
CHARLES: Continuous Homeostatic Adversarial Resilience through Learning and Epistemic Sifting

Original Research Article | Volume 1 | Issue 5 | 2026 | Article Number: 032

Accepted: 21 September 2026 | Published: 10 October 2026 | ISSN: 2979-8582 (Online)



Crossref : <https://doi.org/10.67693/BJCR-8BWGIS2A>



Ishma Deshmukh Britton

Pune University & British Computer Society (BCS), India

 IDB [0009-0008-5068-4320](https://orcid.org/0009-0008-5068-4320)

Correspondence: Ishma Deshmukh Britton, ishmabritton@gmail.com

Abstract

The increasing use of artificial intelligence (AI) and reinforcement learning (RL) in cybersecurity introduces the possibility of autonomous cyber agents capable of observing an environment, learning from interaction, and selecting sequential actions with limited human intervention. This development creates a challenge for conventional defensive strategies, which frequently depend on relatively stable system configurations, predefined detection rules, and human-led response processes. Cyber deception and moving target defense (MTD) provide mechanisms for increasing uncertainty and changing the environment encountered by an attacker, while game theory and reinforcement learning provide mathematical and computational approaches for modelling adaptive attacker-defender interactions. This paper proposes CHARLES (Continuous Homeostatic Adversarial Resilience through Learning and Epistemic Sifting), a conceptual defensive architecture designed to continuously manage the information and environmental conditions available to an autonomous attacker. Rather than attempting only to detect an attack after malicious activity has occurred, CHARLES proposes a feedback-based approach consisting of epistemic sifting, adaptive state transformation, adversarial learning estimation, and adversarial utility management. The framework aims to investigate whether controlled environmental variability and strategically deployed deception can reduce the stability of an attacker's learned policy, increase uncertainty, and alter the expected utility of attack strategies. The paper develops a simple mathematical model for representing attacker utility, environmental uncertainty, and defensive transformation. A simulation-based methodology is proposed to compare static and adaptive defensive environments against learning-based attackers. The work does not claim that continuous transformation universally defeats autonomous attackers; instead, it establishes a testable framework for investigating the relationship between defensive environmental adaptation and autonomous attacker learning.

Keywords: Cybersecurity, Autonomous Cyber Agents, Reinforcement Learning, Cyber Deception, Moving Target Defense, Adversarial Learning, Game Theory, Adaptive Defense

1. INTRODUCTION

Cybersecurity has traditionally relied heavily on mechanisms such as access control, intrusion detection, firewalls, endpoint protection, and human-led incident response. These mechanisms remain fundamental to modern security architectures, but the increasing automation of cyber operations introduces a different type of challenge. An autonomous attacker does not necessarily have to follow a fixed sequence of manually selected actions. A learning-based agent can potentially interact with an environment, observe its state, evaluate the consequences of actions, and modify its future behaviour.

Reinforcement learning provides a general framework for sequential decision-making in which an agent learns through interaction with an environment. Its application to cybersecurity has included automated penetration testing, network defence, attack simulation, and autonomous cyber operations. Research on autonomous cyber environments has also produced platforms such as CyberBattleSim and CybORG for studying machine-driven cyber operations. Walter, Ferguson-Walte and Ridley (2021) , for example, investigated the incorporation of deception into CyberBattleSim specifically to study its effects on reinforcement-learning-based attackers. This creates an important defensive question. If an autonomous attacker learns an effective representation of a defended environment, then the stability of that representation may become an important factor in the success of its attack strategy. A relatively stable environment may provide repeated observations from which an attacker can learn useful relationships between states and actions. Conversely, continuously modifying selected environmental properties or introducing strategically controlled deceptive information may make the attacker's observations less reliable.

Cyber deception already attempts to manipulate attacker perception by introducing misleading information, decoys, or deceptive resources. Research has shown that deception can divert attackers, gather information about their tactics, and potentially increase the operational cost of an attack. Moving target defense similarly changes system configurations or observable properties to increase uncertainty for attackers. The distinction between the two approaches is important: MTD generally changes characteristics of existing resources, whereas deception can introduce false resources or information intended to influence attacker beliefs.

