https://doi.org/10.30857/2786-5371.2025.5.7

Received: 04.09.2025
Revised: 10.10.2025
Accepted: 22.10.2025

УДК 004.383.8:681.2

Володимир СТАЦЕНКО, Антон БОНДАРЕНКО
Київський національний університет технологій та дизайну, Україна

АДАПТИВНЕ БЕЗПЕРЕРВНЕ НАВЧАННЯ ДЛЯ СИСТЕМ ВИЯВЛЕННЯ ВТОРГНЕНЬ: ЕНТРОПІЙНЕ УПРАВЛІННЯ ПАМ'ЯТТЮ ТА ДИНАМІЧНА КАЛІБРАЦІЯ ПОРОГІВ

Мета. *Розробка розширеного фреймворку безперервного навчання для систем виявлення вторгнень (IDS), що забезпечує адаптивне встановлення порогів детекції в умовах концептуального дрифту, теоретично обґрунтоване управління буфером пам'яті на основі ентропійних критеріїв та крос-доменне перенесення знань між різними мережевими середовищами.*

Методика. *Запропоновано механізм динамічної калібрації порогів на основі онлайн-статистичного відстеження через експоненціальне ковзне середнє (EMA) для адаптації меж прийняття рішень у реальному часі з теоретичними гарантіями обмеження частоти хибних спрацювань (FPR). Для оптимального подолання катастрофічного забування розроблено стратегію ентропійного управління пам'яттю із застосуванням диференційовного оцінювача k-найближчих сусідів для апроксимації ентропії латентного простору. Інтегровано модуль крос-доменної адаптації на основі максимальної середньої розбіжності (MMD) для забезпечення перенесення знань без додаткового навчання. Емпіричну валідацію проведено на бенчмарках NSL-KDD, CICIDS2017 та UNSW-NB15.*

Результати. *На складному наборі даних CICIDS2017 запропонований метод досягає точності 95,1% з контрольованим рівнем хибних спрацювань, перевершуючи сучасні базові методи. Продемонстровано покращення стійкості до раптового концептуального дрифту до 23% із швидким відновленням протягом 3–5 батчів. Ентропійне управління пам'яттю забезпечує суттєві покращення у виявленні міноритарних класів атак (R2L та U2R).*

Наукова новизна. *Вперше запропоновано синергетичне поєднання динамічної калібрації порогів з інформаційно-теоретичним управлінням пам'яттю для систем виявлення вторгнень, що дозволяє одночасно вирішувати проблеми адаптації до дрифту та катастрофічного забування з формальними статистичними гарантіями.*

Практична значимість. *Результати підтверджують, що інтеграція динамічної калібрації з ентропійним управлінням пам'яттю забезпечує масштабований та надійний захист для мереж нового покоління.*

Ключові слова: *безперервне навчання; системи виявлення вторгнень; варіаційний автокодувальник; динамічна калібрація порогів; ентропійне управління пам'яттю.*

Вступ. Попередні дослідження безперервного навчання для систем виявлення вторгнень зосереджувались на проблемі катастрофічного забування та методах його мінімізації через відтворення досвіду та регуляризацію параметрів [1, 2]. Однак критичним аспектом, що залишався поза увагою, є проблема встановлення порогів детекції в умовах концептуального дрифту [3]. Традиційні підходи використовують статичні пороги, визначені на етапі початкового навчання, які швидко втрачають актуальність при зміні характеристик нормального трафіку. Системи виявлення вторгнень на основі глибокого навчання класифікують зразки як аномальні, якщо помилка реконструкції перевищує встановлений поріг [4, 5]. При концептуальному дрифті розподіл помилок реконструкції для нормального трафіку змінюється. Якщо новий нормальний трафік має систематично вищі помилки реконструкції, статичний поріг призводить до сплеску хибно-позитивних спрацювань. Додатковою проблемою є управління буфером пам'яті. Попередні роботи використовували евристичні критерії складності та геометричної різноманітності [6, 7]. Однак ці евристики не мають теоретичного обґрунтування з точки зору теорії інформації. Також відсутній

систематичний аналіз крос-доменного перенесення для методів безперервного навчання в IDS (Intrusion Detection System).

