



DOI 10.28925/2663-4023.2026.35.1426

УДК 004.056

Стрельніков Віталій Ігорович

PhD in Computer engineering,

доцент кафедри інформаційної та кібернетичної безпеки імені професора Володимира Бурячка
Київський столичний університет імені Бориса Грінченка, Київ, Україна

ORCID: 0000-0003-3439-3220

v.strelnikov@kubg.edu.ua

Ворохоб Максим Віталійович

PhD in Cybersecurity

доцент кафедри інформаційної та кібернетичної безпеки імені професора Володимира Бурячка
Київський столичний університет імені Бориса Грінченка, м. Київ, Україна

ORCID: 0000-0001-5160-7134

m.vorokhob@kubg.edu.ua

Бржевська Зореслава Михайлівна

PhD in Cybersecurity, доцент,

доцент кафедри інформаційної та кібернетичної безпеки імені професора Володимира Бурячка
Київський столичний університет імені Бориса Грінченка, Київ, Україна

ORCID: 0000-0002-7029-9525

z.brzhevska@kubg.edu.ua

Круглик Владислав Сергійович

доктор педагогічних наук, професор

професор кафедри інформатики і кібернетики

Мелітопольський державний педагогічний

університет імені Богдана Хмельницького, Запоріжжя, Україна

ORCID: 0000-0002-5196-7241

krugvs@mdpu.org.ua

ЕМПІРИЧНИЙ АУДИТ ПРІОРИТИЗАЦІЇ ВРАЗЛИВОСТЕЙ ЗА EPSS V4 І V5 BETA

Анотація. Досліджено, як відрізняються пріоритетні списки вразливостей, отримані за EPSS v4 та EPSS v5 beta, і які обмеження має оцінювання цих відмінностей за наступними включеннями до каталогу CISA Known Exploited Vulnerabilities. Мета роботи — зіставити зміну складу списків із наступними записами KEV та перевірити часову доступність самих файлів оцінок. Використано всі 18 парних дат, наявних у зафіксованому відкритому beta-архіві: від 22 травня до 8 червня 2026 року. Для кожної дати сформовано спільну сукупність CVE, оцінених обома версіями, і вилучено записи, вже включені до KEV. Порівняно списки розміру 100, 500, 1000 і 5000 та охоплення наступних включень до каталогу за 30 і 90 днів. Повторні спостереження одного CVE збережено разом під час кластерного bootstrap; додатково перевірено групування за датою включення до KEV та однакову вагу кожного CVE. У списках top-100 двох версій спільними є лише 21–23 CVE, хоча між сусідніми днями кожної окремої версії зберігається щонайменше 98% цього списку. За 90-денного горизонту 420 позитивних спостережень CVE-дата відповідають лише 35 різним CVE. Для top-5000 частка охоплення з вагами за кількістю дат становить 7,86% у v4 та 15,71% у beta; парна різниця дорівнює 7,86 відсоткового пункту, а розвідувальний 95% кластерний інтервал — від –4,18 до 21,44 пункту. Аудит Git виявив, що поточні байти 12 beta-файлів мають коміт пізніше за дату оцінки, а три файли виправлено без зміни позначень дати та версії моделі в заголовку. На шести датах за критерієм коміту того самого дня оцінка різниці також має високу невизначеність. Результати показують необхідність розділяти зміну пріоритетного списку, доступність оцінок і невизначеність охоплення наступних KEV. Практичним результатом є відтворений аудит із пофайловими хешами та перевірками, придатний для аналізу міграції джерел оцінювання. У чистому тестовому середовищі повторено обчислення, а на трьох раніше не використаних липневих зрізах випущеної v5 одержано повний збіг відбору й метрик у двох програмних реалізаціях. Висновки не переносяться автоматично на



випущену v5, реальну експлуатацію або інвентар конкретної організації.

Ключові слова: EPSS; CVE; CISA KEV; пріоритизація вразливостей; відтворюваність; часова доступність; кластерний bootstrap.

ВСТУП

Постановка проблеми. Оновлення моделі оцінювання вразливостей може змінити склад черги перевірки без зміни її запланованого розміру. FIRST повідомила про початок публікації EPSS v5 15 червня 2026 року та зміни калібрування і джерел ознак [1]. Документація окремо застерігає, що перехід між поколіннями моделі змінює шкалу оцінок [2]. Для команди, яка відбирає фіксовану кількість CVE, важливо з'ясувати, які саме записи залишають чергу та потрапляють до неї. Проте сама заміна записів ще не показує, чи покращилася якість відбору.

Аналіз останніх досліджень і публікацій. Попередні праці доцільно розрізняти за ціллю оцінювання та моментом доступності доказу. Денисюк побудував модель на структурованих даних NVD і мітках KEV, використав стратифікований поділ та дослідив пороги рішення [3]. EPSS розглянуто в огляді цієї праці, а власний експеримент спирається на ознаки NVD і мітки KEV. У препринті Sherif та співавторів порівнюються EPSS і комбінований показник, зокрема за KEV Recall@k; методичний розділ описує історичні оцінки та наступне 30-денне включення до KEV [4]. Комбінування сигналів, фіксований бюджет і майбутня KEV-ціль самі по собі не утворюють внеску цієї роботи.

Koscinski та співавтори порівняли CVSS, SSVC, EPSS і Exploitability Index, дослідили перетин top-N, наслідки однакових оцінок та поведінку EPSS перед входженням CVE до KEV [5]. Порівняння систем проведено на 600 Microsoft CVE; поглиблений аналіз EPSS/KEV у повному тексті описано ширше. Тому наша робота не претендує на запровадження метрики перетину або самого історичного оцінювання. Відмінність повинна впливати з конкретного досліджуваного та перевірених умов його відтворення.