The challenge is therefore not simply to make defensive systems faster. It is to investigate whether defenders can influence the learning environment itself. This paper proposes CHARLES as a conceptual architecture for studying this problem.

CHARLES does not assume that an autonomous attacker can be directly observed internally or that its learning algorithm can be manipulated. Instead, it assumes that the defender can observe external behaviour and modify selected properties of the environment. The central proposition is that a defender may improve resilience by managing what an autonomous attacker can reliably observe, learn, and exploit.

2. BACKGROUND

2.1 Autonomous Cyber Agents

An autonomous cyber agent can be conceptualized as an artificial agent capable of observing a cyber environment, selecting actions, and adapting its behaviour according to feedback. Reinforcement learning is particularly relevant because an RL agent learns a policy that maps observed states to actions based on accumulated experience.

In cybersecurity, RL has been explored for both offensive and defensive applications. Existing work includes autonomous penetration testing and automated cyber operations, while simulation environments such as CyberBattleSim and CybORG provide platforms for studying autonomous agents.

The important characteristic for this research is adaptation. An attacker that learns from environmental observations may depend on the stability and reliability of those observations.

2.2 Cyber Deception

Cyber deception deliberately introduces misleading information, systems, services, or resources into a cyber environment. Honeypots, honeynets, decoy services, fake telemetry, and deceptive network information are examples.

The purpose is not necessarily to prevent an attacker from interacting with the environment. Instead, deception can attempt to make the attacker's decisions less effective while simultaneously generating intelligence about attacker behaviour. Recent cyber-deception literature identifies deception as both a proactive and reactive defensive mechanism.

However, deception introduces a trade-off. High-quality deception can require additional resources and must remain sufficiently believable to an attacker. Therefore, an effective system must consider not only whether deception works but also when, where, and how much deception should be deployed.

2.3 Moving Target Defense

Moving Target Defense changes selected properties of a computing environment to reduce the predictability and persistence of an attack. Examples include changing:

- IP addresses;
- network routes;
- service configurations;
- software variants;
- physical hosts;
- runtime configurations.

ChameleonSoft demonstrated an early form of multidimensional software diversity in which functionally equivalent but behaviourally different software variants could be shuffled during runtime. It also explored temporal and spatial diversity through runtime changes and migration. This establishes an important point for CHARLES: dynamic environmental transformation itself is not a new concept.

The research question is instead whether such transformation can be explicitly evaluated according to its effect on the learning behaviour of autonomous attackers.

2.4 Game Theory and Adversarial Decision-Making

Cybersecurity can be represented as an interaction between attackers and defenders with competing objectives. Game theory provides mathematical tools for representing these interactions through utility functions, strategies, equilibria, and incomplete information. Game-theoretic cyber-deception research commonly attempts to identify defensive strategies that increase attacker uncertainty or encourage attackers to select suboptimal actions. However, game-theoretic models can become difficult to apply to complex, multi-stage cyber operations. Prior surveys have noted that many defensive-deception models simplify attacker actions and environments, while practical attacks may involve multiple stages and multiple interacting entities.

3. RELATED WORK

The existing literature relevant to CHARLES can be grouped into four major areas.

3.1 Autonomous Learning-Based Attackers

Research has investigated reinforcement learning for automated penetration testing and autonomous cyber

operations. (Walter et al., 2021). specifically examined the effects of cyber deception on RL-trained attackers and incorporated deception into CyberBattleSim. Their objective was to quantify how deception affected an attacker trained using reinforcement learning. Other research has developed intelligent penetration-testing frameworks using RL and imitation learning. For example, GAIL-PT applies generative adversarial imitation learning to penetration testing (Chen et al., 2023). These studies establish that autonomous learning-based attackers are a legitimate research subject rather than merely a hypothetical future scenario.