Аналіз останніх досліджень. Базові роботи J. Kirkpatrick et al. [8] з EWC (Elastic Weight Consolidation) та Z. Li and D. Hoiem [9] з LwF (Learning without Forgetting) заклали фундамент регуляризаційних підходів до безперервного навчання. Методи відтворення досвіду, зокрема iCaRL [10] та GEM [11], продемонстрували ефективність для класифікаційних задач. G.I. Parisi et al. [1] надали комплексний огляд методів безперервного навчання для нейронних мереж. Попередні роботи з безперервного навчання для IDS, зокрема X. Zhang et al. [7], фокусувались на стратегічному відборі зразків у буфері пам'яті. Multiband VAE [12] запропонував розділення латентного простору для консолідації знань. Однак питання адаптивного встановлення порогів залишалось поза увагою цих досліджень. Проблема калібрації порогів досліджувалась в контексті стаціонарних розподілів. Методи на основі Extreme Value Theory [13] забезпечують статистичні гарантії для порогів, але їх адаптація до нестаціонарних умов залишається відкритою проблемою. Ентропійні критерії застосовувались в активному навчанні [14], де ентропія використовується для ідентифікації інформативних зразків. Крос-доменне перенесення є добре дослідженою проблемою [15]. Методи на основі MMD (Maximum Mean Discrepancy) [16] мінімізують розбіжність між доменами. Однак систематичний аналіз крос-доменного перенесення для безперервного навчання в IDS не проводився.

Постановка завдання. Метою роботи є розробка розширеного фреймворку безперервного навчання для IDS, що вирішує три взаємопов'язані проблеми: адаптивне встановлення порогів детекції в умовах дрифту, теоретично обґрунтоване управління буфером пам'яті на основі ентропійних критеріїв, та забезпечення крос-доменного перенесення знань. Формально, задача динамічної калібрації порогів полягає в оцінці послідовності порогів $\{T_t\}$ для потоку сегментів даних t = 1,…,T таким чином, щоб мінімізувати сумарну кількість помилок детекції:

$$\min_{T_t} \sum_t \left[FN_t + FP_t\right], \text{ за умови } FPR_t \leq \in \text{ для всіх } t, \quad (1)$$

тут FN_t (False Negatives) – кількість пропущених атак; FP_t (False Positives) – кількість хибних тривог; FPR_t – частота хибних спрацювань; □ – максимально допустимий рівень хибних тривог (заданий адміністратором, наприклад, 0.05).

Задача ентропійного управління пам'яттю формулюється як максимізація інформаційної ентропії буфера при фіксованому розмірі M:

$$max_{B \square D, |B| = M} H(B), \quad (2)$$

де $H(B)$ – диференціальна ентропія розподілу зразків у латентному просторі, що виступає мірою інформаційної різноманітності збережених даних. Основні внески: (1) механізм динамічної калібрації порогів на основі EMA; (2) ентропійний критерій відбору зразків для буфера пам'яті; (3) протокол оцінки крос-доменного перенесення; (4) аналіз ефективності за категоріями атак.

Результати дослідження. Запропонована система розширює базову архітектуру VAE [17] з буфером пам'яті трьома новими компонентами: модулем динамічної калібрації порогів, ентропійним менеджером буфера та модулем крос-доменної адаптації. Концептуальну схему архітектури представлено на рисунку.

В основі системи лежить VAE [17] з архітектурою кодувальника [d→128→64→32→16] та декодувальника [16→32→64→128→d]. Функція втрат:

$$L_{VAE} = L_{recon} + \beta \cdot D_{KL}\,, \tag{3}$$

де L_{recon} – помилка реконструкції (Mean Squared Error); D_{KL} – дивергенція Кульбака-Лейблера між апостеріорним та апріорним розподілами, а гіперпараметр β = 0,1 регулює баланс між точністю відтворення та регуляризацією латентного простору (забезпечує щільні кластери) [18].