Часову коректність оцінювання розглядають Iannone та співавтори: використання пізніше доступної інформації може створити нереалістичний результат моделі, а відсутність позитивного запису потребує обережного трактування [6]. Відкритий протокол Bakare передбачає відновлення історичних записів CVE з Git, прив'язку EPSS до дати рішення, вилучення вже відомих KEV та контроль повноти горизонту спостереження [7]. Ці принципи застосовуємо як методичну основу; власним об'єктом аудиту є дата появи конкретних байтів beta-файлів EPSS.

Найближчий попередній аналіз тих самих даних — відкритий звіт Gamblin / RogoLabs EPSS-Compare [8]. Його код і метадані містять ту саму панель із 18 дат, порівняння розподілів, нахилів траєкторій та мінливості оцінок v4/v5 beta. Вендорський матеріал Haskuity також описує перерозподіл пріоритетів між версіями [9]. Отже, новими не є ані ця панель, ані спостереження зміни оцінок. У перевіреному коді Gamblin (коміт 4428d8ed7120611dbe81be3b23f2a211f373e36a) не реалізовано оцінювання наступних включень до KEV серед CVE поза відомим каталогом, кластерних інтервалів такого охоплення та аудиту найранішого коміту збереженої версії кожного файлу. Саме це поєднання утворює вузьке емпіричне питання роботи. Межі цього внеску визначає перевірений корпус джерел. Огляд обмежено доступними повними текстами; абсолютний пріоритет такого поєднання не встановлено.

Мета статті. Виміряти відмінність пріоритетних списків v4/v5 beta за однакової дати оцінки, визначити їхнє охоплення наступних включень до KEV і встановити, як повторення CVE та історія публікації файлів обмежують інтерпретацію результату. Практична цінність полягає у перевіреному способі аудиту зміни джерела оцінювання перед використанням його для побудови черги.

МЕТОДИКА ДОСЛІДЖЕННЯ

Джерела та фіксація експерименту

Використано відкриті щоденні файли EPSS та зафіксований каталог CISA KEV [2], [10], [11]. На момент заморожування даних, 28 вересня 2026 року, beta-каталог у вибраному коміті містив 18 послідовних дат: 22.05–08.06.2026. Для кожної отримано production-файл v4 того самого дня. У заголовках зазначено моделі v2025.03.14 і v2026.05.12. Фінальний реліз v2026.06.15 розглянуто тільки в окремому описовому зіставленні 14 і 15 червня. Ця пара не входить до основної панелі чи її bootstrap.

Для кожного файлу збережено URL, SHA-256, розмір і заголовки. Репозиторій beta зафіксовано комітом f9b61220a5a289812f54c3c1327b178c51f01d08, KEV — комітом 4d1aff213a881341db8a276c89802e05bf69481c. Вихідний каталог KEV має дату випуску 27.09.2026, тож усі 30- та 90-денні вікна завершуються до неї. Пізні описи CVE або KEV не використовуються як ознаки ранжування. Дані не доповнюються синтетичними вразливостями чи організаційним інвентарем.



Для beta додатково відновлено доступну Git-історію та найраніший коміт, що містить саме збережений blob. Перевірено і SHA-1 Git blob, і SHA-256 файла. Час коміту розглядається як наближення доступності у цьому репозиторії. Він не доводить час публічного push і не виключає іншого каналу розповсюдження. Така межа важлива: дата оцінки в заголовку не є достатнім доказом доступності цих байтів у минулому.

Сукупність CVE та результат спостереження

Для дати t спільна придатна сукупність $U(t)$ складається з CVE, що мають оцінку в обох файлах і за зафіксованим каталогом ще не включені до KEV на t . CVE з dateAdded не пізніше t вилучаються до сортування. У кожній версії записи впорядковуються за спаданням EPSS, а за однакових значень — лексикографічно за CVE ID. Аналізуються бюджети $k = 100, 500, 1000, 5000$. Це обрана після пілоту сітка від вузького до ширшого відбору, а не нормативні потужності SOC чи оптимізовані пороги. Вона показує, як результат залежить від розміру списку. Бюджет задає розмір відбору з глобального переліку; він не вимірює обсяг доступних робіт конкретної організації.

Позитивна подія визначається як включення CVE до KEV у напіввідкритому часовому інтервалі $(t; t + H]$, де H дорівнює 30 або 90 календарним дням. Дату t виключено, кінцеву дату включено. Горизонт 30 днів забезпечує близький календарний масштаб до прогнозу EPSS; 90 днів дає довший період спостереження каталогу й більше різних позитивних CVE. Вибір довшого вікна є розвідувальним і не змінює KEV-подію на ціль експлуатації EPSS. CISA визначає dateAdded як дату внесення запису до каталогу [10]. Вона не є датою першої експлуатації. Невключення до KEV до кінця вікна означає тільки відсутність спостереженого включення у цьому каталозі.

На кожен дату обліковуються всі наступні KEV та чотири взамовиключні категорії доступності EPSS: в обох версіях, лише у v4, лише в beta, у жодній. Охоплення серед $U(t)$ оцінює ранжування порівнянних CVE. Охоплення серед усіх наступних KEV додатково враховує недоступність оцінки. Ці знаменники не можна підміняти один одним. Зокрема, CVE без оцінки на t не зараховується як помилка ранжування серед оцінених.

Метрики та залежність спостережень

Стабільність top- k описують кількість спільних CVE, її частка від k , індекс Жаккара як відношення перетину до об'єднання та кількість заміненних записів k мінус розмір перетину. Для зафіксованої повної сукупності це точні описові величини. Діапазон за 18 датами показує часову варіацію, але не є довірчим інтервалом. Вплив нічий перевіряється через можливі альтернативні склади граничної групи однакових оцінок.

