

**Міскевич Оксана Іванівна**

старший викладач кафедри кібербезпеки
Луцький національний технічний університет, Луцьк, Україна
ORCID: 0000-0002-5009-2391
miskevich87@gmail.com

Сурова Наталія Миколаївна

к.т.н., доцент кафедри кібербезпеки
Луцький національний технічний університет, Луцьк, Україна
ORCID: 0000-0002-7116-0197
n.surova@lntu.edu.ua

Костючко Сергій Миколайович

к.т.н., доцент, завідувач кафедри кібербезпеки
Луцький національний технічний університет, Луцьк, Україна
ORCID: 0000-0002-1262-6268
s.kostiuchko@gmail.com

**СТОХАСТИЧНЕ МОДЕЛЮВАННЯ ТА ОПТИМІЗАЦІЯ ЗАХИСТУ ВІД БАГАТОХОДОВИХ
ФІШИНГОВИХ КАМПАНІЙ**

Анотація. Фішинг залишається одним із провідних векторів первинного доступу до інформаційних систем, а результативність таких атак значною мірою визначається поведінкою користувача. Традиційні бінарні показники на кшталт «клік/не клік» не відображають багатоходової траєкторії, у межах якої користувач переходить між станами доставки повідомлення, залучення, підозри, взаємодії, компрометації або безпечного виходу. Метою дослідження є розроблення та чисельна верифікація процесної ймовірнісної моделі багатоходового фішингового сценарію, що дає змогу оцінювати ймовірність компрометації, очікувану кількість кроків до завершення сценарію та чутливість результату до захисних втручань. Модель побудовано на основі поглинального марківського ланцюга зі скінченним простором станів; сформовано матрицю переходів, визначено транзитні та поглинальні стани, фундаментальну матрицю, матрицю ймовірностей поглинання та вектор очікуваної кількості кроків. Для ілюстративного credential-phishing сценарію аналітично отримано ймовірність компрометації $p = 0,1290625$ та очікувану кількість кроків $t(S_0) = 1,7255208$. Внутрішню чисельну коректність реалізації перевірено методом Монте-Карло на $2 \cdot 10^6$ траєкторій із побудовою 95 % довірчих інтервалів; результати узгоджуються з аналітичним розв'язком. При цьому Монте-Карло використано лише як засіб чисельної верифікації, а не як емпіричну валідацію моделі. Для аналізу захисних заходів розглянуто стохастично допустимі зміни переходів, що відповідають удосконаленню поштової автентифікації, навчанню користувачів і застосуванню фішинг-стійкої багатофакторної автентифікації. Бюджетний вибір заходів виконано жадібною процедурою за критерієм зниження ризику на одиницю вартості та зіставлено з повним перебором. Показано, що жадібне рішення не завжди збігається з глобально оптимальним. Наукова новизна полягає у процесному поданні поведінкової траєкторії користувача як марківського ланцюга та формалізації зв'язку між захисними втручаннями, зміною перехідних ймовірностей і кінцевим ризиком компрометації.

Ключові слова: фішинг; соціальна інженерія; ланцюги Маркова; поглинальний марківський ланцюг; матриця переходів; ймовірність компрометації; багатоходовою атака; аналіз чутливості; оцінювання кіберризiku.

ВСТУП

Фішинг та пов'язані з ним методи соціальної інженерії залишаються одним із провідних векторів первинного доступу до інформаційних систем, причому після подолання периметрових засобів захисту результат атаки значною мірою визначається поведінкою користувача [1], [2]. Поширення генеративних технологій спрощує створення персоналізованого та переконливого контенту, підвищуючи



масштабованість фішингових кампаній [3]. У зв'язку з цим актуальним є кількісне оцінювання процесу переходу користувача від отримання повідомлення до небезпечної дії та визначення ефективності захисних заходів.

Сучасний фішинг доцільно розглядати як багатоходовий процес, у межах якого користувач послідовно проходить стани доставки, залучення, підозри та взаємодії. Особливе значення має перехід зі стану підозри до компрометації або безпечного завершення взаємодії [8], [9]. Водночас більшість поведінкових досліджень зосереджена на бінарних показниках, зокрема факті кліку, і не відображає динаміки проміжних станів [2], [7]. Непередбачуваність індивідуальної реакції користувачів додатково обґрунтовує застосування стохастичного підходу [6], тоді як агреговані результати тренувальних кампаній самі по собі не формують процесної моделі переходів [10].

Марківські, приховані марківські та поглинальні моделі широко застосовуються для кількісного оцінювання кіберризиків та визначення ймовірності досягнення критичних станів [4], [5], [11]-[13]. Проте переважно вони описують мережеве поширення атак і технічні вразливості, а не поведінкову траєкторію користувача. Отже, існує методологічний розрив між поведінковими дослідженнями фішингу та стохастичним моделюванням атак. Його подолання можливе шляхом представлення послідовності «підозра → компрометація» як поглинального марківського ланцюга, що дає змогу оцінити ймовірність компрометації, очікувану кількість кроків до кінцевого стану та чутливість окремих переходів до захисних втручань [14], [15]. Саме розробленню та аналізу такої моделі присвячено подальше дослідження.

Постановка проблеми. Наукова проблема полягає у недостатній опрацьованості процесних імовірнісних моделей, які кількісно пов'язують захисні заходи з імовірністю проходження користувачем шляху від підозри до компрометації у багатоходовому фішинговому сценарії. Традиційне трактування фішингу як бінарної події «клік/не клік» не враховує проміжні стани та переходи між ними, а отже не дає змоги оцінити, на якому етапі захисне втручання є найбільш ефективним.