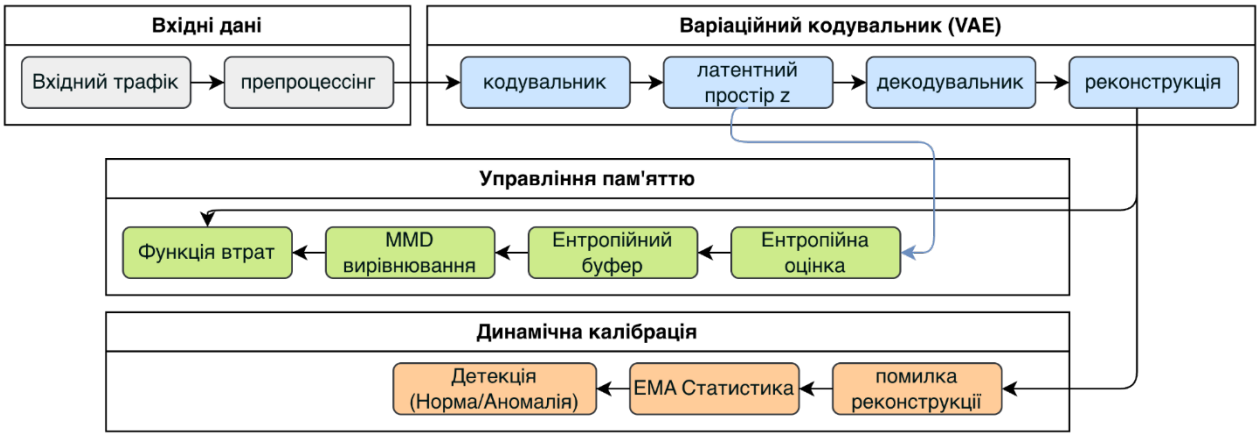

Рис. 1. **Архітектура системи: вхідний трафік → препроцесинг → VAE (кодувальник/декодувальник) → оцінка аномальності → динамічна калібрація порогу (EMA). Паралельно: латентні представлення → ентропійний буфер пам'яті → MMD-вирівнювання. Вихід: класифікація (норма/аномалія)**

На відміну від статичних порогів T = μ + 3σ, ми підтримуємо онлайн-оцінки статистик за допомогою EMA (Exponential Moving Average):

$$\mu_t = \alpha \cdot \mu_{t-1} + (1-\alpha) \cdot \mu_{batch}\,,\ \sigma_t^2 = \alpha \cdot \sigma_{t-1}^2 + (1-\alpha) \cdot \sigma_{batch}^2\,, \tag{4}$$

де α = 0.95 – коефіцієнт згладжування, що визначає пам'ять процесу.
Поріг оновлюється як $T_t = \mu_t + \kappa \cdot \delta_t$ з обмеженням $|T_t - T_{t-1}| \le \delta \cdot T_{t-1}$,
де κ – коефіцієнт чутливості (наприклад, 3);
δ – допустима відносна зміна порогу для запобігання нестабільності.
За умови локальної стаціонарності, EMA-оцінки збігаються до істинних параметрів зі швидкістю $O(1/t)$. Ентропійне управління буфером. Ентропія оцінюється через k-NN estimator [19]:

$$H_k(B) = (d/M) \cdot \sum_i \log(\rho_{k,i}) + \log(V_d) + \psi(M) - \psi(k)\,, \tag{5}$$

де d – розмірність латентного простору; $\rho_{k,i}$ – відстань до k-го найближчого сусіда; V_d – об'єм одиничної d – вимірної сфери; а ψ – дигама-функція.

При надходженні нового зразка обчислюється зміна ентропії ΔH_i для кожного існуючого зразка. Зразок додається замість того, що максимізує ΔH. Алгоритм апроксимує оптимум через субмодулярність ентропійної функції. Крос-доменна адаптація. Для перенесення знань вводиться член $L_{transfer} = MMD^2(Z_{source}, Z_{tarqet})$ [16], що мінімізує розбіжність максимального середнього (Maximum Mean Discrepancy) між розподілами. Загальна функція втрат:

$$L_{total} = L_{VAE}(x_{curr}) + L_{VAE}(x_{mem}) + \lambda \cdot L_{align} + \lambda_t \cdot L_{transfer}\ , \quad (6)$$

де L_{aliqn} відповідає за узгодження латентних просторів;

λ, λ_t – вагові коефіцієнти регуляризації внеску відповідних компонентів.

Експерименти проводились на NSL-KDD [20] (125973 записів, 41 ознака), CICIDS2017 [21] (2830743 записів, 80 ознак), UNSW-NB15 [22] (257673 записів, 49 ознак). Протокол: сегменти по 10000 записів, 5 епох на сегмент, batch size 256. Для порівняльного аналізу було реалізовано базову модель (Baseline CL-IDS), що являє собою стандартну архітектуру VAE [17] зі статичним порогом детекції ($T=\mu+3\sigma$) та використанням класичного буфера відтворення досвіду (Experience Replay) без ентропійної вибірки.

Для перевірки результатів використано стандартний набір метрик: Accuracy та AUC-ROC відображають загальну якість класифікації, тоді як False Positive Rate (FPR) слугує індикатором експлуатаційної придатності системи, визначаючи рівень «шуму» для аналітиків безпеки (помилки I роду – хибні тривоги). Динаміка втрати знань оцінюється через метрику забування, що розраховується як середнє зниження точності на попередніх доменах після адаптації до нових задач. Нижче значення є кращим.