Денну стабільність кожної версії обчислено окремо. Для двох сусідніх дат залишаються лише CVE, наявні в усіх чотирьох відповідних файлах, а вже відомі KEV вилучаються на пізнішу дату. Ранжування повторюється на цій спільній сукупності. Отримана величина задає описовий масштаб щоденного руху списку. Вона не ізолює внесок окремих ознак або алгоритмічних змін EPSS.

Охоплення наступних KEV серед оцінених CVE визначається як кількість позитивних записів у top- k , поділена на кількість позитивних записів у $U(t)$. Для панелі подано також відношення сум таких чисельників і знаменників. Воно зважає CVE за кількістю дат, на яких вони відповідають умовам позитивної події, і не трактується як вибірка незалежних подій. Додатково обчислюється середня за різними CVE парна різниця часток їх відбору; цей варіант надає однакову вагу кожному CVE.

Для оцінювання невизначеності використано парний кластерний bootstrap: повторно вибираються різні позитивні CVE, а всі рядки одного вибраного CVE переносяться разом. Такий принцип збереження кластера спирається на відому методику повторної вибірки залежних даних [12]. Виконано 20000 повторів із фіксованим seed 20260928; межі визначено 2,5-м і 97,5-м процентилями різниці охоплення. Ранги й календарні дати залишаються зафіксованими. Інтервали характеризують умовну чутливість до складу спостережених KEV-подій, а не невизначеність навчання EPSS, усієї експлуатації або випадкового вибору календарного періоду.

Додатковий bootstrap групує всі CVE з однаковою датою включення до KEV, щоб перевірити чутливість до пакетного додавання. Це не повна модель залежності між продуктами чи кампаніями. Для окремих дат у додаткових таблицях наведено інтервали Вілсона для частки охоплення за біноміального припущення; вони не об'єднуються як незалежні оцінки. Якщо всі спостережені парні різниці дорівнюють нулю, bootstrap не може створити неспостережені розбіжності. Вироджений інтервал у такому випадку не використовується як доказ еквівалентності.

Дослідження є розвідувальним: бюджети та часові вікна розширено після пілоту. Формальні заяви про статистичну перевагу або підтвердження наперед зареєстрованої гіпотези не робляться. Розрахунки перевірено незалежною реалізацією відбору й ранжування через pandas; bootstrap додатково повторено вибіркою індексів кластерів з іншим seed і 50000 повторів.



Відтворення у тестовому середовищі

Перевірку виконано у двох окремих режимах. Спочатку у новому віртуальному середовищі Python 3.14.7 на macOS arm64 встановлено зафіксовані залежності: NumPy 2.5.2, pandas 3.0.5 та Matplotlib 3.11.2. У тимчасовий робочий каталог скопійовано вихідні файли й код без готових результатів основного аналізу. Повторно виконано перевірку контрольних сум, аналіз, оцінювання чутливості та незалежний перерахунок; результати звірено побайтово. Це відтворення обчислень на тих самих вхідних даних, а не нова вибірка.

Окремо 12 жовтня 2026 року зафіксовано параметри додаткового ретроспективного тесту до завантаження його файлів: production-зрізи EPSS за 1, 8 і 15 липня 2026 року, горизонт 30 днів і списки розміру 100, 500, 1000 та 5000. Ці дати раніше не входили до пілоту або beta-панелі. Файли отримано з відкритого архіву EPSS [2]; у заголовках усіх трьох зазначено модель v2026.06.15. Використано той самий зафіксований каталог KEV [10], дата випуску якого пізніша за завершення всіх трьох вікон спостереження. Протокол, URL, час отримання та SHA-256 збережено окремо від основних даних. Ця локальна фіксація параметрів не є проспективною пререєстрацією дослідження.

Перед ранжуванням перевірено структуру CSV, відповідність дати заголовка, унікальність ідентифікаторів, скінченність оцінок та їх належність діапазону від 0 до 1. Відомі на дату зрізу KEV вилучено. Реалізацію на стандартній бібліотеці Python зіставлено з окремим читанням, об'єднанням і сортуванням через pandas; для кожної дати перевірено придатну сукупність, множину наступних KEV, повний склад кожного тор-списку та кількість збігів. Нові зрізи є перевіркою виконаності процедури на іншому періоді, а не парним порівнянням v4 і v5: оцінок v4 за ті самі липневі дати тут немає.

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Дата оцінки та доступність збереженого файла

Аудит виявив відмінність між датою оцінки та датою коміту поточних байтів для 12 із 18 beta-файлів. Найбільша різниця календарних дат становить шість днів. Лише файли за 22, 27, 28, 29 травня, 7 і 8 червня мають найраніший коміт поточного blob тієї самої дати. Рис. 1 показує розподіл затримок. Отже, основна панель описує ретроспективно доступні парні оцінки; наявна Git-історія не підтверджує можливості використати саме ці файли того дня для всіх 18 дат. Про їхню доступність через інші канали цей аудит не дає висновку.

