (TTPs), enabling structured reasoning about attacker behaviour across the kill chain [23]. Prior work has demonstrated the value of ATT&CK-aligned threat intelligence for improving analyst workflows, incident understanding, and defensive planning [24,25]. However, in most operational settings, ATT&CK mapping remains retrospective and analyst-driven, limiting its applicability for real-time defence, adaptive engagement, and automated decision-making under uncertainty [26].

Taken together, these observations reveal a persistent and critical gap in contemporary cyber defence: the absence of autonomous, closed-loop architectures that tightly integrate real-time detection, adaptive deception, and structured behavioural attribution. Existing systems predominantly emphasise isolated performance improvements whether in detection accuracy, deception fidelity, or attribution taxonomy without addressing the systemic interaction between these components [16,17,27]. Consequently, defensive actions rarely influence attacker behaviour in a principled or optimised manner. In this study, behavioural intelligence refers to high-fidelity evidence captured during attacker interaction—such as command sequences, tool usage, timing patterns, navigation paths, privilege-escalation attempts, and transitions between MITRE ATT&CK techniques—which can support technique-level attribution, adaptive defence, and analyst decision-making; however, such evidence remains underexploited in existing systems.

In response to these limitations, this paper develops a theory-informed computational and operational framework for autonomous cyber deception. The primary contribution is the framework as a system-level research artefact, not a single monolithic predictive model. In this paper, a model denotes a bounded mathematical or computational representation used to perform a specific function, whereas a framework denotes the higher-order, reusable architecture that organises multiple models and operational services. The framework therefore unifies real-time threat detection, adaptive deception orchestration, behavioural telemetry, and behaviour-based attribution by specifying their component responsibilities, interfaces, data and control flows, decision rules, and feedback relationships within a closed-loop process.

The framework contains and coordinates several specialised component models. Ensemble-learning and temporal-modelling principles are used to construct the hybrid detection model [7,9,10]; sequential decision theory represents deception as a Markov Decision Process, and reinforcement learning is used to learn a deception policy under engagement, intelligence-gain, resource-cost, and defender-side operational and containment-risk objectives [20–22]. In this paper, risk does not mean predictive-model uncertainty or a general enterprise-risk score. It refers specifically to the possibility that a deception action could weaken containment or disrupt defensive operations, including attempted movement beyond the isolated deception environment, misuse of a decoy as a pivot, accidental exposure of production assets or sensitive information, excessive consumption of CPU, memory, or container resources, and failed or unsafe deployment or reconfiguration. Similarity-based and Bayesian models support technique-level attribution from incomplete and uncertain behavioural evidence [25]. Closed-loop control principles organise the recurring detect–act–observe–adapt cycle, while the MITRE ATT&CK framework supplies the domain ontology for expressing inferred tactics and techniques [2]. Thus, the paper builds component-

level models within a broader framework; the models perform individual analytical or decision tasks, whereas the framework governs how those tasks interoperate to produce autonomous defensive behaviour.

Contributions

This paper makes the following key contributions to the design and evaluation of autonomous cyber defence systems:

- Closed-loop autonomous cyber deception architecture: We propose a fully autonomous, closed-loop cyber defence architecture that tightly integrates real-time threat detection, adaptive deception orchestration, behavioural telemetry, and structured adversarial attribution within a single decision-making pipeline. Unlike existing systems that treat detection, deception, and attribution as loosely coupled or sequential processes, the proposed architecture enables continuous feedback between these components, allowing defensive actions to influence attackers' behaviour in real time. By explicitly modelling the interaction between detection outcomes, deception responses, and behavioural evidence, the framework advances cyber defence from static monitoring toward adaptive adversarial engagement aligned with modern threat models.
- Hybrid machine learning-based detection engine for low-latency threat identification: We design and implement a hybrid intrusion detection engine that combines supervised classification, anomaly detection, and temporal sequence modelling to achieve robust and low-latency threat detection. By fusing complementary learning paradigms, the detection engine mitigates the limitations of single-model approaches, improving generalisation to unseen attacks while maintaining operational efficiency. Temporal modelling enables the detection of slow, multi-stage attack patterns that evade snapshot-based analysis, providing a reliable trigger for downstream deception and attribution mechanisms.
- Reinforcement learning-driven adaptive deception controller: We formulate cyber deception as a sequential decision-making problem and introduce a reinforcement learning-based deception controller that dynamically reconfigures deception strategies in response to observed attacker behaviour. The controller optimises deception actions to maximise attacker engagement and intelligence extraction while accounting for resource utilisation and defender-side operational and containment risk. This risk covers loss of isolation or pivoting beyond the decoy environment, exposure of production assets or sensitive data, resource exhaustion, and failed or unsafe orchestration. This adaptive approach addresses the limitations of static and manually configured deception systems, enabling sustained engagement and reducing susceptibility to fingerprinting by skilled adversaries.
- Behaviour-based adversarial attribution aligned with MITRE ATT&CK. We develop a behaviour-based attribution engine that maps deception-derived behavioural evidence to adversarial tactics and techniques defined in the MITRE ATT&CK framework. The attribution process combines similarity-based scoring with Bayesian evidence updating to represent uncertainty and aggregate observations across interaction stages. By operating at the technique level rather than relying only on static indicators, the engine provides a structured mechanism for situational awareness and analyst decision support. The contribution of the present study is the design and integration of this attribution pipeline; a class-level comparative evaluation of similarity-only, Bayesian-only, and combined variants is not claimed.
- Comprehensive experimental evaluation in a controlled cyber-range environment. We evaluate the proposed framework within a controlled cyber-range environment that emulates realistic attacker workflows and enterprise-like network conditions. The evaluation assesses detection accuracy and latency, deception believability and fingerprinting

resistance, attacker engagement duration and interaction depth, and the end-to-end feasibility of generating ATT&CK-aligned attribution outputs from deception-derived evidence. Detection and deception results are reported quantitatively, whereas attribution is treated as a proof-of-concept component because the archived experiment does not retain the per-technique predictions and ablation outputs required for a confusion matrix or baseline comparison. This distinction prevents framework-level feasibility from being misrepresented as complete validation of every component model.

2. Related Work and Conceptual Positioning

This section positions the study at two distinct but connected levels: component-level computational modelling and system-level framework design. The proposed approach is not confined to threat detection. Threat detection is the first analytical component and supplies confidence estimates and trigger signals; however, the overall approach also models adaptive deception as a sequential decision problem, transforms attacker interactions into structured behavioural evidence, and performs probabilistic technique-level attribution. The framework coordinates these functions through closed-loop data and control flows so that detection initiates an action, deception elicits further behaviour, and the resulting evidence supports attribution and subsequent defensive adaptation.

Accordingly, the term model is used in this paper for a bounded mathematical or computational representation that performs a specific function. The framework contains three principal model families: (i) supervised, anomaly-based, and temporal models for threat detection; (ii) a Markov Decision Process and reinforcement-learning policy model for selecting and reconfiguring deception actions; and (iii) similarity-based and Bayesian models for mapping behavioural evidence to MITRE ATT&CK techniques. Behavioural telemetry and orchestration are enabling services that connect these models but are not presented as predictive models.

The literature is therefore reviewed across three domains: (i) AI-based intrusion detection, (ii) adaptive cyber deception, and (iii) behavioural attribution and structured threat modelling. This organisation does not imply that the framework is primarily a threat-detection framework. Rather, Section 2.1 examines the detection models that generate the initial control signal, Section 2.2 examines the decision and adaptation mechanisms used to change the deception environment, and Section 2.3 examines the models used to interpret deception-derived evidence. Section 2.4 then synthesises these strands and identifies the system-level integration gap addressed by the proposed framework.

2.1. AI-Based Intrusion Detection as the Perception Layer

Learning-based intrusion detection systems (IDSs) have been extensively investigated as a means of overcoming the inherent limitations of signature-based approaches, particularly in the detection of previously unseen attacks and the modelling of complex, high-dimensional behaviour patterns [1,2]. Classical supervised learning techniques, including tree-based ensemble models, have demonstrated strong detection performance when trained and evaluated on representative datasets. However, their effectiveness is often constrained by sensitivity to concept drift, dataset bias, and limited generalisation beyond the conditions observed during training, which reduces robustness in dynamic operational environments.

To mitigate these limitations, unsupervised and semi-supervised anomaly detection approaches have been proposed to identify deviations from established baseline behaviour without reliance on labelled attack data. While such methods offer improved coverage for novel and unknown threats, empirical studies consistently report elevated false-positive rates, which complicate operational deployment and reduce analyst trust in

alert outputs [5,10]. The trade-off between sensitivity to anomalous activity and practical usability remains a persistent challenge in anomaly-driven IDS design.

More recently, deep learning techniques have been adopted to capture richer spatial and temporal dependencies in network and host-level data. Temporal architectures, such as recurrent neural networks and long short-term memory (LSTM) models, have demonstrated effectiveness in modelling sequential patterns associated with multi-stage and slow-moving attacks that evade snapshot-based detection [7,8]. In parallel, hybrid deep learning models have been applied to encrypted traffic analysis and complex security classification tasks, further improving detection capability in challenging environments [9]. Bono et al. [8] similarly used an LSTM-based deep-learning architecture for fault detection and remaining-useful-life estimation, illustrating the value of gated memory for temporally dependent condition-monitoring signals.

Despite these advances, a common limitation across much of the IDS literature is the narrow focus on classification performance as the primary evaluation objective. Detection is frequently treated as an isolated analytical task, with limited consideration of how detection outputs can inform adaptive response, adversarial engagement, or higher-level behavioural reasoning. In particular, IDS outputs are rarely exploited as control signals for cyber deception mechanisms, nor are they systematically integrated with real-time attribution pipelines. This disconnect limits the strategic value of detection and highlights the need for architectures in which intrusion detection operates as an active component of an autonomous, closed-loop defence system.

Section 2.1 does not, by itself, describe agentic AI. The intrusion-detection models reviewed here perform a perception and inference function: they classify traffic, estimate anomaly or attack likelihood, and generate confidence scores or trigger signals. They do not independently formulate goals, plan multi-step actions, execute changes in the environment, or learn from an action-observation feedback cycle. Agent-like functionality arises only at the system level when the detection output updates the Markov Decision Process state, the reinforcement-learning controller selects a bounded deception action under an explicit reward function and safety constraints, the orchestration service executes that action, and subsequent telemetry provides feedback for the next decision. The proposed system is therefore characterised as a bounded autonomous decision-making framework with agentic properties, rather than as a general-purpose agentic AI system. In this interpretation, Section 2.1 reviews the perception layer, whereas Section 2.2 and the closed-loop integration described in Sections 2.4 and 3 provide the decision, action, and feedback mechanisms.

2.2. Cyber Deception Systems and Associated Risk Types

Deception mechanisms, including honeypots and decoy systems, have been widely studied as defensive techniques aimed at increasing attacker uncertainty, diverting malicious activity away from production assets, and eliciting high-value intelligence about adversarial behaviour [14]. Honeypot systems are commonly categorised according to their level of interaction, reflecting a fundamental trade-off between behavioural fidelity, scalability, and defender-side operational risk [15,16]. In this subsection, operational risk is used as an umbrella term for four safety and service-related categories: (i) containment risk, namely the possibility that an attacker may communicate with, pivot to, or affect systems beyond the isolated deception environment; (ii) exposure risk, including unintended disclosure or reachability of production assets, real credentials, configurations, or sensitive information; (iii) resource and availability risk, arising when deceptive services consume excessive CPU, memory, network, container, or orchestration capacity; and (iv) configuration and management risk, including failed, inconsistent, or unsafe deployment, monitoring, and reconfiguration of deception assets. High-interaction honeypots provide realistic envi-

ronments that enable detailed observation of attacker tactics and decision-making, but they are more resource-intensive and generally carry greater containment, exposure, and management risk. In contrast, low-interaction honeypots reduce these defender-side risks and operational overhead, but their simplified behaviour increases fingerprinting risk—the likelihood that an attacker recognises the asset as deceptive and disengages. Fingerprinting risk is therefore an effectiveness risk that reduces engagement and intelligence yield, rather than a direct containment or exposure risk.

Recognising the limitations of static deception deployments, prior research has explored adaptive deception strategies that respond dynamically to attacker behaviour. Game-theoretic formulations model deception as a signalling problem under asymmetric information, capturing the strategic interplay between attackers and defenders and providing insight into optimal defensive signalling policies [17,19]. Complementary to this perspective, survey and taxonomy-driven studies have systematically characterised defensive deception techniques, threat models, and deployment considerations, establishing a conceptual foundation for adaptive deception design [17].

