SBSeg 2024
Paper # 241371

Kill Chain Catalyst for Autonomous Red Team Operations in Dynamic Attack Scenarios

Autores: Antonio Horta, Anderson dos Santos e Ronaldo Goldschmidt
IME – RJ

Imagem gerada pelo ideogram.ai

Sumário

Contextualização

Com o avanço da tecnologia, tarefas antes executadas por humanos agora são executadas de forma autônoma por máquinas incorporadas à inteligência artificial. Além de robôs industriais e veículos autônomos, estudos no campo de ataques cibernéticos visam automatizar as operações do **Red Team** (Al-Azzawi et al. 2024; Paudel e Amariucai 2023) para aumentar a segurança de empresas, treinamentos, competições e guerra cibernética.

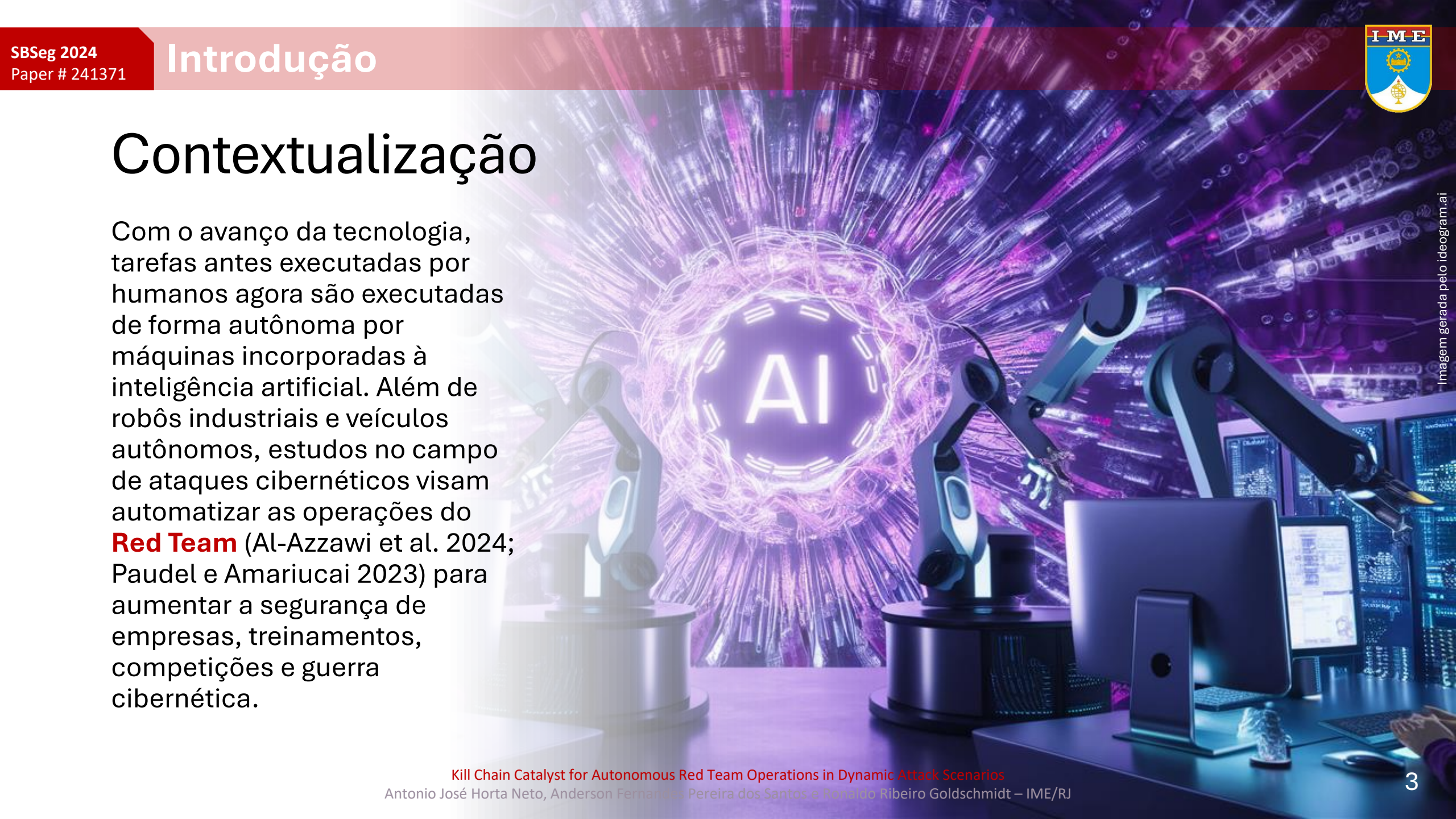


Imagem gerada pelo ideogram.ai

Problema de pesquisa

Como tornar os ataques conduzidos por agentes de **Aprendizado por Reforço**, do inglês Reinforcement Learning (RL) mais **furtivos*** e **resilientes em ambientes dinâmicos?**

* considerando como furtividade a minimização os passos e falhas durante a incursão.