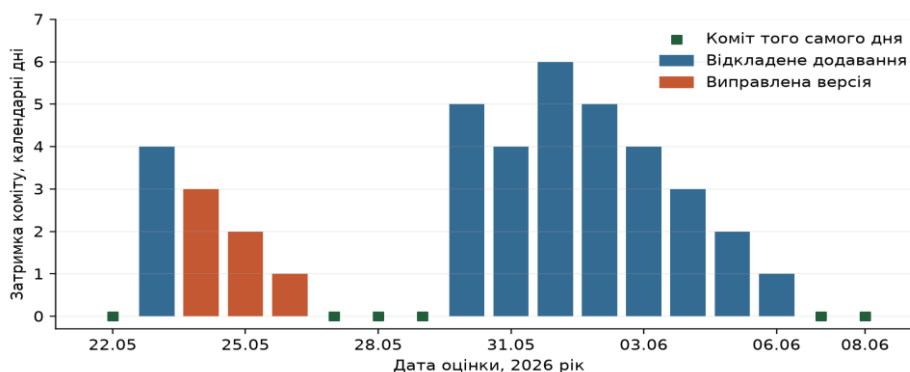


Рис. 1. Різниця календарних дат між `score_date` та найранішим комітом збереженого beta-blob; час коміту є наближенням доступності

Файли за 24–26 травня спочатку додано пізніше від дати оцінки, а 27 травня виправлено. Заголовки `model_version` і `score_date` до та після виправлення однакові. Водночас змінилися і кількість записів, і оцінки частини спільних CVE (табл. 1). Цей результат показує, що ідентифікація вхідних даних лише парою «версія моделі — дата» недостатня для відтворення цього дослід. Потрібна контрольна сума конкретних байтів.

Таблиця 1

Зміни у виправлених beta-файлах

Дата оцінки	CVE до / після виправлення	Змінених оцінок серед спільних CVE	Змінених процентилів
24.05.2026	334676 / 334734	2989	289341



Дата оцінки	CVE до / після виправлення	Змінених оцінок серед спільних CVE	Змінених процентилів
25.05.2026	334676 / 334817	5319	308206
26.05.2026	334676 / 334999	8356	325654

Вплив виправлень перевірено окремо на перетині CVE з трьох файлів: v4, первісної beta та виправленої beta, після вилучення відомих KEV. У top-5000 до й після виправлення спільними залишаються 4996, 4995 і 4995 CVE відповідно; у top-100 — 100, 99 і 99. За всіх чотирьох бюджетів і обох горизонтів кількість наступних KEV на цій зафіксованій сукупності не змінюється. На цьому перетині виправлення не змінюють різниці охоплення; висновок не поширюється на CVE, додані або вилучені під час виправлення. Наявність різних файлів обґрунтовує точну ідентифікацію входів, але не твердження про суттєве викривлення результату.

Міжверсійна та щоденна стабільність

Спільна придатна сукупність збільшується від 332965 до 336883 CVE. У top-100 двох версій спільними залишаються тільки 21–23 записи: 77–79 позицій змінюються. Частка перетину збільшується разом із бюджетом, але навіть у top-5000 становить лише 60,30–60,50% (табл. 2, рис. 2). Ці величини вимірюють зміну складу відбору на спільній сукупності. Кількість відвернених інцидентів у цьому досліді не вимірюється.

Таблиця 2

Стабільність top-k на зафіксованій панелі

k	Спільних CVE між версіями, min–max	Медіана	Денне збереження v4, %	Денне збереження beta, %
100	21–23	23	98,00–100,00	99,00–100,00
500	183–186	184	99,20–100,00	99,40–100,00
1000	422–426	424	99,60–100,00	99,80–100,00
5000	3015–3025	3021	99,74–99,96	99,90–100,00

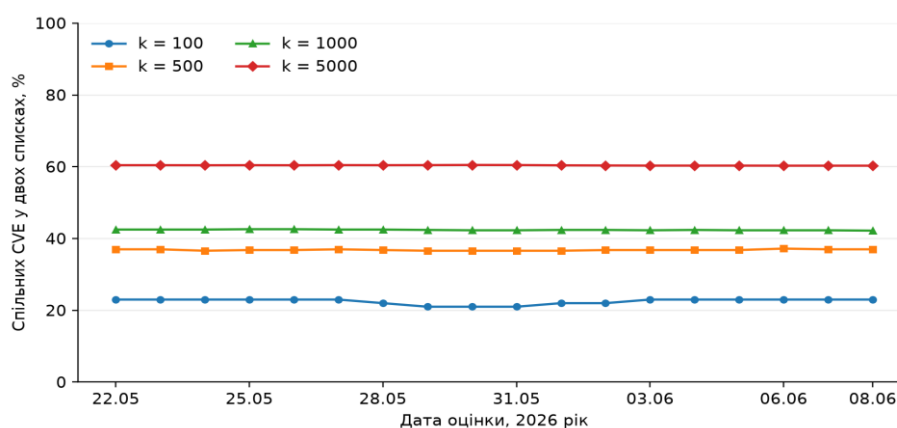


Рис. 2. Частка спільних CVE у пріоритетних списках v4 і v5 beta за однакової дати оцінки

Усередині кожної версії списки набагато сталіші: за сусідніми днями зберігається щонайменше 98 зі 100 CVE у v4 та 99 зі 100 у beta. Отже, велика міжверсійна розбіжність не пояснюється самим типовим денним рухом списків у цьому короткому вікні. Водночас це зіставлення не визначає причину кожної зміни й не дає причинної оцінки алгоритму. Граничні групи однакових оцінок у парній панелі мають не більше п'яти CVE; їх перестановки не змінюють кількість наступних KEV у відповідних top-k. Точний перебір граничних складів змінює міжверсійний перетин лише для $k = 5000$ на двох датах: 1 червня можливі 3019–3020 спільних CVE, 5 червня — 3015–3016. Підсумковий діапазон за панеллю залишається 3015–3025.

Перевірено також самі значення оцінок. Межа top-100 за 18 дат лежить у межах 0,93847–0,93878



для v4 та 0,97105–0,97106 для beta. Проте 8 червня 77 CVE, що залишають top-100 v4, мають у beta медіанний ранг 521, а 77 нових учасників beta мали у v4 медіанний ранг 623. Для першої групи медіани оцінок у відповідних версіях становлять 0,94035 і 0,85784, для другої — 0,91142 і 0,98810. Для 8 червня ці ранги показують переміщення за межі безпосереднього околу порога. Окремий аналіз нічий на всій панелі виключає лексикографічне правило як пояснення основної різниці складу. Водночас міжверсійні дельти балів не тлумачаться як зміна реального ризику, оскільки шкали й калібрування моделей відрізняються [1], [2].

