УДК 004.056.5:004.94:355.01
DOI: https://doi.org/10.17721/ISTS.2025.9.47-53

Дмитро ДЖЕНДЖЕРО, асп.
ORCID ID: 0009-0007-9999-850X
e-mail: dzhendzherod@gmail.com
Київський національний університет імені Тараса Шевченка, Київ, Україна

# МЕТОД МОДЕЛЮВАННЯ ГІБРИДНОГО ВПЛИВУ НА ДЕРЖАВНІ СЕРВІСИ З ВИКОРИСТАННЯМ АГЕНТНО-ОРІЄНТОВАНОГО ПІДХОДУ

**Вступ.** *У контексті гібридної війни зростає роль інформаційного впливу на критичну інфраструктуру держави. Фейкові повідомлення можуть спричиняти хвилі поведінкової активності користувачів, що імітує DDoS-атаки й призводить до перевантаження державних цифрових сервісів навіть без технічного втручання. Виникає потреба у моделюванні таких сценаріїв з урахуванням технічних і соціально-поведінкових чинників.*

**Методи.** *Запропоновано агентно-орієнтовану модель гібридного впливу, що включає користувачів, ботів, джерела фейків і захисні механізми. Модель реалізовано у вигляді математичних співвідношень і сценарного моделювання, включаючи базовий сценарій, сценарій із затриманою реакцією, сценарій із технічним втручанням і сценарій з інформаційним спростуванням. Аналіз здійснено з урахуванням параметрів навантаження, поведінкової активації та реакції захисної системи.*

**Результати.** *Встановлено, що поведінкова активність користувачів може створити навантаження, еквівалентне DDoS-атаці. Найефективнішим виявився комбінований сценарій, що поєднує технічне блокування й інформаційне спростування, який дозволив знизити навантаження до стабільного рівня. Визначено, що класичні моделі кіберзахисту не враховують відкладену реакцію користувачів або хвильову динаміку, що обмежує їхню ефективність у гібридному середовищі.*

**Висновки.** *Запропонована модель дозволить не лише змоделювати гібридний вплив, але й оцінити ефективність тактик реагування на цей вплив. Вона може бути використана в CERT-платформах, симуляціях Red Team / Blue Team, а також у системах раннього попередження. Подальші дослідження можуть зосередитись на адаптації моделі до реального середовища й автоматизованому реагуванні на інформаційні загрози.*

**Ключові слова:** *гібридна війна, агентне моделювання, фейкові повідомлення, цифрова безпека, поведінкове навантаження.*

## Вступ

У сучасних умовах зростаючої цифровізації державні інформаційні ресурси дедалі частіше стають об'єктами складних багатовекторних атак. Одним із проявів таких загроз є гібридний вплив, що поєднує технічні засоби (зокрема, DDoS-атаки – Distributed Denial of Service) з інформаційними кампаніями, спрямованими на формування певної поведінкової реакції користувачів. Внаслідок поширення фейкових повідомлень або викривленої інформації користувачі масово здійснюють запити до сервісів, що призводить до перевантаження навіть за відсутності суто технічної атаки. Схожий ефект описано як “інфодемію” у роботі Matteo et al. (2020), де підкреслено, що соціальні платформи можуть стати тригером пікового навантаження виключно через поведінкову реакцію аудиторії.

Традиційні підходи до виявлення і протидії DDoS-атакам передбачають реагування на трафік, який генерується ботами або автоматизованими системами. Однак у гібридному середовищі значна частина навантаження створюється легітимними користувачами, які реагують на дезінформацію, поширену через месенджери, соціальні мережі чи анонімні канали. Це створює серйозні виклики для систем захисту, оскільки поведінкова активність не є злочинною за суттю, але може мати руйнівні наслідки для стабільності інфраструктури.

Актуальність дослідження посилюється з огляду на реальні інциденти, які відбувалися під час активних фаз гібридної війни в Україні. Наприклад, фейкові повідомлення про “злам” державних цифрових сервісів 2022 р. призводили до хвильових навантажень на інфраструктуру, спричинених масовим входом, змінами паролів або надсиланням запитів підтримки. Це свідчить про необхідність моделювання таких процесів не лише на рівні мережного трафіка, а і з урахуванням поведінки користувачів. Швидкість і масштаб поширення фейкової інформації є викликом не лише для технічного, а і для соціального контролю за інформаційними потоками (David et al., 2018).

В таких умовах виникає потреба у моделюванні гібридного впливу з урахуванням ролей окремих агентів: користувачів, джерел фейків, ботів і захисних систем. Агентно-орієнтоване моделювання (ABM – Agent-Based Modeling) дає змогу реалізувати таку багатошарову систему. Цей підхід дозволяє враховувати індивідуальну реакцію агентів, часову затримку в активності, ефект від спростування фейків і можливість комбінованого навантаження.

