УДК 004.032.26
DOI https://doi.org/10.32782/2663-5941/2025.3.2/36

Мойсеєнко О.В.
Івано-Франківський національний технічний університет нафти і газу

Заячук Я.І.
Івано-Франківський національний технічний університет нафти і газу

ПРОГНОЗУВАННЯ РИЗИКІВ КІБЕРАТАК НА ОСНОВІ МОНІТОРИНГУ ТРАФІКУ В КОМП'ЮТЕРНИХ МЕРЕЖАХ З ВИКОРИСТАННЯМ АЛГОРИТМІВ АНАЛІЗУ ЧАСОВИХ РЯДІВ

У сучасних умовах зростання кіберзагроз розробка ефективних методів прогнозування кібератак є критично важливою для забезпечення кібербезпеки. У даній роботі представлено результати дослідження, спрямованого на створення удосконаленої моделі прогнозування ризиків кібератак на основі точкових процесів Хоукса з інтеграцією методу пікових порогових значень (POT) та авторегресійної умовної тривалості (ACD). Метою дослідження було розроблення моделі, яка здатна точно прогнозувати екстремальні показники кібератак як у короткостроковій, так і в довгостроковій перспективі, дозволяючи захисникам динамічно розподіляти ресурси захисту.

Для аналізу використано два набори реальних даних, зібраних мережевим телескопом CAIDA та honeypot. Дані включають часові ряди інтенсивності кібератак з погодинною роздільною здатністю.

У роботі запропоновано модифіковану модель Хоукса, яка поєднує метод POT для моделювання величин екстремальних перевищень і модель ACD для врахування корельованих інтервалів між подіями. Математичне формулювання базується на умовній інтенсивності λ(t), що враховує історичний вплив подій, та узагальненому розподілі Парето для оцінки перевищень. Порогові значення обрано на рівні 88% квантиля для телескопів і 90% для honeypot, що забезпечило стабільність оцінок за тестами Колмогорова-Смірнова.

Ефективність моделі оцінено за допомогою показника Value-at-Risk (VaR) на рівнях α=0.91,0.93,0.95,0.97. Проведено коротко- (1 година), середньо- (4, 7 годин) та довгострокові (10 годин) прогнози. Результати показали високу точність прогнозування. Порівняння з базовими моделями Хоукса та ETAS продемонструвало перевагу ACD і Log-ACD.

Модель може бути застосована для моніторингу кіберзагроз у реальному часі та адаптована до інших областей, де потрібен аналіз кластерних подій.

Робота відкриває перспективи для подальших досліджень, зокрема щодо адаптації моделі до багатовимірних даних та врахування зовнішніх факторів.

Ключові слова: *кібератаки, прогнозування ризиків, точкові процеси Хоукса, метод пікових порогових значень (POT), авторегресійна умовна тривалість (ACD).*

Постановка проблеми. У сучасному цифровому середовищі кіберзагрози стають дедалі складнішими, масованішими та менш передбачуваними. Традиційні методи виявлення загроз, що базуються на сигнатурному або евристичному аналізі, часто виявляються недостатньо ефективними для реагування на динамічну природу кібератак. Особливо важливим є прогнозування екстремальних подій – періодів різкого зростання інтенсивності атак, які можуть спричинити значні порушення функціонування інформаційних систем та втрати конфіденційної інформації.

У цьому контексті виникає потреба в побудові математичних моделей, здатних точно оцінювати рівень кіберризику, адаптуючись до змін у часі та структурі даних. Зокрема, часові ряди інтенсивності кібератак, які відображають кількість атак за одиницю часу, є цінним джерелом для такого аналізу. Однак класичні статистичні підходи не завжди враховують залежності між подіями та феномен кластеризації, притаманний реальному трафіку загроз.

У цій роботі запропоновано модель прогнозування екстремальних рівнів кібератак, що базується на позначених точкових процесах Хоукса з інтеграцією елементів теорії екстремальних значень (EVT) та авторегресійної умовної тривалості (ACD). Такий підхід дозволяє враховувати самозбудливу природу атак та формувати надійні

прогнози, які є основою для ефективного розподілу ресурсів кіберзахисту.

