УДК 004.056.52

В. П. Коваленко

РОЗЧИНЕННЯ КРИТИЧНИХ ФАКТОРІВ У ЗВАЖЕНИХ МОДЕЛЯХ РИЗИКУ ДОСТУПУ

Вінницький національний технічний університет

Анотація

Розглянуто структурне обмеження поширеного підходу до агрегування ризику запиту на доступ — лінійної зваженої суми з фіксованими вагами. Аналітично показано, що внесок кожного окремого критерію обмежений зверху його ваговим коефіцієнтом. Унаслідок цього критерії, вага яких є меншою за поріг відмови, не здатні самостійно заблокувати запит навіть за критичного значення відповідного показника. Обґрунтовано, що зазначену ваду не можна усунути лише підбором порогового значення. Визначений «ефект розчинення» розглядається як структурна вразливість моделей лінійної компенсаторної згортки та слугує підставою для переходу до некомпенсаторних схем агрегування ризику з градуйованим реагуванням.

Ключові слова: контроль доступу, оцінювання ризику, нульова довіра, компенсаторна агрегація, вето-критерії, багатокритеріальне прийняття рішень.

Abstract

The paper examines a structural limitation of a widely used approach to access request risk aggregation: the linear weighted sum with fixed weights. It is analytically shown that the contribution of each individual criterion is upper-bounded by its weighting coefficient. As a result, criteria whose weights are lower than the denial threshold cannot independently block a request, even when the corresponding indicator reaches a critical value. It is substantiated that this limitation cannot be eliminated merely by adjusting the threshold value. The identified "dilution effect" is considered a structural vulnerability of linear compensatory aggregation models and provides the rationale for transitioning to non-compensatory risk aggregation schemes with graduated response.

Keywords: access control, risk assessment, zero trust, compensatory aggregation, non-compensatory aggregation, veto criteria, multi-criteria decision-making.

Вступ

Архітектура нульової довіри (Zero Trust) вимагає, щоб кожен запит на доступ оцінювався динамічно, з урахуванням багатовимірного контексту [1]. На зміну детермінованим моделям RBAC/ABAC [2] приходить ризик-орієнтований контроль доступу (Risk-Based Access Control), у якому рішення ухвалюється на основі кількісного індексу ризику [3]. Домінуючим способом отримання цього індексу є лінійна зважена згортка нормалізованих факторів ризику з наперед заданими (експертними) вагами. Зокрема, у раніше запропонованій авторами гібридній LLM/Policy-as-Code системі [4] поєднано детермінований шар Policy-as-Code (pre-check) та інтегральний показник ризику

$$R(Q) = W_1 \cdot Sensitivity + W_2 \cdot Anomaly + W_3 \cdot (1 - JustificationQuality),$$

де $W_1 = 0.45$, $W_2 = 0.35$, $W_3 = 0.20$. Така згортка є простою, інтерпретованою та зручною для аудиту, а тому широко вживаною й у системах, що залучають великі мовні моделі (LLM) як аналітичний компонент [5].

Попри високі агреговані показники якості на типових сценаріях, лінійна компенсаторна форма приховує структурну ваду, яка проявляється саме на рідкісних, але найнебезпечніших однофакторно-критичних подіях. Метою цих тез є формальне виявлення та демонстрація цієї вади.

Постановка проблеми

Передусім слід розрізнити два рівні захисту в гібридній моделі [4]. Жорсткі, бінарні інваріанти (заборона доступу контракторів до PII, доступ із санкційних юрисдикцій, беззаперечно шкідливий IP) реалізує детермінований шар Policy-as-Code поза формулою ризику і вони розчиненню не підлягають. Натомість градуйовані, ймовірнісні критичні фактори, які принципово не виражаються бінарним правилом, живуть саме у зваженій сумі: поведінкова аномалія та «неможливе переміщення» (UEBA), стан довіреності пристрою, достовірність підозрюваного IoC. Саме для них, і лише для них — компенсаторна форма створює структурну ваду, причому, на відміну від жорстких правил PaC, окремого механізму, що компенсував би це, у моделі немає.

Лінійна зважена сума є компенсаторною: низьке значення одного критерію врівноважує (компенсує) високе значення іншого, що неявно припускає взаємозамінність факторів. Однак перелічені градуйовані критичні фактори за своєю природою некомпенсаторні — їхнє високе значення має визначати рішення незалежно від решти. Саме цю вимогу компенсаторна модель порушує.

Аналітичний результат

Твердження про обмеженість внеску критерію. Нехай інтегральний ризик обчислюється як $R(Q) = \sum_{i=1}^{n} w_i\, r_i(Q)$, де $r_i(Q) \in [0,1]$, $w_i \geq 0$, $\sum_{i=1}^{n} w_i = 1$. Тоді внесок i-го критерію обмежений зверху його вагою:

$$0 \leq w_i r_i(Q) \leq w_i,$$

і верхня межа досягається лише при $r_i(Q) = 1$.

