УДК 004.056.55

Заверуха О. А.
Салієва О. В.

ПІДВИЩЕННЯ СТІЙКОСТІ СТЕГАНОГРАФІЧНИХ СИСТЕМ ДО ЗЛОВМИСНИХ АТАК

Вінницький національний технічний університет

Анотація. *У роботі досліджено актуальну проблему підвищення стійкості стеганографічних систем до атак та спотворень. Запропоновано новий підхід до генерації стегоконтейнерів, що базується на використанні дифузійних імовірнісних моделей (Diffusion Models). На відміну від методів, що базуються на генеративно-змагальних мережах (GAN), запропонований підхід забезпечує вищу якість, різноманітність генерованих даних та, що найважливіше, підвищену стійкість вбудованого повідомлення до таких атак, як стиснення з втратами, додавання шуму та фільтрація.*

Ключові слова: стеганографія, дифузійні моделі, стійкість стегоконтейнерів, приховування інформації, кібербезпека, глибоке навчання.

Abstract. *This paper investigates the urgent problem of increasing the robustness of steganographic systems against attacks and distortions. A new approach to stegocontainer generation is proposed, based on the use of diffusion probabilistic models (Diffusion Models). Unlike methods based on generative adversarial networks (GANs) [1], the proposed approach provides higher quality, greater diversity of generated data, and, most importantly, enhanced robustness of the embedded message to attacks such as lossy compression, noise addition, and filtering.*

Keywords: steganography, diffusion models, stegocontainer robustness, information hiding, cybersecurity, deep learning

Вступ

Забезпечення конфіденційності інформації в сучасних мережах залишається критичним завданням. Традиційне шифрування захищає вміст, однак сам факт його використання привертає увагу. Стеганографія вирішує цю проблему шляхом приховування факту передачі даних.

Результати досліджень ефективності генеративних моделей у стеганографії фокусувалися на генеративно-змагальних мережах (GAN) [1]. Попри високу якість генерованих контейнерів, GAN-системи продемонстрували суттєві недоліки: нестабільність навчання та, що найважливіше, низьку робастність (стійкість) прихованих даних. Вбудовані повідомлення часто руйнуються під дією поширених операцій обробки даних, таких як стиснення з втратами (MP3, JPEG) або застосування фільтрів.

Для вирішення зазначених проблем у даній роботі пропонується здійснити перехід від GAN до більш досконалих генеративних архітектур – дифузійних моделей (DMs) [2-3], які демонструють надзвичайні результати у генерації високоякісних даних, вирішуючи проблеми нестабільності GAN. У цьому дослідженні дифузійні моделі вперше застосовано цілеспрямовано для створення стійких стегоконтейнерів, що відкриває перспективи для розробки стеганографічних систем нового покоління.

Результати досліджень

На відміну від GAN [1], де генератор змагається з дискримінатором, дифузійні моделі [2] працюють за іншим механізмом. Вони складаються з двох процесів: прямого (поступове зашумлення даних) та зворотного (покрокове відновлення даних із шуму).

В основі запропонованого методу лежить умовна генерація на базі зворотного процесу дифузії. Нейронна мережа, зазвичай архітектури U-Net [4], навчається відновлювати сигнал із шуму. Основна ідея полягає у тому, що секретне повідомлення M виступає як додаткова умова (condition) для нейронної мережі на кожному кроці відновлення $p_\theta(x_{t-1}, x_t, M)$.

Таким чином, модель навчається генерувати стегоконтейнер, в якому прихована інформація M є невід'ємною, "вродженою" частиною самої структури даних, а не поверхнево доданим артефактом (як в методах LSB). Це дозволило інтегрувати біти повідомлення у найбільш значущі компоненти сигналу, перцептивна важливість яких є мінімальною.

Експериментальне дослідження проводилося на аудіо-наборах даних (TIMIT, UME) та продемонструвало наступні результати:

1. Якість генерації. Показники якості (SNR, PESQ [5]) для стегоконтейнерів, згенерованих дифузійною моделлю, підтвердили високу перцептивну якість згенерованих аудіофайлів, що не поступається результатам GAN-систем.
2. Непомітність. Точність виявлення стегоконтейнерів сучасними стегоаналізаторами на основі глибокого навчання (Chen-Net [6], Lin-Net [7]) наблизилась до 50% (рівень випадкового вгадування). Це підтверджує високу непомітність методу, навіть при значних швидкостях вбудовування (bps).
3. Стійкість (Робастність). Після застосування до стегоконтейнерів деструктивних атак (MP3-стиснення з різним бітрейтом, додавання AWGN-шуму), побітова помилка (BER) при вилученні секретного повідомлення виявилася значно нижчою у дифузійного методу, порівняно з контрольними методами на базі GAN та LSBM.

Висновки

У даній роботі запропоновано розробку нової архітектури умовної дифузійної моделі, адаптованої для задачі стеганографії, де секретне повідомлення інтегрується в латентний простір під час генерації, а не додається до готового контейнера. Це вирішує ключові проблеми стійкості та непомітності, які залишаються актуальними для GAN-базованих систем.

Таким чином, вагомість дослідження полягає у побудові стеганографічних систем, здатних витримувати реальні умови передачі даних в мережах (стиснення, шуми, перекодування), що критично важливо для створення надійних прихованих каналів зв'язку.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., et al. Generative adversarial nets. Advances in neural information processing systems, 27. 2014.
2. Ho, J., Jain, A., & Abbeel, P. Denoising diffusion probabilistic models. Advances in neural information processing systems, 33, pp. 6840-6851. 2020.
3. Song, J., Meng, C., & Ermon, S. Denoising diffusion implicit models. arXiv preprint arXiv:2010.02502. 2020.
4. Ronneberger, O., Fischer, P., & Brox, T. U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham. 2015.
5. ITU-T Recommendation P.862. Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs. International Telecommunication Union. 2001.
6. Chen, B., Luo, W., & Li, H. Audio steganalysis with convolutional neural network. Proceedings of the 5th ACM Workshop on Information Hiding and Multimedia Security (pp. 85-90). 2017.
7. Lin, Y., Wang, R., Yan, D., Dong, L., & Zhang, X. Audio steganalysis with improved convolutional neural network. Proceedings of the 7th ACM Workshop on Information Hiding and Multimedia Security (pp. 210-215). 2019.

Заверуха Олександр Андрійович – студент групи 2КІТС-24М, факультет менеджменту та інформаційної безпеки, Вінницький національний технічний університет, Вінниця, e-mail: sashazaverukha@gmail.com

Науковий керівник: ***Салієва Ольга Володимирівна*** – доктор філософії (PhD) за спеціальністю 125 «Кібербезпека», доцент кафедри менеджменту та безпеки інформаційних систем, Вінницький національний технічний університет, Вінниця, e-mail: salieva8257@vntu.edu.ua

Zaverukha Oleksandr A. – student of 2KITS-24m group, Faculty of Management and Information Security, Vinnytsia National Technical University, Vinnytsia, email: sashazaverukha@gmail.com

Supervisor*:* ***Saliieva Olha V.*** – Doctor of Philosophy (PhD) in specialty 125 "Cybersecurity", Associate Professor of the Department of Management and Security of Information Systems, Vinnytsia National Technical University, Vinnytsia, email: salieva8257@ vntu.edu.ua