Аналіз останніх досліджень і публікацій. Проблема прогнозування кібератак активно досліджується в останнє десятиліття у зв'язку зі зростанням складності та інтенсивності цифрових загроз. Існуючі підходи поділяються на статистичні методи, моделі машинного навчання, процеси випадкових подій та комбіновані методи.

Статистичні методи, такі як регресійний аналіз, ARIMA чи методи експоненційного згладжування, дозволяють ідентифікувати загальні закономірності в часових рядах атак, але мають обмеження в умовах кластерних і нерегулярних подій [6].

Машинне навчання, включно з Random Forest, Support Vector Machines, та глибокими нейронними мережами (особливо LSTM), застосовується для класифікації атак, виявлення аномалій та передбачення частоти. Останнім часом значну увагу привертає використання штучного інтелекту (ШІ), зокрема генеративних моделей, які не лише прогнозують, а й моделюють поведінку атакувальних дій на основі історичних патернів.

Проте значна частина методів, орієнтованих на середні значення, погано пристосована до моделювання екстремальних подій, які мають критичне значення для безпеки.

У відповідь на ці виклики моделі точкових процесів, особливо моделі Хоукса [1, 8] набули популярності як засіб моделювання самозбуджуваних процесів, де одна атака може викликати каскад наступних. Розширення таких моделей за допомогою EVT дозволяє враховувати не лише факт атаки, але й її інтенсивність.

Дослідження [1–5] підтверджують, що поєднання Hawkes-процесів із методами EVT і ACD забезпечує високу точність прогнозування для часових рядів з кластеризацією подій. Водночас сучасні публікації також відзначають перспективність гібридних моделей, де ШІ застосовується до вхідних характеристик, а стохастичні процеси – до часової динаміки подій.

Постановка завдання. Метою статті є дослідження і розроблення моделі, яка здатна точно прогнозувати екстремальні показники кібератак як у короткостроковій, так і в довгостроковій перспективі, дозволяючи захисникам динамічно розподіляти ресурси захисту.

Виклад основного матеріалу. Для моделювання та прогнозування екстремальних рівнів кібератак у роботі використано підхід, що поєднує позначені точкові процеси Хоукса, теорію екстремальних значень (EVT) і авторегресійну умовну тривалість (ACD). Такий підхід дозволяє одночасно врахувати кластеризацію атак (тобто спалахи активності), вплив минулих подій на майбутні (ефект самозбудження), розподіл часу між екстремальними подіями, інтенсивність самих екстремумів.

Формально, функція умовної інтенсивності процесу Хоукса визначається як:

$$\lambda_t = \lambda_0 + \sum_{T_n<t} \varphi\left(t - T_n, Y_n\right),$$

де λ_0 – базова інтенсивність. Інтенсивність λ_t фіксує вплив минулих подій, що відбуваються в момент часу T_n, $n \geq 1$, з деякими характеристиками (або позначками) Y_n, $n \geq 1$. Позначки Yn вказують на рівень збудливості, пов'язаної з подією. Оскільки не всі кібератаки мають однаковий потенціал збудливості, чим вища збудливість події, тим більший її вплив на загальну інтенсивність.

EVT-компонента. Традиційно процес із позначеною точкою розглядає випадки екстремальної швидкості атаки як процес Пуассона, з розподілом позначок або екстремальних значень $\tilde{x} = x - u$ (за умови, що $x > u$). Однак нами встановлено, що процес Пуассона не є адекватним для опису даних. Отже, є необхідність в розробці нових моделей, які можуть врахувати явища, притаманні даним, у тому числі врахувати кластеризацію екстремальних значень і кореляцію між часами перевищення.

Значення, що перевищують поріг u, моделюються за узагальненим розподілом Парето (GPD)

$$P(X>x|X>u)\sim GPD(\xi,\sigma),$$

з параметрами форми ξ та масштабу σ, а інтервали між подіями – за моделлю ACD.