More recently, learning-based approaches have been proposed to enable autonomous adaptation of deception strategies in adversarial environments. Reinforcement learning techniques, in particular, have been investigated as a means of optimising deception policies through interaction and feedback, allowing defenders to balance attacker engagement and intelligence gain against operational cost and the defender-side risks defined above [21,22]. In this study, containment, exposure, resource/availability, and configuration/management risks constrain whether a deception action is safe to deploy, whereas fingerprinting resistance is evaluated as a measure of deception effectiveness and believability. While learning-based approaches demonstrate promising results in controlled settings, much of the existing literature evaluates deception mechanisms in isolation or under narrowly defined scenarios. As a result, few systems provide a fully integrated architecture in which real-time intrusion detection informs autonomous deception reconfiguration, and deception-derived behavioural evidence is systematically leveraged for structured attribution. This fragmentation limits the effectiveness and practical applicability of adaptive deception in operational cyber defence.

2.3. Behavioural Attribution and Structured Threat Modelling

Attribution remains one of the most challenging problems in cyber security, owing to widespread adversarial practices such as obfuscation, tool reuse, infrastructure sharing, and deliberate counter-forensic techniques designed to mislead defenders. These factors significantly complicate actor-level attribution, which often requires external intelligence sources and long-term analysis. From an operational perspective, technique-level attribution has therefore emerged as a more practical and actionable objective, as it directly supports incident response, threat hunting, and defensive planning by characterising adversarial behaviour rather than attempting to infer identity.

The MITRE ATT&CK framework has become a widely adopted standard for representing adversarial behaviour through a structured taxonomy of tactics, techniques, and procedures (TTPs), enabling systematic mapping of observed activity to known adversarial patterns [24]. Prior research has demonstrated the value of ATT&CK-aligned threat intelligence for enhancing analyst workflows, improving situational awareness, and supporting consistent reasoning across incidents and organisations [25,26]. By abstracting low-level events into higher-level behavioural constructs, ATT&CK facilitates a shared vocabulary for both human analysts and automated systems.

To address uncertainty and incomplete evidence in attribution, probabilistic and Bayesian reasoning approaches have been explored as a means of aggregating heteroge-

neous indicators and supporting inference under uncertainty [27]. Such methods offer a principled framework for combining multiple behavioural signals, weighting evidence over time, and expressing confidence in attribution outcomes. However, their application in operational settings remains limited, particularly in environments where behavioural evidence is generated dynamically through attacker interaction.

Despite these advances, most ATT&CK-driven attribution pipelines remain largely retrospective and analyst-driven, operating offline and detached from real-time defensive processes. Automated attribution mechanisms that operate in real time and are tightly integrated with deception systems are still rare. In particular, the systematic coupling of deception-derived behavioural evidence with probabilistic attribution models remains underexplored. This gap constrains the ability of current systems to exploit attacker–defender interaction for timely, intelligence-driven defence, underscoring the need for integrated architectures that unify adaptive deception and behaviour-based attribution within a closed-loop defensive framework.

2.4. Synthesis and Positioning of the Framework

The review confirms that the proposed methodological approach is broader than threat detection. Detection provides the initial inference and trigger, but it is only one component of the complete defensive process. The deception component determines how the environment should respond and adapt; the telemetry component captures the attacker's subsequent interactions; and the attribution component converts those interactions into structured, uncertainty-aware evidence about adversarial techniques. These functions address different research problems and are represented by different component-level models. This division also clarifies the relationship to agentic AI: the detection subsystem supplies perception but is not itself an agent; bounded agency emerges from the framework-level coupling of state estimation, goal-directed policy selection, action execution, and feedback.

At the system level, the contribution is the framework that specifies how these models and operational services interact through defined inputs, outputs, decision rules, and feedback paths. The central research gap is therefore not simply insufficient threat-detection accuracy; it is the absence of an autonomous architecture that tightly couples detection with reinforcement learning-driven deception and ATT&CK-aligned probabilistic attribution. The proposed framework addresses this gap and is evaluated both at the component-model level and at the end-to-end system level within the cyber range.

The term 'continuous learning architecture' has been removed because it could incorrectly imply that all models are retrained online after each new event. The proposed system is more accurately described as a closed-loop adaptive architecture. It continuously receives telemetry, updates the current operational state, accumulates behavioural evidence, revises attribution confidence, and selects the next deception action. This operational adaptation must be distinguished from continual machine learning: the detection and attribution models are trained offline and remain fixed during runtime, while the reinforcement-learning policy is learned during a controlled training phase and then used for action selection. Thus, 'continuous' refers to uninterrupted sensing, feedback, evidence accumulation, and decision-making—not to automatic, incremental retraining of every component model. Full continual learning, including drift-aware updates and safeguards against catastrophic forgetting, is outside the scope of this study.

3. Proposed Framework

The proposed framework is a theory-informed computational and operational architecture composed of four tightly coupled and functionally interdependent subsystems: a hybrid machine learning-based detection engine, a reinforcement learning (RL)-driven

deception controller, a behavioural telemetry pipeline, and an attribution engine aligned with the MITRE ATT&CK framework [24,25]. The framework should not be interpreted as a single model. Rather, it is the system-level structure within which separate models are developed, trained, executed, and coordinated. At the component level, the detection engine operationalises ensemble and temporal inference, the deception controller operationalises sequential decision theory through an MDP and an RL policy, and the attribution engine operationalises similarity-based and Bayesian reasoning under uncertainty. At the architectural level, the framework defines component responsibilities, interfaces, data flows, control signals, decision rules, and feedback relationships required for autonomous operation. These subsystems are therefore not independent models assembled ad hoc, but mutually dependent elements of a reusable closed-loop architecture. The overall system follows an event-driven design paradigm in which defensive actions are triggered by observed system and network events rather than static schedules or predefined workflows. Detection outcomes act as control signals that initiate and guide deception strategies. Deception interactions, in turn, generate rich behavioural telemetry capturing attacker actions, temporal dynamics, and interaction context. This evidence is processed by the attribution engine, which updates situational understanding by mapping observed activity to structured adversarial techniques. The resulting attribution insights are fed back into the system to inform ongoing deception and defensive adaptation, thereby establishing a closed-loop interaction between detection, deception, and behavioural analysis.

Figure 1 presents the conceptual, system-level architecture of the proposed autonomous cyber deception framework, whereas Figure 2 presents the corresponding implementation and deployment view. Figure 1 defines the closed-loop relationships among threat detection, adaptive deception, behavioural telemetry, attribution, and analyst reporting. Figure 2 shows how these architectural functions are realised as interacting services within the cyber-range environment, including the detection service, deception orchestrator, attacker-engagement assets, behavioural analytics, attribution engine, external threat-intelligence interfaces, and analyst-facing dashboard.

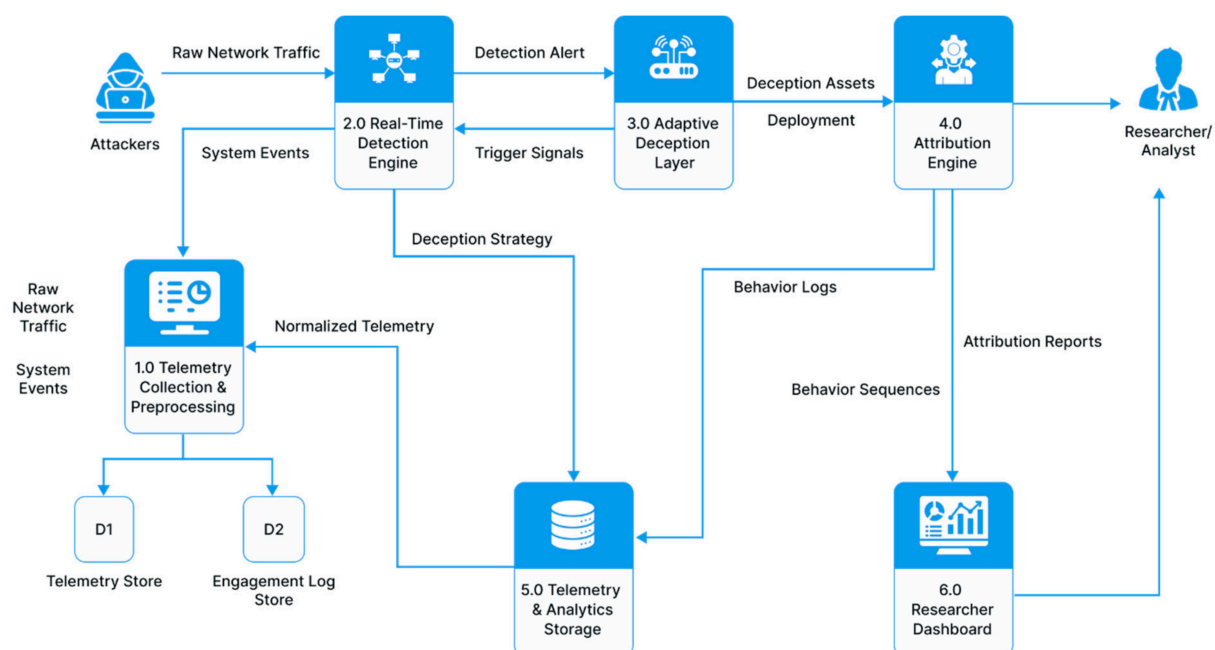


Figure 1. Closed-loop autonomous cyber deception architecture.

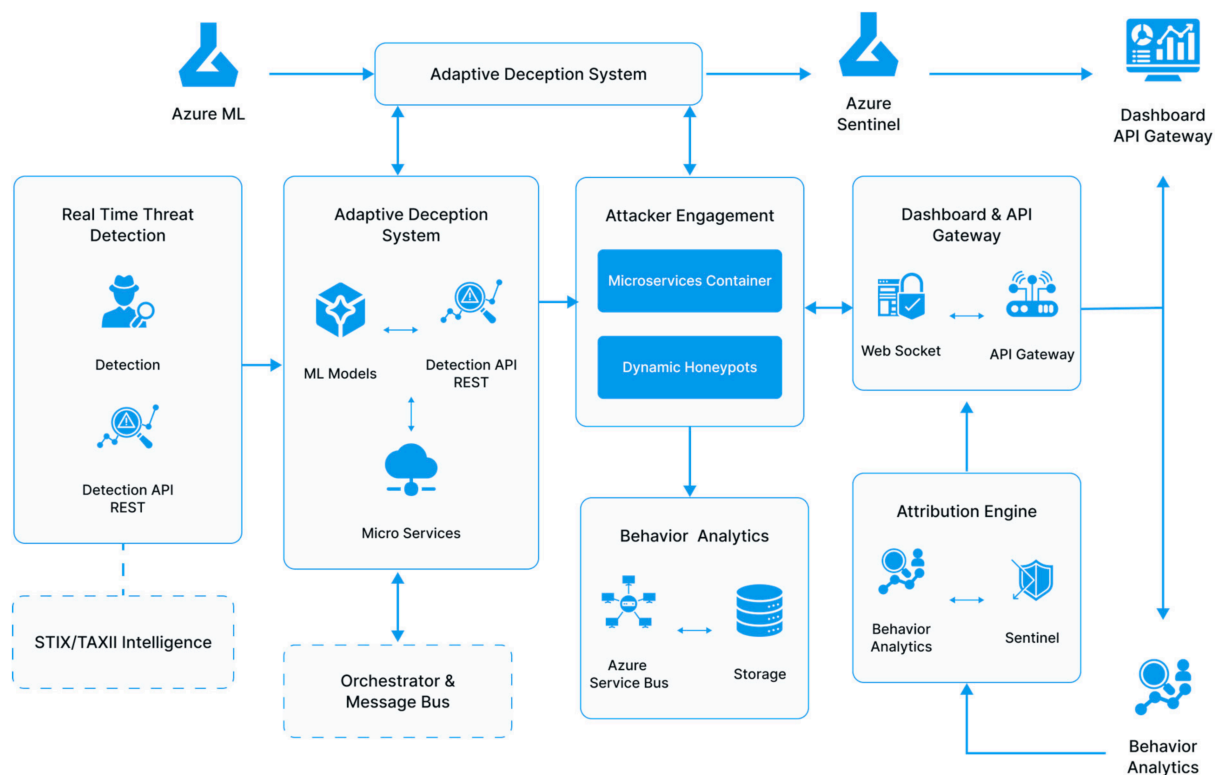


Figure 2. System-level implementation of the proposed framework.