Поведінкові дослідження переважно виявляють чинники сприйнятливості, але не формують процесної моделі переходів [2], [7], [10]. Стохастичні підходи, зокрема марківські та приховані марківські моделі, здебільшого застосовуються до технічних вразливостей і мережевого поширення атак [4], [5], тоді як засоби автоматизованого виявлення не охоплюють подальшу поведінку користувача [16]. Тому актуальним є побудування поглинального марківського ланцюга з виокремленими станами «підозра» і «компрометація», який дає змогу оцінювати ризик, очікувану кількість кроків та чутливість окремих переходів до захисних заходів.

Аналіз останніх досліджень і публікацій. Сучасні дослідження охоплюють три основні напрями. Перший пов'язаний зі сприйнятливістю користувачів до фішингу. Роботи [2], [6]-[10] виявляють поведінкові, ситуативні та психологічні чинники, що впливають на реакцію користувачів, однак переважно не описують динаміку переходів між станами. Зокрема, дослідження [6] засвідчує непередбачуваність індивідуальної реакції, що додатково обґрунтовує стохастичне трактування.

Другий напрям охоплює стохастичне моделювання атак. Поглинальні марківські ланцюги [4], [11], [12], приховані марківські моделі [5] та баєсівські графи атак [13] дають змогу оцінювати ймовірність досягнення критичних станів, однак переважно описують технічні вразливості та мережеві вузли, а не поведінкові рішення користувача.

Третій напрям стосується моделювання соціальної інженерії. Підходи [14], [15] формують структурну та стратегічну основу, проте не доводяться до процесної моделі траєкторії «підозра → компрометація». Отже, наявні роботи або описують поведінкові чинники без процесу, або моделюють процес атаки без урахування людського рішення. Запропоноване дослідження поєднує ці напрями шляхом застосування поглинального марківського ланцюга до поведінкової траєкторії користувача.

Мета роботи. Метою дослідження є розроблення, аналітична та чисельна верифікація процесної ймовірнісної моделі багатоходового фішингового сценарію на основі поглинального марківського ланцюга для оцінювання ймовірності компрометації, очікуваної кількості кроків до поглинання та чутливості результату до захисних втручань.

Окремим завданням є формалізація бюджетного вибору захисних заходів і перевірка меж жадібного алгоритму порівняно з точним перебором. Практична інтерпретація результатів обмежується заданою параметризацією моделі та потребує подальшого калібрування на реальній телеметрії.

МЕТОДОЛОГІЯ ДОСЛІДЖЕННЯ

Загальний підхід. Дослідження поєднує теоретичний аналіз, математичне моделювання, числовий експеримент і порівняльну перевірку. Методологічною основою є марківське наближення першого порядку, за якого наступний стан визначається поточним. Таке припущення використовується як базова



аналітична модель, тоді як накопичення довіри, втома уваги та латентні когнітивні стани розглядаються як обмеження, що потребують складніших розширень.

Математичне моделювання. Фішинговий процес подано як поглинальний марківський ланцюг із транзитними станами доставки, залучення, підозри та взаємодії і поглинальними станами компрометації та безпечного виходу. Числовий приклад обмежено credential-phishing сценарієм. Матрицю переходів зведено до канонічної форми з підматрицями Q і R ; фундаментальну матрицю $N = (I - Q)^{-1}$, матрицю ймовірностей поглинання $B = N \cdot R$ та вектор очікуваної кількості кроків $t = N \cdot \mathbf{1}$ отримано аналітично.

Обчислювальний експеримент і програмна реалізація. Модель реалізовано мовою Python із використанням NumPy. Аналітичний розв'язок отримано з систем $(I - Q)B = R$ та $(I - Q)t = \mathbf{1}$. Для чисельної перевірки виконано $2 \cdot 10^6$ траєкторій Монте-Карло від стану S_0 до поглинання з генератором `numpy.random.default_rng(42)`. Для оцінок ймовірності компрометації та середньої кількості кроків побудовано 95 % довірчі інтервали. Монте-Карло використовується лише для перевірки чисельної реалізації, а не для емпіричної валідації моделі.

Порівняльний аналіз і чисельна перевірка. Перевірено базовий сценарій, три захисні втручання, їх комбінації, бюджетний вибір і три допустимі напрями зміни переходів. У кожному випадку зміна цільового переходу компенсується зміною іншого переходу того самого рядка, що зберігає стохастичність матриці. Жадібне бюджетне рішення порівнюється з повним перебором допустимих підмножин, а чутливість оцінюється через напрямлені похідні вздовж допустимих збурень.

Метрики оцінювання. Основними показниками є ймовірність компрометації $p = b(S_0, C)$, умовні ймовірності $b(s_i, C)$, очікувана кількість кроків t , абсолютна і відносна редукція ризику Δp , ефективність на одиницю вартості $\eta = \frac{\Delta p}{c}$, оцінки Монте-Карло з 95 % довірчими інтервалами та напрямлена похідна $\frac{dp}{d\theta}$. Остання враховує обмеження $\sum_j p_{ij} = 1$ для кожного рядка матриці переходів.

ТЕОРЕТИЧНІ ОСНОВИ ДОСЛІДЖЕННЯ

Базові поняття та позначення. Дискретний марківський ланцюг задається трійкою (S, P, π_0) , де $S = \{s_1, \dots, s_m\}$ – скінченний простір станів, $P = [p_{ij}]$ – стохастична матриця переходів, π_0 – початковий розподіл. Елемент $p_{ij} = P(X_{t+1} = s_j | X_t = s_i)$ є ймовірністю переходу зі стану s_i до s_j за один крок. Модель ґрунтується на марковській властивості

$$P(X_{t+1} = s_j | X_t = s_i, X_{t-1}, \dots, X_0) = P(X_{t+1} = s_j | X_t = s_i) \quad (1)$$

тобто наступний стан залежить лише від поточного, і на умові стохастичності рядків

$$p_{ij} \geq 0, \quad \sum_j p_{ij} = 1, \quad \forall i \quad (2)$$

Стан s_i називають поглинальним, якщо $p_{ii} = 1$; ланцюг є поглинальним, якщо з кожного транзитного стану досяжний хоча б один поглинальний. Ця структура відповідає фішинговому сценарію з двома кінцевими наслідками [4], [13].