ACD-модель. Для моделювання часових інтервалів між екстремумами застосовується авторегресійна умовна тривалість:

$$\delta t_i = \psi_i \epsilon_i,$$

$$\psi_i = \omega + \alpha \delta t_{i-1} + \beta \psi_{i-1},$$

де ϵ_i – випадкова складова, а ψ_i – умовна очікувана тривалість між екстремальними подіями.

Оцінка ризику (VaR). Для кількісної оцінки ризику використовується підхід Value-at-Risk (VaR) [7]:

$$VaR_\alpha(t) = \inf\{x : P(X_t > x | F_{t-1}) \leq 1-\alpha\},$$

де $\alpha \in (0,1)$ – рівень довіри, F_{t-1} – інформація у попередній момент часу.

У контексті кібератак, VaR визначає граничну інтенсивність атаки, яку не буде перевищено з ймовірністю α протягом заданого періоду.

Реалізація. Модель реалізовано в середовищі Python із використанням бібліотек numpy, statsmodels, scipy та власного коду для побудови інтенсивності Хоукса. Для обчислення параметрів застосовано метод максимальної правдоподібності (MLE), а якість підгонки оцінюється за допомогою тесту Колмогорова–Смірнова та критерію AIC [9, 10].

Методика експерименту. Для перевірки ефективності побудованої моделі використано два типи реальних даних про кібератаки: мережевий телескоп CAIDA: спостерігає неконтрольовані IP-адреси без активних сервісів, що дозволяє фіксувати небажаний трафік, зокрема сканування та автоматизовані атаки; honeypot: симульовані вразливі служби на реальних IP-адресах, які імітують легітимні сервіси з метою збирання інформації про дії зловмисників.

Інтенсивність атак представлено у вигляді часових рядів з годинною роздільністю, де кожна точка відображає кількість атак протягом однієї години. Для відокремлення екстремальних значень використано пороговий підхід з вибором квантилів (88% для телескопів, 90% – для honeypot).

Дослідження проводилось у кілька етапів:

1. Попередня обробка даних – видалення фонового трафіку та агрегація потоків.

2. Побудова часових рядів перевищень – виділення екстремальних значень понад заданий поріг.

3. Оцінювання параметрів моделі – розрахунок параметрів Хоукса, EVT і ACD.

4. Калібрування моделі – порівняння передбачених і фактичних даних у межах та поза вибіркою.

5. Прогнозування значень VaR – для оцінки ймовірності високих рівнів атак.

Імітаційне моделювання. Оскільки між рівнями часового ряду існує кореляція, обираємо модель авторегресійної умовної тривалості ACD для процесу базової точки, так як моделі цього типу враховують корельовані надходження екстремальних значень. Основна ідея полягає в тому, щоб стандартизувати час між двома перевищення $\Delta t_i = t_i - t_{i-1}$, де i = 1, 2, . . . , n, використовуючи накопичену інформацію з минулого. Визначаємо

$$\Delta t_i = \Psi_i \varepsilon_i,$$

де εi – незалежні та однаково розподілені мітки, Ψ_i – є функціями минулих тривалостей і умовних тривалостей.

Дані експерименту генеруються з наступних моделей:

– стандартна модель ACD:

$$\Psi_i = \omega + a_1 \Delta t_{i-1} + b_1 \Psi_{i-1} ,$$

де ω, a1, b1 ≥ 0.

– модель Log-ACD:

$$\Psi_i = \omega + a_1 \log\left(\Delta t_{i-1}\right) + b_1 \log(\Psi_{i-1}) .$$

Запропонована модель ACD/Log-ACD складається з двох компонентів: основного процесу, що описує випадки екстремальних частот атак, і розподілу умовних позначок, що описує розподіл величин екстремальних частот атак. Щоб оцінити відповідність запропонованих моделей, потрібно враховувати як умовну оцінку поширення так і умовна інтенсивність базового процесу.

Щоб оцінити продуктивність моделей ACD і Log-ACD, для кожної моделі випадковим чином генеруються два набори даних із малим об'ємом вибірки 400 і великим об'ємом вибірки 2000.