Accordingly, Figure 2 should be read as the implementation-level realisation of Figure 1 rather than as a separate or unrelated framework. It makes the operational data and control flows explicit: detection outputs trigger deception actions, deceptive interactions generate behavioural telemetry, and the resulting evidence is processed for attribution and presented through analyst-facing interfaces.

3.1. Hybrid Machine Learning-Based Detection Engine

To mitigate reliance on a single modelling paradigm, detection is formulated as ensemble inference combining supervised, anomaly-based, and temporal models [2,7,8].

Let C denote the set of classes and let x be a feature vector describing a network flow or session. $P_{RF}(c | x)$ denotes the posterior probability predicted by a Random Forest classifier, $P_{LSTM}(c | x_{1:t})$ denotes the temporal prediction from an LSTM model given a sequence of observations $x_{1:t}$, and $P_{IF}(c | x)$ represents a transformed anomaly-derived score from Isolation Forest.

The final ensemble decision is computed as follows:

$$\hat{y} = \arg \max_{c \in C} (w_1 P_{RF}(c | x) + w_2 P_{LSTM}(c | x_{1:t}) + w_3 P_{IF}(c | x)), \quad (1)$$

subject to:

$$w_1 + w_2 + w_3 = 1, w_i \geq 0. \quad (2)$$

The weights w_i are calibrated on validation data to balance detection performance and latency.

Temporal Modelling with LSTM

The LSTM component captures sequential dependencies indicative of multi-stage attacks [7,8]. Unlike a vanilla Elman recurrent neural network, an LSTM maintains both a hidden state and a separate cell state, with gated transformations controlling information

retention, update, and exposure. Let x_t denote the feature vector at time step t , h_{t-1} the previous hidden state, and c_{t-1} the previous cell state. The standard LSTM recurrence is as follows:

$$\begin{aligned} i_t &= \sigma(W_i x_t + U_i h_{t-1} + b_i), f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f), o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o), \\ g_t &= \tanh(W_g x_t + U_g h_{t-1} + b_g), c_t = f_t \odot c_{t-1} + i_t \odot g_t, h_t = o_t \odot \tanh(c_t), \end{aligned} \quad (3)$$

$$y_t = \text{softmax}(W_y h_t + b_y), \quad (4)$$

Here, i_t , f_t , and o_t are the input, forget, and output gates, respectively, while g_t is the candidate-cell transformation. Together, these four transformations regulate what information is written, retained, exposed, and proposed for the cell state. The vector c_t is the updated cell state and h_t is the exposed hidden state. The logistic function $\sigma(\cdot)$ maps each gate to $[0, 1]$, $\tanh(\cdot)$ forms bounded candidate content, and $\odot$ denotes element-wise multiplication. W , U , and b denote the trainable input-weight, recurrent-weight, and bias parameter families. c_t is the explicit cell-state update that enables selective retention of long-range information, while h_t produces the hidden state. The class-probability distribution y_t is obtained through the softmax output in Equation (4).

3.2. Adaptive Deception Engine

Adaptive deception is treated as sequential decision-making under uncertainty, leveraging reinforcement learning principles [22,28]. The deception controller selects and deploys deception actions to maximise engagement and intelligence yield while controlling operational cost.

3.2.1. Finite MDP Specification

The adaptive-deception controller is formulated as a finite Markov Decision Process (MDP), $M = (S, A, P, R, \gamma)$, in which S is the discrete state space, A is the bounded deception-action set, $P(s' | s, a)$ is the environment transition structure, R is the immediate reward, and γ is the discount factor. A decision step is generated after each five-second telemetry window or immediately after a high-severity event. An episode begins when the malicious-confidence trigger is reached and ends when the attacker disengages, the interaction is isolated or terminated, a containment violation occurs, or 50 decision steps have elapsed.

$$M = (S, A, P, R, \gamma). \quad (5)$$

The state at decision time t is represented by the five-component tuple $s_t = (d_t, p_t, c_t, b_t, u_t)$. The variables encode detection confidence, attack phase, active deception configuration, engagement depth, and resource utilisation, respectively. Table 1 specifies the discretisation used to make the state space finite and technically evaluable.

$$s_t = (d_t, p_t, c_t, b_t, u_t) \in S. \quad (5a)$$

The Cartesian product therefore contains $|S| = 3 \times 4 \times 3 \times 3 = 432$ possible discrete states. The bounded action set contains six actions:

- a_0 : maintain the current deception configuration;
- a_1 : deploy or reconfigure Cowrie for SSH/Telnet interaction;
- a_2 : deploy or reconfigure Dionaea for malware-oriented interaction;
- a_3 : deploy the combined Cowrie–Dionaea deception profile;
- a_4 : mutate deception artefacts, including service banners, decoy credentials, and exposed metadata;
- a_5 : isolate and terminate the interaction when a safety threshold is exceeded.

Table 1. Finite state representation used by the adaptive-deception MDP.

State Element	Symbol	Discrete Categories and Operational Definition	Cardinality
Detection confidence	d_t	Low: 0.50–<0.70; medium: 0.70–<0.90; high: 0.90–1.00. The MDP is activated only after the detector reaches 0.50.	3
Attack phase	p_t	Reconnaissance; credential access; exploitation; post-exploitation, assigned from the current scenario step and ATT&CK-aligned behavioural evidence.	4
Deception configuration	c_t	No active decoy; Cowrie; Dionaea; combined Cowrie–Dionaea profile.	4
Engagement depth	b_t	Shallow: 0–2 distinct behavioural events; moderate: 3–7; deep: ≥ 8 or any privilege-escalation/post-exploitation indicator.	3
Resource utilisation	u_t	Low: <0.40; medium: 0.40–0.70; high: >0.70, using the maximum normalised CPU, memory, active-container, and orchestration-latency ratio.	3

Actions that are infeasible in the current configuration or would exceed the resource or containment threshold are excluded by action masking. The transition structure is partly deterministic and partly stochastic. Conditional on successful orchestration, the deception-configuration component follows the selected action; detection confidence, attack phase, engagement depth, and resource utilisation are recomputed from the next telemetry window and therefore depend on both attacker behaviour and system response. Because Q-learning is model-free, an analytical transition matrix is not assumed in advance. For reproducibility and diagnostic analysis, the transition probability can be estimated from observed cyber-range transitions as follows:

$$\hat{P}(s'|s, a) = N(s, a, s') / \sum_{\tilde{s}} N(s, a, \tilde{s}), \quad (6)$$

where $N(s, a, s')$ is the number of observed transitions from state s to state s' after action a . A deployment failure leaves c_t unchanged and updates u_t and the cost term; attacker disengagement, explicit isolation, a containment/exposure event, or the 50-step limit produces a terminal transition. This formulation makes it explicit that the Markov property is an operational approximation: the selected state variables summarise the information available to the controller, while unobserved attacker intent remains outside the current MDP and motivates the POMDP extension discussed in Section 8.

3.2.2. Q-Learning Policy and Reward Definition

For a non-terminal transition (s_t, a_t, r_t, s_{t+1}) , the controller updates the action-value function using the standard tabular Q-learning rule:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_t + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t)]. \quad (7)$$

Here, α is the learning rate and γ is the discount factor. The finite-horizon controller specification fixes $\gamma = 0.95$ so that delayed engagement and intelligence gains influence action selection while future rewards remain discounted. As explained in Section Reinforcement-Learning Evaluation Scope, the historical 100-episode trace is not retained as reproducible training evidence because its complete α and exploration metadata were not archived; therefore, no convergence claim is derived from Equation (7).

The immediate reward is defined explicitly as a weighted combination of engagement, intelligence gain, and defender-side cost:

$$r_t = \lambda_1 f_engage(t) + \lambda_2 f_intel(t) - \lambda_3 f_cost(t) + r_safe(t), \quad (8)$$

with $\lambda_1 = 0.35$, $\lambda_2 = 0.45$, and $\lambda_3 = 0.20$. These values were fixed a priori to prioritise information gain, followed by sustained engagement, while retaining an explicit penalty for resource use and defender-side risk; they are design parameters and are not claimed to be globally optimal.

The engagement term measures incremental active interaction time during a decision interval and is normalised to $[0, 1]$:

$$f_engage(t) = \min(\Delta\tau_t/30 \text{ s}, 1). \quad (9)$$

The intelligence term combines newly observed behavioural diversity with the reduction in uncertainty of the ATT&CK-technique posterior:

$$f_intel(t) = 0.5 \min(n_new,t/5, 1) + 0.5 \max\{[H(q_{t-1}) - H(q_t)]/\log K, 0\}. \quad (10)$$

Here, n_new,t is the number of previously unseen commands, tools, privilege-escalation indicators, or technique transitions observed since the preceding decision; q_t is the posterior distribution over K candidate ATT&CK techniques; and $H(q) = -\sum_k q(k)\log q(k)$ is Shannon entropy. The first term rewards evidence diversity and the second rewards a reduction in attribution uncertainty.

The cost term combines normalised resource demand, orchestration delay, and operational risk:

$$f_cost(t) = 0.40\rho_t + 0.20\ell_t + 0.40q_t\hat{risk}. \quad (11)$$

The variable ρ_t is the maximum normalised CPU, memory, and active-container utilisation; $\ell_t = \min(L_t/2 \text{ s}, 1)$ is the normalised orchestration latency; and $q_t\hat{risk}$ is 0 under normal containment, 0.5 when a resource/configuration warning threshold is crossed, and 1 when a containment or exposure event is observed. A containment or exposure event additionally sets $r_safe(t) = -1$ and terminates the episode; otherwise $r_safe(t) = 0$. This hard penalty prevents engagement or information gain from compensating for an unsafe action.

The four defender-side risk categories are: (i) containment risk, including attempted communication or pivoting beyond the isolated deception segment; (ii) exposure risk, including unintended reachability or disclosure of production assets, credentials, or sensitive information; (iii) resource and availability risk; and (iv) configuration risk arising from failed or unsafe deployment. These risks are distinct from uncertainty in the detection and attribution models.

3.2.3. Deception Optimisation Objective

The policy objective is to maximise the expected discounted return over an episode:

$$J(\pi) = E\pi[\sum_{t=0}^T \gamma^t r_t], \quad \pi^* = \arg \max_{\pi} J(\pi). \quad (12)$$

Equivalently, substituting Equation (7) gives the explicit multi-objective form:

$$\max_{\pi} E\pi[\sum_{t=0}^T \gamma^t (0.35 f_engage(t) + 0.45 f_intel(t) - 0.20 f_cost(t) + r_safe(t))]. \quad (13)$$

The fixed coefficients make the optimisation target reproducible and interpretable. They are not presented as universally optimal weights; sensitivity analysis across alternative organisational priorities remains necessary before operational deployment.

3.3. Behavioural Telemetry Pipeline

The telemetry pipeline collects raw interaction data from deceptive assets (e.g., session logs, command sequences, timing, and packet captures). Raw telemetry is transformed into a structured behavioural evidence representation $\mathbf{B}$ used for RL state construction and attribution. The pipeline is implemented to support near-real-time ingestion and analysis within controlled experiments [6].

3.4. Attribution Engine

Attribution maps behavioural evidence to MITRE ATT&CK techniques using similarity scoring and Bayesian inference [25,27].

3.4.1. Similarity-Based Technique Scoring

Let T_i denote a candidate technique and $\mathbf{b} = (\mathbf{b}_1, \dots, \mathbf{b}_n)$ an observed behavioural feature vector. Each technique has a signature $\mathbf{t}_i = (t_{i,1}, \dots, t_{i,n})$. The similarity score is computed as follows:

$$\text{Score}(T_i | \mathbf{b}) = \frac{\sum_{k=1}^n \omega_k \text{sim}(\mathbf{b}_k, t_{i,k})}{\sum_{k=1}^n \omega_k}, \quad (14)$$

and the predicted technique is:

$$\hat{T} = \arg \max_{T_i} \text{Score}(T_i | \mathbf{b}). \quad (15)$$

3.4.2. Bayesian Attribution

To incorporate uncertainty:

$$P(T_i | \mathbf{B}) = \frac{P(\mathbf{B} | T_i)P(T_i)}{\sum_j P(\mathbf{B} | T_j)P(T_j)}. \quad (16)$$

Assuming conditional independence:

$$P(\mathbf{B} | T_i) = \prod_{k=1}^n P(\mathbf{b}_k | T_i), \quad (17)$$

and the Bayesian decision rule is:

$$\hat{T}_{\text{Bayes}} = \arg \max_{T_i} P(T_i | \mathbf{B}). \quad (18)$$

3.5. Engagement Duration Modelling

Attacker engagement duration τ is treated as a random variable. The expected engagement time is as follows:

$$\mathbb{E}[\tau] = \sum_{t=0}^{\infty} t P(\tau = t), \quad (19)$$

where effective deception strategies increase $\mathbb{E}[\tau]$ compared to non-deceptive baselines [17,28].