Introdução

A perspectiva do atacante

No domínio dos ataques cibernéticos, particularmente em cenários, como torneios Capture the Flag (CTF), **os invasores não têm conhecimento prévio do ambiente alvo**, o qual, é descoberto durante a campanha (Ortiz-Garces et al. 2023). De acordo com Che Mat et al. (Che Mat et al. 2024), uma estratégia furtiva (stealthy) é essencial e o processo de tomada de decisão sequencial é decisivo para que as ações do ataque tenham **menos etapas e falhas para minimizar a probabilidade do ataque ser exposto**.

Imagem gerada pelo ideogram.ai

Introdução

Imagem gerada pelo ideogram.ai

Objetivo

Esta pesquisa tem objetivo de propor um **algoritmo de RL** que considera a **perspectiva de um atacante, em minimizar o número de passos e falhas**, assim como **resiliente** na execução de um ataque a um **cenário dinâmico e desconhecido**, de modo a minimizar a possibilidade de exposição.

Introdução

A proposta

- O algoritmo **Kill Chain Catalyst (KCC)** será apresentado. O KCC emprega **a lógica da árvore de decisão** para orientar o agente em ataques, aumentando a resiliência em ambientes dinâmicos e a furtividade por meio da **minimização de falhas e passos;**

- Além disso, um **catalisador inspirado no alinhamento genético**, otimiza a busca por um encadeamento mais eficiente no sequenciamento das técnicas de ataque utilizadas;

- A característica de destaque do KCC em problemas de tomada de decisão sequencial para ataques cibernéticos está no uso do **Random Forest como mecanismo RL.**

Imagem gerada pelo ideogram.ai

Trabalhos Relacionados

Table 2. Related works with experiments.

Environment	Algorithm	Analysis	Ref.
Real	A3C/DPPO/GAIL	Steps, Reward, Loss	(Chen et al. 2023)
Real	DQN	Steps, Rewards	(Li et al. 2022)
Real	Random Forest	Rewards	(Holm 2022)
Simulation	NDSPI-DQN dec.	Steps, Rewards	(Zhou et al. 2021)
Simulation	CLAP(PPO+RND)	Steps, Rewards	(Yang and Liu 2022)
Simulation	HA-DQN	Steps, Rewards	(Tran et al. 2021)
Simulation	DQN + LSTM	Steps, Rewards	(Standen et al. 2021)

Fundamentos e conceitos

Alinhamento de sequências

Table 1. Example for sequence align by Needleman-Wunsch algorithm

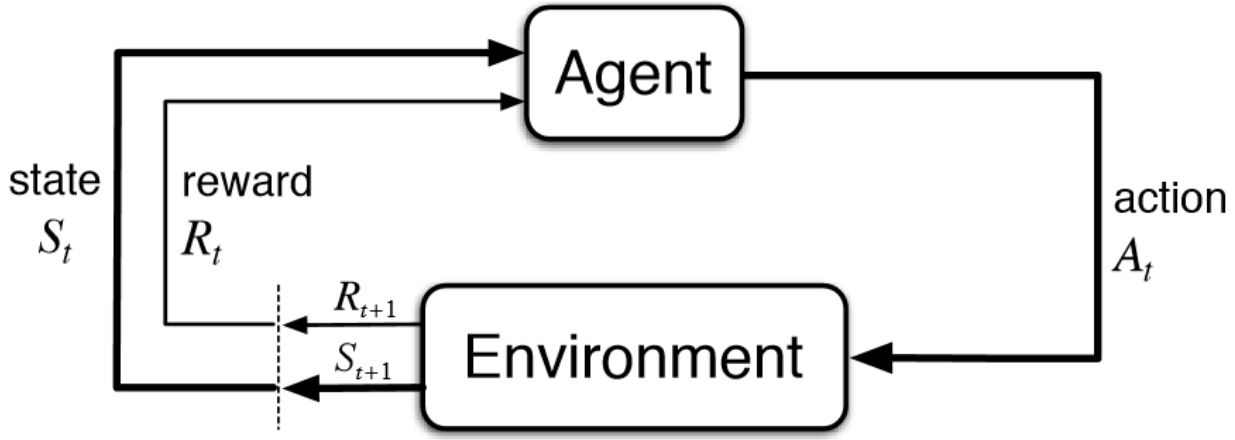
Id	Sequence	Align
Seq1	AGTACGTA	A-GTACGT-A
Seq2	ACTACGTA	AC-TACGT-A
Seq3	ACGTATT	ACGTA--TT-
Seq4	ACGTACGTT	ACGTACGTT-
Seq5	ACGTACGTC	ACGTACGT-C
Consensus	-	ACGTACGTTA

Existem 4 letras principais que representam os nucleotídeos em uma sequência de DNA:

- **A**: Adenina
- **T**: Timina
- **C**: Citosina
- **G**: Guanina

Fundamentos & Conceitos

Aprendizado por reforço



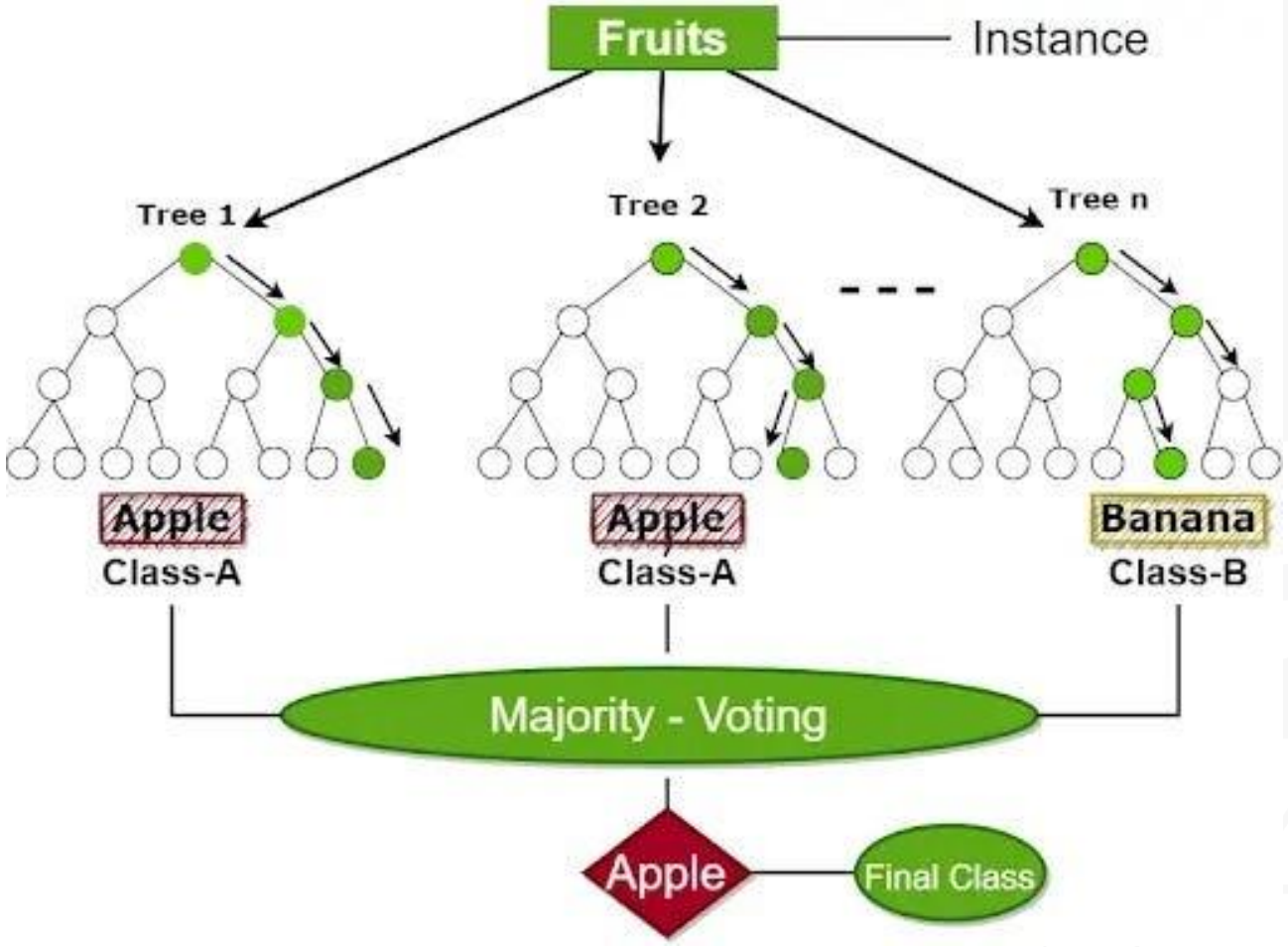
Fonte: Sutton e Barto (Sutton e Barto 2018)

Os algoritmos RL incluem:

- Deep Q-Network (**DQN**) (Mnih et al. 2015);
- Trust Region Policy Optimization (**TRPO**) (Schulman et al. 2015);
- Actor-Critic (**A2C**) (Mnih et al. 2016);
- Proximal Policy Optimization (**PPO**) (Schulman et al. 2017).

Fundamentos & Conceitos

Random Forest



Fonte: https://miro.medium.com

No contexto de tarefas de classificação, o Random Forest emprega medidas como **entropia** ou **impureza de Gini** para determinar o recurso ideal para divisões de nós.