***Метою*** цієї роботи є розроблення й апробація агентно-орієнтованої моделі виявлення і стримування гібридного впливу, зокрема й поведінкової активності користувачів, викликаної інформаційними атаками. Запропонована модель поєднує класичні компоненти аналізу навантаження з реакцією на інформаційні стимули та дозволяє формувати сценарії, в яких навіть невелика кількість фейкових повідомлень може призвести до системного збою. Це створює умови, за яких без технічного втручання цифрові системи можуть зазнавати суттєвого навантаження внаслідок неконтрольованої активності користувачів.

***Огляд літератури.*** Одним із критичних факторів у контексті гібридного впливу є динаміка та тип реакції користувачів на інформаційні повідомлення. У дослідженні Glenski, Weninger, & Volkova (2018) запропоновано нейронну модель, що класифікує реакції користувачів на новини в соціальних мережах за дев'ятьма типами, зокрема такими: запитання, уточнення, вдячність, емоційна оцінка, заперечення або гумор. Така деталізація дозволила авторам виявити відмінності у сприйнятті достовірних і дезінформаційних джерел, що також впливає на швидкість та інтенсивність реакцій.

Аналіз понад 10 млн твітів і 6 млн коментарів Reddit показав, що реакції на недостовірні новини часто

© Дженджеро Дмитро, 2025



ISSN 2707-1758

мають іншу природу – наприклад, у відповідь на фейкові новини користувачі частіше демонструють вдячність або задають уточнювальні запитання, тоді як достовірні джерела викликають глибші аналітичні відповіді чи емоційне схвалення.

Зазначене дослідження підтверджує, що моделювання поведінки користувачів у контексті гібридних загроз не повинно зводитися до бінарної моделі "активовано / не активовано". Натомість, доцільним є використання функціональних імовірнісних моделей, які враховують різноманітність реакцій – саме це і було реалізовано у запропонованій агентно-орієнтованій моделі. Shu et al. (2019) підкреслюють, що ефективне виявлення фейкової інформації на соціальних платформах неможливе без урахування трьох взаємопов'язаних аспектів: змісту, реакцій користувачів і характеристик джерел.

У дослідженні Dhiman Shah, Thirunarayan (2024) запропоновано гібридну модель GBERT (GPT-BERT), яка поєднує архітектури BERT (Bidirectional Encoder Representations from Transformers) і GPT (Generative Pre-trained Transformer) для підвищення точності виявлення фейкових новин. Модель поєднує дві найсучасніші технології: BERT відповідає за глибокий контекстуальний аналіз тексту, а GPT – за виявлення довгострокових семантичних зв'язків і логічної узгодженості повідомлень. Отримане поєднання протестовано на двох реальних наборах даних із Twitter і Reddit, де продемонструвало підвищену точність класифікації фейкових повідомлень порівняно з традиційними архітектурами. Такий підхід підтверджує доцільність поєднання контент-орієнтованого аналізу з мовними моделями у разі виявлення дезінформації – саме це лягло в основу структури поведінкової оцінки в нашій моделі.

Як показують результати дослідження Tafur, & Sarkar (2023), на формування думки користувачів щодо достовірності інформації впливають не лише зміст повідомлення чи джерело публікації, а й сигнали, які надходять від автоматизованих систем верифікації. У проведеному експерименті із 40 учасниками виявлено, що навіть за наявності помилок із боку системи класифікації фейкових новин, користувачі продовжували покладатися на її висновки у понад 57 % випадків. Така поведінка свідчить про високий рівень довіри до алгоритмів і демонструє важливу роль соціального контексту, когнітивних упереджень і настанов у процесі прийняття рішень. Ці висновки важливі для побудови агентно-орієнтованих моделей, у яких реакція користувача моделюється не лише як відповідь на контент, а і як результат взаємодії зі складною інформаційною екосистемою.

Агентно-орієнтовані моделі виявилися ефективними для моделювання складних загроз у кіберпросторі. У роботі Kotenko Stepashkin, & Doynikova (2008) запропоновано архітектуру симуляційного середовища для аналізу атак ботнетів і механізмів їхнього стримування. Автори представили динамічну взаємодію між агентами-атаками (ботами) та агентами-захисниками, що моделювали координацію у межах багаторівневих команд. У фокусі дослідження – протистояння DDoS-атакам із використанням сценаріїв самонавчання, фільтрації та обміну сигнатурами між агентами. Такий підхід близький до реалізованої в нашій моделі логіки: поєднання поведінкових агентів, системи блокування та реагування у складному інформаційному середовищі. Це свідчить про сталість підходу до використання агентних структур у моделюванні кіберзагроз, зокрема і в разі врахування поведінки ботмереж та адаптивних захисних стратегій.

**Методи**

У моделі розглядають сукупність агентів, кожен з яких відіграє окрему роль у сценарії гібридного впливу. Кожен агент має набір поведінкових правил і здатний змінювати стан системи. Основні типи агентів наведено в табл. 1.