Твердження є елементарним наслідком умови $r_i \leq 1$; нетривіальним є її наслідок для прийняття рішень.

Наслідок, некритичність за побудовою. Нехай рішення «відмовити» приймається за умови $R(Q) \geq \theta_{deny}$. Тоді критерій k_i з вагою $w_i < \theta_{deny}$ не здатний самостійно (за нульових значень інших критеріїв) спричинити відмову: навіть за $r_i(Q) = 1$ маємо $R(Q) = w_i < \theta_{deny}$. Назвемо такий критерій невето-здатним у даній моделі. Зауважимо: невето-здатний критерій усе ж робить внесок у відмову в комбінації з іншими — він лише не може спричинити її самотужки, а отже на нього не можна покладатися для критичної однофакторної події.

Застосуємо наслідок до моделі [4]. Оскільки найбільша з ваг становить 0.45, для будь-якого порога $\theta_{deny} > 0.45$ жоден із трьох факторів не є вето-здатним — відмова стає можливою лише за одночасного підвищення щонайменше двох факторів. Зведення наведено в таблиці 1.

Таблиця 1. Вето-здатність факторів моделі

Фактор	Вага w_i	Макс. внесок w_i	Вето-здатний при $\theta_{deny} = 0.5$?
Чутливість ресурсу (Sensitivity)	0.45	0.45	Ні
Аномальність (Anomaly)	0.35	0.35	Ні
Якість обґрунтування $(1 - JQ)$	0.20	0.20	Ні

Принципово, що цю ваду не можна усунути підбором порога. Щоб зробити окремий фактор вето-здатним, поріг θ_{deny} довелося б опустити нижче його ваги — але поріг є глобальним, тож його зниження робить чутливішою всю згортку й різко підвищує частку хибних відмов (False Deny Rate) на рутинному трафіку. Отже, у компенсаторній моделі вето-здатність окремого критерію неможливо забезпечити локально, не погіршивши поведінку на решті запитів. До того ж у [4] маршрутизація багаторівнева (автосхвалення / ручний розгляд / відмова), і розчинення стосується кожного рівня: фактор із малою вагою не здатен самотужки навіть ескалювати запит на ручний розгляд.

Графічно ефект ілюструє рис. 1. За доброякісних інших факторів зважена сума зростає лише до 0.415 навіть за максимального значення критичного фактора ($r_i = 1$) і жодного разу не перетинає поріг відмови. Натомість некомпенсаторна (вето-) агрегація, що трактує критичний фактор як окремий сигнал поза зваженою сумою, перетинає поріг, щойно цей фактор сягає θ_{deny}. Заштрихована область — «сліпа зона», у якій фактор уже критичний, але компенсаторна модель усе ще дозволяє доступ.

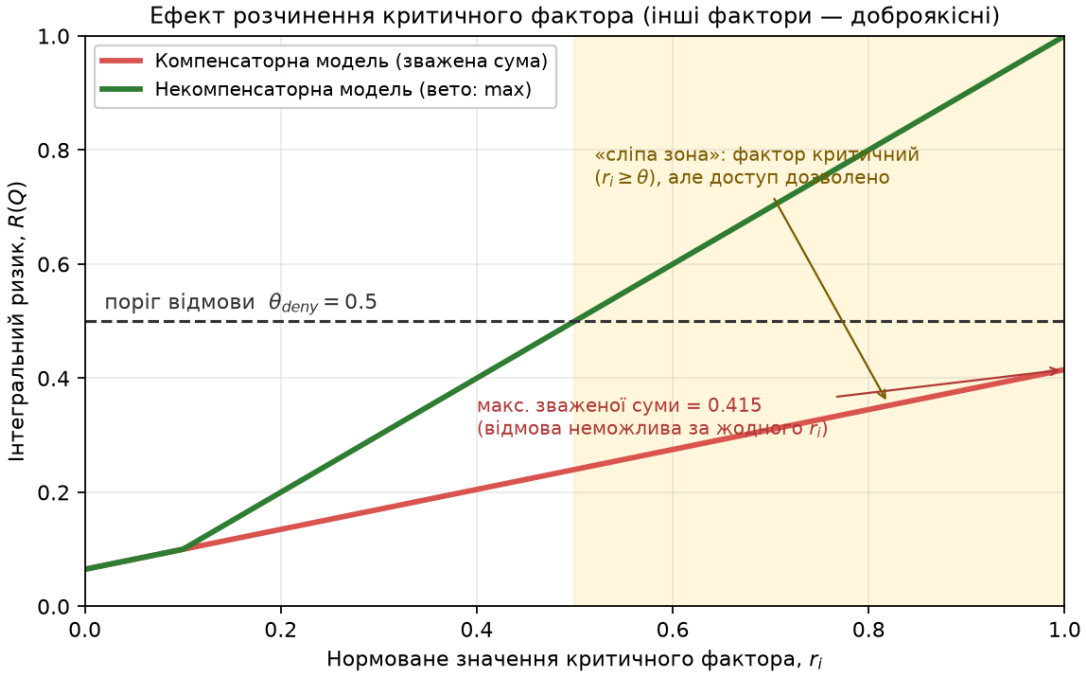

Рис. 1. Ефект розчинення критичного фактора (спрощена аналітична ілюстрація на вагах моделі; некомпенсаторна крива трактує критичний фактор як окремий вето-сигнал).