3.2 Cyber Deception

(Zhu et al., 2021). conducted a survey of game-theoretic and machine-learning-based defensive deception. Their work identifies game theory as a mechanism for modelling strategic attacker-defender interaction and machine learning as a mechanism for generating or managing deceptive resources. Importantly, the survey also identified shortcomings in existing evaluation methods. Many approaches measure defender utility or classifier performance but do not directly quantify the uncertainty or confusion induced in attackers. This observation is particularly relevant to CHARLES because attacker uncertainty is one of the proposed evaluation dimensions.

3.3 Moving Target Defense

Moving target defense has been studied as a proactive approach to reducing attacker predictability. ChameleonSoft, for example, employed multidimensional software diversity and runtime shuffling to continuously alter the characteristics of software execution (Azab et al., 2011). Later research has investigated MTD through game theory and reinforcement learning. A 2023 study of intelligent cyber confrontation described MTD as an important component of proactive defence against intelligent attacks and discussed the combination of game theory and deep reinforcement learning (Cho et al., 2020). These studies demonstrate that adaptive system transformation is already an established research area.

3.4 Learning-Based Dynamic Deception

Recent research has moved toward combining reinforcement learning with active deception. D3O-IIoT, for example, uses a Dueling Deep Q-Network to dynamically coordinate honeypots, moving target defense, fake telemetry injection, and node isolation in industrial IoT environments (Wushishi et al., 2026). D3O-IIoT is especially important to the positioning of CHARLES because it demonstrates that AI-driven deception and MTD can already be combined in a practical architecture. Therefore, CHARLES should not be presented as the first framework to combine these technologies. Instead, CHARLES focuses on a different question: how does continuous defensive transformation affect the learning process and stability of an autonomous attacker?

4. RESEARCH GAP

The literature demonstrates substantial progress in autonomous cyber agents, cyber deception, moving target defense, reinforcement learning, and game-theoretic cybersecurity. However, several issues remain. First, existing defensive-deception research has frequently evaluated systems using defender utility, detection accuracy, or attack mitigation rather than directly measuring changes in attacker uncertainty and behaviour. Second, existing MTD research demonstrates that systems can dynamically change configurations, software variants, and other environmental characteristics. However, the effect of such changes on the learning stability of an autonomous attacker remains a useful experimental dimension. Third, recent learning-based deception systems such as D3O-IIoT demonstrate sophisticated adaptive orchestration of deception and MTD. Their primary objectives, however, include attack mitigation, false-positive control, deception engagement, and resource efficiency. Consequently, this study

proposes the following research gap:

There is a need for controlled experimental investigation of how adaptive environmental transformation and strategically managed deception influence the learning stability, uncertainty, and expected utility of autonomous learning-based cyber attackers.

CHARLES is therefore proposed not as a replacement for existing MTD or cyber-deception systems, but as a framework for investigating this particular interaction.

5. RESEARCH QUESTIONS

RQ1

How does continuous defensive transformation of observable system states affect the ability of an autonomous learning-based attacker to establish a stable attack policy?

RQ2

To what extent can strategically controlled deception increase uncertainty in an autonomous attacker's inferred representation of a defended environment?

RQ3

Can adaptive defensive transformations reduce the expected utility of selected autonomous attack strategies while maintaining acceptable system-performance constraints?

RQ4

Does a combined strategy of environmental transformation and deception provide greater resilience against learning-based attackers than static defensive configurations under equivalent experimental conditions?

6. THREAT MODEL

The proposed threat model considers an autonomous learning-based attacker operating against a defended cyber environment. The attacker is assumed to have:

- the ability to observe selected environmental states;
- the ability to interact with exposed services or resources;
- an adaptive decision-making mechanism;
- the ability to learn from previous interactions;
- a defined operational objective;
- limited knowledge of the complete defensive architecture.

The attacker is **not** assumed to have access to the defender's internal control logic, defensive policies, or internal model parameters. The defender is assumed to have:

- visibility into selected network and system telemetry;
- the ability to deploy controlled deception;
- the ability to modify selected observable system properties;
- the ability to monitor attacker behaviour;
- a mechanism for deciding when environmental transformations should occur.