Gini Impurity-Based Weighted Random Forest

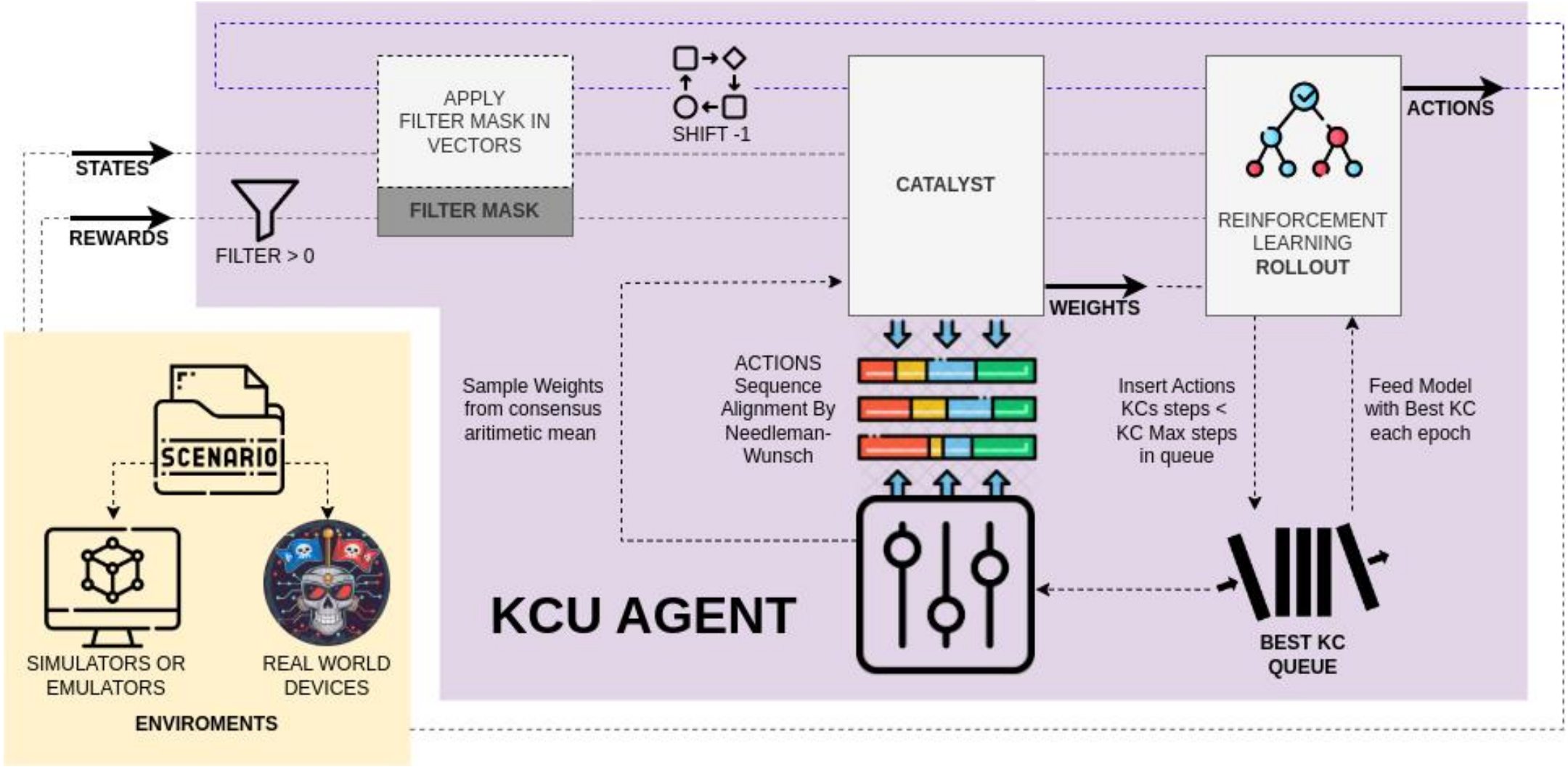
- De acordo com Disha e Waheed (Disha e Waheed 2022), Random Forest é um classificador *ensemble* construído a partir de várias árvores de decisão, incorporando várias métricas de importância de recursos.

- Nesse sentido, a **Gini Impurity-Based Weighted Random Forest (GIWRF)** é uma abordagem para seleção de recursos. Uma dessas métricas envolve derivar a pontuação de importância por meio do treinamento do classificador.