Формалізація задачі. Простір станів фішингового діалогу поділяємо на множину транзитних станів $T = \{S_0, S_1, S_2, S_3\}$ та множину поглинальних $\mathcal{A} = \{C, X\}$ (Таблиця 1). Упорядкувавши стани так, що транзитні передують поглинальним, зводимо P до канонічної форми

$$P = \begin{bmatrix} Q & R \\ 0 & I \end{bmatrix} \quad (3)$$

Таблиця 1.

Простір станів багатогодового фішингового сценарію

Стан	Позначення	Опис	Тип
S0	Доставка / периметрова обробка	Фішингове повідомлення надходить до поштової інфраструктури організації та проходить периметрову перевірку	транзитний
S1	Залучення	Повідомлення доставлено до скриньки; користувач відкриває та читає його	транзитний
S2	Підозра	У жертви виникли сумніви щодо легітимності	транзитний



Стан	Позначення	Опис	Тип
S3	Взаємодія	Користувач переходить за посиланням та вводить облікові дані на фішинговій сторінці	транзитний
C	Компрометація	Обліковий запис фактично скомпрометовано: зломисник успішно автентифікувався або отримав активну сесію	поглинальний
X	Безпечний вихід	Повідомлення видалено, проігноровано або повідомлено про інцидент	поглинальний

де $Q \in \mathbb{R}^{4 \times 4}$ описує переходи між транзитними станами, $R \in \mathbb{R}^{4 \times 2}$ – переходи з транзитних станів у поглинальні, I – одинична матриця. Оскільки з кожного транзитного стану досягне поглинання, матриця $(I-Q)$ є оборотною, і визначено фундаментальну матрицю

$$N = (I - Q)^{-1} = I + Q + Q^2 + \dots \quad (4)$$

елемент n_{ij} якої дорівнює очікуваній кількості відвідувань стану s_j за умови старту в s_i . З неї виводяться матриця ймовірностей поглинання та вектор очікуваної кількості кроків:

$$B = N \cdot R, \quad b_{ik} = P(\text{поглинання в стані } k \mid \text{старт у } i) \quad (5)$$

$$t = N \cdot \mathbf{1}, \quad t_i \text{ — очікувана кількість кроків до поглинання зі стану } i \quad (6)$$

де $\mathbf{1}$ – одиничний вектор-стовпець.

Основні параметри моделі. Параметрами є елементи Q і R , тобто ймовірності переходу між станами. У разі емпіричного застосування рядок $S0$ має калібруватися за телеметрією поштового шлюзу, результатами SPF/DKIM/DMARC-перевірок та фільтрації; рядок $S1$ – за фактом відкриття/переходу; стан $S2$ є латентним і не може бути надійно відновлений лише з агрегованих логів; рядок $S3$ для credential phishing має оцінювати ймовірність фактичного захоплення облікового запису після введення даних з урахуванням типу автентифікації. У поточному числовому експерименті використано ілюстративні переходи з Таблиці 2. Посилання [8]–[10] обґрунтовують правдоподібність поведінкових механізмів і діапазонів, але не є статистичним калібруванням конкретних значень матриці; тому отримані числові оцінки не слід переносити на реальні організації без власної телеметрії.

Цільова функція та критерій ефективності. Ключовим показником ризику є ймовірність компрометації за старту з доставки

$$\rho = b(S_0, C) = e(S_0)^T \cdot N \cdot R \quad (\text{стовпець } C) \quad (7)$$

Атакувальний і захисний боки мають протилежно спрямовані цілі щодо величини ρ : атакувальний бік прагне її збільшення через персоналізацію, багатоконтактність і згенерований контент [3], тоді як захисний – зменшення [15]. У цій роботі формально розв'язується саме задача захисника. Нехай задано множину доступних втручань K , кожне з яких k змінює певні елементи P (переводячи їх у $P^{(k)}$) і має вартість c_k , а бюджет становить B . Тоді

$$\min \rho(P^{(K')}) \quad \text{за умови} \quad \sum_{k \in K'} c_k \leq B, \quad K' \subseteq K \quad (8)$$

Критерієм ефективності окремого втручання слугує гранична редукція ризику на одиницю вартості

$$\eta_k = \frac{\Delta \rho_k}{c_k}, \quad \Delta \rho_k = \rho(P) - \rho(P^{(k)}) \quad (9)$$

що дозволяє ранжувати заходи за їхнім внеском у зниження ρ .

Алгоритм. На основі критерію (9) будується жадібна процедура бюджетного вибору, яка на кожному кроці додає допустиме втручання з найбільшою поточною ефективністю η . Процедура є евристичною: без доведеної субмодулярності цільової функції вона не гарантує глобального оптимуму. Тому для малого $|K|$ її результат доцільно перевіряти повним перебором усіх допустимих підмножин, а



для більших задач – застосовувати точні або апроксимаційні методи комбінаторної оптимізації залежно від структури витрат та взаємодії заходів.

Застосування на практиці. Модель може використовуватися як інструмент підтримки рішень лише після калібрування. Спочатку з телеметрії реальних тренувальних кампаній і поштової інфраструктури оцінюють спостережні переходи, а латентний стан підозри – за допомогою окремих поведінкових вимірювань або прихованої марківської моделі. Далі обчислюють базовий ризик та очікувану довжину траєкторії, після чого оцінюють зміни переходів, пов'язані з конкретними контролюями. У межах поточної credential-phishing постановки периметрові засоби впливають на $S0 \rightarrow S1$, навчання і контекстні попередження – на перерозподіл виходів зі $S2$, а фішинг-стійка автентифікація – на $S3 \rightarrow C$. Спільний ефект визначається повторним розрахунком матриці, а не припущенням про адитивність. Усі практичні пріоритети є організаційно-залежними і не можуть встановлюватися лише за ілюстративною матрицею.