*Таблиця 1*

**Основні типи агентів у моделі**

| Агент | Позначення | Поведінка |
|---|---|---|
| Користувач | Ac | Реагує на фейкову інформацію, генерує запити до сервісу |
| Telegram-канал | At | Поширює фейкові повідомлення |
| Бот/DDoS-агент | Ab | Імітує масові запити на перевантаження |
| Державний сервіс | As | Обробляє запити, має обмежену пропускну здатність |
| Захисна система | Az | Реагує на навантаження, блокує джерела трафіка |

Імовірність активації користувача розраховують як

$$P_c = \alpha \cdot D_t + \beta \cdot V_t,$$

де $P_c$ – імовірність того, що користувач здійснить дію (перехід, запит), $D_t$ – ступінь довіри до джерела, $V(t)$ – інтенсивність поширення повідомлення.

Розглянемо формулу

$$\alpha, \beta \in [0; 1], \quad \alpha + \beta = 1.$$

Ця формула дозволяє змоделювати ймовірність того, що користувач повірить у фейкову інформацію та здійснить активну дію (напр., перехід на сервіс). Параметри $D_t$ та $V_t$ відповідають за рівень довіри до джерела повідомлення та його видимість у соціальному середовищі відповідно. Значення $\alpha$ та $\beta$ задають вагу кожного чинника. Таке представлення дозволяє моделювати поведінкову залежність користувача в умовах інформаційного тиску.

Загальне навантаження на сервіс визначають як

$$L(t) = N_u \cdot p(t) \cdot r_u + N_b \cdot r_b,$$

де $N_u$ – кількість користувачів, $N_b$ – кількість ботів, $p(t)$ – імовірність активації користувачів у момент часу $(t)$, $r_u$, $r_b$ – інтенсивність запитів від користувача та бота відповідно.

Критичне навантаження, за якого сервіс відмовляє:

$$L(t) > L_{\text{крит}} \Rightarrow \text{сервіс у стані відмови.}$$



ISSN 2707-1758

Це рівняння моделює загальне навантаження на державний сервіс у конкретний момент часу. Воно враховує як запити від звичайних користувачів (параметри $N_u$, $r_u$, $p(t)$), так і від ботів, які виконують DDoS-атаки ($N_b$, $r_b$). Під час перевищення критичного порогу навантаження $L_{\text{крит}}$, система може перейти в нестабільний стан або припинити оброблення запитів. Такий підхід дозволяє виявляти порогові умови відмови в межах змодельованого сценарію.

Модель дає змогу відтворити сценарії, за яких внаслідок активності користувачів відбувається різке зростання запитів до сервісу. У табл. 2 представлено один із таких сценаріїв: у початковій фазі (0–19 хв) система функціонує стабільно, однак після активізації фейку (20–30 хв) навантаження зростає, що свідчить про ризик перевантаження навіть без технічної атаки з боку ботів.

*Таблиця 2*

**Сценарій зміни навантаження**

| Час, хвилини | Активовані користувачі, % | Запити від користувачів | Запити від ботів | Загальне навантаження | Стан сервісу |
|---|---|---|---|---|---|
| 0–19 | 12 | 1200 | 2000 | 3200 | Стабільна робота |
| 20+ | 25 | 2500 | 2000 | 4500 | Підвищене навантаження |
| 30+ | 30 | 3000 | 2000 | 5000 | Підвищене навантаження |

Наступний сценарій моделює ситуацію, у якій після початкового навантаження на сервіс відбувається активація захисної системи. Агент Az виявляє зростання трафіка та починає блокування частини джерел, що підозрюються у генерації фейкових запитів. У моделі враховується як вплив бота, так і реакція користувачів на фейкове повідомлення. Таблиця 3 ілюструє, як змінюється рівень загального навантаження та стан сервісу після втручання системи захисту.

*Таблиця 3*

**Сценарій із реакцією захисної системи**

| Час, хвилини | Активовані користувачі, % | Запити від користувачів | Запити від ботів | Заблоковані джерела | Загальне навантаження | Стан сервісу |
|---|---|---|---|---|---|---|
| 0–19 | 12 | 1200 | 2000 | 0 | 3200 | Стабільна робота |
| 20+ | 25 | 2500 | 2000 | 500 | 4000 | Підвищене навантаження |
| 30+ | 30 | 3000 | 2000 | 1000 | 4000 | Підвищене навантаження |

У базовій моделі користувачі реагують на фейкову інформацію миттєво після її отримання. Проте на практиці реакція часто є відкладеною: деякі користувачі спершу спостерігають, очікують підтвердження з інших джерел або активізуються лише після повторного контакту з фейком. З огляду на цей ефект, у модель додано сценарій із часовою затримкою активації.