4. Methodological Approach and Experimental Design

This study adopts a theory-informed engineering design-and-evaluation methodology in which the framework is treated as the principal research artefact. The methodological process comprises five linked stages: (i) identifying the operational problem and deriving

design requirements from the limitations established in the literature; (ii) synthesising the closed-loop architecture and specifying component responsibilities and interfaces; (iii) developing the separate computational models required by the detection, deception, and attribution subsystems; (iv) implementing and integrating these models and services as an executable prototype; and (v) evaluating both component performance and end-to-end framework behaviour in a controlled cyber-range environment. This sequence makes it clear that mathematical modelling is one stage within the construction of the broader framework, rather than an alternative description of the framework itself.

To ensure reproducibility and methodological transparency, all experiments are conducted within a controlled cyber-range environment that enables repeatable attacker–defender interactions under consistent network and system configurations. The experimental setup mirrors realistic enterprise scenarios while allowing fine-grained control over attacker behaviour, deception deployment, and telemetry collection. This design enables precise measurement of system responses to diverse attack strategies and facilitates direct comparison across experimental runs.

Detection performance is evaluated using standard classification metrics alongside latency measurements, capturing both analytical accuracy and real-time operational viability. Deception effectiveness is assessed through measures of believability and resistance to fingerprinting, as well as the system’s ability to sustain meaningful attacker interaction over time. Attacker engagement is analysed by examining interaction depth and duration, reflecting the framework’s capacity to delay adversarial progress and elicit behavioural evidence. For attribution, scripted attack steps provide ATT&CK-aligned ground-truth labels and the pipeline is evaluated for end-to-end generation of technique hypotheses. However, the archived evaluation preserves only aggregate summaries and not the per-technique predictions, posterior vectors, session-duration strata, or similarity-only baseline outputs required for class-level metrics, confusion matrices, calibration analysis, or ablation. Accordingly, attribution is evaluated as implementation feasibility rather than as a fully validated classification result.

The experimental design is explicitly aligned with the component-level mathematical formulations presented in Section 3. Detection metrics evaluate the hybrid ensemble model, engagement and intelligence measures evaluate the reinforcement learning policy and reward formulation, and attribution outcomes evaluate the probabilistic and similarity-based attribution models. System-level measures, including orchestration reliability, reconfiguration success, latency, and end-to-end engagement, evaluate whether these component models operate coherently when integrated within the framework. By separating component-model validation from framework-level evaluation, the methodology provides an evidence-driven assessment of both analytical performance and practical closed-loop operation. Risk-related evaluation is correspondingly interpreted as a testbed-level safety and operational-cost assessment rather than as a comprehensive financial, legal, or enterprise-risk analysis.

To provide a clear overview of the experimental process, Figure 3 illustrates the end-to-end methodology and evaluation workflow used in this study.

Figure 3 illustrates the end-to-end experimental workflow used to evaluate the proposed autonomous cyber deception framework. The workflow captures the interaction between the controlled cyber-range environment, execution of adversarial scenarios, real-time detection, adaptive deception, behavioural telemetry collection, attribution, and performance evaluation across repeated experimental runs. This representation demonstrates how the experimental design aligns with the closed-loop operation of the proposed framework and supports systematic, reproducible evaluation.

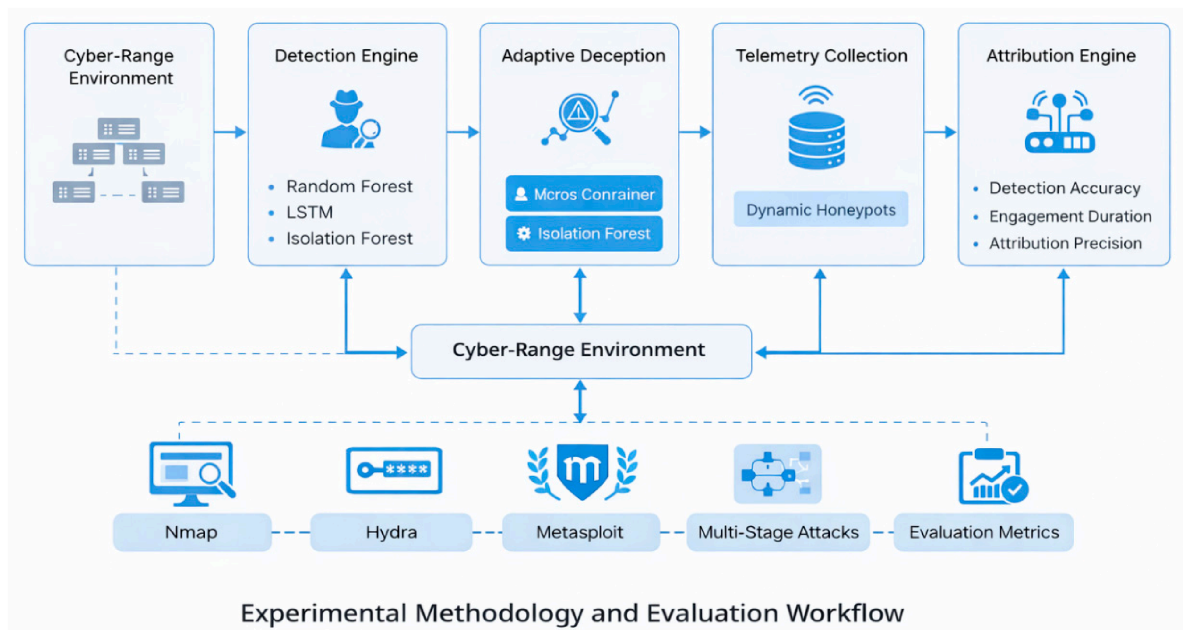


Figure 3. Experimental methodology and evaluation workflow.

4.1. Experimental Environment and System State Formalisation

Experiments were conducted in a controlled cyber-range environment [6]. The environment emulates a small enterprise-like topology using VirtualBox virtual machines interconnected through segmented virtual networks. Deception services were deployed using containerised Cowrie and Dionaea instances, enabling controlled instantiation, mutation, and isolation.

At every controller decision point, raw telemetry was converted to the finite state tuple defined in Equation (6) and Table 1. Detection confidence was obtained from the hybrid detector; the attack phase was assigned from the current scripted scenario step and ATT&CK-aligned evidence; the deception-configuration variable was read from the orchestrator; engagement depth was derived from cumulative distinct behavioural events; and resource utilisation was calculated from the maximum normalised CPU, memory, container, and orchestration-latency measurement. This mapping ensures that every observed interaction state has an explicit discrete representation within the 432-state design space.

The controller selected one of the six actions defined in Section 3.2.1. The orchestrator returned action success or failure, after which the next telemetry window generated s_{t+1} and the reward components in Equation (8). Transition counts $N(s,a,s')$ were retained for diagnostic estimation of Equation (5), although tabular Q-learning did not require an explicit transition matrix during action-value updating.

4.2. Offline Benchmark Data and Component-Model Development

The NSL-KDD, CICIDS2017, UNSW-NB15, and ToN-IoT datasets were used as offline sources for component-model development because they provide complementary traffic characteristics and attack categories [11–13]. These datasets were not concatenated into a common test set, and the 95.4% accuracy reported in Section 5.2 was not calculated by pooling records from them.

The study followed a two-stage evaluation design. Stage 1 used the benchmark datasets for offline training, feature development, and model selection at the component level. Stage 2 evaluated the integrated detector–deception–attribution prototype using labelled events generated during repeated cyber-range attack scenarios. The quantitative accuracy, precision, recall, F1-score, false-positive rate, and latency reported in this paper

refer to Stage 2. Accordingly, the manuscript does not present the 95.4% value as a cross-dataset average or as evidence of uniform performance on all four benchmark datasets. Dataset-specific benchmark generalisation would require a separate per-dataset evaluation and is outside the evidential scope of the reported integrated-system result.

- Supervised components were trained on labelled samples.
- Anomaly detectors were trained on benign traffic.
- Temporal models were trained on sequentially ordered samples to preserve attack progression. The implementation followed the gated recurrence in Equation (4), maintaining separate hidden and cell states across each ordered sequence, with the softmax layer in Equation (4) producing class probabilities.

Reinforcement-Learning Evaluation Scope

The revised manuscript separates the reproducibility of the controller design from the evidential status of the historical training trace. Section 3.2 now completely specifies the finite state space (432 states), six-action set, transition semantics, terminal conditions, discount factor, reward components, normalisation constants, safety override, and coefficients $\lambda_1 = 0.35$, $\lambda_2 = 0.45$, and $\lambda_3 = 0.20$. These details make the MDP and reward design technically evaluable and implementable.

The prototype was originally explored through a 100-episode single-run trajectory. Because the archived record does not contain the complete learning-rate and exploration metadata needed to reproduce that trajectory, the curve is excluded from the quantitative results and is not used to claim convergence, stability, optimality, or robustness. The state/action/reward specification reported above is therefore a transparent controller-design specification, whereas the retained numerical outcomes in Section 5.4 evaluate orchestration feasibility rather than learned-policy performance. A future policy-learning evaluation must report α , the complete epsilon-greedy schedule, random seeds, multi-run uncertainty, a prespecified stability criterion, and held-out policy performance.

4.3. Attack Scenarios and Adversarial Workloads

Evaluation scenarios emulate common attacker workflows, including:

- Nmap reconnaissance and service enumeration;
- Hydra brute-force authentication attempts;
- Metasploit exploitation and post-exploitation;
- Scripted multi-stage intrusion combining reconnaissance, exploitation, and lateral-movement actions.

The fourth workload was the scripted multi-stage intrusion described above; no separate Cobalt Strike scenario was evaluated. Each scenario was executed repeatedly under identical initial conditions to support aggregation and reduce variance from stochastic attacker behaviour.

4.4. Evaluation Metrics

The detection performance for the integrated prototype was assessed from the labelled cyber-range evaluation events using accuracy, precision, recall, F1-score, false-positive rate, and detection latency. True-positive, false-positive, true-negative, and false-negative counts were derived from the scripted scenario ground truth and the detector outputs recorded during the repeated cyber-range runs; they were not obtained from a pooled four-dataset benchmark test set.

Let TP, FP, TN, FN denote true positives, false positives, true negatives, and false negatives. Metrics are defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \text{Precision} = \frac{TP}{TP + FP}, \text{Recall} = \frac{TP}{TP + FN} \dots \quad (20)$$

Deception effectiveness was evaluated using:

- Attacker engagement duration τ (Equation (15));
- Interaction depth (e.g., command diversity, privilege escalation attempts);
- Fingerprinting resistance (the fraction of sessions in which the attacker did not disengage due to detection of deception artefacts).

Attribution performance was measured by comparing inferred techniques $\hat{T}_{\text{Bayes}}$ (Equation (14)) against ground-truth technique labels derived from scripted attack steps and known TTP mappings.

4.5. Experimental Protocol

For each run:

1. The environment was reset to a clean baseline state.
2. Detection models operated continuously.
3. Upon detecting malicious activity, the deception controller selected and executed deception actions.
4. Behavioural telemetry was collected throughout each session, transformed into structured behavioural evidence, and used for attribution once sufficient evidence was observed.
5. The results were aggregated across repeated runs to ensure stability and reliability of evaluation.

Benchmark-dataset development and cyber-range evaluation were kept analytically separate. No records from NSL-KDD, CICIDS2017, UNSW-NB15, or ToN-IoT were counted as cyber-range test observations, and the integrated-system confusion counts were computed only from the labelled cyber-range events.

The 20 repeated-run latency observations constitute one experimental dataset. The same dataset is used to discuss both absolute detection latency and run-to-run consistency; it is not treated as two independent experiments.

5. Results and Analysis

This section analyses empirical results in terms of detection robustness, adaptive deception orchestration, engagement outcomes, and attribution reliability under uncertainty. Rather than reporting metrics in isolation, the results are interpreted with respect to the optimisation objectives and probabilistic formulations introduced in Section 3.