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Умови експерименту та платформа. Дослідження має характер числового експерименту над формалізованою моделлю, а не польового експерименту з користувачами. Аналітичні значення обчислено в Python із NumPy; для внутрішньої чисельної верифікації виконано $2 \cdot 10^6$ траєкторій Монте-Карло від $S0$ до поглинання з генератором `numpy.random.default_rng(42)`. Вхідну матрицю переходів наведено в Таблиці 2. Оскільки вона не оцінена з конкретної реальної кампанії, Монте-Карло перевіряє коректність обчислювальної реалізації для заданої P , але не підтверджує емпіричну валідність самої P . Зведені параметри експерименту наведено в Таблиці 3.

Таблиця 2.

Ілюстративна матриця ймовірностей переходу (за один крок)

Зі стану \ У стан	S0	S1	S2	S3	C	X
S0 (Доставка)	–	0,35	–	–	–	0,65
S1 (Залучення)	–	–	0,40	0,45	–	0,15
S2 (Підозра)	–	0,10	–	0,35	–	0,55
S3 (Взаємодія)	–	–	–	–	0,60	0,40

Примітка. Символ «–» позначає нульову ймовірність переходу. Поглинальні стани C і X мають $p(C, C) = 1$ та $p(X, X) = 1$. Значення є ілюстративними: вони сформовані так, щоб відтворити правдоподібну багатоходову траєкторію та узгоджуватися за порядком величин із поведінковими спостереженнями [8]–[10], але не є статистичними оцінками, отриманими з цих робіт або з телеметрії конкретної організації.

Таблиця 3.

Параметри числового експерименту

Параметр	Значення
Програмна платформа	Python, NumPy
Аналітичний метод	розв'язання лінійних систем $(I - Q)B = R$; $(I - Q)t = 1$
Чисельна верифікація	Монте-Карло, $2 \cdot 10^6$ траєкторій; не емпірична валідація
Зерно генератора	<code>numpy.random.default_rng(42)</code> , один $U(0,1)$ на активну траєкторію за крок
Вхідні дані	матриця переходів (Таблиця 2); 4 транзитні + 2 поглинальні стани
Початковий стан	$S0$ (периметрова доставка)
Тестові сценарії	базовий; 3 одиничні втручання; комбінації; бюджетний benchmark; 3 допустимі напрями чутливості
Метрики	ρ ; $b(s_i, C)$; t ; $\Delta\rho$; η ; $\hat{\rho}_{MC}$ і 95%ДІ; $\bar{t}_{MC}$ і 95%ДІ; $\frac{d\rho}{d\theta}$



Параметр	Значення
Статистична оцінка Монте-Карло	$SE(\hat{\rho}_{MC}) = \sqrt{\frac{\hat{\rho}_{MC}(1 - \hat{\rho}_{MC})}{M}}; \quad 95\% \text{ ДІ: } \hat{\rho}_{MC} \pm 1,96 SE(\hat{\rho}_{MC})$ для $\bar{t}_{MC}$: $\bar{t}_{MC} \pm 1,96 \frac{s_t}{\sqrt{M}}$

Набір тестів і метрики. Окрім базового сценарію та трьох одиничних втручань перевірено всі їх комбінації, жадібний бюджетний вибір для кількох значень нормованого бюджету та три допустимі одновимірні напрями зміни матриці. Для кожного збурення сума відповідного рядка Р зберігається рівною одиниці. Монте-Карло оцінюється не через довільну «відносну похибку» однієї реалізації, а через стандартну похибку та 95 % довірчий інтервал, що є статистично коректнішим способом зіставлення з аналітичним значенням.

Результати та їх чисельна верифікація. Аналітичний розв'язок дає $\rho = 0,1290625$ та $t(S_0) = 1,7255208$. Для $M = 2 \cdot 10^6$ траєкторій реалізація Монте-Карло з генератором `default_rng(42)` дала $\hat{\rho}_{MC} = 0,12943$. Стандартна похибка становить приблизно 0,0002374, а 95 % довірчий інтервал для ймовірності компрометації – [0,12896; 0,12990]; аналітичне значення 0,1290625 потрапляє до цього інтервалу. Для кількості кроків отримано $\bar{t}_{MC} = 1,725861$, $s_t \approx 1,07005$ та 95 % довірчий інтервал [1,72438; 1,72734], який також містить аналітичне $t(S_0)$. Отже, симуляція підтверджує внутрішню чисельну узгодженість двох способів розрахунку для заданої матриці Р. Вона не є незалежною емпіричною валідацією моделі: і аналітичний розрахунок, і Монте-Карло використовують ті самі задані перехідні ймовірності.

Таблиця 4.

Базові результати: аналітичний розв'язок і чисельна верифікація Монте-Карло			
Початковий стан	$b(\cdot, C)$	$b(\cdot, X)$	t (кроків)
S0 (доставка)	0,12906	0,87094	1,72552
S1 (залучення)	0,36875	0,63125	2,07292
S2 (підозра)	0,24688	0,75312	1,55729
S3 (взаємодія)	0,60000	0,40000	1,00000
Монте-Карло, старт S0	0,12943 [0,12896; 0,12990]	0,87057	1,72586 [1,72438; 1,72734]

