УДК 004.056.5

Я.Б Бобрович
В.В Карпінець

МАТЕМАТИЧНА ТА АЛГОРИТМІЧНА МОДЕЛЬ ПРОГРАМНОГО МОДУЛЯ ГЕНЕРАТИВНОЇ ЛІНГВІСТИЧНОЇ СТЕГАНОГРАФІЇ НА БАЗІ МАЛИХ МОВНИХ МОДЕЛЕЙ

Вінницький національний технічний університет

Анотація

У роботі розглядається математична та алгоритмічна модель побудови прихованих каналів передачі даних за допомогою генеративної лінгвістичної стеганографії. Обґрунтовано перехід від класичних криптографічних підходів до імовірнісного керування авторегресійною генерацією на базі малих мовних моделей (SLM), зокрема Phi-3 Mini. Запропонована модель інтегрує оптимізацію розподілів (OD-Stega) за допомогою Лагранжевого формалізму для максимізації ентропійної ємності з жорстким контролем дивергенції Кульбака–Лейблера. Для усунення проблеми токенізаційної невідповідності (Tokenization Inconsistency) алгоритмів субслівного розбиття впроваджено метод покрокової верифікації (Stepwise Verification). Доведено, що такий підхід гарантує абсолютну бієктивність ентропійного мапінгу та асимптотично зводить ефективність нейромережевого стегоаналізу до рівня випадкового вгадування.

Ключові слова: генеративна стеганографія, оптимізація розподілів, OD-Stega, арифметичне кодування, токенізаційна невідповідність, малі мовні моделі, аналітика поведінки користувачів (UEBA).

Abstract

This paper examines a mathematical and algorithmic model for constructing covert data transmission channels using generative linguistic steganography. It substantiates the transition from classical cryptographic approaches to probabilistic control of autoregressive generation based on small language models (SLMs), in particular Phi-3 Mini. The proposed model integrates distribution optimization (OD-Stega) using the Lagrangian formalism to maximize entropy capacity while maintaining strict control over Kullback–Leibler divergence. To address the problem of tokenization inconsistency in subword segmentation algorithms, a Stepwise Verification method is introduced. It is demonstrated that this approach guarantees absolute bijectivity of entropy mapping and asymptotically reduces the effectiveness of neural network-based steganalysis to the level of random guessing.

Keywords: generative steganography, distribution optimization, OD-Stega, arithmetic coding, tokenization inconsistency, small language models, user and entity behavior analytics (UEBA).

Вступ

Глобальна цифровізація та повсюдне впровадження архітектури нульової довіри радикально трансформували методики інспекції мережевого трафіку [1]. Сучасні платформи виявлення мережевих загроз та системи аналітики поведінки користувачів і сутностей відійшли від статичного сигнатурного аналізу, зосередившись на виявленні статистичних і поведінкових аномалій [2]. У таких умовах класична інкапсуляція даних у зашифровані бінарні контейнери стає неефективною: наближення зашифрованого масиву до ідеально випадкової послідовності генерує максимальну ентропію Шеннона, що фіксується системами глибокого аналізу як індикатор компрометації [3].

З огляду на це, актуальним стає перехід до генеративної лінгвістичної стеганографії на базі малих мовних моделей [4]. Використання компактних архітектур, таких як Phi-3 Mini, дозволяє локалізувати процес генерації, уникаючи мережевих слідів від звернень до зовнішніх API, та мінімізувати обчислювальні накладні витрати [5]. На відміну від класичної синонімічної заміни, яка руйнує дистрибутивну семантику тексту [6], генеративний підхід формує стего-текст безпосередньо в процесі авторегресійного вибору токенів, зберігаючи природний імовірнісний профіль [7].

Результати дослідження

Основою генеративної стеганографії є керований дискретний стохастичний процес [7]. На кожному часовому кроці мовна модель генерує вектор логітів для словника токенів V, який після

застосування функції Softmax з параметром температури T трансформується у природний умовний розподіл P [7].

Пряме використання розподілу P критично обмежує пропускну здатність прихованого каналу. Тому обчислювальне ядро моделі застосовує метод Optimized Distributions Steganography (OD-Stega), який зводиться до задачі опуклого програмування: необхідно максимізувати ентропію стего-розподілу $H(Q)$ для збільшення обсягу вбудованих даних, не виходячи за межі статистичної непомітності, що контролюється через дивергенцію Кульбака–Лейблера $D_{KL}(Q \parallel P) \leq$ [7].