5.1. Threat Detection Performance, Latency, and Ensemble Behaviour

The hybrid detection engine defined in Equation (1) demonstrated strong and consistent performance across the evaluated attack scenarios. Figure 4 presents the single set of detection-latency measurements obtained from 20 repeated experimental runs. The mean detection latency was 85 ms, and every observed run remained below 0.1 s, indicating that the detector can provide a prompt control signal for downstream deception orchestration under the controlled cyber-range conditions.

Figure 4 is analysed jointly in terms of latency magnitude and run-to-run consistency; these are two interpretations of the same repeated-run experiment rather than evidence from separate experiments. The limited dispersion across the 20 observations indicates stable inference and messaging-pipeline execution within the evaluated workload. La-

tency magnitude and consistency are therefore reported together in this subsection and represented by a single figure. This evidence supports near-real-time triggering in the testbed, but it should not be interpreted as a production-scale latency distribution under high concurrency or substantially different network-load conditions.

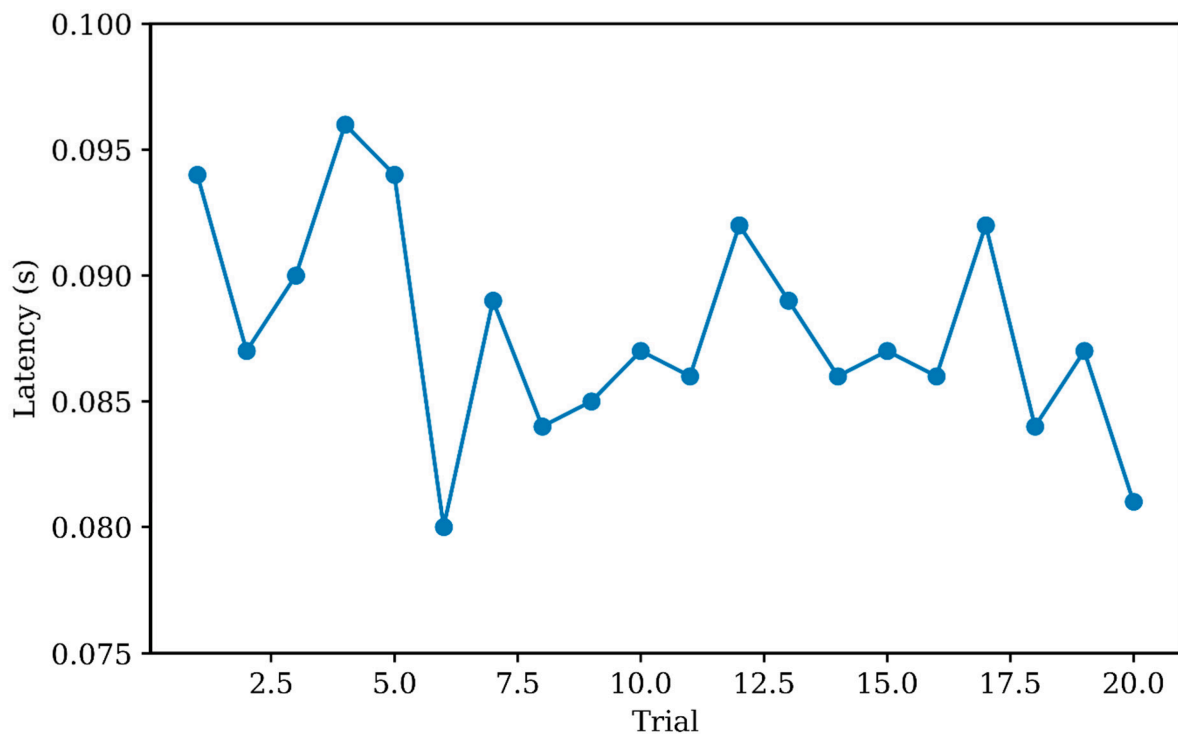


Figure 4. Detection latency across 20 repeated experimental runs. The single series is used to evaluate both absolute latency and run-to-run stability of the hybrid detection engine.

5.2. Integrated Cyber-Range Detection Performance

Table 2 reports the classification performance of the integrated detection engine on the labelled cyber-range evaluation events. The values should not be interpreted as a pooled result across NSL-KDD, CICIDS2017, UNSW-NB15, and ToN-IoT. Those datasets supported offline component-model development, whereas the metrics in Table 2 quantify end-to-end detection under the reconnaissance, brute-force, exploitation, and multi-stage scenarios executed in the controlled cyber range. The 4.6% false-positive rate is operationally relevant because false alerts may trigger unnecessary deception actions, while the mean detection latency of 85 ms indicates that the inference and messaging pipeline can provide a sufficiently prompt trigger for downstream orchestration under the evaluated testbed conditions.

Table 2. Integrated cyber-range detection performance metrics.

Metric	Value
Accuracy	0.954
Precision	0.936
Recall	0.947
F1-Score	0.941
False-Positive Rate	0.046
Mean Time to Detection	0.085 s

On the integrated cyber-range evaluation set, the proposed hybrid ensemble achieved 95.4% accuracy, 93.6% precision, 94.7% recall, and a 94.1% F1-score. These values summarise detection across the labelled scenario events generated in the controlled testbed and are distinct from benchmark-dataset validation results. The findings indicate that combining supervised classification, anomaly detection, and temporal modelling provided a robust trigger signal across the evaluated attack workflows. However, because the four offline benchmark datasets differ in feature spaces, class distributions, and attack coverage, this cyber-range result must not be interpreted as demonstrating an identical performance on each benchmark dataset.

Figure 5 visualises the integrated prototype's classification performance on the labelled cyber-range evaluation events. It is a graphical representation of the same accuracy, precision, recall, and F1-score values reported in Table 2 and does not represent a separate cross-dataset experiment.

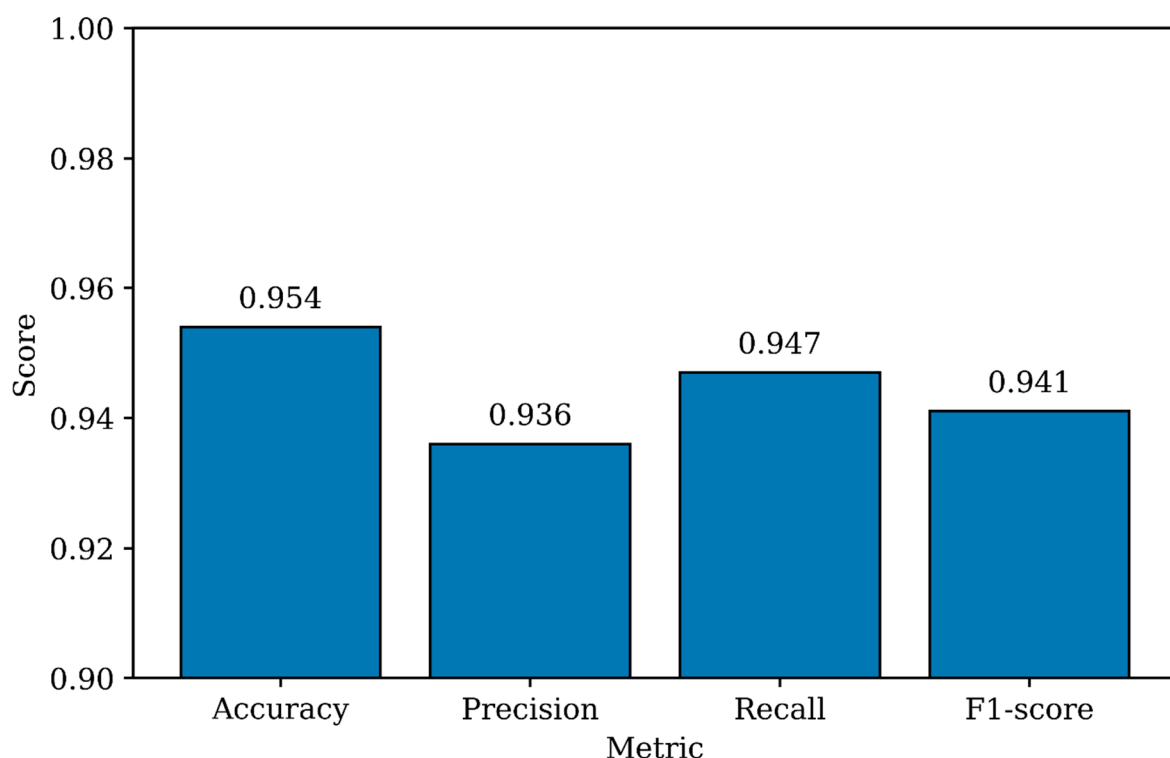


Figure 5. Integrated cyber-range detection performance in terms of accuracy, precision, recall, and F1-score.

5.3. Temporal Sensitivity in Multi-Stage Detection

The LSTM component contributed to detection in sequential and multi-stage scenarios by increasing confidence as context accumulated [7,8]. Empirically, temporal modelling reduced misclassification during transient benign anomalies while maintaining responsiveness to progressive attack sequences.

5.4. Adaptive Deception Controller: Operational Feasibility

The revised evaluation does not claim formal reinforcement-learning convergence. Instead, it assesses whether the bounded deception controller could execute policy-selected actions reliably within the controlled cyber-range environment. Operationally, the system achieved:

- 92% dynamic reconfiguration success;
- 100% honeypot deployment reliability.

These outcomes demonstrate the implementation feasibility of automated deception orchestration under the evaluated testbed conditions. They are execution-level measures of the policy interface and orchestrator, not evidence that the Q-learning policy converged or outperformed alternative policies. Policy-learning efficacy requires the fully controlled, multi-seed protocol described as future work in Section Reinforcement-Learning Evaluation Scope.

5.5. Attacker Engagement Dynamics and Intelligence Yield

Observed mean engagement durations were:

- Nmap reconnaissance: 5.5 s;
- Hydra brute-force: 28 s;
- Metasploit exploitation: 140 s;
- Multi-stage intrusion: 280 s.

These results indicate a substantial increase in expected engagement time $\mathbb{E}[\tau]$ (Equation (15)) relative to static or non-deceptive baselines. Engagement gains were most pronounced for higher-sophistication attacks, consistent with the hypothesis that adaptive deception is more effective against attackers capable of identifying static decoys [17,28].

Longer engagements also increased behavioural diversity (commands, timing patterns, privilege escalation attempts), supporting higher discriminative power for attribution.

Figure 6 illustrates the average attacker engagement duration for the four workloads specified in Section 4.3: Nmap reconnaissance, Hydra brute-force, Metasploit exploitation, and the scripted multi-stage intrusion. The fourth bar corresponds to the multi-stage workflow and not to Cobalt Strike. Engagement time increases substantially for the multi-stage scenario, demonstrating the ability of adaptive deception to sustain interaction and elicit richer behavioural evidence under the evaluated conditions.

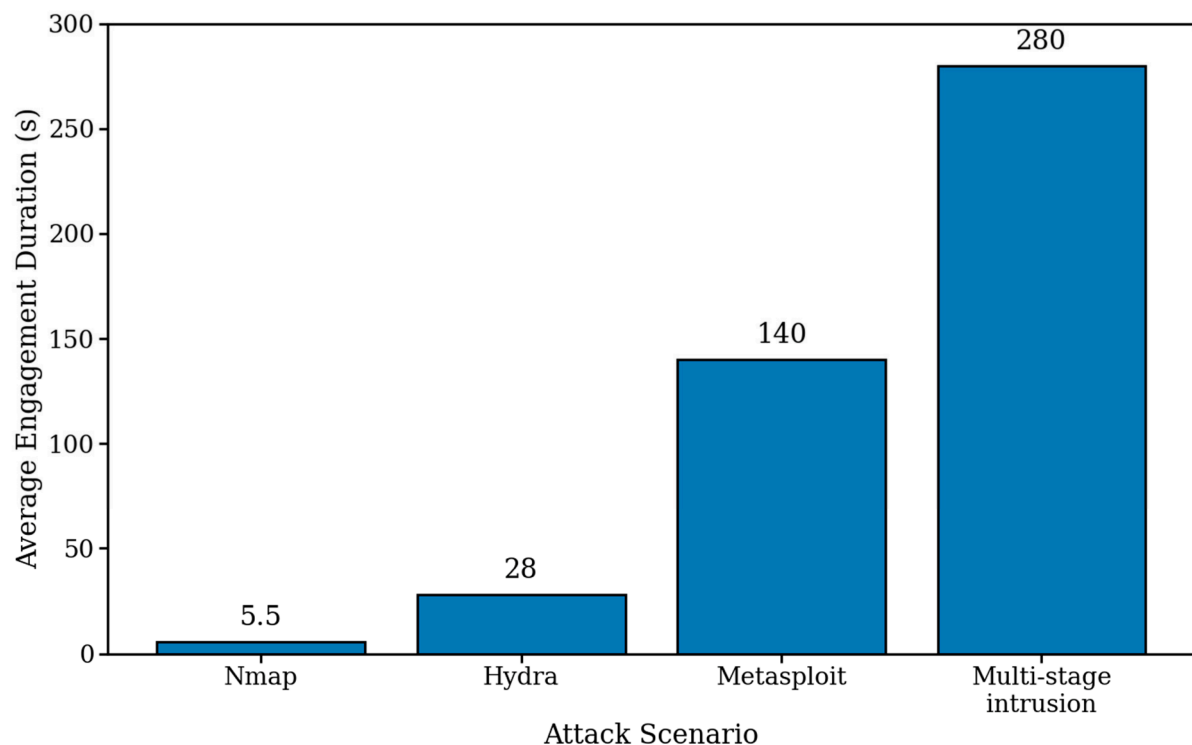


Figure 6. Average attacker engagement duration for the four evaluated workloads: Nmap reconnaissance, Hydra brute-force, Metasploit exploitation, and the scripted multi-stage intrusion.