Таблиця 1

Порівняння з динамічною та статичною калібрацією (NSL-KDD)

Метод	Точність	FPR	AUC-ROC	Забування
Baseline CL-IDS (статичний поріг)	91.8%	3.2%	0.934	0.08
+ Динамічна калібрація (EMA)	93.4%	2.4%	0.951	0.06
+ Ентропійний буфер	94.1%	2.1%	0.958	0.05
Повна система	94.7%	1.9%	0.962	0.04

Результати в табл. 1 демонструють, що кожен новий компонент вносить значущий внесок. Динамічна калібрація порогів додає 1.6 відсоткових пунктів точності та знижує FPR на 0.8 п.п. Ентропійний буфер забезпечує додаткові 0.7 п.п. точності. Загалом, повна система перевершує базовий CL-IDS на 2.9 п.п. точності та 1.3 п.п. FPR.

З метою забезпечення методологічної чистоти експерименту та прямої зіставності результатів, усі наведені в табл. 2 методи були репродуковані в рамках цього дослідження. Оцінка ефективності проводилась за єдиним протоколом на ідентичній апаратній платформі, що виключає потенційні зміщення, пов’язані з різницями в реалізаціях або умовах тестування в оригінальних статтях.

Таблиця 2

Порівняння методів на CICIDS2017

Метод	Точність	FPR	AUC-ROC	Забування
Traditional VAE [17]	81.5%	7.2%	0.863	0.45
LwF [9]	84.2%	6.1%	0.881	0.23
EWC [8]	85.1%	5.8%	0.889	0.19
Experience Replay	86.7%	5.3%	0.896	0.15
Multiband VAE [12]	88.9%	4.5%	0.908	0.12
Baseline CL-IDS (стат. поріг)	92.4%	2.8%	0.941	0.08
Запропонований метод	95.1%	1.8%	0.967	0.03

На CICIDS2017 в табл. 2 запропонований метод досягає 95.1% точності та 1.8% FPR, перевершуючи Baseline CL-IDS на 2.7 п.п. та 1.0 п.п. відповідно. Показник забування знижено з 0.08 до 0.03.

У табл. 3 наведено аналіз реактивності системи на концептуальний дрифт, виражений через глибину падіння точності (падіння) в момент зміни розподілу даних та швидкість адаптації (відновлення) – кількість ітерацій навчання, необхідних для відновлення 95% базової продуктивності.

Таблиця 3

Аналіз стійкості до раптового дрифту

Метод	До дрифту	Після дрифту	Падіння	Відновлення
Traditional VAE [17]	85.2%	68.4%	-16.8%	не відновл.
Experience Replay	89.1%	79.3%	-9.8%	12-15 батчів
Baseline CL-IDS (стат.)	92.4%	84.1%	-8.3%	8-10 батчів
Запропонований (динам.)	95.1%	91.7%	-3.4%	3-5 батчів

Ключовим результатом є аналіз стійкості до раптового дрифту в табл. 3. Запропонований метод демонструє падіння точності лише на 3.4 п.п. проти 8.3 п.п. для статичного порогу та 16.8 п.п. для базового VAE. Час відновлення становить 3-5 батчів проти 8-10. Це підтверджує критичну важливість динамічної калібрації.

Деталізація результатів за класами атак в табл. 4 дозволяє оцінити ефективність методу для різних векторів загроз: від об'ємних атак, що формують виразні кластери (DoS, Probe), до складних для виявлення рідкісних подій (R2L, U2R), які часто маскуються під нормальний трафік і становлять найбільшу складність для методів на основі щільності.

Таблиця 4

Ефективність за категоріями атак (NSL-KDD)

Категорія	Traditional VAE	Baseline CL-IDS	Запропонований	Δ
DoS	92.3%	96.8%	97.9%	+1.1
Probe	88.7%	94.5%	96.1%	+1.6
R2L	71.2%	85.3%	89.7%	+4.4
U2R	58.3%	76.8%	83.2%	+6.4

Аналіз за категоріями атак в табл. 4 виявляє, що метод найбільш ефективний для складних атак R2L та U2R: +4.4 та +6.4 п.п. порівняно з Baseline CL-IDS. Це пояснюється тим, що ентропійний буфер краще зберігає рідкісні патерни нормального трафіку.

Оцінка здатності до узагальнення знань проводилась через експерименти з крос-доменного перенесення в табл. 5. Модель навчалась на джерельному наборі даних і тестувалась на цільовому без додаткового донавчання (zero-shot evaluation). Метрикою слугує точність класифікації на цільовому домені, що демонструє стійкість вивчених ознак до зміни мережевого середовища.