Metodologia

Metodologia

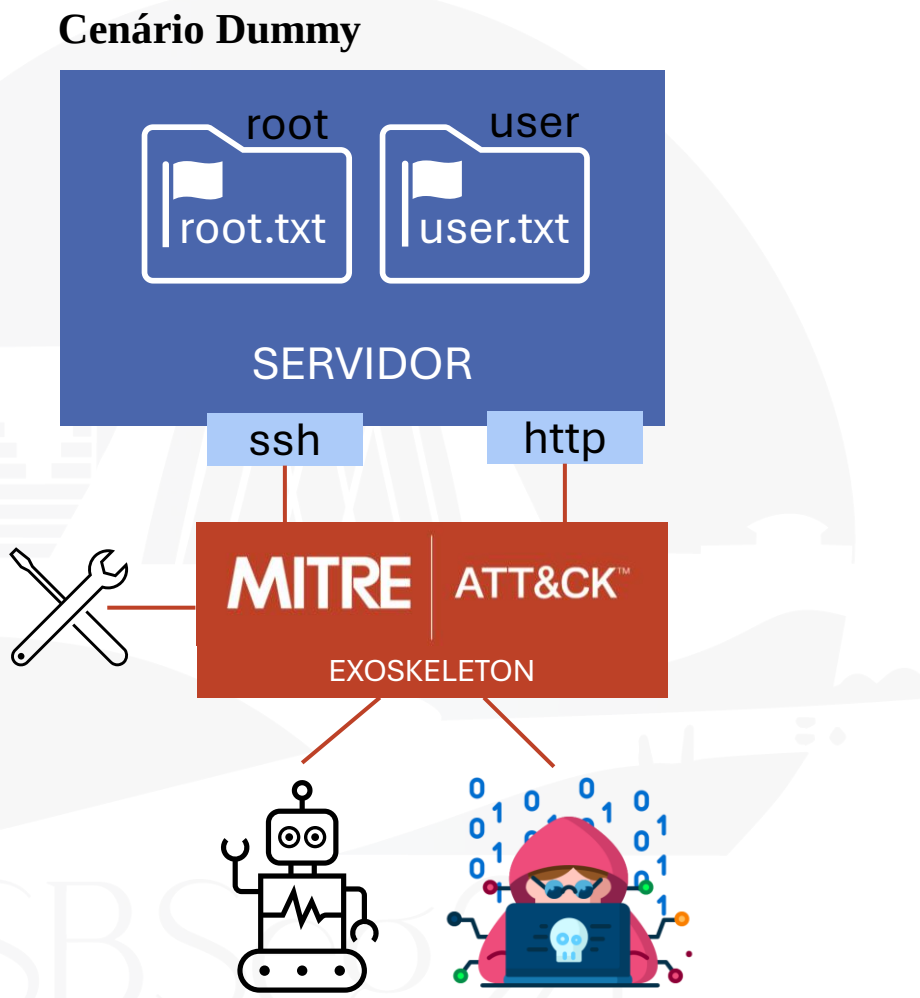
Kill Chain Unscrambler



Metodologia

Ambiente

- Um servidor foi configurado para um CTF, usando imagens do **vunlhub[4]**:
- Para integração do aprendizado por reforço, a **interface Exoskeleton[2]** foi desenvolvida. Essa interface opera com um baixo nível de abstração, permitindo que algoritmos de RL ou humanos, interajam com o servidor. O Exoesqueleto[2] conta com **9 técnicas do MITRE ATT&CK[5] e respectivas ferramentas** que podem ser executadas contra o servidor.



[2]https://gitlab.com/antonio50/exoskeleton
[4]https://github.com/vulhub/vulhub/tree/master/libssh/CVE-2018-10933
[5]https://attack.mitre.org/matrices/enterprise/

Metodologia

Espaço de ação

Table 3. Exoskeleton's action space

Tactic id	Tactic name	Technique id	Technique name	Cmd or Tool
TA0043	Reconnaissance	T1595.001	Active Scanning: Scanning IP Blocks	nmap
TA0043	Reconnaissance	T1595.003	Active Scanning: Wordlist Scanning	dirb
TA0043	Reconnaissance	T1589.002	Gather Victim Identity Information: Email	script parser
TA0001	Initial Access	T1078.003	Valid Accounts: Local Accounts	hydra
TA0001	Initial Access	T1190	Exploit Public-Facing Application	libssh exploit
TA0006	Credential Access	T1110.001	Brute Force: Password Guessing	ssh
TA0007	Discovery	T1083	File and Directory Discovery	bash
TA0004	Privilege Escalation	T1548.001	Abuse Elevation Ctrl Mechanism: Setuid & Setgid	find
TA0009	Collection	T1005	Data from Local System	cat

O **Exoskeleton** adota um modelo de recompensa baseado em falhas. **Ações com sucesso** que resultam em modificações de estado (observável) **são recompensadas com +1**, **ações executadas sem alterar o estado (observável)** recebem uma **recompensa de 0** (por exemplo, comandos repetidos) **e falhas de execução**, seja devido a problemas de conexão ou falhas de serviço, resultam em uma **penalidade de -1**. **Ao alcançar ambas as bandeiras** dentro do cenário recebe uma **recompensa de 100**.