5.6. Attribution Evaluation Scope and Evidence Limitations

The attribution pipeline transforms deception-derived commands, timing patterns, tool use, and interaction transitions into structured evidence and maps that evidence to MITRE ATT&CK technique hypotheses through similarity scoring and Bayesian updating. Scripted attack steps provide the technique-level ground truth used during the controlled cyber-range experiments.

The present attribution evaluation was designed primarily to verify the end-to-end functional integration of the attribution pipeline rather than to provide a class-resolved benchmark. The retained experimental evidence confirms that deception-derived telemetry can be transformed into structured behavioural features, mapped to MITRE ATT&CK technique hypotheses, and incorporated into the closed-loop decision process. However, session-level predicted labels, complete posterior probability distributions, and technique-specific ground-truth annotations were not available in a form that permitted reconstruction of a confusion matrix or calculation of per-technique precision, recall, and F1-score.

In addition, outputs from a matched similarity-only comparator were unavailable. Consequently, the independent contribution of Bayesian updating relative to similarity-based scoring could not be quantified. The attribution findings should therefore be interpreted as evidence of architectural feasibility and functional integration, rather than as a definitive assessment of technique-level classification performance or proof that the Bayesian stage provides a measurable improvement.

A more comprehensive evaluation should retain session-level predictions, posterior probabilities, ground-truth technique labels, attack-stage information, and interaction duration. Future experiments should compare similarity-only, Bayesian-only, and combined attribution pipelines under identical experimental conditions and report per-technique precision, recall, F1-score, confusion matrices, calibration measures, confidence intervals, and performance stratified by attack stage and session duration.

The evidence retained from the present prototype supports a narrower conclusion: deception telemetry can be processed through the complete attribution pathway, matched to ATT&CK-labelled scripted behaviours, and returned to the closed-loop architecture as a structured technique hypothesis. This verifies functional integration and data flow, but it does not establish which techniques are reliably recognised, which are systematically confused, or how performance changes with attack stage and interaction duration.

A technically adequate follow-up evaluation must retain session-level predictions and posterior vectors and report, for every technique, support, precision, recall, F1-score, and a complete confusion matrix. The results should be stratified by attack stage and session-duration band, and uncertainty should be assessed through confidence intervals and probability-calibration measures. The attribution models should also be compared under matched conditions using at least three variants: similarity scoring alone, Bayesian inference alone, and the combined pipeline. A paired significance test or bootstrap comparison should be used to determine whether any observed improvement is attributable to the Bayesian evidence-update step rather than to sampling variation.

Accordingly, the attribution engine is retained as a framework component and methodological contribution, but its present empirical status is explicitly limited to proof-of-concept integration. It is not used as a headline performance outcome in the Abstract, Section 6, or Section 8.

5.7. Honeypot Deployment Reliability

Table 3 summarises the reliability and responsiveness of the deception engine. All scheduled honeypot deployments completed successfully, with a mean activation time of 1.2 s, indicating that the container-based services could be instantiated promptly under

the evaluated cyber-range conditions. The 88% fingerprinting-resistance rate indicates that most evaluated sessions did not terminate because the attacker identified a deception artefact, thereby supporting the believability of the deployed assets. The 92% dynamic-reconfiguration success rate further indicates that the orchestrator adapted the deception configuration reliably as attacker behaviour changed during extended sessions. These findings demonstrate operational feasibility within the testbed; they do not establish equivalent reliability under production-scale concurrency or previously unseen workloads.

Table 3. Honeypot deployment outcomes.

Parameter	Value
Deployment Success	1
Activation Time	1.2 s
Fingerprinting Resistance	0.88
Dynamic Reconfiguration	0.92

5.8. Comparative Evaluation Against Baseline Models

Figure 7 compares the hybrid detector with the Random Forest, support vector machine (SVM), and K-means baselines. Under the evaluated cyber-range conditions, the hybrid approach achieved higher detection accuracy and lower latency than the individual baselines. This improvement is consistent with the complementary roles of supervised classification, statistical anomaly detection, reconstruction-based evidence, and temporal sequence modelling. The comparison supports the use of ensemble fusion for real-time triggering in the prototype, although broader dataset-specific and production-scale evaluation is required before generalising the observed advantage.

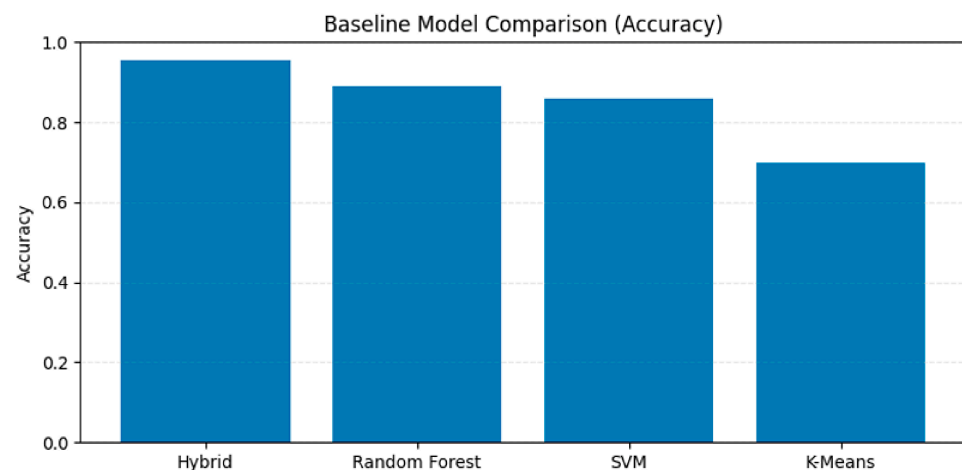


Figure 7. Comparative detection performance against baseline models.

Figure 7 compares the proposed hybrid detection model with baseline approaches, including Random Forest, SVM, and K-means. The hybrid model demonstrates superior detection accuracy and reduced latency, highlighting the effectiveness of integrating supervised, anomaly-based, and temporal detection techniques.

5.9. Attack Timeline Reconstruction

Figure 8 reconstructs the chronology of a representative multi-stage experimental run from time-stamped detection, orchestration, honeypot, and attribution logs. Its purpose is not to report an additional classification metric, but to examine whether the independently evaluated components operate in the required causal order as a single closed loop.

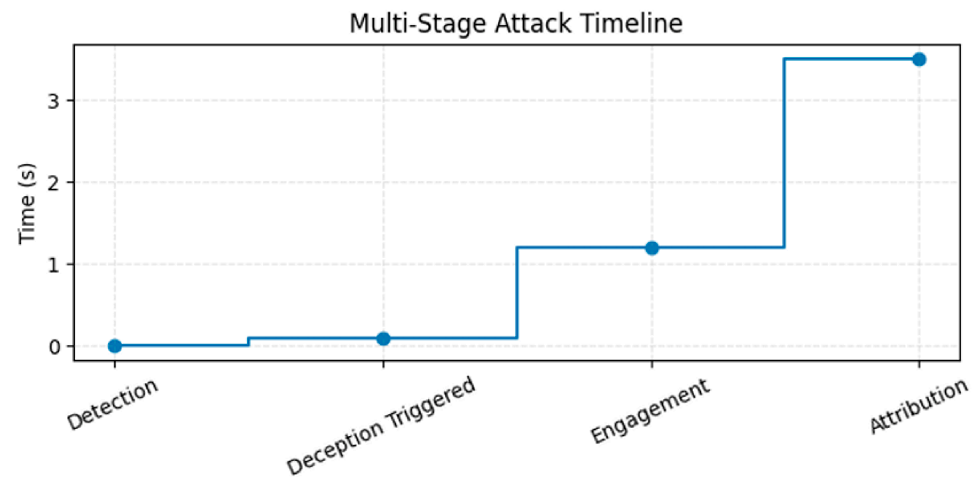


Figure 8. Representative reconstructed attack timeline and closed-loop system response.

The plotted milestones show an alert at $t = 0$, a deception trigger at approximately 0.1 s, observable attacker engagement with the deployed decoy at approximately 1.2 s, and an attribution output at approximately 3.5 s. The first transition is consistent with the mean detection time of 85 ms reported in Table 2, while the engagement milestone is consistent with the 1.2 s honeypot activation time reported in Table 3. This agreement between component-level and end-to-end measurements supports the internal consistency of the experimental pipeline.

Operationally, the timeline separates three different forms of delay. Detection latency determines how quickly malicious activity is recognised; orchestration and activation latency determine how quickly the selected deception action becomes available; and evidence-accumulation latency determines when sufficient behavioural observations exist for technique-level attribution. In this representative run, the latter two stages account for most of the end-to-end time. This indicates that further reductions in detector latency alone would yield limited system-level benefit unless deception deployment and behavioural processing are also optimised.

The interval between engagement and attribution also exposes a time–information trade-off. Producing an attribution decision immediately after redirection relies on sparse evidence and may reduce posterior confidence, whereas waiting for additional commands, tool-use patterns, timing observations, and interaction transitions increases evidential richness but delays the analyst-facing result. A practical implementation should therefore employ a calibrated confidence-based stopping rule that releases an attribution result when the posterior probability and evidence coverage exceed validated thresholds, rather than relying only on a fixed observation window.

Figure 8 should be interpreted as a representative feasibility trace rather than as a statistical latency distribution or proof that every attack is contained before lateral movement. It validates temporal ordering and end-to-end coordination under the controlled cyber-range conditions. Generalisation to operational networks would require latency distributions and tail percentiles, concurrent-session stress testing, and explicit analysis of delayed, failed, or incomplete deception and attribution events.

5.10. System-Level Closed-Loop Validation

Beyond component-level metrics, the primary finding is the closed-loop synergy: low-latency detection enabled early deception activation; adaptive deception increased engagement and behavioural diversity; and richer behavioural evidence improved attribution confidence. These results support the central thesis that defensive utility is maximised

through integrated optimisation of detection fidelity, engagement, and attribution under constraints.

6. Discussion

6.1. Detection Performance in Relation to Intrusion-Detection Research

The detection results should be interpreted against the long-standing intrusion-detection literature reviewed in Section 2.1. Foundational work by Axelsson and the NIST guidance on intrusion detection emphasised that an IDS must balance sensitivity, false alarms, and operational usability rather than optimise classification accuracy in isolation [1,4]. The 95.4% accuracy, 94.1% F1-score, and 4.6% false-positive rate obtained on the labelled cyber-range evaluation set indicate a favourable balance under the controlled test conditions. These values are not a macro-average or pooled score across NSL-KDD, CICIDS2017, UNSW-NB15, and ToN-IoT; the benchmark datasets were used for offline component development. Consequently, the result supports the detector's role as a perception and triggering layer in the integrated testbed, but it does not establish uniform generalisation across the four heterogeneous benchmark datasets. The false-positive rate also remains operationally meaningful because every erroneous trigger may initiate unnecessary deception deployment, consume resources, or expose the presence of defensive controls. This qualification is consistent with the anomaly-detection concerns regarding alert burden and environmental drift identified by Gates and Taylor and by collaborative IDS research [5,6].