Таблиця 5

Крос-доменне перенесення

Напрямок	Trad. VAE	Baseline	Запропонований	Δ
NSL-KDD → CICIDS2017	62.3%	71.8%	78.4%	+6.6
CICIDS2017 → NSL-KDD	65.1%	74.3%	80.7%	+6.4
CICIDS2017 → UNSW-NB15	67.8%	76.1%	82.3%	+6.2
UNSW-NB15 → CICIDS2017	64.5%	73.9%	79.6%	+5.7
Середнє	64.9%	74.0%	80.3%	+6.2

Результати крос-доменного перенесення в табл.5 показують середнє покращення на 6.2 п.п. Найкраще перенесення – CICIDS2017 → UNSW-NB15 (82.3%), що пояснюється подібністю сучасних датасетів.

Таблиця 6

Аналіз абляції (CICIDS2017)

Конфігурація	Точність	FPR	Забування
Baseline CL-IDS	92.4%	2.8%	0.08
+ Динамічна калібрація	94.2%	2.2%	0.06
+ Ентропійний буфер	94.9%	1.9%	0.04
Повна система	95.1%	1.8%	0.03

Аналіз абляції в табл. 6 підтверджує внесок компонентів: динамічна калібрація додає 1.8 п.п. точності, ентропійний буфер – 0.7 п.п. Компоненти є взаємодоповнюючими.

Оцінка обчислювальної ефективності в табл. 7 включає вимірювання часу навчання на пакет (для оцінки пропускної здатності при оновленні моделі), затримки інференсу (latency), що є критичною для роботи в реальному часі, та пікового споживання пам'яті, що визначає вимоги до апаратного забезпечення.

Таблиця 7

Обчислювальна ефективність

Метод	Час навчання (ms/батч)	Inference (ms/sample)	Пам'ять (MB)
Baseline CL-IDS	130	2.6	225
+ Динамічна калібрація	135	2.6	230
+ Ентропійний буфер	145	2.6	240
Повна система	150	2.7	245

Обчислювальні витрати в табл. 7 зростають помірно: час навчання збільшується на 15% (150 ms проти 130 ms через обчислення ентропії). Однак inference latency залишається практично незмінною (2.7 ms), що критично для застосувань реального часу.

Висновки. У статті представлено розширений фреймворк безперервного навчання для IDS з трьома ключовими інноваціями: механізмом динамічної калібрації порогів на основі EMA, ентропійним критерієм управління буфером пам'яті, та протоколом крос-доменного перенесення. Експериментальна валідація на трьох датасетах продемонструвала: покращення точності на 2.7–2.9 п.п. порівняно з методами зі статичними порогами; зниження FPR до 1.8%; підвищення стійкості до раптового дрифту до 23%; покращення крос-доменного перенесення на 6.5 п.п. Аналіз за категоріями атак виявив, що метод особливо ефективний для складних атак R2L та U2R завдяки кращому збереженню рідкісних патернів нормального трафіку. Напрямки подальших досліджень: адаптивний вибір параметра α для EMA на основі детекції дрифту; застосування conformal prediction для статистичних гарантій порогів; федеративне навчання для розподіленого оновлення моделей.

References

1. Parisi, G. I., Kemker, R., Part, J. L., Kanan, C., & Wermter, S. (2019). Continual lifelong learning with neural networks: A review. *Neural Networks*, 113, 54–71. DOI: https://doi.org/10.1016/j.neunet.2019.01.012.

2. Chaudhry, A., Ranzato, M., Rohrbach, M., & Elhoseiny, M. (2019). Efficient Lifelong Learning with A-GEM. *International Conference on Learning Representations* (ICLR 2019). DOI: https://doi.org/10.48550/arXiv.1812.00420.

3. Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift

Література

1. Parisi G. I., Kemker R., Part J. L., Kanan C., Wermter S. Continual lifelong learning with neural networks: A review. *Neural Networks.* 2019. Vol. 113. P. 54–71, 2019. DOI: https://doi.org/10.1016/j.neunet.2019.01.012.

2. Chaudhry A., Ranzato M., Rohrbach M., Elhoseiny M. Efficient Lifelong Learning with A-GEM. *International Conference on Learning Representations* (ICLR 2019). 2019. DOI: https://doi.org/10.48550/arXiv.1812.00420.

3. Gama J., Žliobaite I., Bifet A., Pechenizkiy M., Bouchachia A. A survey on concept drift adaptation.

adaptation. *ACM Computing Surveys*, 46(4), Art. 44. DOI: https://doi.org/10.1145/2523813.