Імовірність переходу користувача до активних дій описується логістичною функцією

$$p(t) = \frac{1}{1 + e^{-k(t-\tau)}},$$

де $t$ – час затримки, а $k$ – коефіцієнт швидкості реагування.

У сценарії прикладу активність користувачів поступово зростає впродовж перших 24 хв: від 10 % до 25 %. Це створює умови для поступового накопичення навантаження, яке досягає критичного рівня не відразу, а лише в кінці інтервалу. Такий сценарій дозволяє моделювати ефект уповільненого піку навантаження та необхідність більш ранньої реакції системи. Детальну динаміку навантаження в цьому сценарії представлено у табл. 4.

*Таблиця 4*

**Сценарій затриманої активації**

| Час, хвилини | Активовані користувачі, % | Запити від користувачів | Запити від ботів | Загальне навантаження | Стан сервісу |
|---|---|---|---|---|---|
| 0–9 | 10 | 1000 | 2000 | 3000 | Стабільна робота |
| 10–14 | 15 | 1500 | 2000 | 3500 | Стабільна робота |
| 15–19 | 20 | 2000 | 2000 | 4000 | Підвищене навантаження |
| 20–24 | 25 | 2500 | 2000 | 4500 | Підвищене навантаження |



ISSN 2707-1758

Ще одним сценарієм, реалізованим у моделі, є інформаційне втручання з боку держави після виявлення пікового навантаження. Офіційне спростування може бути реалізоване через повідомлення на державному порталі, офіційні сторінки у соціальних мережах або push-сповіщення у мобільному додатку. Цей сценарій враховує зміну функції активації користувачів у часі. До 20-ї хвилини ймовірність $p(t)$ зберігається на рівні 0.3, після чого знижується до 0.1 у результаті втручання. Це моделюється як кускова функція:

$$p(t) = \begin{cases} 0.3, & \text{якщо } t < 20, \\ 0.1, & \text{якщо } t \geq 20. \end{cases}$$

У моделі також враховано дію агента Az, який реалізує технічне блокування підозрілих джерел після моменту втручання. Таблиця 5 ілюструє, як змінюється навантаження: спершу воно досягає порога перевантаження, але згодом стабілізується завдяки одночасному зниженню поведінкової активності користувачів і технічному блокуванню джерел.

Зазначений підхід дозволяє враховувати не лише технічні механізми захисту, але й вплив інформаційного середовища. За даними ENISA (2023), оперативне інформування громадян і контрнаративи з боку офіційних джерел можуть суттєво знизити поведінкову реакцію на шкідливі інформаційні кампанії. Відповідно, інтеграція подібної логіки у технічні моделі є обґрунтованим кроком для сценарного аналізу гібридних загроз.

***Таблиця 5***

**Зміна активності користувачів після спростування**

| Час, хвилини | Активовані користувачі, % | Запити від користувачів | Запити від ботів | Заблоковані джерела | Загальне навантаження | Стан сервісу |
|---|---|---|---|---|---|---|
| 0–9 | 30 | 3000 | 2000 | 0 | 5000 | Стабільна робота |
| 10–14 | 30 | 3000 | 2000 | 0 | 5000 | Підвищене навантаження |
| 15–19 | 30 | 3000 | 2000 | 0 | 5000 | Підвищене навантаження |
| 20–24 | 10 | 1000 | 2000 | 500 | 2500 | Стабілізація |

**Результати**

Сценарне моделювання дозволило оцінити, як різні варіанти реагування користувачів і системи захисту впливають на навантаження державного інформаційного сервісу. Загалом змодельовано чотири сценарії: базовий без втручання, базовий із втручанням захисної системи, сценарій із затриманою реакцією користувачів і сценарій з інформаційним спростуванням.

У базовому сценарії (табл. 2) користувачі миттєво реагують на фейкове повідомлення, що призводить до поступового зростання навантаження. Навіть якщо DDoS-трафік залишається незмінним (2000 запитів), збільшення частки активованих користувачів до 30 % створює додаткове навантаження у 3000 запитів, що в сукупності становить 5000 – критичний поріг для сервісу. Це свідчить про те, що навіть без технічної атаки система може зазнати відмови внаслідок поведінкової активності легітимних користувачів.

У сценарії з активною захисною системою (табл. 3), після виявлення перевищення навантаження агент Az блокує частину джерел запитів. Блокування 1000 одиниць трафіка дозволяє знизити загальне навантаження до 4000, що стабілізує роботу сервісу. Це підтверджує, що навіть обмежене технічне реагування здатне стабілізувати систему, особливо якщо воно здійснюється до досягнення критичного рівня.