The contribution of temporal modelling is also consistent with the literature on slow and multi-stage attack detection. Ring et al. showed that slow scans and temporally distributed activity can evade snapshot-oriented analysis, while Gwon et al. demonstrated the value of LSTM-based sequence modelling for network intrusion detection [2,7,8]. In the present study, the LSTM component increased confidence as attack context accumulated and reduced misclassification during transient benign anomalies. The result supports the rationale for combining sequence information with supervised and anomaly-based evidence rather than relying on a single observation. It should not, however, be interpreted as proof that one architecture will perform uniformly across operational networks. The benchmark datasets used in this study—UNSW-NB15, CICIDS2017, and related AI-security corpora—support reproducible comparison [11–13], but they cannot fully reproduce production traffic, evolving legitimate behaviour, or adaptive evasion. Accordingly, the detection findings establish component feasibility and low-latency operation, while cross-site and drift-aware validation remains necessary for external generalisation.

A further implication is that the value of the detector arises from its position within the wider framework. Much of the IDS literature terminates at the alert-generation stage [1,4,6]. Here, the detection output is converted into a state variable and control signal that initiates a defensive action. This changes the interpretation of detection latency: 85 ms is important not merely because it is computationally efficient, but because it leaves time for the orchestration layer to deploy or reconfigure a deception asset before the adversary progresses. The results therefore reinforce the distinction made in Sections 1 and 2 between a component model and a framework: the hybrid detector performs bounded inference, whereas the framework determines how that inference changes the environment and produces additional evidence.

6.2. Adaptive Deception, Engagement, and Defender-Side Risk

The deception findings extend the honeypot literature reviewed in Section 2.2. Provos and Holz and the surveys by Nawrocki et al. describe the central trade-off between interaction fidelity, scalability, finger printability, and containment or management burden [14–16].

The observed 100% deployment reliability and 1.2 s activation time show that containerised deceptive services can be orchestrated rapidly in the experimental environment. The 88% fingerprinting resistance and 92% dynamic-reconfiguration success further suggest that adaptation can preserve believability more effectively than a purely static configuration. These values should nevertheless be interpreted carefully: 88% resistance also means that deception artefacts were recognised or followed by disengagement in a non-negligible proportion of sessions. The framework therefore reduces, but does not eliminate, fingerprinting risk, and its effectiveness will depend on the realism and diversity of the deployed services.

The reinforcement-learning formulation provides an operational bridge between earlier game-theoretic accounts of deception and executable adaptive control. Pawlick et al. conceptualise defensive deception as signalling under asymmetric information, while Alpcan and Başar formalise cyber defence as strategic decision-making in which defender and attacker choices are interdependent [17,19,20]. More recent work surveys how machine learning can operationalise these strategic ideas [21]. In the proposed framework, the reward function translates this theory into measurable objectives by combining engagement duration, information gain, resource cost, and defender-side risk. However, the original 100-episode single-run trajectory was insufficient to establish convergence and did not provide the complete hyperparameter and multi-seed evidence required for reproducibility. It has therefore been removed from the revised results rather than defended through retrospective parameter assignment. The present findings support the operational feasibility of executing bounded deception actions, as reflected in deployment reliability and dynamic-reconfiguration success, but they do not demonstrate policy convergence or universal optimality. A dedicated follow-up study must report the learning-rate and exploration schedules, discount factor, reward coefficients, state and action cardinalities, random seeds, convergence criterion, uncertainty intervals, and held-out greedy-policy performance. The revised finite MDP specification further makes this bridge technically explicit: a 432-state representation summarises the observable attacker–system context, six bounded actions define the controller’s authority, empirical transition counts characterise environment response, and the reward gives fixed numerical priority to intelligence gain, engagement, and defender-side cost. These design choices enable independent technical evaluation while the manuscript continues to avoid an unsupported convergence claim.

The engagement results provide the most direct evidence that deception serves as more than a diversion mechanism. Mean engagement increased from 5.5 s for reconnaissance to 28 s for brute-force activity, 140 s for exploitation, and 280 s for the multi-stage scenario. This progression indicates that richer attack workflows create more opportunities for the controller to present contextually plausible deceptive responses and to collect diverse commands, tool-use sequences, timing patterns, and privilege-escalation attempts. This interpretation is consistent with defensive-deception surveys that describe deception as a means of increasing attacker uncertainty and eliciting intelligence [17,18], and with the design-oriented view that deception should be constructed around attacker expectations and interaction pathways [28]. In the present study, longer engagement is therefore not treated as an objective in isolation. Its value lies in the additional behavioural evidence generated before the attacker disengages, provided that containment, exposure, resource, and configuration risks remain within the defined operational bounds.

This risk–information trade-off is central to the framework. High-interaction assets can yield detailed behavioural evidence but carry greater containment and resource burdens, whereas low-interaction assets are safer and cheaper but are easier to fingerprint [14–16]. The reward formulation makes this trade-off explicit by penalising unsafe or costly actions rather than maximising dwell time unconditionally. This is important because an apparently successful deception strategy could be operationally unacceptable if it permits pivoting,

exposes production information, or degrades availability. The reported results show that the controller maintained reliable deployment in the cyber range, but they should be understood as testbed-level evidence of operational feasibility, not as a complete enterprise risk assessment.

6.3. Behavioural Attribution and ATT&CK-Aligned Intelligence

The attribution component should be interpreted in relation to the structured threat-modelling literature rather than through an unsupported aggregate score. MITRE ATT&CK provides a common vocabulary for representing tactics and techniques, while cyber-threat-intelligence taxonomies emphasise the transformation of heterogeneous low-level observations into consistent and actionable knowledge [24,25]. The proposed pipeline operationalises this transformation by converting deception-derived commands, timing relationships, tool use, and interaction transitions into ATT&CK-aligned technique hypotheses. This design is conceptually compatible with ATTACK2VEC, which highlights the value of temporal and contextual relationships for understanding evolving attack behaviour [26].

The Bayesian layer remains theoretically motivated because attribution evidence is incomplete and accumulates over time; posterior updating provides a principled means of representing changing confidence under uncertainty [27]. Nevertheless, the present experiment does not provide the class-level predictions, probability vectors, calibration results, or similarity-only comparator required to show that this layer improves empirical performance. The manuscript therefore distinguishes architectural rationale from demonstrated effectiveness. It also retains the important boundary between technique-level and actor-level attribution: the former concerns classification of observed behaviour, whereas the latter would require broader infrastructure, identity, and longitudinal intelligence that are outside the scope of this framework.

The link between deception and attribution remains a framework-level contribution because the architecture uses deception to elicit evidence that can subsequently be represented in ATT&CK terms, rather than limiting attribution to retrospective analysis [24–27]. However, the current results support only the feasibility of this information flow. A rigorous claim about attribution reliability requires a retained session-level evaluation set, a per-technique confusion analysis, stratification by attack stage and engagement duration, and a controlled ablation comparing similarity-only, Bayesian-only, and combined inference. These requirements are now stated explicitly rather than inferred from a single aggregate value.

6.4. Closed-Loop Integration and the Framework-Level Contribution

At the framework level, the findings support the intelligence-driven defence principle that defensive decisions should be informed by the observed progression of an adversary campaign rather than by isolated events [3]. Figure 8 is important in this respect because it demonstrates the temporal ordering of the complete detect–deceive–observe–attribute cycle. The alert at approximately $t = 0$, deception triggering near 0.1 s, observable engagement near 1.2 s, and attribution output near 3.5 s show that the independently evaluated components can operate as a coordinated pipeline. This sequence is the empirical basis for describing the contribution as a closed-loop framework rather than a collection of models: detection changes the state, the policy changes the environment, the attacker generates new telemetry, and the attribution output updates situational understanding for subsequent action.

The timeline also reveals that end-to-end performance cannot be inferred from detection metrics alone. Once detection latency is reduced to tens of milliseconds, orchestration and evidence accumulation account for most of the delay. This produces a time–information trade-off that is not captured by conventional IDS accuracy measures. Immediate attri-

bution may be fast but weakly supported; delayed attribution may be more reliable but less timely. A practical implementation should therefore use calibrated evidence-coverage and posterior-confidence thresholds to determine when an analyst-facing conclusion is sufficiently credible. This interpretation extends the literature reviewed in Section 2 by showing that the operational objective is not simply faster detection or longer engagement, but coordinated optimisation of time-to-action, evidence quality, attribution confidence, and defender-side risk.

The results also clarify the bounded nature of autonomy in the proposed system. The detection subsystem provides perception, the reinforcement-learning controller selects an action from a predefined set, orchestration executes that action, and telemetry supplies feedback. This arrangement exhibits agentic properties at the framework level, but it does not create goals autonomously, perform unrestricted planning, or retrain all models continuously during runtime. The empirical findings therefore support the more precise description adopted in Section 2.4: a closed-loop adaptive architecture with bounded autonomous decision-making, rather than a general-purpose agentic AI or fully continual-learning system.

6.5. Operational Implications and Interpretation Boundaries

For security operations, the combined findings suggest that the framework's strongest contribution is the conversion of alerts into structured opportunities for action and intelligence collection. Low-latency detection enables timely redirection; adaptive deception can sustain interaction; and ATT&CK-aligned attribution can translate the resulting evidence into a vocabulary already used for threat hunting and defensive planning [24,25]. This creates a potential path from automated sensing to analyst decision support. Nevertheless, operational deployment should preserve human oversight for policy definition, risk thresholds, attribution review, and exceptional cases. The framework is best viewed as augmenting analysts with timely evidence and bounded automated actions, rather than replacing judgement in high-consequence security decisions.

Several interpretation boundaries follow from the literature comparison. First, the strong cyber-range detection performance does not resolve the dataset-shift and false-positive concerns documented in IDS research [5,6,11–13]. Second, the deception results do not remove the containment and fingerprinting trade-offs identified in honeypot studies [14–18]. Third, the present operational controller results do not constitute evidence of reinforcement-learning convergence or robustness against strategic policy manipulation and non-stationary attackers [20–23]. Fourth, the attribution pipeline has been demonstrated functionally but has not yet been validated with per-technique metrics, confusion matrices, probability calibration, or a similarity-only baseline; it therefore cannot support a claim of Bayesian improvement or uniform technique-level reliability [24–27]. Finally, Figure 8 is a representative feasibility trace rather than a distribution of end-to-end latencies. These boundaries motivate the ablation studies, concurrent-session stress testing, external-network validation, adversarial robustness assessment, and analyst-in-the-loop evaluation identified in Sections 7 and 8.

Overall, the results substantiate the central research proposition while also refining it. The evidence does not show that integration eliminates the limitations of detection, deception, or attribution. Instead, it shows that their limitations can be managed more coherently when the components are linked through explicit state, action, telemetry, risk, and feedback relationships. Relative to the key literature reviewed in Section 2, the framework's novelty lies less in any single learning algorithm than in the operational synthesis of established intrusion-detection, decision-theoretic deception, reinforcement-learning, and ATT&CK-aligned inference principles into an experimentally evaluated closed loop [3,17,21,24,27,28].

7. Limitations

The benchmark datasets and the cyber-range evaluation serve different methodological purposes. The current paper reports end-to-end cyber-range detection metrics but does not provide a per-dataset benchmark performance table or per-class confusion matrices for NSL-KDD, CICIDS2017, UNSW-NB15, and ToN-IoT. Therefore, no claim of uniform cross-dataset generalisation is made. A dedicated comparative benchmark study with dataset-specific preprocessing, held-out partitions, class-level metrics, and confusion matrices is required before such a claim can be supported.

Several limitations of this study should be acknowledged. First, the attribution evaluation is incomplete. The archived experiment did not retain session-level predicted labels, full posterior probability vectors, or outputs from a similarity-only comparator. Consequently, the former aggregate attribution result cannot be decomposed into per-technique precision, recall, F1-score, support, or a confusion matrix, and the contribution of Bayesian updating cannot be isolated through ablation. The current evidence therefore establishes only end-to-end attribution-pipeline feasibility. In addition, behaviour-based attribution depends on the availability and richness of observable evidence; short-lived, low-noise, or deliberately misleading attacks may produce sparse evidence and uncertain posteriors.

Second, the assumption of full observability adopted in the modelling framework represents an approximation of real-world attacker–defender interactions. In practice, such interactions are not strictly Markovian, and the use of a Markov Decision Process (MDP) serves as a tractable abstraction of an underlying Partially Observable Markov Decision Process (POMDP). While this simplification enables efficient learning and decision-making, partial observability may limit optimality guarantees, particularly in highly adaptive or deceptive adversarial environments.