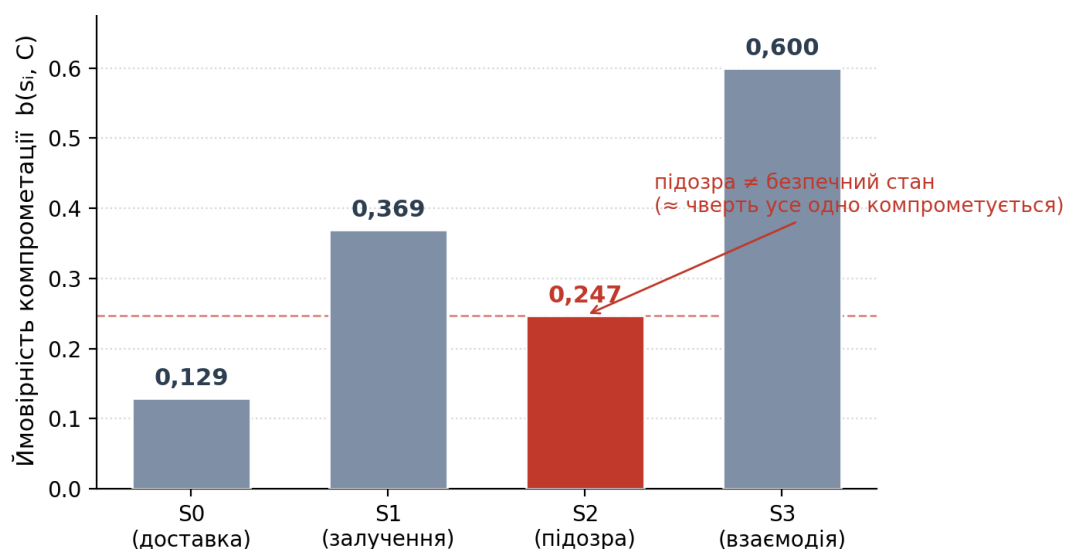


Рис. 1. Ймовірність поглинання у стані компрометації залежно від початкового стану жертви

Ключовим є результат для стану S2: за обраної ілюстративної параметризації $b(S_2, C) = 0,246875$, тобто приблизно 24,7 %. Це означає лише те, що в заданій матриці стан підозри не є поглинальним і частина траєкторій після нього може завершитися компрометацією. Значення 24,7 % не є оцінкою частки реальних користувачів: воно прямо визначається заданими переходами $S2 \rightarrow S1$, $S2 \rightarrow S3$ та $S2 \rightarrow X$.



Емпіричне оцінювання цього результату потребує спостережень, які дають інформацію про латентний стан підозри, або прихованої марківської моделі.

Порівняння втручань із базовим підходом. Кожне втручання задається як стохастично допустима зміна одного рядка матриці: зменшення $S0 \rightarrow S1$ компенсується збільшенням $S0 \rightarrow X$; збільшення $S2 \rightarrow X$ – зменшенням $S2 \rightarrow S3$ за незмінного $S2 \rightarrow S1$; зменшення $S3 \rightarrow C$ – збільшенням $S3 \rightarrow X$. Таким чином, ефекти не моделюються як незалежні зміни окремих p_{ij} . Таблиця 5 показує результат саме для цих сценарних змін, а не емпірично встановлену ефективність відповідних технологій.

Таблиця 5.

Порівняння захисних втручань із базовим сценарієм

Сценарій	Цільовий перехід	ρ	$\Delta\rho$	$\Delta\rho$ (%)
Базовий	–	0,12906	–	–
Автентифікація пошти	$S0 \rightarrow S1$: 0,35 $\rightarrow$ 0,20	0,07375	0,05531	–42,9 %
Навчання / попередження	$S2 \rightarrow X$: 0,55 $\rightarrow$ 0,75	0,11156	0,01750	–13,6 %
Фішинг-стійка MFA	$S3 \rightarrow C$: 0,60 $\rightarrow$ 0,30	0,06453	0,06453	–50,0 %
Комбіноване (усі три)	–	0,03188	0,09719	–75,3 %

Перевірка бюджетної оптимізації. Щоб продемонструвати роботу без удаваної економічної точності, для трьох заходів задано нормовані ілюстративні вартості: $c_{mail} = 3$, $c_{train} = 2$, $c_{MFA} = 5$ умовних одиниць. Для бюджетів $B = 5, 7, 8, 10$ жадібний результат зіставлено з точним повним перебором усіх допустимих підмножин. Такий benchmark принципово важливий, оскільки критерій $\frac{\Delta\rho}{c}$ не гарантує глобального оптимуму.

Таблиця 6.

Жадібний бюджетний вибір і точний benchmark для ілюстративних нормованих витрат

Бюджет B	Жадібний набір	ρ_{greedy}	Точний оптимум	ρ_{opt}	Висновок
5	пошта + навчання	0,06375	пошта + навчання	0,06375	збіг
7	пошта + навчання	0,06375	навчання + MFA	0,05578	не оптимальний
8	пошта + MFA	0,03688	пошта + MFA	0,03688	збіг
10	усі три	0,03188	усі три	0,03188	збіг

Результат при $B = 7$ демонструє принципову межу евристики: після вибору поштового контролю за найбільшим початковим $\frac{\Delta\rho}{c}$ залишку бюджету недостатньо для MFA, тому жадібний набір дає $\rho = 0,06375$, тоді як точна комбінація «навчання + MFA» дає $\rho = 0,05578$.

Аналіз впливу основних параметрів. Оскільки елементи одного рядка P не є незалежними, чутливість оцінюється вздовж явно заданих допустимих напрямів. Для $S3$ параметр θ збільшує $S3 \rightarrow C$ і на ту саму величину зменшує $S3 \rightarrow X$; для $S0$ збільшує $S0 \rightarrow S1$ та зменшує $S0 \rightarrow X$; для $S2$ збільшує $S2 \rightarrow X$ і зменшує $S2 \rightarrow S3$, залишаючи $S2 \rightarrow S1$ сталим. У цих одновимірних сценаріях залежність ρ від θ є лінійною, а напрямлені похідні $\frac{d\rho}{d\theta}$ становлять відповідно +0,2151, +0,36875 та –0,0875. Найбільша локальна чутливість спостерігається для допустимої зміни на межі доставки, але підсумковий ефект конкретного контролю визначається і величиною реально досяжного збурення. Позначення $\frac{\partial\rho}{\partial p}$ на Рис. 2 слід читати як скорочене позначення цієї одновимірної напрямленої похідної, а не як похідну за незалежним елементом стохастичної матриці.