The attacker may adapt its strategy when environmental changes are detected. The defender therefore faces an adaptive opponent rather than a fixed attacker.

Out of Scope

The model does not assume:

1. unrestricted access to the attacker's neural-network weights;
2. perfect knowledge of the attacker's internal state;
3. unlimited defensive resources;
4. that every attack can be made impossible;
5. that environmental changes can occur without operational cost.

These restrictions are important because the purpose of CHARLES is to provide a testable defensive hypothesis, not to claim universal dominance over autonomous attackers.

7. CHARLES ARCHITECTURE

CHARLES stands for:

Continuous Homeostatic Adversarial Resilience through Learning and Epistemic Sifting

The architecture is based on a continuous feedback loop:

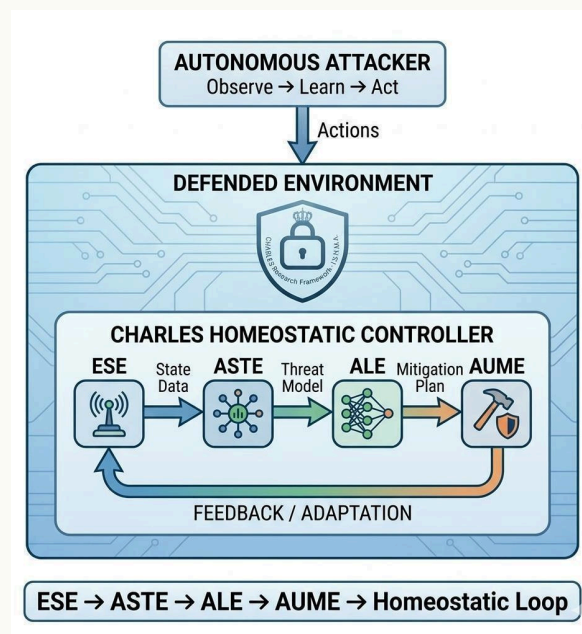
Observe → Estimate → Sift → Transform → Evaluate → Adapt → Repeat

It consists of four primary components.

7.1 Epistemic Sifting Engine (ESE)

The Epistemic Sifting Engine manages the information available to the attacker. Observable information can conceptually be classified into:

- **Real** — information corresponding to genuine system properties;
- **Deceptive** — intentionally misleading information;
- **Uncertain** — information whose reliability cannot easily be established.



The purpose is not to hide everything from the attacker. Excessive concealment may itself be suspicious or operationally impractical.

Instead, the objective is to control the reliability of the attacker's environmental observations.

7.2 Adaptive State Transformation Engine (ASTE)

ASTE modifies selected environmental properties according to current defensive requirements. Potential transformation targets include:

- network addressing;
- service exposure;
- routing;
- software variants;
- virtualized resources;
- selected configuration parameters;
- decoy placement.

The transformation should be controlled rather than indiscriminate.

A key design principle is therefore:

Change the environment enough to disrupt attacker learning while preserving legitimate system functionality. This distinction separates CHARLES from the simplistic idea of continuously changing everything as quickly as possible.

7.3 Adversarial Learning Estimator (ALE)

The defender generally cannot observe an attacker's internal neural-network weights. Therefore, CHARLES does not require such access. Instead, ALE estimates the attacker's adaptation using observable behaviour. Potential indicators include:

- repeated action sequences;
- action-selection consistency;
- successful attack-path repetition;
- exploration versus exploitation behaviour;
- changes in attack strategy following environmental transformations;
- convergence toward particular actions.

This produces an estimated attacker adaptation state.

7.4 Adversarial Utility Management Engine (AUME)

AUME evaluates whether the current defensive state changes the expected attractiveness of an attack strategy. The defender can attempt to:

- reduce expected attack value;
- increase operational cost;
- increase uncertainty;
- increase probability of defensive interruption.