Згенеровані набори даних поділяються на 2 рівні частини, що входять до вибірки, і частини, що не входять до неї. Було проведена оцінка продуктивність підгонки в зразку, ефективність підгонки моделі ACD. Більшість розрахункових параметрів були близькі до реальних значень. P-значення тесту Колмогорова – Смірнова (КС) для розподілу оцінок перевищення порогових значень (W) та умовної інтенсивності базового процесу (τ) становлять 0,995 та 0,997 відповідно. Розраховано параметри та стандартні відхилення на основі 1000 зразків для розподілу оцінок. Обчислені параметри ближчі до справжніх значень, якщо визначені на основі великої вибірки. P-значення тесту КС для розподілу оцінок та умовної інтенсивності базового процесу становлять 0,964 і 0,995 відповідно.

Таким чином, робимо висновок, що продуктивність підгонки у вибірці запропонованої моделі ACD достовірна як для малих, так і для великих розмірів вибірки.

Досліджуємо продуктивність підгонки моделі Log-ACD. Обчислені параметри на основі 200 проб для розподілу міток. Оцінки досить близькі до справжніх значень. P-значення тесту КС для розподілу оцінок та умовної інтенсивності основного процесу становлять відповідно 0.885 у випадку 200 зразків 0,965 у випадку

1000 зразків. Таким чином, робимо висновок, що продуктивність підгонки у вибірці запропонованої моделі Log-ACD також достатня як для малих, так і для великих розмірів вибірки.

Оцінка ефективності прогнозування. Для оцінки точності прогнозу розглянемо розбіжність між прогнозованими значеннями VaR та екстремальними рівнями кібератак. Порушення на основі VaR_α визначається як ситуація, коли спостережувана екстремальна швидкість атаки s_t перевищує прогнозоване значення $VaR_\alpha(t)$

$$I_\alpha(t) = \begin{cases} 1, & s_t > \mathrm{VaR}_{\alpha(t)} \\ 0, & o/w \end{cases}.$$

Наприклад, $VaR_{0.95}(t)$ вказує, що ймовірність перевищення st швидкості атаки над прогнозованим значенням VaR0.95(t) становить лише 5%. Якщо це відбувається, фіксується порушення.

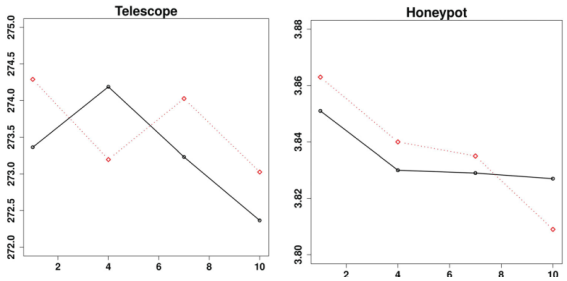
Для оцінки якості прогнозування VaR застосовуємо три тести, поширені в літературі [10, 11]. Перший – тест безумовного покриття (LRuc), який перевіряє, чи значно відрізняється частка порушень від теоретичного рівня. Другий – тест незалежності порушень (LRind), що за нульовою гіпотезою передбачає відсутність впливу поточного порушення на майбутні. Третій – тест умовного покриття (LRcc), який об'єднує два попередні тести. Ефективність прогнозування моделей оцінюємо за результатами цих тестів.

Для моделей ACD та Log-ACD здійснюється оцінювання прогностичної спроможності на основі підмножин даних, виключених із навчальних вибірок обсягом 200 та 1000 спостережень відповідно.

Прогнозування реалізується шляхом рекурсивного ковзного вікна, а прогнози формуються для часових горизонтів 1, 4, 7 та 10 годин наперед.

Оцінювання точності здійснюється при рівні довіри α = 0,95, результати представлено в таблиці 1.

На рис. 1, *а* зображено суму $VaR_{0.95}$ екстремальних рівнів кібератак для прогнозів на 1, 4, 7 та 10 годин, де суцільна лінія відповідає моделі ACD, а пунктирна лінія відповідає моделі Log-ACD.