У сценарії із затриманою реакцією (табл. 4) активація користувачів відбувається не миттєво, а поступово – від 10 до 30 % упродовж 30 хв. Це веде до відкладеного накопичення навантаження: початкові фази є відносно безпечними, але у фінальній фазі сукупна кількість запитів досягає того самого рівня 5000, що й у базовому сценарії. Водночас повільне зростання навантаження ускладнює виявлення загрози для систем, які орієнтовані на миттєві DDoS-сплески. Така поведінка вимагає гнучкішого реагування захисних систем на поведінкову динаміку.

У сценарії з інформаційним втручанням (табл. 5) реакція користувачів також досягає пікових значень до 20-ї хвилини, однак подальше публічне спростування знижує $p(t)$ до 0.1. Одночасно із цим агент Az блокує частину джерел, що спричиняє швидке зменшення навантаження – з 5000 до 3000 запитів. У результаті система стабілізується навіть без потреби в глибокому технічному втручанні. Такий сценарій підтверджує практичну ефективність комбінації інформаційної протидії та точкового блокування для захисту від гібридних атак.

Порівняльний аналіз усіх чотирьох сценаріїв дозволяє зробити висновок, що найефективнішим із погляду стримування навантаження є комбінований підхід – технічне блокування джерел разом з інформаційним втручанням (табл. 5). Саме в цьому випадку вдалося знизити кількість запитів до 3000, тоді як у базовому сценарії цей показник сягав критичного рівня 5000. Навпаки, сценарій із затриманою реакцією (табл. 4) продемонстрував, що навіть без ботатаки, поступове зростання $p(t)$ може призвести до подібного навантаження. Це доводить, що не лише технічна, але й поведінкова динаміка має ключове значення для стабільності сервісу. Водночас сценарій із частковим втручанням (табл. 3) забезпечує лише часткове зниження – до 4000 запитів, що є межовим. Відповідно, саме мультикомпонентні сценарії демонструють найкращі результати з погляду підтримання стабільної роботи сервісу.

Отже, результати демонструють, що агентно-орієнтована модель дозволяє не лише імітувати поведінкові й технічні загрози, а й протестувати ефективність різних тактик реагування. Усі сценарії підтверджують, що швидкість і форма відповіді системи захисту мають критичне значення для запобігання відмові сервісу.

***Обговорення.*** Отримані результати свідчать про критичний вплив поведінкової активності користувачів на навантаження цифрових сервісів публічного сектору. У всіх сценаріях (табл. 2–5) підтверджено, що навіть без



ISSN 2707-1758

змін у DDoS-компоненті моделі (кількість ботів лишалася сталою), реакція користувачів на фейкові повідомлення здатна вивести систему за межі її працездатності. Це демонструє необхідність урахування поведінкових механізмів в оцінюванні кіберзагроз.

Зокрема, сценарії із затриманою реакцією (табл. 4) й інформаційним втручанням (табл. 5) демонструють складнішу динаміку навантаження. У першому випадку – система перебуває в зоні ризику через кумулятивне зростання запитів, у другому – стабілізації досягають лише за умови одночасного впливу інформаційного спростування та блокування джерел. Ці сценарії підтверджують, що класичні DDoS-захисти, орієнтовані на миттєві сплески, можуть бути не ефективними у випадках поведінкових хвиль або комбінованих атак.

Агентно-орієнтований підхід дозволив імітувати взаємодію різних елементів – користувачів, атакуючих, каналів поширення та систем захисту – у єдиній моделі. Завдяки цьому змодельовано не лише технічні, а й інформаційно-поведінкові впливи. Таке поєднання є особливо актуальним у контексті гібридної війни, де атака може відбуватися без жодного коду чи вірусу – лише через маніпуляцію сприйняттям.

Реалістичність змодельованих сценаріїв підтверджується на практиці. У 2022 р. в українському інформаційному просторі поширився фейк про нібито злам додатку “Дія” та витікання даних. Заклики негайно перевірити облікові записи спричинили різкий сплеск активності, який не був пов'язаний із реальним технічним оновленням. Лише після публічного спростування з боку держави навантаження нормалізувалося. Цей кейс демонструє, що навіть без атак ботів система може зазнати DDoS-подібного навантаження виключно через поведінку користувачів.

На відміну від класичних моделей DDoS-аналізу, що фокусуються виключно на технічних метриках (напр., кількість запитів, обсяг трафіка або швидкість генерації навантаження), запропонована модель враховує динаміку поведінки користувачів як окремий чинник впливу. Це дозволяє виявити сценарії, які є малопомітними в класичній аналітиці, але можуть становити не меншу загрозу для стабільності системи. Зокрема, сценарій із затриманою реакцією демонструє, що навіть за відсутності технічної атаки, поступове накопичення активності може призвести до перевантаження.

У практичному контексті це означає, що системи моніторингу, орієнтовані лише на пікові навантаження, можуть виявитися неефективними за наявності поведінкових хвиль. Агентно-орієнтована модель, навпаки, дозволяє адаптувати стратегії захисту до різних форм дестабілізації – від миттєвих фальстартів до довготривалих інформаційних кампаній.