The purpose is not necessarily to make every possible attack impossible. Instead, the goal is to make important attack strategies less reliable or less economically/operationally attractive.

8.5 Homeostatic Control Loop

The four components operate within a continuous feedback mechanism:

1. **Observe** attacker and environment behaviour.
2. **Estimate** attacker adaptation.

3. **Sift** observable information.
4. **Transform** selected environmental properties.
5. **Evaluate** the resulting attacker behaviour.
6. **Adapt** defensive parameters.

Repeat.

The term *homeostatic* refers to maintaining an acceptable defensive state despite changes in the adversarial environment.

Algorithm 1: CHARLES Continuous Homeostatic Loop Execution

Input: System Telemetry Data (State Data), Maximum Performance Latency Bound (L_{max})

Output: Environment Mutated State Configuration (S_{t+1})

```

1: Initialize engines: ESE, ASTE, ALE, AUME
2: while SystemIsActive do
3:   Telemetry ← ObserveSystemActivity()
4:   AttackerBehavior ← TrackNetworkTelemetry(Telemetry)
5:
6:   // Step 1: Adversarial Learning Estimation (Section 8.3)
7:   AttackerState_Est ← ALE.EstimateAdaptation(AttackerBehavior)
8:
9:   // Step 2: Information Sifting & Utility Management (Section 8.1 & 8.4)
10:  Entropy_H ← ESE.ComputeEntropyState(Telemetry)
11:  ExpectedUtility ←

```

```

AUME.EvaluateAttractiveness(AttackerState_Est)
12:
13: // Step 3: Optimization &
Homeostatic Bound Evaluation (Section
9.2)
14: if CurrentLatency <= L_max then
15:   MutationPolicy ←
OptimizeDefensiveState(Entropy_H,
ExpectedUtility)
16:   S_t1 ←
ASTE.ExecuteTransformation(Mutation
Policy) // State change
17: else
18:   S_t1 ← KeepStaticStateConfig()
// Preserve operational health
19: end if
20: end while

```

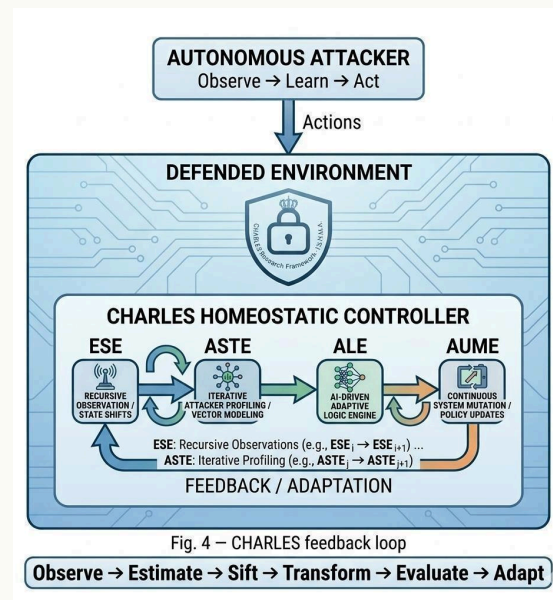


Fig. 4 – CHARLES feedback loop

The defender therefore attempts to avoid two extremes:

Too little adaptation → attacker learns a stable environment.

Too much adaptation → legitimate system performance suffers.

The desired region is an adaptive equilibrium between security and operational stability.

8. MATHEMATICAL MODEL

We can keep this **simple** for the first paper. Let the cyber environment at time t be represented by:

$$S_t$$

where S_t represents the observable system state.

The attacker selects an action:

$$A_t$$

and receives an outcome based on the current environment. The defender selects a transformation:

$$D_t$$

which changes part of the observable environment.

Therefore:

$$S_{t+1} = f(S_t, A_t, D_t)$$

where f represents the evolution of the environment.

8.1 Attacker Utility

We define the attacker's expected utility as:

$$U_A = V_A - C_A - R_A$$

where:

V_A = expected value of successfully achieving the attack objective;

C_A = operational/computational cost of executing the attack;

R_A = expected cost or risk imposed by defensive mechanisms.