Доступність оцінки та охоплення наступних KEV

Чисельність майбутніх записів значно менша за чисельність оцінених CVE. За 90 днів на окрему дату припадає 70–83 наступних включень до KEV, але лише 19–28 із них належать до спільної придатної сукупності. Наприклад, для 08.06.2026 каталог містить 81 наступний запис у 90-денному вікні; 27 CVE оцінено обома версіями, а 54 не мають оцінки в жодній. Причини відсутності оцінок для цих 54 CVE за використаними файлами не встановлено; їх збережено в обліку всіх наступних KEV, але не в знаменнику охоплення серед оцінених.

На цій даті top-5000 охоплює два з 27 оцінених майбутніх KEV у v4 і чотири у beta. Відносно всіх 81 наступних включень відповідні частки становлять 2,47% та 4,94%. Обидва представлення потрібні: перше описує ранг серед порівнянних записів, друге — досягне охоплення всіх зафіксованих наступних включень за такої доступності оцінок. Низька друга частка не може бути повністю приписана помилкам ранжування.

Усі 18 дат разом дають 420 позитивних рядків за 90 днів, але лише 35 різних CVE. Табл. 3 подає відношення сум збігів і позитивних рядків, а не 420 незалежних спроб. Для top-5000 воно дорівнює 33/420 у v4 і 66/420 у beta. Ці збіги стосуються лише трьох різних CVE у v4 і п'яти у beta; два з них хоча б раз відібрані обома версіями. У top-500 усі 12 позитивних рядків beta належать CVE-2021-23758. Для top-5000 множина збігів v4 складається з CVE-2021-23758, CVE-2023-49105 і CVE-2024-21182; у beta це CVE-2019-1068, CVE-2021-23758, CVE-2024-21182, CVE-2026-21962 та CVE-2026-39808. Події, дати й ранги наведено в таблиці future-kev-events.csv електронного додатка. Збільшення чисельника вдвічі тут не доводить подвоєння ефективності: ті самі CVE повторюються на різних датах, а інтервали мають широкий розкид.

Таблиця 3

Охоплення наступних KEV за 90 днів: 18 дат, 35 різних позитивних CVE

k	Збігів v4 / beta з 420 рядків	Охоплення v4 / beta, %	Різниця, в. п.	95% інтервал різниці, CVE-кластери
100	0 / 0	0,00 / 0,00	0,00	Вироджений; не доказ рівності
500	0 / 12	0,00 / 2,86	2,86	[0,00; 9,35]
1000	11 / 12	2,62 / 2,86	0,24	[-7,67; 8,53]
5000	33 / 66	7,86 / 15,71	7,86	[-4,18; 21,44]

За 30-денного горизонту є 154 позитивні рядки та 17 різних CVE. У списках $k \leq 1000$ обидві версії мають нуль збігів; за $k = 5000$ — однакові 10/154, або 6,49%. Усі спостережені парні різниці тут нульові. Bootstrap повертає нульовий розкид різниці, бо повторно вибирає тільки наявні події. З цього не випливає еквівалентність моделей для іншого періоду чи неспостереженої експлуатації.

Невизначеність і перевірки чутливості

Для 90-денного горизонту та $k = 5000$ парна різниця охоплення дорівнює +7,86 в. п., а CVE-кластерний інтервал — [-4,18; 21,44] в. п. Він сумісний із відмінностями обох знаків. Групування за датою включення до KEV дає близький інтервал [-4,26; 21,35] в. п. Однакова вага кожного CVE змінює оцінювану величину: середня різниця часток його відбору становить +5,71 в. п. з інтервалом [-5,71; 17,14]. Ці оцінки не дають підстав для висновку про перевагу beta за охопленням наступних KEV; відсутність такого висновку не є доказом відсутності ефекту.

Додаткове обмеження шістьма датами, для яких поточний beta-blob має коміт того самого дня, залишає 142 позитивні рядки й ті самі 35 різних CVE. Охоплення дорівнює 11/142 (7,75%) для v4 і 23/142 (16,20%) для beta; різниця +8,45 в. п., інтервал [-2,48; 21,64] в. п. Рис. 3 зіставляє ці варіанти. Вони перевіряють чутливість одного досліджу; їх не слід трактувати як незалежні підтвердження.

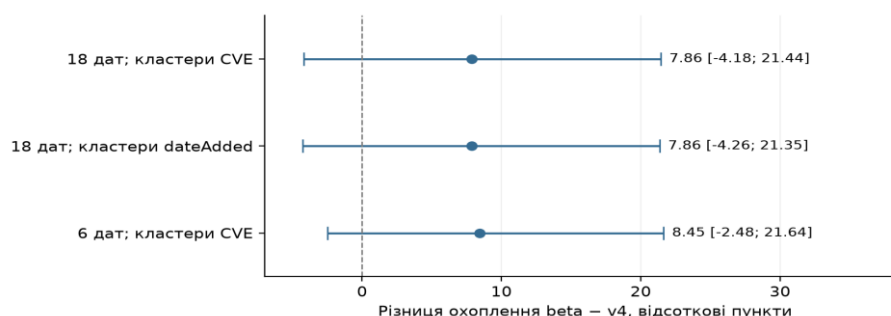


Рис. 3. Різниця охоплення beta мінус v4 для $k = 5000$ та 90-денного горизонту; розвідувальні 95% перцентильні інтервали

Окрема пара на межі релізу охоплює 338584 порівнянні CVE та 30 наступних KEV за 90 днів. Спільними у top-100 є 23 CVE; top-5000 має два збіги у v4 за 14 червня і чотири у випущеній v5 за 15 червня. Описова схожість із beta-пілотом не робить їх тим самим експериментом: одночасно змінилися модель і дата надходження сигналів. Висновок про ефект фінальної v5 із цієї пари не ідентифікується.