Подібний підхід реалізовано в гібридній моделі CSI (Ruchansky, Seo, & Liu, 2017), яка інтегрує контент повідомлення, реакції користувачів і характеристики джерел. Це підтверджує ефективність багатофакторного моделювання для виявлення фейкової активності без необхідності графового аналізу.

Крім дослідницького потенціалу, модель може бути інтегрована в прикладні інструменти оцінювання ризиків і навантаження. Зокрема, вона може використовуватись в аналітичних модулях платформ типу CERT-UA для симуляцій поведінкової активності користувачів під час фейкових інформаційних кампаній. Іншим можливим напрямом є використання моделі як компонента в системах раннього попередження про аномальне навантаження – з урахуванням не лише технічних показників, а й інформаційного контексту. Крім того, модель може стати основою для сценаріїв Red Team / Blue Team навчань у сфері кібербезпеки, де важливо протестувати і технічну стійкість систем, і комунікаційну реакцію на загрозу.

Попри переваги агентно-орієнтованого підходу, запропонована модель має певні обмеження, що варто враховувати при її інтерпретації та подальшому застосуванні. По-перше, у моделюванні використано узагальнені параметри $p(t)$, які наближено відображають поведінку користувачів, але не враховують міжіндивідуальні відмінності, наприклад, вплив демографічних чи психологічних характеристик. По-друге, модель передбачає фіксовану кількість ботагентів і не враховує можливість їхнього саморозширення або адаптації, що властиво сучасним ботнетам. Крім того, інформаційне середовище змодельовано як умовно однорідне, без диференціації за типами каналів або рівнем довіри. Подальші дослідження можуть зосередитись на розширенні моделі за рахунок факторів соціального впливу, багатоканального поширення та реалізації агентів з адаптивною логікою поведінки.

Одним із факторів, який потенційно підсилює вплив фейкової інформації, є соціальний резонанс – ефект, коли користувач не реагує на повідомлення у першому контакті, але активізується після того, як аналогічну інформацію побачить у кількох джерелах. Цей ефект важко моделювати у класичних технічних системах, але його можна відтворити через багатофазну поведінку агентів. Повторне зіткнення з повідомленням, посилене авторитетом джерела або соціальним тиском, може істотно змінити траєкторію поведінки. У роботі Starbird, Arif, Wilson (2019) підкреслено, що дезінформація поширюється не лише як технічний вплив, а і як колективна діяльність аудиторії, яка неусвідомлено підтримує життєвий цикл шкідливого контенту.

У контексті моделі, що розроблена у цьому дослідженні, подібна логіка може бути реалізована завдяки розширенню функції активації $p(t)$ залежно від кількості повторів або "відлуння" повідомлення. Такий підхід дозволяє наблизити симуляцію до реального інформаційного середовища, де не лише контент, але і частота контакту з ним визначає ймовірність дії. Це є потенційним напрямом подальшого розвитку моделі – з урахуванням циклів повторення та міжагентного поширення інформації.

Крім симуляційного аналізу, запропоновану модель можна адаптувати для використання у системах моніторингу в режимі реального часу. Один із перспективних напрямів – адаптація логіки агентної моделі для систем попередження інформаційних аномалій. Це особливо актуально у випадках, коли поширення дезінформації не супроводжується класичними технічними індикаторами (напр., DDoS або мережними скануваннями), але створює високий рівень навантаження через реакцію аудиторії.

Аналіз динаміки поведінки користувачів, зокрема і різких змін у $p(t)$, дозволяє розпізнавати хвильові аномалії, які можуть указувати на запуск скоординованої фейкової кампанії. У таких випадках система може ініціювати автоматизоване спростування або блокування поширювачів – навіть до того, як



ISSN 2707-1758

з'являться технічні наслідки. Отже, модель може функціонувати не лише як інструмент для планування і навчання, але і як практичний компонент у системах кіберзахисту, де швидкість реакції має вирішальне значення. Подальші дослідження можуть бути спрямовані на визначення оптимальних порогових значень $p(t)$ для запуску таких механізмів залежно від типу інформаційної платформи та характеристик аудиторії.

Крім того, модель може бути використана для виявлення потенційно вразливих інформаційних сегментів наприклад, груп користувачів із високим $p(t)$, які найімовірніше піддадуться впливу фейкових повідомлень. Це відкриває можливості для превентивного інформування, спрямованого на посилення кіберстійкості цільових аудиторій.

Запропонована модель дозволяє не лише відтворити загрозу, але й перевірити ефективність різних тактик реагування. Це робить її корисною не лише для наукового аналізу, а й для прикладного використання – наприклад, у межах Red Team / Blue Team навчань, для оцінювання ризиків у CERT-платформах або підготовки до інформаційних кампаній. Модель також може бути розширена: з урахуванням ефекту масового повторення, затримок у поширенні або впливу ключових лідерів думок. Такі модифікації створять основу для адаптивних систем прогнозування поведінки користувачів під час криз.