4. An, J., & Cho, S. (2015). Variational autoencoder based anomaly detection using reconstruction probability. *Special Lecture on IE*, 2(1), 1–18. URL: http://dm.snu.ac.kr/static/docs/TR/SNUDM-TR-2015-03.pdf.

5. Mirsky, Y., Doitshman, T., Elovici, Y., & Shabtai, A. (2018). Kitsune: An ensemble of autoencoders for online network intrusion detection. *Network and Distributed Systems Security Symposium* (*NDSS* 2018). DOI: https://doi.org/10.48550/arXiv.1802.09089.

6. Sener, O., & Savarese, S. (2018). Active Learning for Convolutional Neural Networks: A Core-Set Approach. *International Conference on Learning Representations* (ICLR 2018). URL: https://openreview.net/pdf?id=H1aIuk-RW.

7. Zhang, X., Zhao, R., Jiang, Z., Chen, H., Ding, Y., Ngai, E. C., & Yang, S. H. (2025). Continual Learning with Strategic Selection and Forgetting for Network Intrusion Detection. *IEEE International Conference on Computer Communications* (INFOCOM 2025). DOI: https://doi.org/10.48550/arXiv.2412.16264.

8. Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., Hassabis, D., Clopath, C., Kumaran, D., & Hadsell, R. (2017). Overcoming catastrophic forgetting in neural networks. *Proc. Natl. Acad. Sci.* (PNAS, U.S.A.), 114(13), 3521–3526. DOI: https://doi.org/10.1073/pnas.1611835114.

9. Li, Z., & Hoiem, D. (2018). Learning without Forgetting. *IEEE transactions on pattern analysis and machine intelligence*, 40(12), 2935–2947, Art. 8107520. DOI: https://doi.org/10.1109/TPAMI.2017.2773081.

10. Rebuffi, S.-A., Kolesnikov, A., Sperl, G., Lampert, C. H. (2017). iCaRL: Incremental Classifier and Representation Learning. *Computer Vision and Pattern Recognition* (CVPR 2017). DOI: https://doi.org/10.48550/arXiv.1611.07725.

11. Lopez-Paz, D., & Ranzato, M. (2017). Gradient Episodic Memory for Continual Learning. *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*. P. 6470–6479. DOI: https://doi.org/10.48550/arXiv.1706.08840.

12. Deja, K., Wawrzyński, P., Masarczyk, W., Marczak, D., & Trzciński, T. (2022). Multiband VAE: Latent Space Alignment for Knowledge Consolidation in Continual Learning. *Proceedings of the Thirty-First*

ACM Computing Surveys. 2014. Vol. 46, Iss. 4. Art. 44. DOI: https://doi.org/10.1145/ 2523813.

4. An J., Cho S. Variational autoencoder based anomaly detection using reconstruction probability. *Special Lecture on IE*. 2015. No. 2(1), P. 1–18. URL: http://dm.snu.ac.kr/static/docs/TR/SNUDM-TR-2015-03.pdf.

5. Mirsky Y., Doitshman T., Elovici Y., Shabtai A. Kitsune: An ensemble of autoencoders for online network intrusion detection. *Network and Distributed Systems Security Symposium* (*NDSS* 2018). 2018. DOI: https://doi.org/10.48550/arXiv.1802.09089.

6. Sener O., Savarese S. Active Learning for Convolutional Neural Networks: A Core-Set Approach. *International Conference on Learning Representations* (ICLR 2018). 2018. URL: https://openreview.net/pdf?id=H1aIuk-RW.

7. Zhang X., Zhao R., Jiang Z., Chen H., Ding Y., Ngai E. C., Yang S. H. Continual Learning with Strategic Selection and Forgetting for Network Intrusion Detection. *IEEE International Conference on Computer Communications* (INFOCOM 2025). 2025. DOI: https://doi.org/10.48550/arXiv.2412.16264.

8. Kirkpatrick J., Pascanu R., Rabinowitz N., Veness J., Desjardins G., Rusu A. A., Milan K., Quan J., Ramalho T., Grabska-Barwinska A., Hassabis D., Clopath C., Kumaran D., Hadsell R. Overcoming catastrophic forgetting in neural networks. *Proc. Natl. Acad. Sci.* (PNAS, U.S.A.). 2017. Vol. 114, Iss. 13. P. 3521–3526. DOI: https://doi.org/10.1073/pnas.1611835114.