Таблиця 7.

Чутливість ρ уздовж стохастично допустимих одновимірних збурень

Допустимий напрям зміни	ρ при зміні цільового переходу	$\frac{d\rho}{d\theta}$
$S3 \rightarrow C \uparrow$, $S3 \rightarrow X \downarrow$; $p(C)$: 0,30 $\rightarrow$ 0,90	$0,065 \cdot 0,097 \cdot 0,129 \cdot 0,161 \cdot 0,194$	+0,2151
$S0 \rightarrow S1 \uparrow$, $S0 \rightarrow X \downarrow$; $p(S1)$: 0,20 $\rightarrow$ 0,50	$0,074 \cdot 0,101 \cdot 0,129 \cdot 0,157 \cdot 0,184$	+0,3688



Допустимий напрям зміни	ρ при зміні цільового переходу	$\frac{d\rho}{d\theta}$
$S2 \rightarrow X \uparrow, S2 \rightarrow S3 \downarrow; p(X):$ $0,35 \rightarrow 0,75$	$0,147 \cdot 0,138 \cdot 0,129 \cdot 0,120 \cdot 0,112$	$-0,0875$

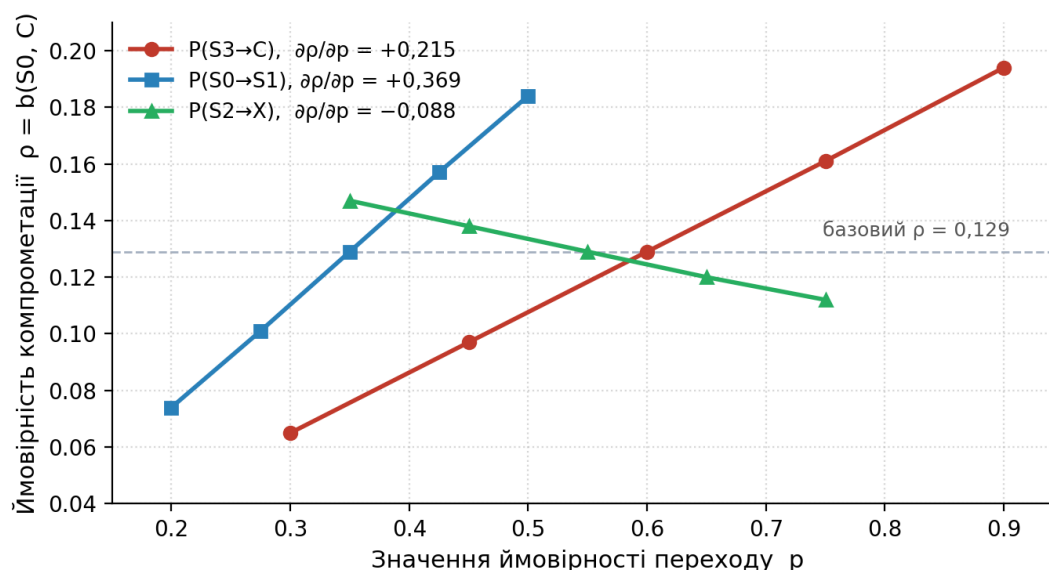


Рис. 2. Чутливість ймовірності компрометації до зміни ключових переходів ланцюга

Отримані результати слід розглядати як демонстрацію властивостей запропонованої моделі, а не як універсальні характеристики фішингових кампаній. У межах заданої матриці ризик визначається сукупністю переходів між станами, тому захисні заходи доцільно трактувати як керування цими ймовірностями. Найбільше зниження ρ у розглянутому сценарії забезпечують зміни на етапах периметрової доставки та фінальної автентифікації, однак реальний пріоритет заходів має визначатися за телеметрією конкретної організації, фактичними витратами та оціненим впливом контролів. Стан «підозра» є латентним, тому переходи з нього не можуть надійно оцінюватися лише за стандартними логами без додаткових поведінкових вимірювань або прихованої марківської моделі.

Модель має перевагу у прозорому зв'язуванні переходів, ризику та захисних втручань, підтримує аналіз чутливості й бюджетне порівняння заходів, а Монте-Карло та повний перебір використовуються для перевірки чисельної реалізації й жадібної евристики.

ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

У роботі побудовано процесну модель багатоходового credential-phishing сценарію у формі поглинального марківського ланцюга та показано, як відокремити проміжні поведінкові стани від фактичної компрометації облікового запису. Для заданої ілюстративної матриці аналітично отримано $\rho = 0,1290625$ і $t(S_0) = 1,7255208$. Внутрішню чисельну коректність реалізації перевірено на $2 \cdot 10^6$ траєкторій Монте-Карло: $\hat{\rho}_{MC} = 0,12943$ з 95 % ДІ $[0,12896; 0,12990]$, а $\bar{t}_{MC} = 1,725861$ з 95 % ДІ $[1,72438; 1,72734]$; обидва аналітичні значення містяться у відповідних інтервалах. Цей збіг підтверджує узгодженість двох обчислювальних реалізацій для однієї P , але не є емпіричною валідацією, оскільки перехідні ймовірності не оцінені з реальної кампанії. Аналіз чутливості переформульовано через напрямлені похідні, що зберігають стохастичність рядків. Для бюджетної процедури введено точний benchmark: при $B = 7$ жадібний набір дає $\rho = 0,06375$, тоді як глобально краща з перевірених комбінацій дає $\rho = 0,05578$, що демонструє межу евристичного алгоритму. Основними умовами є калібрування P на телеметрії, окреме моделювання латентного стану підозри, оцінювання реальних витрат та ефектів контролів і зовнішня валідація на відкладених кампаніях. Практична цінність запропонованої постановки полягає не в універсальних відсотках, а в прозорому зв'язуванні станів, переходів, захисних втручань і кінцевого ризику.



СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. ENISA. (2024). ENISA threat landscape 2024. European Union Agency for Cybersecurity.
2. Zhuo, S., Biddle, R., Koh, Y. S., Lottridge, D., & Russello, G. (2023). SoK: Human-centered phishing susceptibility. *ACM Transactions on Privacy and Security*, 26(3), Article 24. <https://doi.org/10.1145/3575797>
3. Falade, P. V. (2023). Decoding the threat landscape: ChatGPT, FraudGPT, and WormGPT in social engineering attacks. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 9(5), 185-198.
4. Cheimonidis, P., & Rantos, K. (2025). A proactive and time-sensitive cyber risk assessment model integrating Markov chains and Bayesian networks. *IEEE Access*, 13, 96911-96932. <https://doi.org/10.1109/ACCESS.2025.3575070>
5. Theodosiadou, O., Chatzakou, D., Tsikrika, T., Vrochidis, S., & Kompatsiaris, I. (2023). Real-time threat assessment based on hidden Markov models. *Risk Analysis*, 43(10), 2069-2081.
6. Sommestad, T., & Karlzén, H. (2024). The unpredictability of phishing susceptibility: Results from a repeated measures experiment. *Journal of Cybersecurity*, 10(1), tyae021. <https://doi.org/10.1093/cybsec/tyae021>
7. Ribeiro, L., Guedes, I. S., & Cardoso, C. S. (2024). Which factors predict susceptibility to phishing? An empirical study. *Computers & Security*, 136, 103558. <https://doi.org/10.1016/j.cose.2023.103558>
8. Gan, C. L., Lee, Y. Y., & Liew, T. W. (2024). Fishing for phishy messages: Predicting phishing susceptibility through the lens of cyber-routine activities theory and heuristic-systematic model. *Humanities and Social Sciences Communications*, 11(1), Article 1552. <https://doi.org/10.1057/s41599-024-04083-1>
9. Klütsch, J., Schwab, J., Böffel, C., Zimmermann, V., & Schlittmeier, S. J. (2024). Friend or phisher: How known senders and fear of missing out affect young adults' phishing susceptibility on social media. *Humanities and Social Sciences Communications*, 11(1). <https://doi.org/10.1057/s41599-024-03412-8>
10. Sutter, T., Bozkir, A. S., Gehring, B., & Berlich, P. (2022). Avoiding the hook: Influential factors of phishing awareness training on click-rates and a data-driven approach to predict email difficulty perception. *IEEE Access*, 10, 100540-100565. <https://doi.org/10.1109/ACCESS.2022.3207272>
11. Cheimonidis, P., & Rantos, K. (2023). Dynamic risk assessment in cybersecurity: A systematic literature review. *Future Internet*, 15(10), 324. <https://doi.org/10.3390/fi15100324>
12. Cheimonidis, P., & Rantos, K. (2025). A novel proactive and dynamic cyber risk assessment methodology. *Computers & Security*, 154, 104439. <https://doi.org/10.1016/j.cose.2025.104439>
13. Chen, L., Zhang, T., Ma, Y., Li, Y., Wang, C., He, C., Lv, Z., & Li, N. (2024). A Bayesian-attack-graph-based security assessment method for power systems. *Electronics*, 13(13), 2628. <https://doi.org/10.3390/electronics13132628>
14. Li, T., Song, C., & Pang, Q. (2023). Defending against social engineering attacks: A security pattern-based analysis framework. *IET Information Security*, 17(4), 703–726. <https://doi.org/10.1049/ise2.12125>
15. Sridhar, A. P. (2025). Analyzing social engineering attack patterns using behavioral psychology and AI-driven defense mechanisms. *Journal of Internet Services and Information Security*, 15(1), 502-519.
16. Tamal, M. A., Islam, M. K., Bhuiyan, T., Sattar, A., & Prince, N. U. (2024). Unveiling suspicious phishing attacks: Enhancing detection with an optimal feature vectorization algorithm and supervised machine learning. *Frontiers in Computer Science*, 6, Article 1428013. <https://doi.org/10.3389/fcomp.2024.1428013>
17. ISO/IEC. (2022). Information security, cybersecurity and privacy protection – Guidance on managing information security risks (ISO/IEC 27005:2022). International Organization for Standardization.

**Miskevych Oksana**

Senior Lecturer of Cybersecurity Department
Lutsk National Technical University, Lutsk, Ukraine
ORCID: 0000-0002-5009-2391
miskevich87@gmail.com

Surova Natalia

Ph.D., associate professor of Cybersecurity Department
Lutsk National Technical University, Lutsk, Ukraine
ORCID: 0000-0002-7116-0197
n.surova@lntu.edu.ua

Kostiuchko Serhii

Ph.D., associate professor, head of Cybersecurity Department
Lutsk National Technical University, Lutsk, Ukraine
ORCID: 0000-0002-1262-6268
s.kostiuchko@gmail.com