**Дискусія і висновки**

Запропонована агентно-орієнтована модель виявлення та стримування гібридного впливу дозволяє комплексно моделювати взаємодію користувачів, ботагентів, джерел дезінформації та захисних систем. На відміну від класичних DDoS-моделей, вона враховує поведінкову динаміку й інформаційні ефекти, зокрема й затримку реакції користувачів і вплив офіційного спростування.

Сценарне моделювання показало, що навіть легітимна поведінка користувачів, спровокована фейком, здатна створити критичне навантаження на державний сервіс. Водночас своєчасне інформаційне втручання або точкове блокування джерел дозволяють стабілізувати ситуацію без масштабного технічного реагування. Це підкреслює важливість гібридного підходу до кіберзахисту – поєднання технічних і комунікаційних засобів реагування.

Модель узгоджується із сучасними підходами до виявлення фейкової активності: концепції багатофакторного аналізу, реалізовані в CSI, поведінкове розгалуження реакцій користувачів (Glenski et al.). Заявлене вище, а також вплив довіри до верифікаційних систем підтверджують доцільність інтегрованого моделювання поведінки, контенту та технічної динаміки.

Практична цінність моделі полягає в можливості її використання для аналізу навантаження в системах CERT, симуляцій Red Team / Blue Team, а також в оцінюванні ризиків під час дезінформаційних кампаній. Її можна адаптувати до різних конфігурацій: як для окремих сервісів (напр., електронних реєстрів), так і для масштабніших цифрових платформ.

У перспективі модель може бути доповнена модулями довіри, соціального поширення, впливу лідерів думок та алгоритмами прогнозування реакцій у часі. Це розширить її функціональність і дозволить використовувати як компонент в адаптивних системах інформаційного захисту.

Отже, ключовим елементом у формуванні ризику перевантаження виявляється не сам факт атаки, а форма сприйняття та реакція аудиторії. Навіть мінімальна технічна активність, поєднана з панікою або дезінформуванням, може мати наслідки, еквівалентні повномасштабній DDoS-атаці. Це підкреслює необхідність урахування поведінкових чинників до моделей ризику в інформаційній безпеці.

**Джерела фінансування.** Це дослідження не отримало жодного гранту від фінансової установи в державному, комерційному або некомерційному секторах.

**Список використаних джерел**

Matteo, C., Walter, Q., Alessandro, G., Valensise, Carlo, M., Emanuele, B., Schmidt, Ana, L., Paola, Z., Zollo F. & Scala, A. (2020). The COVID-19 social media infodemic. *Scientific Reports*, 10, 16598. https://doi.org/10.1038/s41598-020-73510-5

Dhiman, D., Shah, R., Thirunarayan, K. (2024). Detecting fake news using a hybrid GBERT model. *Procedia Computer Science,* 217, 1427–1436. https://doi.org/10.1016/j.procs.2022.12.341

ENISA. (2023). *Threat Landscape for Disinformation Campaigns.* European Union Agency for Cybersecurity. https://www.enisa.europa.eu/publications/enisa-threat-landscape-for-disinformation-campaigns

Glenski, M., Weninger, T., & Volkova, S. (2018). Identifying and understanding user reactions to deceptive and trusted social news sources. In K. Bharat, R. Jim, & G. Kathleen (Eds.). *Proceedings of the 27th ACM International Conference on Information and Knowledge Management* (pp. 1063–1072). Association for Computing Machinery. https://doi.org/10.1145/3269206.3271774

Kotenko, I., Stepashkin, M., & Doynikova, E. (2008). Agent-based modeling and simulation of botnets and botnet defense. In A. Smith, & B. Johnson (Eds.). *Proceedings of the 2008 16th Euromicro Conference on Parallel, Distributed and Network-Based Processing* (pp. 71–78). IEEE Press. https://doi.org/10.1109/PDP.2008.71

David M. J., Lazer, Matthew A., Baum, Yochai, B., Adam J. B., Kelly M. Greenhill, F. M., Miriam J. Metzger, B., N., Gordon P., & Jonathan L., Z. (2018). The science of fake news. *Science*, 359(6380). https://doi.org/10.1126/science.aao2998

Ruchansky, N., Seo, S., & Liu, Y. (2017). CSI: A hybrid deep model for fake news detection. In J. Doe, & A. Smith (Eds.). *Proceedings of the 2017 ACM Conference on Information and Knowledge Management* (pp. 797–806). Association for Computing Machinery. https://doi.org/10.1145/3132847.3132877

Shu, K., Sliva, A., Wang, S., Tang, J., Liu, H. (2019). Fake news detection on social media: A data mining perspective. *SIGKDD Explorations*, 19(1), 22–36. https://doi.org/10.1145/3137597.3137600