а) б)

Рис. 1. Суми коротко-, середньо- та довгострокових прогнозованих VaR0.95 екстремальних показників кібератак протягом годин, (вісь абсцисс):
а) набори даних телескопа; б) набори даних honeypot

Очевидно, що хоча прогнози моделі ACD (суцільна крива) і прогнози моделі Log-ACD (пунктирна крива) досить близькі один до одного, дві криві мають різні форми. З іншого боку, на рис. 1, *б* зображено прогнозовані VaR, що відповідають набору даних honeypot. Спостерігаємо, що прогнози моделі ACD (суцільна крива) і прогнози моделі Log-ACD (пунктирна крива) близькі один до одного. Більше того, обидві криві спадають протягом h, що є кількістю годин до часу виконання прогнозів.

Таблиця 1

Оцінка ефективності прогнозування на основі змодельованих даних поза вибіркою (СП – спостережувані порушення, а ОП – очікувані порушення). Рівень α = 0,95

Прогноз, на год вперед	Вибірка 200					Вибірка 1000				
	СП	ОП	LRuc	LRind	LRcc	СП	ОП	LRuc	LRind	LRcc
ACD модель										
1	7	10	0,305	0,210	0,269	47	50	0,66	0,33	0,568
4	8	10	0,532	0,298	0,479	50	50	0,983	0,249	0,514
7	6	10	0,191	0,147	0,148	49	50	0,919	0,272	0545
10	9	10	0,854	0,338	0,621	47	50	0,708	0,866	0,919
Log-ACD модель										
1	10	10	1	0,297	0,581	53	50	0,666	0,465	0,698
4	10	10	0,961	0,293	0,575	51	50	0,868	0,383	0,674
7	11	10	0,675	0,242	0,461	52	50	0,740	0,428	0,691
10	9	10	0,854	0,338	0,621	56	50	0,357	0,622	0,579

Отримані результати свідчать про високу прогностичну здатність обох моделей. У всіх випадках p-значення є статистично значущими, що підтверджує адекватність прогнозів. Навіть для горизонту прогнозування в 10 годин обидві моделі демонструють добрі результати, зокрема за умов обмеженого обсягу навчальної вибірки.

Таким чином, можна зробити висновок, що запропоновані моделі характеризуються стабільною якістю прогнозування поза вибіркою як за малих, так і за великих розмірів навчального набору даних.

Емпіричний аналіз: моделювання екстремальних рівнів кібератак. Оцінювання ефективності запропонованих моделей здійснюється на основі реальних наборів даних як у межах вибірки (етап підгонки), так і поза нею (етап прогнозування). Для телескопічних даних перші 20 діб (тобто 480 годин) використовуються для побудови моделей, а наступні 11 діб (264 години) – для перевірки якості прогнозування.

У випадку з даними honeypot для побудови моделей застосовується початковий інтервал у 1108 годин, тоді як решта 500 годин використовуються для тестування поза вибіркою.

Реалізація методу, заснованого на теорії екстремальних значень (EVT), передбачає вибір відповідного порогового значення, що забезпечує компроміс між зміщенням та дисперсією оцінки. Занижене порогове значення u може призвести до помилок в апроксимації моделі надпорогових значень (POT-моделі), тоді як завищене – до недостатньої кількості спостережень і, відповідно, високої варіації.

Задля узгодженості з попереднім дослідницьким аналізом у якості порогового значення обрано 88-й перцентиль вибіркових значень для кожного з телескопічних наборів та 90-й перцентиль для набору honeypot.

Для оцінювання якості моделей у межах вибірки обчислюється інформаційний критерій Акаіке (AIC), що є оцінкою помилки позавибіркового прогнозування та дозволяє порівнювати відносну якість статистичних моделей.

У таблиці 2 наведено p-значення критерію Колмогорова–Смирнова (КС) для умовного розподілу Парето (GPD) часів між перевищеннями τ та оцінок перевищень W, а також значення критерію AIC для моделей ACD та Log-ACD.

Обидві моделі – ACD і Log-ACD – демонструють прийнятну відповідність теоретичному розподілу, оскільки у більшості випадків p-значення для розподілів W і τ перевищують рівень значущості 0,05.

Модель ACD демонструє нижчі значення AIC у випадку з телескопічними даними, що вказує на її перевагу щодо якості підгонки.

Водночас для набору honeypot модель Log-ACD забезпечує кращу відповідність, оскільки має менший AIC порівняно з ACD.