The defender seeks to reduce U_A for selected attack strategies.

For example:

$$U_A < 0$$

would indicate that, under the assumptions of the model, the expected cost and risk of a particular strategy exceed its expected value. This does not mean all attacks become impossible.

8.2 Environmental Uncertainty

We can represent the uncertainty of the attacker's observations using an entropy-like quantity:

$$H(S_t)$$

where higher $H(S_t)$ represents greater uncertainty about the current environmental state.

The conceptual objective of CHARLES is:

$$H(S_t) \uparrow$$

while maintaining acceptable operational performance. The framework therefore seeks an optimization of:

$$\max[H(S_t) + \lambda(-U_A)]$$

subject to:

$$L_t \leq L_{max}$$

where:

L_t = operational performance cost/latency;

L_{max} = maximum acceptable performance degradation;

λ = weighting factor between uncertainty and attacker utility.

This gives us a very simple mathematical interpretation of CHARLES:

Increase attacker uncertainty and reduce attack utility without exceeding an acceptable operational cost.

That is enough mathematics for the conceptual version. We don't need to throw 14 equations at the reader just to look clever.

9. METHODOLOGY

The proposed study should use a controlled simulation rather than immediately attempting deployment against real systems.

9.1 Experimental Environment

A simulated cyber environment can contain:

- multiple hosts;
- network services;
- vulnerabilities;
- legitimate resources;
- decoy resources;
- configurable environmental states.

An autonomous RL-based attacker can interact with this environment.

9.2 Experimental Conditions

At minimum, four conditions should be compared:

Condition A — Static Defense

The environment remains relatively stable.

Condition B — Moving Target Defense

Selected environmental properties are periodically transformed.

Condition C — Cyber Deception

The attacker encounters controlled deceptive resources or information.

Condition D — CHARLES

Adaptive environmental transformation and deception are coordinated according to observed attacker behaviour. This allows the contribution of each mechanism to be separated.

9.3 Attacker

A reinforcement-learning-based attacker should be trained under controlled conditions. Possible measurements include:

- attack success rate;
- cumulative reward;
- number of successful attack steps;
- policy convergence;
- action-selection entropy;
- time to achieve the attack objective;
- number of environmental observations required;
- adaptation following environmental transformations.

9.4 Defender

The CHARLES defender would observe attacker behaviour and select defensive transformations. The defender should not have direct access to the attacker's internal model. Instead, it uses observable behaviour to estimate attacker adaptation.

9.5 Evaluation Metrics

The study should measure at least:

Security

- attack success rate;
- attack completion time;
- successful progression through attack stages.

Learning

- attacker policy stability;
- convergence rate;
- action-selection entropy;
- adaptation frequency.

Deception

- interaction with deceptive resources;
- deviation from optimal attack path;
- attacker uncertainty.

Operational Cost

- latency;
- resource consumption;
- number of transformations;
- disruption to legitimate activity.

This metric selection follows an important observation in defensive-deception research: existing studies have often emphasized defender utility or detection accuracy while giving less attention to directly measuring attacker uncertainty, deviation, and deception cost.

9.6 Ablation Study

The experiment should also remove individual CHARLES components. For example:

- CHARLES without deception;
- CHARLES without state transformation;
- CHARLES without attacker-learning estimation;
- complete CHARLES.

This would help determine which component actually contributes to the observed effect.