Для програмного знаходження оптимального розв'язку використовується метод множників Лагранжа. Функція Лагранжа для цієї системи набуває вигляду:

$$\mathcal{L}(Q, \lambda, u, \omega) = -\sum_{i=1}^{N} Q_i \log_2 Q_i + u\left(\sum_{i=1}^{N} Q_i \log_2 \quad \frac{Q_i}{P_i} - \delta\right) + \lambda\left(\sum_{i=1}^{N} Q_i - 1\right) - \sum_{i=1}^{N} \omega_i Q_i$$

Аналітичне розв'язання цієї задачі через умови Каруша–Куна–Таккера дозволяє динамічно перераховувати ймовірності для кожного токена [8]. Вплив множників на оптимізацію простору зведено у таблицю 1.

Таблиця 1 – Множники Лагранжа для задачі оптимізації OD-Stega та їх алгоритмічна роль

Множник	Зв'язане математичне обмеження	Функціональне призначення в алгоритмі керування
$u \geq 0$	$D_{KL}(Q \parallel P) \leq$	Динамічний балансир між ентропійною ємністю та семантичною природністю тексту.
$\lambda \in R$	$\sum_{i=1}^{N} Q_i = 1$	Забезпечує строге нормування вихідного розподілу перед етапом кодування.
$\omega_i \geq 0$	$Q_i \geq 0$	Запобігає виникненню від'ємних імовірностей у циклі авторегресії.

Оптимізація цього простору гарантує, що при мінімізації відхилення δ до конфігураційних значень (наприклад, 0.05), повна варіаційна відстань між природним та стего-текстом звужується настільки, що ймовірність успішного виявлення каналу системами UEBA асимптотично наближається до 0.5 (рівень випадкового вгадування) [7].

Оптимальний розподіл Q виступає базисом для вбудовування секретного повідомлення M через арифметичне кодування (Arithmetic Coding) [7]. Субслівні межі є динамічними і залежать від контексту. Якщо згенерований токен на стороні отримувача "зливається" з сусіднім (Token Fusion) або розщеплюється (Token Fission), послідовність L' не збігається з еталоном $L_o \oplus w$ [9]. Оскільки арифметичне декодування є чутливим до найменшого зсуву, одинична помилка парсингу викликає лавиноподібну десинхронізацію всього каналу.

Для математичного гарантування бієктивності ($f^{-1}(f(M)) = M$) в архітектуру імплементовано алгоритм покрокової верифікації (Stepwise Verification) [10]. Обчислювальне ядро здійснює Top-K фільтрацію розподілу Q і проводить віртуальну інспекцію кожного кандидата перед його вбудовуванням.

Таблиця 2 – Логічна матриця алгоритму покрокової верифікації (Stepwise Verification)

Фаза інспекції	Математична модель	Аналітична інтерпретація та дія
Look-ahead симуляція	$s_{temp} = \text{Detok}(L_o \oplus w)$	Формування тимчасового рядка, що імітує проходження тексту через відкриті канали.
Інверсна перевірка	$L' = \text{Tok}(s_{temp})$	Повторний парсинг рядка через алгоритм BPE мовної моделі.
Прийняття рішення	$L' \neq L_o \oplus w$	Маркування кандидата як деструктивного. Його ймовірність обнуляється.

Цей структурний фільтр створює так звану "ентропійну вартість бієктивності" ($\Delta H = H(Q) - H(Q')$), що призводить до незначної втрати пропускної здатності (у середньому $0.15 - 0.25$ біт/токен). Однак цей компроміс повністю усуває ризики десинхронізації, дозволяючи декодеру зі стовідсотковою точністю вилучити секретні дані шляхом аналізу збігу старших бітів регістрів [10].

Висновки

Розроблена математична та алгоритмічна модель програмного модуля доводить можливість безпечного транзиту конфіденційної інформації в умовах жорсткого поведінкового моніторингу архітектури нульової довіри (Zero Trust Architecture). Застосування малих мовних моделей (SLM) у комбінації з алгоритмами оптимізації розподілів (OD-Stega) дозволяє формувати семантично зв'язні текстові повідомлення, інформаційна ємність яких алгоритмічно оптимізована під встановлені межі статистичної непомітності.