Third, the reinforcement-learning evaluation is limited to operational proof-of-concept evidence. The original 100-episode single-run trajectory was too short to establish convergence and lacked the complete hyperparameter, random-seed, and stopping-criterion record required for independent reproduction. The trajectory has therefore been removed from the revised Results Section, and no convergence or optimality claim is made. Future validation should use longer training horizons, multiple independent seeds, prespecified stability criteria, confidence intervals, and exploration-disabled held-out evaluation.

The finite MDP also introduces modelling limitations. The 432-state discretisation and fixed reward coefficients (0.35, 0.45, and 0.20) make the controller interpretable and reproducible, but they compress continuous attacker behaviour and encode a particular defender preference. The present study does not provide a sensitivity analysis demonstrating that the selected thresholds or weights remain optimal under different workloads, organisations, or threat models.

Fourth, the current prototype does not implement continual online retraining of every component model. The detection and attribution models remain fixed during experimental evaluation, and reinforcement-learning parameter updates occur during the controlled training phase; at run time, adaptation consists of updating the observed state and behavioural evidence and selecting deception actions under the learned policy. Drift-aware incremental retraining, safeguards against catastrophic forgetting, and validated model-update policies are therefore reserved for future work.

Fifth, the experimental evaluation was conducted within a controlled cyber-range environment. Although this setting allows for reproducibility and precise measurement, absolute performance metrics may differ in real-world deployments due to background noise, environmental uncertainty, and diverse legitimate user behaviour. Nevertheless, the relative improvements achieved through the proposed closed-loop optimisation framework are expected to generalise beyond the experimental setup.

8. Conclusions and Future Work

This paper presented an autonomous AI-driven cyber deception framework that integrates real-time threat detection, adaptive deception, and behaviour-based attribution within a unified closed-loop architecture. By tightly coupling detection outputs with downstream deception and attribution processes, the framework advances beyond traditional reactive defence mechanisms toward proactive and intelligence-driven adversarial engagement. Empirical results demonstrate that the hybrid detection engine achieves high detection accuracy with low operational latency, enabling timely activation of deception strategies. The bounded deception controller executed automated deployment and reconfiguration actions with high operational reliability, while the engagement mechanisms increased the behavioural evidence available for analysis. These operational outcomes should not be interpreted as proof of reinforcement-learning convergence or policy optimality. The attribution component demonstrated functional generation of ATT&CK-aligned technique hypotheses within the closed loop, but the present archived evaluation does not establish per-technique reliability or improvement from Bayesian updating; those claims require a retained class-level dataset and controlled ablation study.

Future work will focus on extending the proposed framework along several technically substantive directions. A primary avenue involves reformulating the deception control problem from the current Markov Decision Process abstraction to an explicit Partially Observable Markov Decision Process formulation. While the MDP-based approach offers computational efficiency, it assumes full state observability that rarely holds in realistic adversarial environments. A POMDP-based model would enable belief-state reasoning over latent attacker intent and partially observable system states, allowing deception policies to be optimised under uncertainty. Such an extension is expected to improve robustness against adaptive and strategic adversaries, although it introduces increased computational complexity, motivating the investigation of approximate, hierarchical, or factored solution methods suitable for real-time operation.

Scalability constitutes another critical direction for future research. Although the current implementation targets a bounded cyber-range environment, practical deployment necessitates support for distributed, cloud-native, and heterogeneous enterprise infrastructures. Future work will explore decentralised and hierarchical control architectures, multi-agent reinforcement learning, and policy-sharing mechanisms to enable large-scale orchestration of deceptive assets across dynamically evolving environments. Particular emphasis will be placed on reproducible multi-seed training, explicit disclosure of learning and exploration schedules, prespecified policy-stability criteria, held-out evaluation, and low-latency coordination across distributed components.

Improving deception realism is also a key objective. Manually engineered or static decoys are increasingly susceptible to detection by sophisticated adversaries employing reconnaissance and fingerprinting techniques. The integration of generative approaches for automated decoy artefact generation, adaptive service emulation, and dynamic content synthesis represents a promising direction for increasing believability. Conditioning such generative mechanisms on observed attacker behaviour and inferred intent may enable context-aware and continuously evolving deception environments, thereby prolonging attacker engagement and enriching the behavioural evidence available for attribution.

A priority for future work is a fully reproducible attribution benchmark. Each cyber-range session will retain the ground-truth technique, predicted technique, similarity scores, posterior probability vector, attack stage, interaction duration, and random seed. The evaluation will report per-technique support, precision, recall, F1-score, confusion matrices, confidence intervals, and probability-calibration measures. Matched ablations will compare similarity-only, Bayesian-only, and combined inference, with paired statistical testing to

quantify the incremental contribution of Bayesian updating. The resulting outputs will then be integrated with Security Operations Centre tooling for response prioritisation, threat hunting, analyst-in-the-loop validation, and mapping of inferred techniques to response playbooks.

Author Contributions: M.S.: Conceptualization, Methodology, Software, Implementation, Data Curation, Formal Analysis, Investigation, Visualisation, Writing—Original Draft, and Validation. M.H.S.: Conceptualization, Supervision, Validation, Methodology, Formal Analysis, Investigation, Visualisation, Writing—Review and Editing, and Project Administration. R.U.: Supervision, Validation, Methodology, Formal Analysis, Resources, Investigation, Writing—Review and Editing, and Technical Guidance. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data used to support the findings of this study are generated within a controlled cyber-range environment as described in the paper. Any supporting datasets or experimental configurations are available from the corresponding author upon reasonable request.

Acknowledgments: The authors acknowledge the School of Computing, Engineering and Physical Sciences at the University of the West of Scotland for providing the infrastructure used in this research. The manuscript underwent a complete author-led technical and language review. Grammarly and a generative-AI language model were used only as supplementary aids for grammar, clarity, and response drafting; they were not used to generate experimental data, execute analyses, or determine scientific conclusions. The authors reviewed and verified every revision and take full responsibility for the accuracy and integrity of the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Axelsson, S. Intrusion detection systems: A survey and taxonomy. *Comput. Netw.* **1999**, *31*, 713–733.
2. Ring, M.; Landes, D.; Hotho, D. Detection of slow port scans: A survey. *IEEE Commun. Surv. Tutor.* **2018**, *20*, 231–259.
3. Hutchins, E.; Cloppert, M.; Amin, R. Intelligence-driven computer network defense informed by analysis of adversary campaigns. In *Leading Issues in Information Warfare & Security Research*; Academic Publishing International Ltd.: Reading, UK, 2011.
4. Scarfone, K.; Mell, P. *Guide to Intrusion Detection and Prevention Systems (IDPS)*; NIST SP 800-94; National Institute of Standards and Technology: Gaithersburg, MD, USA, 2007.
5. Gates, C.; Taylor, J. Challenging the anomaly detection paradigm: A provocative discussion. In *Proceedings of the New Security Paradigms Workshop 2006*, Wadern, Germany, 19–22 September 2006.
6. Vasilomanolakis, S.; Karuppayah, M.; Mühlhäuser, M. Taxonomy and survey of collaborative intrusion detection. *ACM Comput. Surv.* **2015**, *47*, 1–33. [[CrossRef](#)]
7. Gwon, H.; Lee, C.; Keum, R.; Choi, H. Network Intrusion Detection based on LSTM and Feature Embedding. *arXiv* **2019**, arXiv:1911.11552. [[CrossRef](#)]
8. Bono, F.M.; Cinquemani, S.; Chatterton, S.; Pennacchi, P. A deep learning approach for fault detection and RUL estimation in bearings. In *Proceedings of the SPIE Smart Structures and Materials + Nondestructive Evaluation, 2022, Long Beach, CA, USA, 6–11 March 2022*; SPIE: Bellingham, WA, USA, 2022; Volume 12049, p. 1204908. [[CrossRef](#)]
9. Wang, W.; Zhu, M.; Wang, J.; Zeng, X.; Yang, Z. End-to-end encrypted traffic classification with deep learning. *IEEE Trans. Inf. Forensics Secur.* **2018**, *13*, 2526–2539.
10. Creech, G.; Hu, J. A semantic approach to host-based intrusion detection systems using contiguous and discontiguous system call patterns. *IEEE Trans. Comput.* **2014**, *63*, 807–819. [[CrossRef](#)]
11. Moustafa, N.; Slay, J. UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). In *Proceedings of the 2015 Military Communications and Information Systems Conference (MilCIS)*, Canberra, Australia, 10–12 November 2015.
12. Sharafaldin, I.; Habibi Lashkari, A.; Ghorbani, A.A. Toward generating a new intrusion detection dataset and intrusion traffic characterization. In *Proceedings of the ICISSP 2018*, Madeira, Portugal, 22–24 January 2018.
13. Moustafa, N. A new distributed architecture for evaluating AI-based security systems. *IEEE Access* **2020**, *8*, 165660–165678.
14. Provos, N.; Holz, T. *Virtual Honeypots: From Botnet Tracking to Intrusion Detection*; Addison-Wesley: Hoboken, NJ, USA, 2007.

15. Nawrocki, M.; Wählisch, M.; Schmidt, T.C.; Keil, C.; Schönfelder, J. A Survey on HoneyPot Software and Data Analysis. *arXiv* **2016**, arXiv:1608.06249.
16. Pawlick, J.; Colbert, E.; Zhu, Q. A Game-Theoretic Taxonomy and Survey of Defensive Deception for Cybersecurity and Privacy. *ACM Comput. Surv.* **2019**, *52*, 1. [[CrossRef](#)]
17. Fraunholz, D.; Duque Antón, S.; Lipps, C.; Reti, D.; Krohmer, D.; Pohl, F.; Tammen, M.; Schotten, H.D. Demystifying Deception Technology: A Survey. *arXiv* **2018**, arXiv:1804.06196. [[CrossRef](#)]
18. Pawlick, A.; Zhu, Q. Deception by design: Evidence-based signaling games for network defense. In Proceedings of the ACM CCS, Denver, CO, USA, 12–16 October 2015.
19. Alpcan, T.; Başar, T. *Network Security: A Decision and Game-Theoretic Approach*; Cambridge University Press: Cambridge, UK, 2010.
20. Zhu, M.; Anwar, A.H.; Wan, Z.; Cho, J.-H.; Kamhoua, C.; Singh, M.P. Game-Theoretic and Machine Learning-based Approaches for Defensive Deception: A Survey. *arXiv* **2021**, arXiv:2101.10121. [[CrossRef](#)]
21. Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*; MIT Press: Cambridge, MA, USA, 2018.
22. Schlette, D.; Caselli, M.; Pernul, G. A Comparative Study on Cyber Threat Intelligence: The Security Incident Response Perspective. In *IEEE Communications Surveys & Tutorials*; IEEE: New York, NY, USA, 2021; Volume 23, pp. 2525–2556. [[CrossRef](#)]
23. MITRE Corporation, MITRE ATT&CK Framework. Available online: <https://attack.mitre.org> (accessed on 1 June 2026).
24. Irfan, A.N.; Chuprat, S.; Mahrin, M.N.; Ariffin, A. Taxonomy of Cyber Threat Intelligence Framework. In *2022 13th International Conference on Information and Communication Technology Convergence (ICTC)*; IEEE: Piscataway, NJ, USA, 2022. [[CrossRef](#)]
25. Shen, Y.; Stringhini, G. ATTACK2VEC: Leveraging Temporal Word Embeddings to Understand the Evolution of Cyberattacks. In Proceedings of the 28th USENIX Security Symposium (USENIX Security '19), Santa Clara, CA, USA, 14–16 August 2019; pp. 905–921. Available online: <https://www.usenix.org/conference/usenixsecurity19/presentation/shen> (accessed on 1 June 2026).
26. Husák, M.; Čermák, M.; Jirsík, T.; Čeleda, P. Survey of attack projection, prediction, and forecasting in cyber security. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 640–660.
27. Ashenden, D.; Black, R.; Reid, I.D.; Henderson, S. Design Thinking for Cyber Deception. In Proceedings of the 54th Hawaii International Conference on System Sciences (HICSS), Kauai, HI, USA, 5–8 January 2021; pp. 1958–1967. [[CrossRef](#)]
28. Kaelbling, L.P.; Littman, M.L.; Cassandra, A.R. Planning and acting in partially observable stochastic domains. *Artif. Intell.* **1998**, *101*, 99–134. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.