УДК 004.056.55

Є. В. Котелянець
В. В. Карпінець

ОЦІНКА ЕФЕКТИВНОСТІ ГЕНЕРАТИВНОЇ СТЕГАНОГРАФІЇ В ЛАТЕНТНОМУ ПРОСТОРІ У ПОРІВНЯННІ З КЛАСИЧНИМИ МЕТОДАМИ ВБУДОВУВАННЯ ДАНИХ

Вінницький національний технічний університет

Анотація

У роботі проведено порівняльний аналіз традиційних методів приховування інформації в цифрових зображеннях (LSB, DCT, DWT) та генеративних підходів на основі латентного простору. Розглянуто проблему вразливості класичних методів до сучасних систем нейромережевого стегоаналізу. Визначено поняття латентного простору та описано механізми функціонування генеративної стеганографії на базі дифузійних моделей. На основі аналізу літературних джерел показано, що методи з використанням латентного простору забезпечують вищу стеганографічну ємність, кращу візуальну якість, підвищену стійкість до канальних спотворень та більшу безпеку щодо машинного виявлення порівняно з традиційними підходами.

Ключові слова: Генеративна стеганографія, латентний простір, LSB, DCT, DWT, стеганографія без вбудовування (SWE), стегоаналіз.

Abstract

The paper presents a comparative analysis of traditional information hiding methods in digital images (LSB, DCT, DWT) and generative approaches based on latent space. The vulnerability of classical methods to modern neural network steganalysis systems is considered. The concept of latent space is defined, and the mechanisms of generative steganography based on diffusion models are described. Based on the analysis of literature sources, it is shown that methods using latent space provide higher steganographic capacity, better visual quality, increased robustness to channel distortions, and greater security against machine detection compared to traditional approaches.

Keywords: Generative steganography, latent space, LSB, DCT, DWT, steganography without embedding (SWE), steganalysis.

Вступ

Стрімка глобальна цифровізація сформувала запит на надійні системи захисту конфіденційної інформації. Традиційні криптографічні методи здатні лише зашифрувати вміст повідомлення, однак сам факт передачі зашифрованих даних залишається відкритим для систем моніторингу мережі. Цифрова стеганографія вирішує цю проблему шляхом приховування самого факту комунікації у звичайних медіафайлах.

Історично розвиток алгоритмів стеганографії зображень базувався на модифікації існуючих файлів-контейнерів. Однак сучасні системи нейромережевого стегоаналізу здатні виявляти статистичні аномалії, які неминуче виникають при такому вбудовуванні. Це робить традиційні методи вразливими та недостатньо ефективними для забезпечення прихованої комунікації.

Альтернативним підходом є генеративна стеганографія, де контейнер створюється «з нуля» одночасно з вбудовуванням секретних даних. Особливо перспективними є методи, що функціонують у латентному просторі – стисненому математичному вимірі ознак, отриманому за допомогою варіаційних автокодувальників (VAE). Робота дифузійних моделей, зокрема Stable Diffusion, у латентному просторі дозволяє інтегрувати приховану інформацію на етапі синтезу зображення, що унеможливлює пряме порівняння з оригінальним контейнером.

Метою даного дослідження є проведення порівняльного аналізу традиційних методів приховування інформації (LSB, DCT, DWT) та генеративних підходів на основі латентного простору за такими критеріями: стеганографічна ємність, візуальна якість, стійкість до канальних спотворень, обчислювальна ефективність та безпека щодо машинного виявлення.

Результати дослідження

Традиційні методи приховування інформації в цифрових зображеннях концептуально базуються на підході прямої модифікації попередньо існуючого зображення-контейнера. Найвідомішим просторовим алгоритмом є метод заміни найменш значущого біта (LSB), який функціонує шляхом безпосередньої заміни останніх (молодших) бітів інтенсивності пікселя на бінарні значення секретного повідомлення [1]. Хоча метод LSB забезпечує обчислювальну простоту та візуальну непомітність для ока, він є вкрай крихким: будь-яка компресія формату JPEG чи легке масштабування безповоротно знищує приховані дані [2]. Для подолання цієї крихкості розроблено частотні алгоритми, такі як дискретне косинусне перетворення (DCT) та дискретне вейвлет-перетворення (DWT). Наприклад, метод DCT розбиває зображення на матричні блоки і впроваджує дані у середньочастотні коефіцієнти [1]. Ці методи значно краще протистоять стисненню, проте зіштовхуються з жорсткими математичними обмеженнями обсягу даних, які можна сховати [2]. Головним фундаментальним недоліком усіх без винятку методів традиційного приховування є математична неминучість порушення природної статистики зображення. Сучасні нейромережеві системи стегоаналізу (такі як Spatial Rich Models та CNN-аналізатори) виокремлюють ці структурні аномалії з точністю понад 95%, що робить традиційні методи непридатними для серйозного захисту даних у сучасних умовах [3].