9. Li Z., Hoiem D. Learning without Forgetting. *IEEE transactions on pattern analysis and machine intelligence*. 2018. Vol. 40, Iss. 12. P. 2935–2947. Art. 8107520. DOI: https://doi.org/10.1109/TPAMI.2017.2773081.

10. Rebuffi S.-A., Kolesnikov A., Sperl G., Lampert C. H. iCaRL: Incremental Classifier and Representation Learning. *Computer Vision and Pattern Recognition* (CVPR 2017). 2017. DOI: https://doi.org/10.48550/arXiv.1611.07725.

11. Lopez-Paz D., Ranzato M. Gradient episodic memory for continual learning. *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*. P. 6470–6479. DOI: https://doi.org/10.48550/arXiv.1706.08840.

12. Deja K., Wawrzyński P., Masarczyk W., Marczak D., Trzciński T. Multiband VAE: Latent Space Alignment for Knowledge Consolidation in Continual Learning. *Proceedings of the Thirty-First*

International Joint Conference on Artificial Intelligence (pp. 2902–2908). DOI: https://doi.org/10.24963/ijcai.2022/402.

13. Siffer, A., Fouque, P. A., Termier, A., & Largouet, C. (2017). Anomaly Detection in Streams with Extreme Value Theory. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (KDD '17), Association for Computing Machinery, New York, NY, USA (pp. 1067–1075). DOI: https://doi.org/10.1145/3097983.3098144.

14. Settles, B. (2009). Active Learning Literature Survey. Computer Sciences Technical Report 1648, University of Wisconsin-Madison. URL: http://digital.library.wisc.edu/1793/60660.

15. Wang, M., & Deng, W. (2018). Deep Visual Domain Adaptation: A Survey. *Neurocomputing*, 312, 135–153. DOI: https://doi.org/10.1016/j.neucom.2018.05.083.

16. Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., & Smola, A. J. (2012). A kernel two-sample test. *Journal of Machine Learning Research*, 13(1), 723–773. URL: https://www.jmlr.org/papers/volume13/gretton12a/gretton12a.pdf.

17. Kingma, D. P., & Welling, M. ()2014. Auto-encoding variational bayes. *International Conference on Learning Representations* (ICLR 2014). DOI: https://doi.org/10.48550/arXiv.1312.6114.

18. Higgins, I., Matthey, L., Pal, A., Burgess, C. P., Glorot, X., Botvinick, M. M., Mohamed, S., & Lerchner, A. (2016). beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework. *International Conference on Learning Representations* (ICLR 2017). URL: https://openreview.net/forum?id=Sy2fzU9gl.

19. Kozachenko, L. F., & Leonenko, N. N. (1987). Sample estimate of the entropy of a random vector. *Problems of Information Transmission*, 23(2), 95–101. URL: https://dmitripavlov.org/scans/kozachenko-leonenko.pdf.

20. Tavallaee, M., Bagheri, E., Lu, W., & Ghorbani, A. A. (2009). A detailed analysis of the KDD CUP 99 data set. *IEEE Symposium on Computational Intelligence for Security and Defense Applications*. URL: https://www.ee.torontomu.ca/~bagheri/papers/cisda.pdf.

21. Sharafaldin, I., Habibi, Lashkari, A., & Ghorbani, A. A. (2018). Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization. In *Proceedings of the 4th International Conference on Information Systems Security and Privacy* (ICISSP 2018), Vol. 1 (pp. 108–116). DOI: https://doi.org/10.5220/0006639801080116.

International Joint Conference on Artificial Intelligence (IJCAI-22). 2022. P. 2902–2908. DOI: https://doi.org/10.24963/ijcai.2022/402.

13. Siffer A., Fouque P. A., Termier A., Largouet C. Anomaly Detection in Streams with Extreme Value Theory. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (KDD '17), Association for Computing Machinery, New York, NY, USA, 2017. (pp. 1067–1075). DOI: https://doi.org/10.1145/3097983.3098144.

14. Settles B. Active Learning Literature Survey. Computer Sciences Technical Report 1648, University of Wisconsin-Madison, 2009. URL: http://digital.library.wisc.edu/1793/60660.

15. Wang M., Deng W. Deep Visual Domain Adaptation: A Survey. *Neurocomputing*. 2018. Vol. 312. P. 135–153. DOI: https://doi.org/10.1016/j.neucom.2018.05.083.

16. Gretton A., Borgwardt K. M., Rasch M. J., Schölkopf B., Smola A. J. A kernel two-sample test. *Journal of Machine Learning Research.* 2012. No. 13(1). P. 723–773. URL: https://www.jmlr.org/papers/volume13/gretton12a/gretton12a.pdf.