10. LIMITATIONS

This study has several limitations. First, a simulation cannot completely reproduce the complexity of a real enterprise or industrial environment. Cybersecurity environments contain heterogeneous systems, unpredictable human behaviour, incomplete information, and operational constraints. Second, reinforcement-learning attackers are models of autonomous attackers rather than perfect representations of real-world AI systems. Results may therefore depend on the selected learning algorithm, reward function, environment, and training procedure. Third, environmental transformation can introduce operational overhead. Excessive changes may negatively affect legitimate users, application availability, latency, or system reliability. Fourth, deception is not guaranteed to remain effective. A sufficiently capable attacker may eventually identify patterns in deceptive resources or distinguish genuine and artificial information. Fifth, the mathematical model presented in this paper is intentionally simplified. It represents attacker utility and environmental uncertainty conceptually rather than providing a complete formal proof of equilibrium. Finally, this research does not assume that CHARLES can defeat every autonomous attacker. Its objective is to experimentally investigate whether controlled environmental adaptation can influence attacker learning and decision-making.

11. FUTURE WORK

Several directions can extend this research.

11.1 Advanced Autonomous Attackers

Future experiments should evaluate multiple attacker architectures, including:

- DQN-based agents;
- policy-gradient agents;
- actor-critic agents;
- multi-agent attackers;
- imitation-learning-based attackers.

This would determine whether the proposed effects are specific to one learning algorithm or general across different autonomous agents.

11.2 Adaptive Attackers

A particularly important extension is to allow the attacker itself to recognize defensive transformations and modify its strategy. This creates a genuine adaptive attacker-defender environment:

Attacker↔DefenderAttacker \leftrightarrow Defenderrather than treating the attacker as a fixed RL policy.

11.3 Multi-Agent Environments

Future work could investigate multiple attackers and defenders operating simultaneously. This would make the problem closer to a multi-agent Markov game and allow investigation of coordination, collusion, deception, and strategic adaptation.

11.4 Realistic Cyber Environments

The framework should eventually be tested in more realistic environments such as:

- CybORG;
- CyberBattleSim;
- emulated enterprise networks;
- software-defined networks;
- industrial IoT testbeds.

11.5 Formal Game-Theoretic Analysis

The simple utility model could eventually be expanded into a dynamic or stochastic game. Future research could investigate whether stable defensive policies or equilibrium conditions exist under different attacker capabilities.

11.6 Human Oversight

Autonomous defence should not necessarily remove humans from the security loop. Future versions of CHARLES could investigate a human-on-the-loop architecture where the system autonomously performs low-risk environmental transformations while escalating high-impact decisions to human operators.

12. CONCLUSIONS

The increasing application of artificial intelligence and reinforcement learning to cyber operations creates a need to reconsider how defensive systems respond to adaptive attackers. An autonomous attacker may depend on repeated observations of an environment to develop an effective strategy. This creates an opportunity for defenders to influence not only the attack itself but also the environment from which the attacker learns. This paper proposed CHARLES — Continuous Homeostatic Adversarial Resilience through Learning and Epistemic Sifting as a conceptual framework for investigating this problem. CHARLES combines four principal mechanisms: epistemic sifting, adaptive state transformation, adversarial learning estimation, and adversarial utility management. Together, these mechanisms form a continuous feedback loop intended to manage attacker uncertainty and reduce the effectiveness or expected utility of selected attack strategies. The existing literature demonstrates that cyber deception, moving target defense, reinforcement learning, and game-theoretic defence are individually well-established research areas. Recent systems have also demonstrated combinations of reinforcement learning, deception, and MTD. Therefore, the contribution proposed here is not the simple combination of these technologies. Instead, CHARLES focuses on experimentally examining whether continuous defensive environmental transformation can disrupt the learning stability and decision-making effectiveness of autonomous learning-based attackers. The proposed mathematical model provides a simple representation of this objective by relating attacker utility, environmental uncertainty, and operational cost. A controlled simulation methodology can subsequently determine whether the hypothesis holds under defined experimental conditions. Ultimately, CHARLES should be treated as a research hypothesis and

experimental architecture, rather than a claim of universal protection. Its scientific value will depend on rigorous experimentation, comparison with existing approaches, transparent limitations, and reproducible evaluation.