На противагу звичайному вбудовуванню, інноваційний підхід реалізується через генеративне приховування інформації безпосередньо в латентному просторі, застосовуючи концепцію стеганографії без вбудовування (SWE), яка дозволяє ховати повідомлення без використання оригінального зображення-контейнера [4]. Для розуміння цього процесу необхідно детально пояснити, що таке латентний простір. Це висококомпресований, абстрактний математичний вимір, у якому штучний інтелект зберігає лише найважливіші, семантичні та структурні властивості даних, повністю відкидаючи зайвий візуальний шум. Наприклад, звичайне RGB-зображення розміром 512 на 512 пікселів стискається варіаційним автокодувальником (VAE) у десятки разів до компактного латентного тензора розміром 64 на 64 [5]. У цьому стисненому просторі алгоритми не бачать конкретних пікселів, вони оперують сутностями та їхніми взаємозв'язками. З математичної точки зору, процеси в латентному просторі керуються стохастичними рівняннями. Прямий процес дифузії на кроці t визначається формулою:

$$x_t = \sqrt{\alpha_t} x_0 + \sqrt{1 - \alpha_t \epsilon_t}$$

де x_0 – початковий чистий сигнал, α_t– коефіцієнт дисперсії, а ϵ_t– стандартний гаусівський шум [6]. Процес приховування інформації відбувається під час генерації зображення «з нуля», де під керуванням варіаційних автокодувальників (VAE) секретне повідомлення ідеально інкапсулюється саме у вектор стартового шуму x_T або тензорні ознаки. Коли штучний інтелект розгортає цей стиснений тензор у повноцінне піксельне зображення, секретні дані стають невіддільною, природною частиною згенерованої картинки, не залишаючи жодних слідів редагування [4].

Безпосереднє порівняння цих двох підходів демонструє суттєві відмінності між ними за ключовими метриками. Розглядаючи стеганографічну ємність, звичайні частотні та просторові методи швидко досягають своєї межі, дозволяючи надійно сховати лише 1–3 біти інформації на один піксель (bpp) [1]. Натомість багатовимірне кодування під час генерації в латентному просторі забезпечує ефективну ємність близько 0.06 біт на піксель (до 16 КБ на зображення) [2]. Щодо візуальної якості, при спробах збільшити обсяг вбудованих даних традиційними алгоритмами зображення стрімко покривається артефактами. Ця деградація об'єктивно вимірюється за допомогою метрики пікового відношення сигнал/шум (PSNR), формула якої має вигляд [1]:

$$PSNR = 10 \cdot \log_{10}\left(\frac{MAX^2}{MSE}\right)$$

де MAX – максимальне значення яскравості пікселя, а MSE – середньоквадратична помилка. У той час як для класичних методів при високому навантаженні значення PSNR критично падає, зображення, згенеровані з латентного простору, забезпечують стабільну фотореалістичність із показниками PSNR понад 57 дБ (наприклад, 58.24 дБ) та надвисоким індексом структурної подібності SSIM [8]. Метрика SSIM, що оцінює структурну цілісність, розраховується як [1]:

$$SSIM(X,\hat{X}) = \frac{(2\mu_x\mu_{\hat{x}} + C_1)(2\sigma_{x\hat{x}} + C_2)}{(\mu_x^2 + \mu_{\hat{x}}^2 + C_1)(\sigma_x^2 + \sigma_{\hat{x}}^2 + C_2)}$$

де μ та σ^2 – середні значення та дисперсії відповідних зображень. Моделі типу Stable Diffusion стабільно демонструють значення SSIM на рівні 0.99, що підтверджує ідеальну візуальну якість [8].

Порівняння стійкості (робастності) до канальних атак виявляє ще одну фундаментальну перевагу. Під час передачі через мережу Інтернет зображення неминуче піддаються компресії чи обрізанню. Алгоритми LSB повністю руйнуються від таких завад [2]. У системах латентного простору варіаційний автокодувальник діє як потужний математичний фільтр: він розпізнає ці пошкодження каналу передачі як несемантичний шум і відфільтровує їх під час декодування, що дозволяє відновити дані з високою точністю (за результатами окремих досліджень – до 98%) [4]. Іншим критично важливим аспектом є обчислювальна ефективність. Оскільки генеративні моделі переносять обчислення у стиснений вимір (працюючи з тензорами меншої розмірності замість повноцінних піксельних масивів), операції в латентному просторі є до 20 разів швидшими порівняно з піксельними методами, займаючи близько 3 мілісекунд на зображення [4]. Нарешті, вирішальним критерієм є безпека від виявлення: змінені файли методів традиційного приховування легко детектуються. На противагу цьому, стегоконтейнери з латентного простору статистично близькі до природних зображень, через що показник їх успішного виявлення (AUC) наближається до 0.5, що відповідає випадковому вгадуванню [8].