Наслідки для безпеки

Виявлена властивість є структурною — вона притаманна функціональній формі незалежно від виміряної точності класифікації, а отже не спростовується високими агрегованими показниками. Агрегована точність, отримана на переважно рутинному трафіку, не суперечить наявності вади, бо однофакторно-критичні події становлять мізерну частку вибірки — проте саме вони є цільовими для зловмисника. По-друге, додавання нового градуйованого критичного критерію до компенсаторної суми лише знижує його відносну вагу, посилюючи розчинення. По-третє, у гібридних системах, де частину критеріїв постачає LLM із властивою їй невизначеністю [6], компенсаторне усереднення додатково маскує поодинокі тривожні сигнали. Отже, коректність реакції на критичні події є окремим, недостатньо врахованим у літературі виміром надійності ризик-орієнтованих IAM-систем, який не зводиться до середньої якості класифікації.

Напрям подолання

Усунення ефекту розчинення потребує відмови від суто компенсаторної форми агрегування на користь некомпенсаторних (вето-) схем, у яких визначено підмножину критичних критеріїв, високе значення яких визначає рішення незалежно від решти. Водночас наївне жорстке вето має власну ціну — крихкість і зростання хибних відмов за хибнопозитивних критичних сигналів. Тому практичний механізм має бути градуйованим (поріг спрацювання та проміжний стан «ручний розгляд» замість безумовної відмови) і узгодженим з адаптивним зважуванням компенсаторних критеріїв та контрольованим залученням LLM. Конкретну таку конструкцію автори розвивають у

подальшій роботі; ці тези обмежуються формальним обґрунтуванням самої проблеми та вимог до її розв’язання.

Висновки

Аналітично доведено, що в лінійній зваженій моделі оцінювання ризику доступу внесок окремого критерію обмежений його вагою, а тому критерій із вагою, меншою за поріг відмови, є невето-здатним за побудовою; до того ж цю властивість не можна усунути підбором глобального порога, не погіршивши частку хибних відмов. На вагах раніше запропонованої авторами моделі [4] жоден із трьох факторів не здатний самостійно заблокувати запит за реалістичних порогів, що ілюструє «ефект розчинення» градуйованих критичних факторів (рис. 1, табл. 1). Це обґрунтовує перехід від компенсаторної зваженої суми до некомпенсаторних схем агрегування з градуйованим реагуванням.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Zero Trust Architecture [Електронний ресурс] / S. Rose, O. Borchert, S. Mitchell, S. Connelly. — Gaithersburg, MD : National Institute of Standards and Technology, 2020. — 59 p. — (NIST Special Publication ; 800-207). — Режим доступу: https://doi.org/10.6028/NIST.SP.800-207
2. Penelova M. Access Control Models [Електронний ресурс] / M. Penelova // Cybernetics and Information Technologies. — 2021. — Vol. 21, No. 4. — P. 77–104. — Режим доступу: https://doi.org/10.2478/cait-2021-0044
3. Risk-Based Access Control Model: A Systematic Literature Review [Електронний ресурс] / H. F. Atlam [et al.] // Future Internet. — 2020. — Vol. 12, No. 6. — Article 103. — Режим доступу: https://doi.org/10.3390/fi12060103
4. Коваленко В. П. Розробка гібридної архітектури інтелектуальної системи керування контролем доступу на базі AI-агентів та великих мовних моделей [Електронний ресурс] / В. П. Коваленко, О. О. Ковалюк // Наукові праці ВНТУ. — 2026. — № 1. — Режим доступу: https://doi.org/10.31649/2307-5376-2026-1-22-30
5. Fleming C. Uncertainty-Aware, Risk-Adaptive Access Control for Agentic Systems using an LLM-Judged TBAC Model [Електронний ресурс] / C. Fleming, A. Kundu, R. Kompella // arXiv preprint. — 2025. — arXiv:2510.11414. — Режим доступу: https://doi.org/10.48550/arXiv.2510.11414
6. OWASP Top 10 for Large Language Model Applications [Електронний ресурс]. — OWASP Foundation, 2025. — Режим доступу: https://owasp.org/www-project-top-10-for-large-language-model-applications/

Коваленко Володимир Петрович — аспірант кафедри комп’ютерних систем управління Вінницький національний технічний університет, Вінниця, e-mail: digit.vova@gmail.com.

Науковий керівник: ***Ковалюк Олег Олександрович*** — канд. техн. наук, доцент кафедри комп’ютерних систем управління, Вінницький національний технічний університет, Вінниця.