References

- Azab, M., Hassan, R., Eltoweissy, M., 2011. ChameleonSoft: A Moving Target Defense System. In: *Proceedings of the International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom)*, pp. 1–10.
- Chen, J., Hu, S., Zheng, H., Xing, C., Zhang, G., 2023. GAIL-PT: An Intelligent Penetration Testing Framework with Generative Adversarial Imitation Learning. *Computers & Security* 126, 103055.
- Cho, J.H., Sharma, D.P., Alavizadeh, H., Yoon, S., Ben-Asher, N., Moore, T.J., Kim, D.S., Lim, H., Xiang, C., 2020. Toward Proactive, Adaptive Defense: A Survey on Moving Target Defense. *IEEE Communications Surveys & Tutorials* 22(1), 709–745.
- Jajodia, S., Ghosh, A.K., Swarup, V., Wang, C., Wang, X.S. (Eds.), 2012. *Moving Target Defense II: Application of Game Theory and Adversarial Modeling*. Springer, New York.
- López, P.B., Pérez, M.G., Nespoli, P., 2024. Cyber Deception: State of the Art, Trends, and Open Challenges. *arXiv preprint arXiv:2409.07194*.
- Sutton, R.S., Barto, A.G., 2017. *Reinforcement Learning: An Introduction*, 2nd ed. MIT Press, Cambridge, MA.
- Walter, E., Ferguson-Walter, K., Ridley, A., 2021. Incorporating Deception into CyberBattleSim for Autonomous Defense. In: *Proceedings of the 1st International Workshop on Adaptive Cyber Defense (ACD 2021) at IJCAI-21*.
- Wushishi, M., Hussain, A., Khalid, M.I., Hussain, N., Jamjoom, M., Ullah, Z., 2026. D3O-IIoT: Deep Reinforcement Learning-Driven Dynamic Deception Orchestration for Industrial IoT Security. *Scientific Reports* 16, 2389.
- Yang, Q., Wang, Y., Xu, X., 2023. Multi-agent Markov-game-based intelligent cyber defence study. *Entropy* 25(4), 605.
- Zhu, M., Anwar, A.H., Wan, Z., Cho, J.-H., Kamhoua, C., Singh, M.P., 2021. Game-Theoretical and Machine Learning-based Approaches for Defensive Deception: A Survey. *arXiv preprint arXiv:2101.10121*.

Glossary

<u>Term</u>	<u>Definition</u>
Autonomous Cyber Agent	A software agent capable of observing an environment, making decisions, and executing actions with limited human intervention.
Autonomous Attacker	A cyber agent that independently observes, learns from, and selects actions against a target environment.

Adversarial Learning	Learning in an environment where another intelligent agent actively influences or opposes the learner.
Cyber Deception	Defensive techniques that deliberately present misleading information or systems to influence or study an attacker.
Moving Target Defence (MTD)	A defensive approach that changes selected system properties over time to reduce the usefulness of an attacker's accumulated knowledge.
Epistemic Sifting	Within CHARLES, the proposed process of managing which environmental information is reliable, deceptive, or uncertain from the attacker's perspective.
Environmental Transformation	Controlled modification of observable system or network properties to alter the environment encountered by an attacker.

Policy Stability	The degree to which an attacker's learned decision policy remains consistent and effective under changing conditions.
Attacker Utility	The expected value an attacker obtains from an attack after considering its potential benefit, operational cost, and defensive risk.
Homeostatic Control	A feedback-based process that continuously adjusts system behaviour to maintain desired conditions despite changing external conditions.
Reinforcement Learning (RL)	A machine-learning paradigm in which an agent learns action-selection behaviour through interaction with an environment and feedback such as rewards or penalties.
State Representation	The information used by a learning agent to represent the relevant condition of its environment at a particular time.
Deception Environment	A deliberately constructed environment containing misleading or decoy resources intended to influence attacker behaviour.
Adaptive Defence	A defence that changes its behaviour in response to observed environmental or adversarial conditions.
Threat Model	A formal description of the assumed attacker capabilities, knowledge, objectives, constraints, and interactions with the defended system.



© 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY 4.0) licence (<https://creativecommons.org/licenses/by/4.0/>)