Висновки

Проведений порівняльний аналіз показує, що традиційні методи приховування інформації в існуючих цифрових контейнерах (алгоритми LSB, DCT, DWT) є недостатньо ефективними перед викликами сучасного світу. Швидкий розвиток систем машинного стегоаналізу зробив таємну комунікацію за допомогою класичних методів вкрай небезпечною, адже будь-яке пряме втручання в піксельну чи частотну структуру залишає сліди і гарантовано розпізнається з точністю понад 95% [3]. Крім того, традиційні алгоритми страждають від критичних технологічних вразливостей: вкрай обмеженої ємності, швидкої втрати візуальної якості та слабкої стійкості до геометричних спотворень під час передачі через інтернет [2]. Нова епоха прихованих комунікацій ґрунтується на інноваційному підході генерації в латентному просторі. Використання стисненого семантичного виміру для інтеграції даних на етапі створення зображення усуває класичний компроміс між обсягом інформації та її непомітністю. Цей метод забезпечує високу стійкість до статистичного аналізу (детектування наближається до випадкового вгадування з AUC ≈ 0.5 [8], гарантує високу ємність передачі даних (до 16 КБ на зображення, що відповідає ≈0.06 біт/піксель) з можливістю мульти-приховування, стійкість до канального шуму (точність відновлення > 98%), високу обчислювальну швидкість генерації та високу якість згенерованих зображень (PSNR > 57 дБ, SSIM ≈ 0.99) [8]. З огляду на результати порівняння, перехід від модифікації файлів до приховування інформації в латентному просторі генеративних систем є перспективним та науково обґрунтованим напрямом для забезпечення надійного кіберзахисту в сучасних умовах.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. High-Resolution Image Synthesis with Latent Diffusion Models URL: https://arxiv.org/abs/2112.10752 (дата звернення: 01.04.2026)
2. LDStega: Practical and Robust Generative Image Steganography based on Latent Diffusion Mo URL: https://www.researchgate.net/publication/385306369_LDStega_Practical_and_Robust_Generative_Image_Steganography_based_on_Latent_Diffusion_Models (дата звернення: 01.04.2026)
3. Deep Learning-Based Image Steganography with Latent Space Embedding and Smart Decoder Selection URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC12731544/ (дата звернення: 01.04.2026)
4. Double-Flow-Based Steganography Without Embedding for Image-to-Image Hiding URL: https://www.mdpi.com/2079-9292/14/21/4270 (дата звернення: 02.04.2026)
5. Denoising Diffusion Implicit Models URL: https://arxiv.org/abs/2010.02502 (дата звернення: 03.04.2026)
6. StyleGAN2-Stego: Secure Coverless Image Steganography via Latent Space Encoding URL: https://eastpublication.com/index.php/ejcs/article/view/188 (дата звернення: 03.04.2026)
7. Comparison Between DCT and DWT Steganography Algorithms. URL: https://www.ijaist.com/wp-

content/uploads/2018/08/ComparisonBetweenDCTandDWTSteganographyAlgorithms.pdf (дата звернення: 05.04.2026)
8. A Comprehensive Survey of Digital Image Steganography and Steganalysis URL: https://www.emerald.com/atsip/article/13/1/1/1331386/A-Comprehensive-Survey-of-Digital-Image (дата звернення: 05.04.2026)

Котелянець Євгеній В'ячеславович **–** студент групи 1КІТС-22б, Факультет менеджменту та інформаційної безпеки, Вінницький національний технічний університет, м. Вінниця, e-mail: evgen.kotelyanecz@gmail.com
Карпінець Василь Васильович **–** завідувач кафедри МБІС, кандидат технічних наук, доцент, Вінницький національний технічний університет, м. Вінниця, e-mail: karpinets@vntu.edu.ua

Kotelyanets Evgeniy V. – student of group 1KITS-22b, Faculty of Management and Information Security, Vinnytsia National Technical University, Vinnytsia, e-mail: evgen.kotelyanecz@gmail.com.
Karpinets Vasyl V. – Head of the Department of Management and Security of Information Systems, Candidate of Technical Sciences, Associate Professor, Vinnytsia National Technical University, Vinnytsia, e-mail: karpinets@vntu.edu.ua