Результати тестового відтворення

Чистий повторний прогін завершив усі 12 етапів без помилок. У журналі reproduction.json електронного додатка наведено команди, коди завершення й пофайлові SHA-256. Для 15 вихідних артефактів, зокрема числових результатів, таблиць і PNG-графіків, отримано збіг SHA-256 із зафіксованою версією перед редактурою статті. Подальше доповнення тексту змінює метадані рукопису, але не ці числові результати. Перевірка підтверджує обчислювальну відтворюваність цього пакета у зазначеному середовищі; вона не відтворює навчання моделей EPSS.

Додатковий тест перевіряє реалізацію процедури, а не нову наукову гіпотезу про моделі. Липневі файли пройшли перевірки якості даних. Протокол, вхідний маніфест, таблиця подій і звіт results.json містяться в каталозі test-environment електронного додатка. Усі 30 точних звірок між двома реалізаціями дали однакові результати: по дві перевірки сукупностей і по дві перевірки для кожного з чотирьох бюджетів на кожну дату. Результати для найбільшого списку наведено в табл. 4. Для списків розміру 100, 500 та 1000 збігів із наступними 30-денними KEV не було на жодній із трьох дат.

Таблиця 4

Додатковий тест на липневих зрізах випущеної EPSS v5

Дата зрізу	Оцінено CVE	Придатно CVE	Наступні KEV: оцінені / усі	Топ-5000: збіги / охоплення оцінених
01.07.2026	344852	343221	12 / 25	1 / 8,33%
08.07.2026	346176	344541	10 / 27	1 / 10,00%
15.07.2026	348601	346957	8 / 21	2 / 25,00%

Охоплення серед усіх наступних KEV для top-5000 становить відповідно 4,00%, 3,70% та 9,52%. Відмінність від останнього стовпця табл. 4 пояснюється відсутністю оцінок для 13, 17 і 13 майбутніх записів каталогу на відповідні дати. Вікна перекриваються, тому ці числа не підсумовуються як незалежні події. Отриманий результат засвідчує успішний тестовий запуск на раніше не використаних зрізах і відтворення метрик окремою реалізацією. Невелика кількість позитивних записів, спільний ретроспективний каталог і відсутність парних оцінок v4 не дозволяють із цього тесту визначити ефект переходу на v5.

Обмеження та практичне значення

Найсильніша частина результату — точний опис зміни списків і походження файлів. Натомість оцінка охоплення спирається на малу кількість позитивних CVE, неповний каталог і дату його поповнення. KEV-ціль відрізняється від прогнозованої EPSS події експлуатації протягом 30 днів. Тому подані частки не перевіряють калібрування EPSS щодо його власної цілі й не вимірюють частку відвернених атак. Затримка між експлуатацією та включенням до каталогу може переміщувати спостережувану подію між вікнами; її тривалість у цьому дослідді невідома.

Обмеження дослідження стосуються також часової реконструкції. Проаналізовано збережену Git-



історію, але не час публічного push, усі можливі канали розповсюдження чи історію навчальних даних. Повну історію виправлень і видалень записів KEV не реконструйовано; мітки відновлюються за одним пізнішим знімком каталогу. Набір навчальних даних та всі вхідні сигнали EPSS у цьому досліді не перевірялись; відсутність витоку на рівні самого навчання тут не засвідчується. Підмножина шести дат зменшує конкретну проблему запізнених комітів, але не перетворює дослід на доведене відтворення оперативного процесу.

Кластерний bootstrap зберігає повтори CVE, проте передбачає достатню обмінність різних кластерів для умовного узагальнення. Можлива залежність CVE за спільним продуктом або кампанією не вимірювалася. У самих даних 35 позитивних CVE мають 25 різних дат включення до KEV, що мотивує додаткове групування за цією датою; внутрішню процедуру роботи CISA з цього не виводимо. Перевірка за датою KEV охоплює лише частину цієї залежності. Вісімнадцять днів не є репрезентативною вибіркою всіх міграцій EPSS, а великі числа оцінених CVE не компенсують 17 або 35 позитивних подій. Вибір після пілоту та кілька бюджетів додатково обмежують силу інференційних висновків.

Для практики дослід пропонує перевіряти окремо чотири речі: ідентичність байтів даних, порівнянність сукупності CVE, зміну черги та охоплення подальшого спостережуваного результату. Перші три можуть бути встановлені навіть тоді, коли позитивних подій замало для надійної оцінки якості. Перенесення такого аудиту до організації вимагатиме її реального інвентарю й контексту активів; у цій роботі їх не підмінено вигаданим портфелем.

ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

На зафіксованій панелі v4/v5 beta зміна версії супроводжується значною перебудовою пріоритетного списку: top-100 має лише 21–23 спільні CVE, за значно вищої денної стабільності в межах кожної версії. Цей результат стосується складу глобального відбору після видалення вже відомих KEV. Він не доводить поліпшення захищеності.

Аудит походження файлів виявив 12 випадків пізнішого коміту та три виправлення при незмінних заголовках дати й версії. Отже, відтворюване порівняння міграції повинно ідентифікувати конкретні байти й розділяти дату оцінки та підтверджену доступність файла. Це емпіричне доповнення до попереднього аналізу тих самих beta-даних, у якому вже вивчалися траєкторії оцінок.

