УДК 004.056

# RL-ADMIT З ПІДТРИМКОЮ ПРЕТРЕЙНІНГУ: ГІБРИДНИЙ ПІДХІД ДО БЕЗПЕКИ KUBERNETES ЗІ ЗБЕРЕЖЕННЯМ АДАПТИВНОСТІ

Мазепа А.Д., Фісенко Д. М., Пантєлєєв В.О
e-mail: artem.mazepa@nure.ua, dmytro.fisenko@nure.ua, vadym.pantielieiev@nure.ua
Науковий керівник – д.т.н., проф. Радівілова Т.А.
Харківський національний університет радіоелектроніки, каф. ІКІ ім. В.В. Поповського
м. Харків, Україна

Kubernetes dominates container orchestration, yet its dynamic architecture creates significant control-plane security challenges. Static admission controllers like OPA Gatekeeper and Kyverno rely on fixed policies, yielding high false positives and slow adaptation to novel threats. RL-Admit models webhook decisions as a POMDP and applies Proximal Policy Optimization (PPO) for online training, with PAC guarantees of convergence to an ε-optimal policy. The cold-start issue—random initialization causing excessive false denials or undetected attacks—is mitigated by offline pretraining. The hybrid framework supports safe shadow/audit deployment, reduces initial false positives by 15–20×, halves convergence time, and enhances chained and zero-day attack detection.

З усіх сучасних хмарних рішень, Kubernetes виступає центральною частиною у більшості інфраструктур, проте його гнучка та динамічна природа породжує велику кількість проблем контролю доступу та безпеки. Традиційні рішення забезпечення такого контролю, як наприклад, OPA Gatekeeper та Kyverno, були завжди основані на статичних правилах, які не дозволяли реагувати на динамічні загрози, які виходять за обмеження цих правил. Це призводить до хибних спрацювань, та повільної реакції на загрози для кластера. Для вирішення цієї проблеми була запропонована система RL-Admit. Вона є новим підходом до забезпечення контролю доступу на основі навчання з підкріпленням (RL), який моделює задачу Validating/Mutating Webhook як частково спостережуваний марковський процес прийняття рішень (POMDP) [1] .

У POMDP-формалізації стан S включає вектор ознак запиту (userinfo, resource type, namespace, operation, RBAC-ролі [2], історія запитів, метрики подів/нод), нормалізованих за допомогою MinMaxScaler та відібраних за mutual information. Простір дій A обмежений двома варіантами: allow або deny (з логуванням пояснення). Функція спостереження $O \approx S$ (з можливим шумом), перехід T оцінюється емпірично з replay buffer, а фактор дисконтування $\gamma \approx 0.99$ враховує довгострокові наслідки (наприклад, ескалацію привілеїв через 5–10 запитів). Початковий розподіл станів $\rho$ відповідає реальному трафіку з audit logs перших 24 годин.

Для навчання, модель використовує Proximal Policy Optimization (PPO) [3]. Нижче наведена формула алгоритму:

$$L^{CLIP}(\theta) = \hat{E}_t\left[ \left(r_t(\theta)\hat{A}_t, clip(r_t(\theta),1-\epsilon,1+\epsilon)\hat{A}_t\right)\right]$$

Також для усунення проблеми холодного страту, після розгортання нової версії моделі, був введений процес претрейнінгу, перед кожним розгортанням нової версії моделі. RL-Admit з претрейнінгом складається з: офлайн-фази (на 50k запитів з audit logs + симульовані атаки), гібридного переходу (shadow mode) та онлайн-файнтюнінгу.

Проведено порівняльний розрахунок для production-кластера (трафік ≈1200 запитів/хв). Результати покращення наведені на таблиці 1.

Табл. 1. Порівняння проценту хибно-позитивних спрацювань з холодним стартом, та з претрейнінгом

| Кількість онлайн запитів | RL-Admit з претрейнінгом | Оригінальний RL-Admit |
|---|---|---|
| 5 000 | 3% | 48% |
| 10 000 | 1.8% | 20% |
| 50 000 | >0.9% | 1.4% |

Список використаних джерел:

1. Zhang, X., Zheng, K., Wang, C., Chen, J., & Qi, H. (2025). A novel deep reinforcement learning for POMDP-based autonomous ship collision decision-making. Neural Computing and Applications, 37(21), 15963-15977.

2. Gu, Y., Tan, X., Zhang, Y., Gao, S., & Yang, M. (2025, May). Epscan: Automated detection of excessive rbac permissions in kubernetes applications. In 2025 IEEE Symposium on Security and Privacy (SP) (pp. 3199-3217). IEEE.

3. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347.