Starbird, K., Arif, A., Wilson, T. (2019). Disinformation as collaborative work: Surfacing the participatory nature of strategic information operations. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW). https://doi.org/10.1145/3359229

**References**

Matteo, C., Walter, Q., Alessandro, G., Valensise, Carlo, M., Emanuele, B., Schmidt, Ana, L., Paola, Z., Zollo F. & Scala, A. (2020). The COVID-19 social media infodemic. *Scientific Reports*, 10, 16598. https://doi.org/10.1038/s41598-020-73510-5

Dhiman, D., Shah, R., Thirunarayan, K. (2024). Detecting fake news using a hybrid GBERT model. *Procedia Computer Science,* 217, 1427–1436. https://doi.org/10.1016/j.procs.2022.12.341

ENISA. (2023). *Threat Landscape for Disinformation Campaigns.* European Union Agency for Cybersecurity. https://www.enisa.europa.eu/publications/enisa-threat-landscape-for-disinformation-campaigns

Glenski, M., Weninger, T., & Volkova, S. (2018). Identifying and understanding user reactions to deceptive and trusted social news sources. In K. Bharat, R. Jim, & G. Kathleen (Eds.). *Proceedings of the 27th ACM International Conference on Information and Knowledge Management* (pp. 1063–1072). Association for Computing Machinery. https://doi.org/10.1145/3269206.3271774

Kotenko, I., Stepashkin, M., & Doynikova, E. (2008). Agent-based modeling and simulation of botnets and botnet defense. In A. Smith, & B. Johnson (Eds.). *Proceedings of the 2008 16th Euromicro Conference on Parallel, Distributed and Network-Based Processing* (pp. 71–78). IEEE Press. https://doi.org/10.1109/PDP.2008.71

David M. J., Lazer, Matthew A., Baum, Yochai, B., Adam J. B., Kelly M. Greenhill, F. M., Miriam J. Metzger, B., N., Gordon P., & Jonathan L., Z. (2018). The science of fake news. *Science*, 359(6380). https://doi.org/10.1126/science.aao2998

Ruchansky, N., Seo, S., & Liu, Y. (2017). CSI: A hybrid deep model for fake news detection. In J. Doe, & A. Smith (Eds.). *Proceedings of the 2017 ACM Conference on Information and Knowledge Management*



ISSN 2707-1758

(pp. 797–806). Association for Computing Machinery. https://doi.org/10.1145/3132847.3132877

Shu, K., Sliva, A., Wang, S., Tang, J., Liu, H. (2019). Fake news detection on social media: A data mining perspective. *SIGKDD Explorations*, 19(1), 22–36. https://doi.org/10.1145/3137597.3137600

Starbird, K., Arif, A., Wilson, T. (2019). Disinformation as collaborative work: Surfacing the participatory nature of strategic information operations. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW). https://doi.org/10.1145/3359229

**Отримано редакцією журналу / Received: 01.06.25**
**Прорецензовано / Revised: 27.06.25**
**Схвалено до друку / Accepted: 30.06.25**

**Dmytro DZHENDZHERO, PhD Student**
**ORCID ID: 0009-0006-7394-4995**
**e-mail: dzhendzherod@gmail.com**
**Taras Shevchenko National University of Kyiv, Kyiv, Ukraine**

# A METHOD FOR MODELING HYBRID INFLUENCE ON GOVERNMENT SERVICES USING AN AGENT-BASED APPROACH

**Background.** ***In the context of hybrid warfare, the role of information influence on critical infrastructure is growing. Fake news can trigger waves of user behavioral activity that mimic DDoS attacks, overloading government digital services even without direct technical interference. This creates the need to model such scenarios considering both technical and socio-behavioral factors.***

**Methods.** ***An agent-based model of hybrid influence is proposed, including users, bots, fake news sources, and protective mechanisms. The model is implemented using mathematical expressions and scenario simulation, including a baseline scenario, delayed response scenario, technical intervention scenario, and official rebuttal scenario. The analysis considers system load parameters, behavioral activation probability, and the system's response.***

**Results.** ***It was found that user behavior alone can generate critical system load equivalent to a DDoS attack. The most effective scenario combined technical blocking and information rebuttal, reducing load to a stable level. Classic cybersecurity models often fail to account for delayed reactions or wave-like behavioral patterns, limiting their effectiveness in hybrid environments.***

**Conclusions.** ***The proposed model enables not only simulation of hybrid influence but also evaluation of response tactics. It can be applied in CERT platforms, Red Team / Blue Team exercises, and early-warning systems. Further research may focus on adapting the model to real-world environments and developing automated response mechanisms to information threats.***

**Keywords:** ***hybrid warfare, agent-based modeling, fake news, digital security, behavioral load.***

Автор заявляє про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

The author declares no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; in the decision to publish the results.



ISSN 2707-1758