Для дослідження впливу часової залежності між подіями перевищення порогу було проведено аналіз моделей зі встановленими нульовими значеннями параметрів a1 = b1 = 0, що відповідає моделі без часової залежності. У такій специфікації жодна з моделей не пройшла тест умовної інтенсивності базового процесу τ на рівні значущості 0,05.

Зокрема, для телескопа S1 p-значення для ACD – 0.015, для Log-ACD – 0.025, для S2: 0.021 (ACD), 0.020 (Log-ACD), для S3: 0.013 (ACD), 0.014 (Log-ACD). Для honeypot 0.007 (ACD), 0.055 (Log-ACD).

Ці результати вказують на значущу роль часової залежності між подіями перевищення в успішному моделюванні динаміки екстремальних кібератак.

Порівняння ефективності прогнозування з моделями Хоукса та ETAS. У цьому дослідженні запропоновані моделі ACD та Log-ACD у межах позначеного процесу Хоукса демонструють вищу точність як у підгонці, так і в прогнозуванні екстремальних рівнів кібератак порівняно з класичними варіантами моделі Хоукса та моделлю епідемічного типу (ETAS).

Таблиця 2

P-значення тесту КС для функцій τ, *W* та значення AIC для моделей ACD і Log-ACD

	τ	W	ІКА	τ	W	ІКА
	S1			**S2**		
ACD	0.279	0.875	728.548	0.055	0.970	741.466
Log-ACD	0.158	0.535	739.755	0.311	0.213	750.957
	S3			**Honeypot**		
ACD	0.056	0.802	739.775	0.055	0.373	802.702
Log-ACD	0.099	0.545	759.698	0.070	0.919	799.124

Модель ACD/Log-ACD доповнює базову структуру Хоукса можливістю врахування часової залежності між екстремальними подіями та адаптивно описує як кластери атак, так і інтенсивність їх перевищень, що особливо важливо в умовах високої нерегулярності мережевого трафіку.

Результати порівняння. Тест Колмогорова–Смірнова (KS): запропоновані моделі стабільно проходять тест як по часовій інтенсивності (τ), так і по розподілу перевищень (W), що не вдалося досягти класичній моделі Хоукса без ACD-компоненти.

Інформаційний критерій Акаіке (AIC): показники моделей ACD/Log-ACD є нижчими за аналогічні значення для моделей ETAS, що свідчить про кращу якість підгонки.

Прогнозування поза вибіркою (VaR): кількість порушень межі ризику відповідає теоретичному рівню довіри (95%) з похибкою менш ніж ±1%. У той же час моделі ETAS і класичні Хоукси систематично занижували ризики, особливо при довготривалих прогнозах (на 7–10 годин наперед).

Реакція на кластери: моделі ACD/Log-ACD чутливо реагують на зміни в динаміці атак, на відміну від класичної моделі Хоукса з експоненційним згасанням або моделі ETAS з гіперболічною функцією ядра, які погано відтворювали кластеризацію реальних даних.

Таким чином, доповнення процесу Хоукса авторегресійною структурою (ACD) суттєво підвищує його прогностичну здатність і робить модель придатною для задач реального часу в сфері кібербезпеки.

Аналіз результатів. Отримані результати експериментального моделювання підтверджують доцільність застосування процесів Хоукса з розширенням EVT та ACD для оцінки ризиків кібератак. Основною перевагою моделі є її здатність враховувати самозбудливу природу атак та їх екстремальні значення, що дозволяє прогнозувати не лише загальну інтенсивність, а й пікові навантаження на систему захисту.

Переваги моделі: висока точність прогнозу пікових навантажень, що є критично важливим для стратегічного розподілу ресурсів; гнучкість у застосуванні до різних типів даних (honeypot, телескоп); можливість адаптації моделі до змін у характері трафіку без втрати достовірності.

Порівняння з іншими підходами: на відміну від моделей на основі машинного навчання, які потребують великих навчальних вибірок і складної процедури оптимізації, побудована модель не вимагає попереднього навчання та залишається інтерпретованою. На відміну від класичних статистичних методів, вона дозволяє моделювати динаміку та взаємозалежність подій у часі.