Metodologia

Espaço de ação

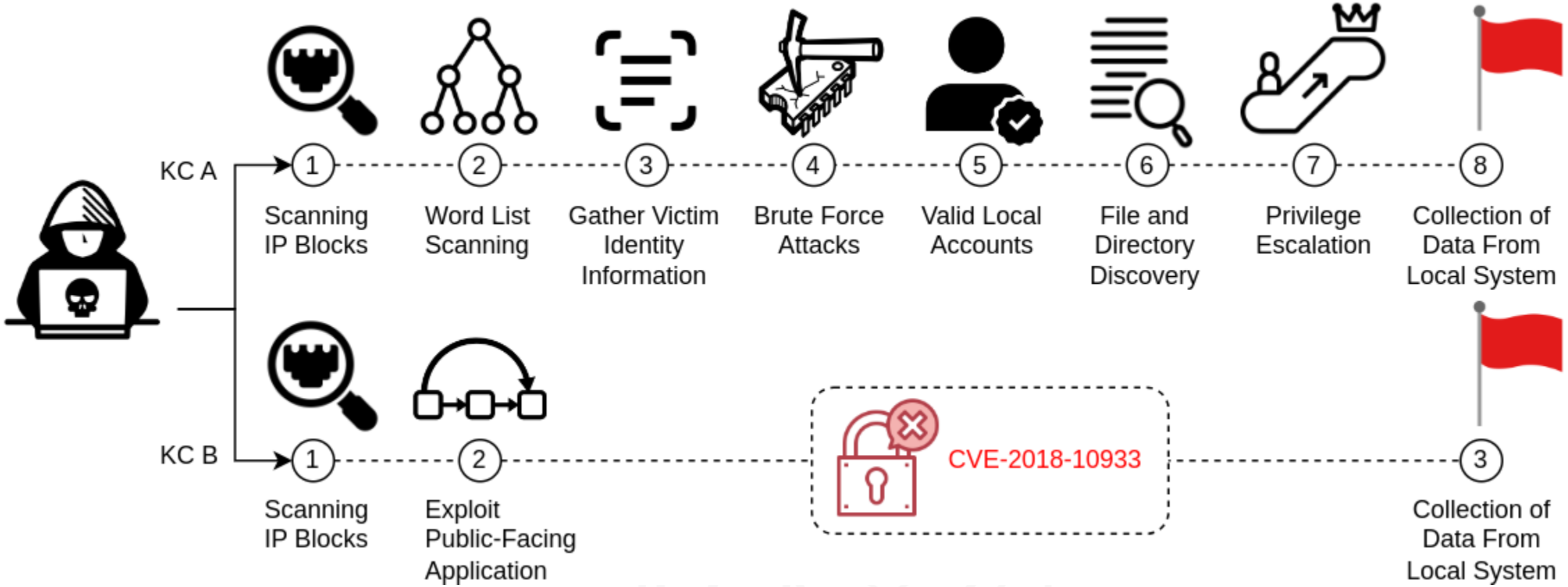


Figure 2. Kill chains available in Exoskeleton's Dummy Scenario.

Experimentos & Resultados

Experimentos e Resultados

Experimentos

Table 5. Essays' parameters

Agent	Parameter	Essay 1	Essay 2
KCC, PPO, TRPO, A2C, DQN	epochs	50	50
KCC, PPO, TRPO, A2C, DQN	max steps per epoch	50	50
KCC, PPO, TRPO, A2C, DQN	timesteps	2500	2500
KCC	epsilon (e-greedy)	0.25	0.25
KCC	decay_rate (e-greedy decay rate)	0.01	0.05
KCC	kcc (catalyst)	True	True
KCC	seed (random seed)	1	1
KCC	buffer (size of the best KCs QUEUE)	20	20
KCC	n_estimators	50	50
KCC	min_samples_leaf	1	1
KCC	dynamic_scenario	False	True

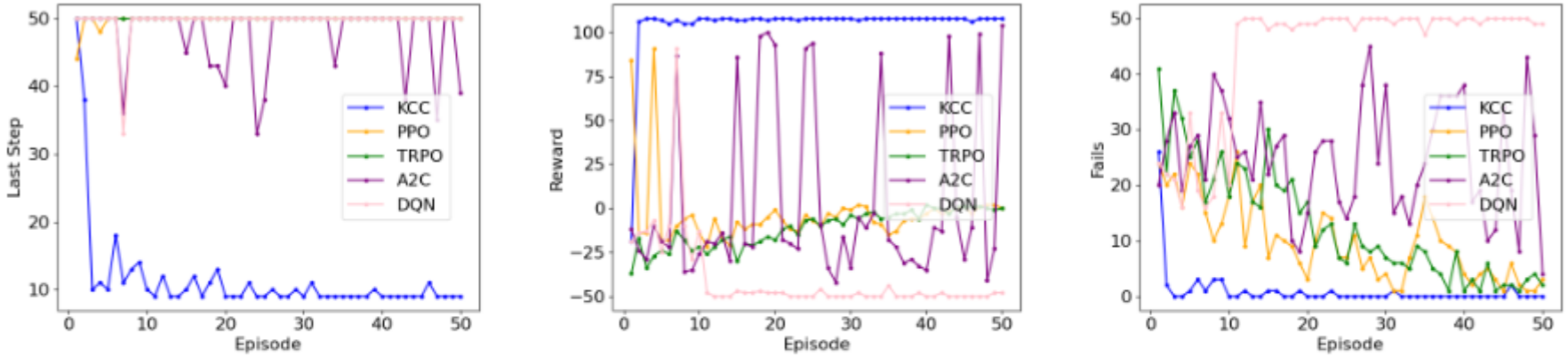
Os experimentos envolvendo o KCC serão avaliados juntamente com quatro algoritmos: DQN, PPO, TRPO e o A2C. A avaliação se concentrará nas principais métricas, como curva de aprendizado, recompensas totais, total de passos e total de falhas.

Para analisar o desempenho do KCC em operações de ataque, considerando a curva de aprendizado para experiências de ataque limitadas e capacidade de lidar com cenários dinâmicos, o experimento foi estruturado em dois ensaios independentes:

- Um ensaio com objetivo comparar a curva de aprendizado do KCC, com os quatro algoritmos de RL: PPO, TRPO, A2C e DQN;
- O segundo para demonstrar a resiliência do KCC a cenários dinâmicos. Neste ensaio final, as vulnerabilidades são corrigidas no meio do ciclo de aprendizado de cada algoritmo, levando os algoritmos convergentes a explorar novos caminhos de ataque.