17. Kingma D. P., Welling M. Auto-encoding variational bayes. *International Conference on Learning Representations* (ICLR 2014). 2014. DOI: https://doi.org/10.48550/arXiv.1312.6114.

18. Higgins I., Matthey L., Pal A., Burgess C. P., Glorot X., Botvinick M. M., Mohamed S., Lerchner A. beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework. International Conference on Learning Representations (ICLR 2017). 2017. URL: https://openreview.net/forum?id=Sy2fzU9gl.

19. Kozachenko L. F., Leonenko N. N. Sample estimate of the entropy of a random vector. *Problems of Information Transmission.* 1987. No. 23(2). P. 95–101. URL: https://dmitripavlov.org/scans/kozachenko-leonenko.pdf.

20. Tavallaee M., Bagheri E., Lu W., Ghorbani A. A. A detailed analysis of the KDD CUP 99 data set. *IEEE Symposium on Computational Intelligence for Security and Defense Applications*. 2009. URL: https://www.ee.torontomu.ca/~bagheri/papers/cisda.pdf.

21. Sharafaldin I., Habibi Lashkari A., Ghorbani A. A. Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization. In *Proceedings of the 4th International Conference on Information Systems Security and Privacy* (ICISSP 2018). 2018. Vol. 1. P. 108–116. DOI: https://doi.org/10.5220/0006639801080116.

22. Moustafa, N., & Slay, J. (2015). UNSW-NB15: A comprehensive data set for network intrusion detection systems. *Military Communications and Information Systems Conference* (MilCIS 2015). DOI: https://doi.org/10.1109/MilCIS.2015.7348942.

22. Moustafa N., Slay J. UNSW-NB15: A comprehensive data set for network intrusion detection systems. *Military Communications and Information Systems Conference* (MilCIS 2015). 2015. DOI: https://doi.org/10.1109/MilCIS.2015.7348942.

STATSENKO VOLODYMYR
Doctor of Technical Sciences, Professor,
Department of Computer Engineering and Electromechanics,
Kyiv National University of Technologies and Design, Ukraine
https://orcid.org/0000-0002-3932-792X
Scopus Author ID: 57210344190
Researcher ID: C-3646-2017
E-mail: statsenko.v@knutd.edu.ua

BONDARENKO ANTON
PhD student,
Department of Computer Engineering and Electromechanics,
Kyiv National University of Technologies and Design, Ukraine
https://orcid.org/0009-0007-5087-6173
E-mail: anton.bondarenko.ua@gmail.com

Volodymyr STATSENKO, Anton BONDARENKO
Kyiv National University of Technologies and Design, Ukraine

ADAPTIVE CONTINUAL LEARNING FOR INTRUSION DETECTION SYSTEMS: ENTROPY-BASED MEMORY MANAGEMENT AND DYNAMIC THRESHOLD CALIBRATION

Objective*: To develop an extended continuous learning framework for intrusion detection systems (IDS) that provides adaptive detection threshold setting under conceptual drift, theoretically grounded memory buffer management based on entropy criteria, and cross-domain knowledge transfer between different network environments.*

Methodology*: A dynamic threshold calibration mechanism based on online statistical tracking via exponential moving average (EMA) is proposed to adapt decision boundaries in real time with theoretical guarantees of limiting false positive rate (FPR). To optimally overcome catastrophic forgetting, an entropy memory management strategy is developed using a differential k-nearest neighbor estimator to approximate the latent space entropy. A cross-domain adaptation module based on maximum mean divergence (MMD) is integrated to ensure knowledge transfer without additional training. Empirical validation was performed on the NSL-KDD, CICIDS2017, and UNSW-NB15 benchmarks.*

Results*. On the complex CICIDS2017 dataset, the proposed method achieves an accuracy of 95.1% with a controlled false positive rate, outperforming state-of-the-art baseline methods. Improved robustness to sudden conceptual drift up to 23% with fast recovery within 3–5 batches is demonstrated. Entropy memory management provides significant improvements in detecting minority attack classes (R2L and U2R).*

Scientific novelty*. For the first time, a synergistic combination of dynamic threshold calibration with information-theoretic memory management for intrusion detection systems is proposed, which allows simultaneously solving the problems of drift adaptation and catastrophic forgetting with formal statistical guarantees.*

Practical significance*. The results confirm that integrating dynamic calibration with entropy memory management provides scalable and robust protection for next-generation networks.*

Keywords: *continual learning; intrusion detection systems; variational autoencoders; dynamic threshold calibration; entropy-based memory management.*