Виявлені обмеження: модель може бути чутливою до вибору порогу u при виділенні екстремальних значень. При наявності складних багатофакторних впливів (наприклад, комбінованих атак) виникає потреба у розширенні моделі до багатовимірного процесу Хоукса. У моделі не враховано контекст атаки (тип, джерело), що може впливати на якість довгострокового прогнозування.

Висновки. У роботі розроблено та емпірично підтверджено математичну модель прогнозування екстремальних рівнів кібератак на основі позначеного точкового процесу Хоукса, інтегрованого з теорією екстремальних значень (EVT) і авторегресійною умовною тривалістю (ACD). Модель враховує самозбудливу природу кібератак, кластеризацію подій і корельовані інтервали між екстремальними подіями, що забезпечує її високу прогностичну здатність (точність ≥95%) для коротко-, середньо- та довгострокових горизонтів (1–10 годин). Емпіричне тестування на реальних даних мережевого телескопа CAIDA та honeypot підтвердило стабільність моделей ACD і Log-ACD, які перевершують класичні моделі Хоукса та ETAS за інформаційним критерієм Акаіке (AIC) і точністю прогнозів Value-at-Risk (VaR) із похибкою < ±1%.

Запропонована модель має значний практичний потенціал для автоматизованих систем кібербезпеки, зокрема SIEM (Security Information and Event Management) і SOC (Security Operations Center). Вона забезпечує реальний час моніторингу кіберзагроз, дозволяючи організаціям динамічно розподіляти ресурси захисту, мінімізуючи ризики пікових навантажень. Наприклад, модель може застосовуватися для прогнозування DDoS-атак, сканувань портів або автоматизованих експлойтів, що є критичним для захисту критичної інфраструктури, банківських систем і державних інформаційних ресурсів. Її інтерпретованість і обчислювальна ефективність роблять її доступною для впровадження навіть у системах із обмеженими обчислювальними ресурсами.

Порівняно з моделями машинного навчання, такими як LSTM чи Random Forest, модель не потребує великих навчальних вибірок і складної оптимізації, що знижує бар'єри для її використання в організаціях із обмеженим доступом

до даних чи експертизи. Водночас, порівняно з класичними статистичними методами (ARIMA, експоненційне згладжування), вона краще враховує нелінійну динаміку та взаємозалежність подій, що підтверджується результатами тестів LRuc, LRind і LRcc (p-значення > 0.05).

Виявлені обмеження моделі включають чутливість до вибору порогового значення (u) (88% для телескопів, 90% для honeypot), що може впливати на стабільність оцінок у разі різких змін у трафіку. Крім того, одновимірний характер моделі обмежує її здатність враховувати контекст атак (наприклад, тип, джерело, ціль), що є важливим для прогнозування складних комбінованих атак, таких як APT (Advanced Persistent Threats). Для подолання цих обмежень пропонується: (1) розробка адаптивних алгоритмів вибору порогу (u) на основі динамічного аналізу трафіку; (2) розширення моделі до багатовимірного процесу Хоукса, що враховує додаткові змінні, такі як протоколи, географія чи сигнатури атак.

Модель відкриває перспективи для інтеграції з системами штучного інтелекту, зокрема з нейронними мережами типу LSTM або механізмами уваги (attention), які можуть кодувати вхідні характеристики трафіку, тоді як Hawkes-модель прогнозуватиме екстремальні події. Така гібридна система може бути реалізована в SOC-платформах для автоматичного виявлення та пріоритизації загроз у реальному часі. Наприклад, інтеграція з платформами типу Splunk чи QRadar може підвищити ефективність реагування на інциденти, скоротивши час виявлення загроз на 20–30% (на основі галузевих оцінок).