Ключовим досягненням розробки є імплементація алгоритму покрокової верифікації (Stepwise Verification), що повністю усуває вразливості субслівних токенізаторів (BPE). Це інженерне рішення перетворює стохастичну авторегресійну генерацію тексту на детермінований і надійно оборотний криптографічний транспорт. Теоретичний та алгоритмічний аналіз показав, що запропонований архітектурний підхід звужує повну варіаційну відстань між стего-текстом та природною мовою настільки, що ефективність виявлення прихованого каналу сучасними системами інспекції трафіку (NDR/UEBA) стає статистично невідрізненною від випадкового вгадування.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Deep Packet Inspection Using AI for Threat Detection [Електронний ресурс] / ResearchGate // ResearchGate – 2026 – Режим доступу до ресурсу: https://www.researchgate.net/publication/391857632_Deep_Packet_Inspection_Using_AI_for_Threat_Detection
2. User and Entity Behavior Analytics (UEBA) Reference [Електронний ресурс] / Microsoft // Microsoft Learn – 2026 – Режим доступу до ресурсу: https://learn.microsoft.com/en-us/azure/sentinel/ueba-reference
3. A Mathematical Theory of Communication [Електронний ресурс] / C. E. Shannon // The Bell System Technical Journal – 1948 – Режим доступу до ресурсу: https://people.math.harvard.edu/~ctm/home/text/others/shannon/entropy/entropy.pdf
4. Exploiting Language Model for Efficient Linguistic Steganography [Електронний ресурс] / ArXiv // arXiv preprint – 2024 – Режим доступу до ресурсу: https://arxiv.org/pdf/2409.01780
5. Phi-3 Technical Report: A Highly Capable Language Model Locally on Your Phone [Електронний ресурс] / Microsoft // Microsoft Research – 2024 – Режим доступу до ресурсу: https://www.microsoft.com/en-us/research/publication/phi-3-technical-report-a-highly-capable-language-model-locally-on-your-phone/
6. A Comprehensive Survey on Linguistic Steganography [Електронний ресурс] / ResearchGate // ResearchGate – 2025 – Режим доступу до ресурсу: https://www.researchgate.net/publication/398488225_A_Comprehensive_Survey_on_Linguistic_Steganography_Methods_Countermeasures_Evaluation_and_Challenges
7. Optimized Distributions Steganography (OD-Stega) [Електронний ресурс] / ACL Anthology // EACL – 2026 – Режим доступу до ресурсу: https://aclanthology.org/2026.eacl-long.36
8. Lagrange Multipliers in Generative Steganography Optimization [Електронний ресурс] / ArXiv // arXiv preprint – 2025 – Режим доступу до ресурсу: https://arxiv.org/abs/2508.20718
9. Towards Robust Generative Steganography: Tokenization Inconsistency [Електронний ресурс] / ArXiv // arXiv preprint – 2024 – Режим доступу до ресурсу: https://arxiv.org/html/2410.04328v1
10. Stepwise Verification for Bijective Generative Steganography [Електронний ресурс] / ACL Anthology // EMNLP – 2025 – Режим доступу до ресурсу: https://aclanthology.org/2025.emnlp-main.361.pdf

Бобрович Ярослав Борисович - студент групи 2КІТС-22б, факультет менеджменту та інформаційної безпеки, Вінницький національний технічний університет, м. Вінниця, e-mail: yaroslavbobr987@gmail.com

Науковий керівник: ***Карпінець Василь Васильович*** - кандидат технічних наук, доцент, завідувач кафедри менеджменту та безпеки інформаційних систем, Вінницький національний технічний університет, м. Вінниця, e-mail: karpinets@vntu.edu.ua

Bobrovych Yaroslav B. – student of group 2KITS-22b, Faculty of Management and Information Security, Vinnytsia National Technical University, Vinnytsia, e-mail: yaroslavbobr987@gmail.com

Supervisor*:* ***Karpinets Vasyl V.*** – PhD, Associate Professor, Head of the Department of Management and Security of Information Systems, Vinnytsia National Technical University, Vinnytsia, e-mail: karpinets@vntu.edu.ua