STOCHASTIC MODELING AND OPTIMIZATION OF PROTECTION AGAINST MULTI-STAGE PHISHING CAMPAIGNS

Abstract. Phishing remains one of the primary vectors for gaining initial access to information systems, and the success of such attacks depends largely on user behavior. Traditional binary metrics such as "click" versus "no click" fail to capture the multi-stage trajectory in which a user moves through states including message delivery, engagement, suspicion, interaction, compromise, or safe exit. This study aims to develop and numerically verify a probabilistic process model of a multi-stage phishing scenario, enabling the assessment of compromise probability, the expected number of steps to scenario completion, and the sensitivity of the outcome to defensive interventions. The model is constructed using an absorbing Markov chain with a finite state space; a transition matrix is formulated, and transient and absorbing states, the fundamental matrix, the absorption probability matrix, and the vector of expected step counts are defined. For an illustrative credential-phishing scenario, the compromise probability ($\rho = 0.1290625$) and the expected number of steps ($t(S_0) = 1.7255208$) were derived analytically. The internal numerical correctness of the implementation was verified using the Monte Carlo method over 2×10^6 trajectories, with 95% confidence intervals constructed; the results are consistent with the analytical solution. The Monte Carlo method was employed solely for numerical verification rather than as an empirical validation of the model. To analyze security measures, the study considers stochastically admissible changes to transition probabilities corresponding to improvements in email authentication, user training, and the implementation of phishing-resistant multi-factor authentication. A cost-constrained selection of measures was performed using a greedy procedure based on the criterion of risk reduction per unit cost and compared with an exhaustive search. It is shown that the greedy solution does not always coincide with the globally optimal one. The scientific novelty lies in representing the user's behavioral trajectory as a Markov chain and formalizing the relationship between security interventions, changes in transition probabilities, and the final risk of compromise.

Keywords: phishing; social engineering; Markov chains; absorbing Markov chain; transition matrix; probability of compromise; multi-stage attack; sensitivity analysis; cyber risk assessment.

REFERENCES

1. ENISA. (2024). ENISA threat landscape 2024. European Union Agency for Cybersecurity.
2. Zhuo, S., Biddle, R., Koh, Y. S., Lottridge, D., & Russello, G. (2023). SoK: Human-centered phishing susceptibility. *ACM Transactions on Privacy and Security*, 26(3), Article 24. <https://doi.org/10.1145/3575797>
3. Falade, P. V. (2023). Decoding the threat landscape: ChatGPT, FraudGPT, and WormGPT in social engineering attacks. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 9(5), 185-198.



4. Cheimonidis, P., & Rantos, K. (2025). A proactive and time-sensitive cyber risk assessment model integrating Markov chains and Bayesian networks. *IEEE Access*, 13, 96911-96932. <https://doi.org/10.1109/ACCESS.2025.3575070>
5. Theodosiadou, O., Chatzakou, D., Tsikrika, T., Vrochidis, S., & Kompatsiaris, I. (2023). Real-time threat assessment based on hidden Markov models. *Risk Analysis*, 43(10), 2069-2081.
6. Somestad, T., & Karlzén, H. (2024). The unpredictability of phishing susceptibility: Results from a repeated measures experiment. *Journal of Cybersecurity*, 10(1), tyae021. <https://doi.org/10.1093/cybsec/tyae021>
7. Ribeiro, L., Guedes, I. S., & Cardoso, C. S. (2024). Which factors predict susceptibility to phishing? An empirical study. *Computers & Security*, 136, 103558. <https://doi.org/10.1016/j.cose.2023.103558>
8. Gan, C. L., Lee, Y. Y., & Liew, T. W. (2024). Fishing for phishy messages: Predicting phishing susceptibility through the lens of cyber-routine activities theory and heuristic-systematic model. *Humanities and Social Sciences Communications*, 11(1), Article 1552. <https://doi.org/10.1057/s41599-024-04083-1>
9. Klütsch, J., Schwab, J., Böffel, C., Zimmermann, V., & Schlittmeier, S. J. (2024). Friend or phisher: How known senders and fear of missing out affect young adults' phishing susceptibility on social media. *Humanities and Social Sciences Communications*, 11(1). <https://doi.org/10.1057/s41599-024-03412-8>
10. Sutter, T., Bozkir, A. S., Gehring, B., & Berlich, P. (2022). Avoiding the hook: Influential factors of phishing awareness training on click-rates and a data-driven approach to predict email difficulty perception. *IEEE Access*, 10, 100540-100565. <https://doi.org/10.1109/ACCESS.2022.3207272>
11. Cheimonidis, P., & Rantos, K. (2023). Dynamic risk assessment in cybersecurity: A systematic literature review. *Future Internet*, 15(10), 324. <https://doi.org/10.3390/fi15100324>
12. Cheimonidis, P., & Rantos, K. (2025). A novel proactive and dynamic cyber risk assessment methodology. *Computers & Security*, 154, 104439. <https://doi.org/10.1016/j.cose.2025.104439>
13. Chen, L., Zhang, T., Ma, Y., Li, Y., Wang, C., He, C., Lv, Z., & Li, N. (2024). A Bayesian-attack-graph-based security assessment method for power systems. *Electronics*, 13(13), 2628. <https://doi.org/10.3390/electronics13132628>
14. Li, T., Song, C., & Pang, Q. (2023). Defending against social engineering attacks: A security pattern-based analysis framework. *IET Information Security*, 17(4), 703-726. <https://doi.org/10.1049/ise2.12125>
15. Sridhar, A. P. (2025). Analyzing social engineering attack patterns using behavioral psychology and AI-driven defense mechanisms. *Journal of Internet Services and Information Security*, 15(1), 502-519.
16. Tamal, M. A., Islam, M. K., Bhuiyan, T., Sattar, A., & Prince, N. U. (2024). Unveiling suspicious phishing attacks: Enhancing detection with an optimal feature vectorization algorithm and supervised machine learning. *Frontiers in Computer Science*, 6, Article 1428013. <https://doi.org/10.3389/fcomp.2024.1428013>
17. ISO/IEC. (2022). Information security, cybersecurity and privacy protection – Guidance on managing information security risks (ISO/IEC 27005:2022). International Organization for Standardization.

Отримано редакцією журналу / Received: 15.06.26

Прорецензовано / Revised: 28.06.26

Схвалено до друку / Accepted: 26.10.26



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