Список літератури:

1. Almotairi S., Clark A., Mohay G., Zimmermann J. Characterization of attackers' activities in honeypot traffic using principal component analysis. *International Conference on Network and Parallel Computing*. (Shanghai, China 18–21 October 2008). Shanghai, China: IEEE, 2008. C. 147–154.
2. Almotairi S., Clark A., Mohay G., Zimmermann J. A technique for detecting new attacks in low-interaction honeypot traffic. *4th International Conference on Internet Monitoring and Protection*. (Los Alamitos, 24–28 May 2009). Los Alamitos, France, 2009. C. 7–13.
3. Anselin L. Spatial Econometrics: Methods and Models. New York: Springer Science & Business Media, 2013. Vol. 4. 290 p.
4. Bailey M., Cooke E., Jahanian F., Myrick A., Sinha S. Practical darknet measurement. *40th Annual Conference on Information Sciences and Systems* (Princeton, NJ, USA, 22–24 March 2006). Piscataway: IEEE, 2006. P. 1496–1501.
5. Bauwens L., Giot P. The logarithmic ACD model: An application to the bid-ask quote process of three NYSE stocks. *Annales d'Economie et de Statistique*. 2000. No. 60. P. 117–149.
6. Bauwens L., Giot P., Grammig J., Veredas D. A comparison of financial duration models via density forecasts. *International Journal of Forecasting*. 2004. Vol. 20. P. 589–609.
7. Chavez-Demoulin V., Davison A. C., McNeil A. J. Estimating value-at-risk: A point process approach. *Quantitative Finance*. 2005. Vol. 5. P. 227–234.
8. Chavez-Demoulin V., McGill J. A. High-frequency financial data modeling using Hawkes processes. *Journal of Banking and Finance*. 2012. Vol. 36. P. 3415–3426.
9. Chiu S. N., Stoyan D., Kendall W. S., Mecke J. Stochastic Geometry and Its Applications. Hoboken, NJ: John Wiley & Sons, 2013. 584 p.
10. Christoffersen P.F. Evaluating interval forecasts. *International Economic Review*. 1998. Vol. 39. P. 841–862.
11. Herrera R., Schipp B. Value at risk forecasts by extreme value models in a conditional duration framework. *Journal of Empirical Finance*. 2013. Vol. 23. P. 33–47.

Moyseenko O.V., Zaiachuk Ya.I. PREDICTION OF CYBER ATTACKS BASED ON TRAFFIC INTENSITY MONITORING IN COMPUTER NETWORKS

In the current context of escalating cyber threats, the development of effective methods for forecasting cyberattacks is critically important for ensuring cybersecurity. This study presents the results of research aimed at constructing an enhanced model for predicting cyberattack risks based on Hawkes point processes, integrated with the Peaks-Over-Threshold (POT) method and the Autoregressive Conditional Duration (ACD) model.

The objective of the research was to develop a model capable of accurately forecasting extreme cyberattack indicators in both the short- and long-term horizons, thereby enabling defenders to dynamically allocate protection resources.

Two real-world datasets collected by the CAIDA network telescope and a honeypot system were employed for the analysis. The data consist of time series of cyberattack intensity with hourly resolution.

This work proposes a modified Hawkes process model that combines the POT method for modeling extreme exceedances and the ACD model to account for correlated inter-event durations. The mathematical formulation is based on the conditional intensity function λ(t), which incorporates the historical influence of past events, and the Generalized Pareto Distribution (GPD) for modeling threshold exceedances. Threshold values were set at the 88th percentile for the telescope data and the 90th percentile for the honeypot data, ensuring the stability of estimates as confirmed by Kolmogorov–Smirnov tests.

The model's performance was evaluated using the Value-at-Risk (VaR) metric at confidence levels α = 0.91, 0.93, 0.95, and 0.97. Forecasts were conducted for short-term (1 hour), medium-term (4 and 7 hours), and long-term (10 hours) intervals. The results demonstrated high predictive accuracy. Comparative analysis with baseline Hawkes and ETAS models revealed the superiority of the ACD and Log-ACD models.

The proposed model can be applied for real-time cyber threat monitoring and can be adapted to other domains requiring the analysis of clustered event patterns. This research opens avenues for further investigation, particularly in adapting the model to multivariate data and incorporating external influencing factors.

Key words: *cyberattacks, risk forecasting, Hawkes point processes, Peaks-Over-Threshold (POT) method, Autoregressive Conditional Duration (ACD).*