Experimentos e Resultados

Ensaio 1 – Experiência limitada



(a) Last Step x Episode (b) Reward x Episode (c) Fails x Episode

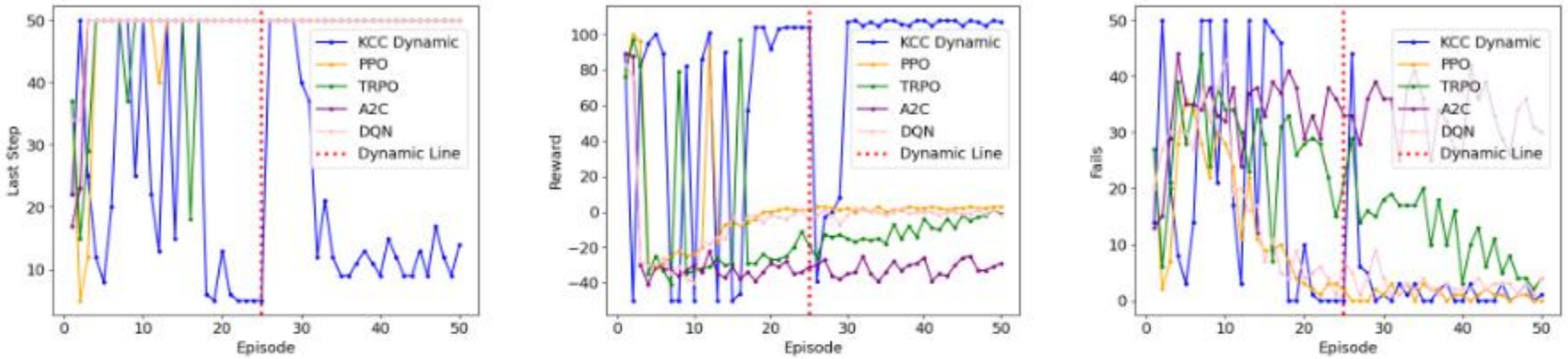
Figure 3. Limited experiences for KCC, A2C, PPO, TRPO and DQN

Table 6. Cumulative Values and Percentage Differences Relative to KCC

Serie	Steps	Diff Steps	Rewards	Diff Rewards	Fails	Diff Fails
KCC	**567**	**0.00%**	**5252**	**0.00%**	**47**	**0.00%**
PPO	2492	339.51%	-162.0	103.08%	496	955.32%
TRPO	2500	340.92%	-567.0	110.80%	640	1261.70%
A2C	2383	320.28%	184.0	96.50%	1241	2540.43%
DQN	2483	337.92%	-2011.0	138.29%	2202	4585.11%

Experimentos e Resultados

Ensaio 2 – Cenário dinâmico



(a) Last Step x Episode (b) Reward x Episode (c) Fails x Episode

Figure 4. Comparing KCC, A2C, PPO, TRPO and DQN in a dynamic scenario

Conclusão e Trabalhos Futuros

- O KCC, utilizando a lógica da árvore de decisão e um catalisador inspirado no alinhamento genético, mostrou-se eficaz na superação das limitações encontradas em algoritmos tradicionais baseados em redes neurais, como PPO, TRPO, A2C e DQN, particularmente em ambientes com experiências de aprendizado limitadas.

- Os resultados mostraram diferenças de até **340,92% para passos**, **138,29% para recompensas** e **4585,11% para falhas** ao realizar ataques usando KCC em comparação com outros algoritmos RL tradicionais.

- Além disso, a capacidade do KCC de se adaptar rapidamente às mudanças ambientais, como a correção de vulnerabilidades, destacou sua resiliência e eficácia em cenários dinâmicos, característica não observada em outros algoritmos testados.

- Os resultados ressaltam que abordagens baseadas em árvore de decisão e o uso de um catalisador de sequencias podem melhorar o desempenho da RL em ataques cibernéticos.

- A pesquisa indica que estudos futuros devem se concentrar em refinar a fase de exploração do algoritmo, principalmente devido à sua natureza estocástica, pois isso é importante para o tipo de estudo realizado.

Referências

[Mnih et al. 2015] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., et al. (2015). Human-level control through deep reinforcement learning. *nature*, 518(7540):529–533.

[Ortiz-Garces et al. 2023] Ortiz-Garces, I., Gutierrez, R., Guerra, D., Sanchez-Viteri, S., and Villegas-Ch., W. (2023). Development of a platform for learning cybersecurity using capturing the flag competitions. *Electronics*, 12(7).

[Paudel and Amariucai 2023] Paudel, B. and Amariucai, G. (2023). Reinforcement learning approach to generate zero-dynamics attacks on control systems without state space models. In *European Symposium on Research in Computer Security*, pages 3–22. Springer.

[Poinsignon et al. 2023] Poinsignon, T., Poulain, P., Gallopin, M., and Lelandais, G. (2023). Working with omics data: An interdisciplinary challenge at the crossroads of biology and computer science. In *Machine Learning for Brain Disorders*, pages 313–330. Springer.