Після обліку повторів CVE доказовість охоплення майбутніх KEV залишається обмеженою. Для top-5000 за 90 днів точкова різниця на користь beta становить 7,86 в. п., але розвідувальні інтервали допускають обидва знаки. Результат стійкий до перевірених способів групування в сенсі відсутності підстав оголосити перевагу; він не є доказом рівності моделей. Далі доцільно зафіксувати протокол до отримання нового періоду, архівувати оцінки разом із перевіреним часом отримання, зібрати більше різних позитивних подій і повторити оцінювання на реальному інвентарі або іншому незалежному відкритому джерелі експлуатаційних сигналів. Додатковий тест випущеної v5 на трьох липневих зрізах підтвердив виконуваність процедури та збіг двох реалізацій. Для оцінювання переваги цієї версії потрібний окремий порівняльний дизайн; тест виконання його не замінює.

Доступність даних і коду. Підготовлено локальний пакет із зафіксованими відкритими даними, маніфестами URL/версій/SHA-256, пофайловою Git-історією, кодом отримання й перевірки, таблицею наступних KEV для кожної дати, результатами та графіками. README описує порядок повторного обчислення; доступність віддалених копій не підмінює перевірку контрольних сум. Усі 18 пар і контрольну межу релізу перевірено окремою програмною реалізацією. До пакета додано протокол, три липневі зрізи, маніфест і результати тестового прогону з незалежною програмною звіркою. Комплект підготовлено як електронний додаток supplementary-materials.zip: код, зафіксовані входи та звіти дозволяють повторити обчислення за README. Журнал основного повторного прогону збережено в reports/reproduction.json, додаткового — у test-environment/results.json. Публічного репозиторію або DOI архіву немає; локальна підготовка додатка не означає його передачі редакції.

Використання штучного інтелекту. Під час підготовки статті використано OpenAI Codex (GPT-6) для пошуку й зіставлення джерел, уточнення питання, генерації коду аналізу, тексту, таблиць та англomовних метаданих. Числа й графіки отримано виконанням коду на реальних зафіксованих даних; додатковий перерахунок виконано окремою програмною реалізацією. Під час додаткової перевірки Codex використано для підготовки коду тестового прогону, звірки двох реалізацій та редактури опису отриманих результатів.



СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Jacobs, J. (2026). *New version of EPSS went live June 15th!* FIRST. <https://www.first.org/newsroom/releases/202606-036>
2. FIRST. (2026). *Get the data.* <https://www.first.org/epss/data>
3. Denysiuk, V. (2025). Predicting CVE exploitation based on open NVD and KEV data for risk-based prioritization. *Cybersecurity: Education, Science, Technique*, 2(30), 428–444. <https://doi.org/10.28925/2663-4023.2025.30.906>
4. Sherif, E., Yevseyeva, I., Basto-Fernandes, V., & Cook, A. (2026). *Bridging the gap between security metrics and key risk indicators: An empirical framework for vulnerability prioritization* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2603.12450>
5. Koscinski, V., Nelson, M., Okutan, A., Falso, R., & Mirakhorli, M. (2025). *Conflicting scores, confusing signals: An empirical study of vulnerability scoring systems* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2508.13644>
6. Iannone, E., Sellitto, G., Iaccarino, E., Ferrucci, F., De Lucia, A., & Palomba, F. (2024). Early and realistic exploitability prediction of just-disclosed software vulnerabilities: How reliable can it be? *ACM Transactions on Software Engineering and Methodology*, 33(6), 1–41. <https://doi.org/10.1145/3654443>
7. Bakare, A. (2026). *Point-in-time vulnerability data: Reproducibility materials for historical vulnerability reconstruction* (Version 1) [Data set]. Mendeley Data. <https://doi.org/10.17632/ftkrdrpckk.1>
8. Gamblin, J. (2026). *EPSS-Compare* [Computer software]. GitHub. <https://github.com/jgamblin/EPSS-Compare>
9. Hackuity. (2026). *EPSS v5 is here. TRS just got sharper.* <https://www.hackuity.ai/blog/epss-v5-is-here-trs-just-got-sharper>
10. Cybersecurity and Infrastructure Security Agency. (2026). *Known Exploited Vulnerabilities Catalog* [Data set; GitHub repository snapshot, commit 4d1aff213a881341db8a276c89802e05bf69481c]. GitHub. <https://github.com/cisagov/kev-data/tree/4d1aff213a881341db8a276c89802e05bf69481c>
11. Empirical Security. (2026). *EPSS scores* [Data set; GitHub repository snapshot, commit f9b61220a5a289812f54c3c1327b178c51f01d08]. GitHub. https://github.com/empiralsec/epss_scores/tree/f9b61220a5a289812f54c3c1327b178c51f01d08
12. Field, C. A., & Welsh, A. H. (2007). Bootstrapping clustered data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(3), 369–390. <https://doi.org/10.1111/j.1467-9868.2007.00593.x>

**Strelnikov Vitalii**

PhD in Computer engineering
Associate Professor of the Department of
Information and Cyber Security named after Professor Volodymyr Buriachok
Borys Grinchenko Kyiv Metropolitan University, Kyiv, Ukraine
ORCID: 0000-0003-3439-3220
v.strelnikov@kubg.edu.ua

Vorokhob Maksym

PhD in Cybersecurity
Associate Professor of the Department of
Information and Cyber Security named after Professor Volodymyr Buriachok
Borys Grinchenko Kyiv Metropolitan University, Kyiv, Ukraine
ORCID: 0000-0001-5160-7134
m.vorokhob@kubg.edu.ua

Brzhevska Zoreslava

PhD in Cybersecurity, Associate Professor
Associate Professor of the Department of
Information and Cyber Security named after Professor Volodymyr Buriachok
Borys Grinchenko Kyiv Metropolitan University, Kyiv, Ukraine
ORCID: 0000-0002-7029-9525
z.brzhevska@kubg.edu.ua

Kruhlyk Vladyslav

Doctor of Sciences in Pedagogy, Full Professor
Professor of the Department of Informatics and Cybernetics
Bohdan Khmelnytsky Melitopol State Pedagogical University, Zaporizhzhia, Ukraine
ORCID: 0000-0002-5196-7241
krugvs@mdpu.org.ua

EMPIRICAL AUDIT OF PRIORITIZATIONS FOR EPSS V4 AND V5 BETA

Abstract. This study examines differences between vulnerability priority lists produced by EPSS v4 and EPSS v5 beta and the limits of evaluating these differences against subsequent additions to the CISA Known Exploited Vulnerabilities catalog. The objective is to relate changes in list composition to later KEV entries while auditing the temporal availability of the score files themselves. The analysis uses all 18 matching dates available in a frozen public beta archive, from 22 May to 8 June 2026. For each date, the eligible population consists of CVEs scored by both versions and absent from KEV on that date. Lists of 100, 500, 1,000, and 5,000 CVEs are compared, together with coverage of subsequent KEV additions over 30 and 90 days. Repeated observations of each CVE are retained as a cluster during paired bootstrap resampling. Sensitivity analyses group events by KEV addition date, give equal weight to each distinct CVE, and restrict the panel according to repository commit history. The two versions share only 21–23 CVEs in their top-100 lists, although at least 98% of each version's list is retained between adjacent days. At the 90-day horizon, 420 positive CVE-date observations represent only 35 distinct CVEs. For top-5,000, pooled coverage is 7.86% for v4 and 15.71% for beta. The paired difference is 7.86 percentage points, with an exploratory 95% CVE-cluster percentile interval from –4.18 to 21.44 points. The Git audit shows that the retained bytes of 12 beta files have a commit date later than their score date. Three files were revised without changing the model-version or score-date header. Restricting the analysis to six dates whose retained files first appear in same-day commits also leaves the coverage difference highly uncertain. The study contributes a reproducible audit that links file provenance, changes in priority lists, score availability, and uncertainty in subsequent KEV coverage. Commit timestamps are only proxies for public availability, and KEV addition dates are not dates of first exploitation. A clean-environment replay reproduced the frozen outputs. An additional execution test on three previously unused July snapshots of released v5 achieved exact agreement between two implementations for eligible populations, subsequent KEV sets, top lists, and hit counts. This test establishes execution reproducibility, not a comparative advantage of released v5. The main



comparative results describe a short retrospective beta panel and should not be generalized to the released v5 model, exploitation prediction accuracy, or organizational remediation outcomes.

Keywords: EPSS; CVE; CISA KEV; vulnerability prioritization; reproducibility; temporal availability; cluster bootstrap.

REFERENCES (TRANSLATED AND TRANSLITERATED)

1. Jacobs, J. (2026). *New version of EPSS went live June 15th!* FIRST. <https://www.first.org/newsroom/releases/202606-036>
2. FIRST. (2026). *Get the data.* <https://www.first.org/epss/data>
3. Denysiuk, V. (2025). Predicting CVE exploitation based on open NVD and KEV data for risk-based prioritization. *Cybersecurity: Education, Science, Technique*, 2(30), 428–444. <https://doi.org/10.28925/2663-4023.2025.30.906>
4. Sherif, E., Yevseyeva, I., Basto-Fernandes, V., & Cook, A. (2026). *Bridging the gap between security metrics and key risk indicators: An empirical framework for vulnerability prioritization* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2603.12450>
5. Koscinski, V., Nelson, M., Okutan, A., Falso, R., & Mirakhorli, M. (2025). *Conflicting scores, confusing signals: An empirical study of vulnerability scoring systems* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2508.13644>
6. Iannone, E., Sellitto, G., Iaccarino, E., Ferrucci, F., De Lucia, A., & Palomba, F. (2024). Early and realistic exploitability prediction of just-disclosed software vulnerabilities: How reliable can it be? *ACM Transactions on Software Engineering and Methodology*, 33(6), 1–41. <https://doi.org/10.1145/3654443>
7. Bakare, A. (2026). *Point-in-time vulnerability data: Reproducibility materials for historical vulnerability reconstruction* (Version 1) [Data set]. Mendeley Data. <https://doi.org/10.17632/ftkdrdpckk.1>
8. Gamblin, J. (2026). *EPSS-Compare* [Computer software]. GitHub. <https://github.com/jgamblin/EPSS-Compare>
9. Hackuity. (2026). *EPSS v5 is here. TRS just got sharper.* <https://www.hackuity.ai/blog/epss-v5-is-here-trs-just-got-sharper>
10. Cybersecurity and Infrastructure Security Agency. (2026). *Known Exploited Vulnerabilities Catalog* [Data set; GitHub repository snapshot, commit 4d1aff213a881341db8a276c89802e05bf69481c]. GitHub. <https://github.com/cisagov/kev-data/tree/4d1aff213a881341db8a276c89802e05bf69481c>
11. Empirical Security. (2026). *EPSS scores* [Data set; GitHub repository snapshot, commit f9b61220a5a289812f54c3c1327b178c51f01d08]. GitHub. https://github.com/empiralsec/epss_scores/tree/f9b61220a5a289812f54c3c1327b178c51f01d08
12. Field, C. A., & Welsh, A. H. (2007). Bootstrapping clustered data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(3), 369–390. <https://doi.org/10.1111/j.1467-9868.2007.00593.x>

Отримано редакцією журналу / Received: 12.08.26

Прорецензовано / Revised: 24.08.26

Схвалено до друку / Accepted: 26.10.26



This work is licensed under Creative Commons Attribution-noncommercial-sharealike 4.0 International License.