[Pozdniakov et al. 2020] Pozdniakov, K., Alonso, E., Stankovic, V., Tam, K., and Jones, K. (2020). Smart security audit: Reinforcement learning with a deep neural network approximator. In *2020 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA)*, pages 1–8.

[Schulman et al. 2015] Schulman, J., Levine, S., Abbeel, P., Jordan, M., and Moritz, P. (2015). Trust region policy optimization. In *International conference on machine learning*, pages 1889–1897. PMLR.

[Schulman et al. 2017] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., and Klimov, O. (2017). Proximal policy optimization algorithms.

[Standen et al. 2021] Standen, M., Lucas, M., Bowman, D., Richer, T. J., Kim, J., and Marriott, D. (2021). Cyborg: A gym for the development of autonomous cyber agents. In *IJCAI-21 1st International Workshop on Adaptive Cyber Defense*. arXiv.

[Sutton and Barto 2018] Sutton, R. S. and Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press, second edition.

[Tran et al. 2021] Tran, K., Akella, A., Standen, M., Kim, J., Bowman, D., Richer, T., and Lin, C.-T. (2021). Deep hierarchical reinforcement agents for automated penetration testing. In *IJCAI-21 1st International Workshop on Adaptive Cyber Defense*. arXiv.

[Yang and Liu 2022] Yang, Y. and Liu, X. (2022). Behaviour-diverse automatic penetration testing: A curiosity-driven multi-objective deep reinforcement learning approach.

[Zhou et al. 2021] Zhou, S., Liu, J., Hou, D., Zhong, X., and Zhang, Y. (2021). Autonomous penetration testing based on improved deep q-network. *Applied Sciences*, 11(19).

[Al-Azzawi et al. 2024] Al-Azzawi, M., Doan, D., Sipola, T., Hautamäki, J., and Kokkonen, T. (2024). Artificial intelligence cyberattacks in red teaming: A scoping review. In *World Conference on Information Systems and Technologies*, pages 129–138. Springer.

[Breiman 2001] Breiman, L. (2001). Random forests. *Machine learning*, 45:5–32.

[Che Mat et al. 2024] Che Mat, N. I., Jamil, N., Yusoff, Y., and Mat Kiah, M. L. (2024). A systematic literature review on advanced persistent threat behaviors and its detection strategy. *Journal of Cybersecurity*, 10(1):tyad023.

[Chen et al. 2023] Chen, J., Hu, S., Zheng, H., Xing, C., and Zhang, G. (2023). Gail-pt: An intelligent penetration testing framework with generative adversarial imitation learning. *Computers Security*, 126:103055.

[Disha and Waheed 2022] Disha, R. A. and Waheed, S. (2022). Performance analysis of machine learning models for intrusion detection system using gini impurity-based weighted random forest (giwrf) feature selection technique. *Cybersecurity*, 5(1):1.

[Farouk et al. 2024] Farouk, M., Sakr, R. H., and Hikal, N. (2024). Identifying the most accurate machine learning classification technique to detect network threats. *Neural Computing and Applications*, 36(16):8977–8994.

[Gancheva and Stoev 2023] Gancheva, V. and Stoev, H. (2023). An algorithm for pairwise dna sequences alignment. In *International Work-Conference on Bioinformatics and Biomedical Engineering*, pages 48–61. Springer.

[Gangupantulu et al. 2021] Gangupantulu, R., Cody, T., Rahma, A., Redino, C., Clark, R., and Park, P. (2021). Crown jewels analysis using reinforcement learning with attack graphs. In *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1–6.

[Holm 2022] Holm, H. (2022). Lore a red team emulation tool. *IEEE Transactions on Dependable and Secure Computing*, 1:1–1.

[Horta Neto et al. 2024] Horta Neto, A. J., dos Santos, A. F. P., and Goldschmidt, R. R. (2024). Evaluating the stealth of reinforcement learning-based cyber attacks against unknown scenarios using knowledge transfer techniques. *Journal of Computer Security*, (Preprint):1–19.

[Ibrahim et al. 2024] Ibrahim, M. K., Yusof, U. K., Eisa, T. A. E., and Nasser, M. (2024). Bioinspired algorithms for multiple sequence alignment: A systematic review and roadmap. *Applied Sciences*, 14(6):2433.

[Janisch et al. 2023] Janisch, J., Pevnỳ, T., and Lisỳ, V. (2023). Nasimemu: Network attack simulator & emulator for training agents generalizing to novel scenarios. In *European Symposium on Research in Computer Security*, pages 589–608. Springer.

[Li et al. 2022] Li, L., El Rami, J.-P. S., Taylor, A., Rao, J. H., and Kunz, T. (2022). Enabling a network ai gym for autonomous cyber agents. In *2022 International Conference on Computational Science and Computational Intelligence (CSCI)*, pages 172–177. IEEE.

[Mnih et al. 2016] Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., Silver, D., and Kavukcuoglu, K. (2016). Asynchronous methods for deep reinforcement learning. In *International conference on machine learning*, pages 1928–1937. PMLR.

Obrigado!


Antonio Horta
Principal Cyber Research Scientist,
MSc, DSc Candidate | Speaker | CISO...


SBSeg24
SÃO JOSÉ DOS CAMPOS