Annual of Konstantin Preslavsky University of Shumen
Faculty of Mathematics and Informatics
vol. XXVII C, 2026, pp. 93 – 102, https://doi.org/10.46687/WSOX2294

EXPERIMENTAL EVALUATION OF TEXT SIZE AND ENCODING EFFECTS ON LSB EMBEDDING OF ENCRYPTED QR DATA

PAVEL E. YORDANOV, HRISTO I. PARASKEVOV

ABSTRACT: This paper examines how text length and UTF-8 encoding affect an image-steganography pipeline that encrypts a message with Fernet, represents it as a QR code, and embeds it in a 24-bit RGB image using LSB modification. Latin and Cyrillic payloads are tested under identical carrier and key parameters. We compare QR version, embedded bits, carrier utilization, MSE, PSNR, SSIM and BER. Cyrillic reaches QR capacity at shorter character lengths because UTF-8 uses more bytes. In all valid tests, BER is 0%, PSNR exceeds 55 dB, and SSIM exceeds 0.9989. Results show that load depends on byte size after encoding and encryption, not character count alone.

KEYWORDS: *image steganography; LSB; QR code; UTF-8; payload capacity; PSNR; SSIM.*

2010 Math. Subject Classification: 68U10, 68P30

ЕКСПЕРИМЕНТАЛНА ОЦЕНКА НА ВЛИЯНИЕТО НА РАЗМЕРА И КОДИРАНЕТО НА ТЕКСТА ВЪРХУ LSB ВГРАЖДАНЕТО НА КРИПТИРАНИ QR ДАННИ*

ПАВЕЛ Е. ЙОРДАНОВ, ХРИСТО ПАРАСКЕВОВ

* Статията е частично финансирана по проект № РД-08-88/02.02.2026

1 Въведение

Стеганографията в цифрови изображения използва информационната излишност на носителя, за да скрие допълнителни данни при минимално визуално изменение. При класическите LSB техники полезният товар се записва чрез промяна на най-малко значимите битове на пикселните стойности. Простотата на този подход го прави подходящ за експериментални и приложни системи, но ефективността му се определя от компромиса между капацитет, визуална незабележимост, надеждност на извличането и устойчивост към последваща обработка [1-3].

При хибридните решения криптографската и стеганографската обработка се комбинират, така че скритите данни да не бъдат достъпни в открит вид дори при извличане от носителя. QR кодът може да се използва като междинно структурирано представяне на криптирания резултат. Подобни подходи са разглеждани в системи, които съчетават криптиране, QR представяне и скриване в изображение [4, 5], а по-нови изследвания показват, че самата структура на QR кодовете също предоставя възможности за информационно скриване [6]. Следователно самото комбиниране на QR код и стеганография не представлява достатъчно основание за научна новост; от значение е как конкретната обработка влияе върху капацитета и качеството на стего-изображението.

Особен практически фактор е начинът, по който текстът се представя на байтово ниво. Два текста с еднакъв брой символи могат да имат различен размер след UTF-8 кодиране. Разликата се пренася към криптирания токен, след това към QR матрицата и накрая към броя LSB позиции, необходими за вграждане. Поради това използването само на броя символи като мярка за полезния товар може да бъде подвеждащо при многоезични текстове.

Целта на настоящото изследване е количествено да се оцени влиянието на дължината и кодирането на текстовото съобщение върху QR представянето и върху качеството на стего-изображението. Сравняват се текстове на латиница и кирилица при

еднакви условия. Проследяват се версията на QR кода, броят битове за вграждане, използваният процент от пикселите, средноквадратичната грешка MSE, пиковото съотношение сигнал/шум PSNR, индексът на структурно сходство SSIM [7] и честотата на битовите грешки BER. Изследването е насочено към капацитета и визуалното въздействие; криптографската устойчивост на използваната ключова схема не е предмет на настоящата статия.

2 Организация на експеримента

Експериментите са проведени с реализирана хардуерно-софтуерна система, в която входното съобщение първо се преобразува в UTF-8 байтова последователност и се криптира чрез Fernet. Криптираният токен се кодира като QR изображение с ниво L за корекция на грешки. Получената двоична матрица се преобразува в последователност от битове и се вгражда в най-малко значимите битове на RGB каналите на изображението носител. Позициите се избират възпроизводимо чрез псевдослучайна последователност, инициализирана от параметър Kpos. В тестваната реализация е използван Mersenne Twister [8]; тук той служи единствено за разпръснато и повторяемо позициониране, а не се оценява като криптографски генератор.

За да се изолира влиянието на текстовия полезен товар, при всички сравнения са запазени едни и същи параметри на ключовата схема и едно и също изображение носител: BMP, 24-битов RGB модел, разделителна способност 640×426 пиксела и размер приблизително 800 KB. Изследвани са две групи съобщения - съставени с латински символи и с кирилски символи. За всеки валиден тест се генерира Fernet токен, автоматично се определя минималната QR версия, извършва се LSB вграждане, след което се реализира обратният процес до възстановяване на съобщението.



Фигура 1. Последователност на експерименталната обработка

Размерът на данните, постъпващи към QR генератора, не съвпада с броя символи на първоначалното съобщение. След UTF-8 преобразуването криптирането допълва данните до блокова граница и добавя служебна информация, а Base64 представянето увеличава обема приблизително с една трета. В използваната реализация служебните компоненти на Fernet токена са 57 байта. Следователно увеличаването на байтовата дължина на текста се пренася нелинейно към QR версията поради дискретните размери на QR матриците.

Качеството на полученото стего-изображение е оценено чрез MSE, PSNR и SSIM. MSE измерва средната квадратна разлика между оригиналните и модифицираните пикселни стойности; по-ниска стойност означава по-малко изменение. PSNR изразява съотношението сигнал/грешка в децибели и нараства при по-малки изменения. SSIM оценява структурното сходство и приема стойности, близки до 1 при висока визуална и структурна близост [7]. BER се използва за проверка на коректността на възстановената битова последователност.

3 Влияние върху QR представянето и използвания капацитет

Таблица 1 представя зависимостта между броя символи, избраната QR версия, броя битове за вграждане и относителния дял от използваните пиксели. При еднакъв брой символи

кирилските съобщения водят до по-голям QR код, тъй като UTF-8 представянето им съдържа повече байтове.

Таблица 1.

Зависимост между текстовия полезен товар и QR/LSB натоварването

Кодиране	Символи	QR версия	Битове	Използвани пиксели
Латиница	200	12/40	48 009	5.87%
Латиница	400	17/40	77 889	9.52%
Латиница	800	24/40	131 817	16.10%
Латиница	1000	27/40	159 249	19.47%
Латиница	1500	34/40	233 337	28.52%
Латиница	2000	39/40	294 897	36.05%
Латиница	2100	40/40	308 073	37.66%
Латиница	2200	над 40	-	-
Кирилица	200	16/40	71 337	8.72%
Кирилица	400	25/40	123 249	15.06%
Кирилица	800	33/40	221 889	27.12%
Кирилица	1000	37/40	269 409	32.93%
Кирилица	1100	39/40	294 897	36.05%
Кирилица	1200	над 40	-	-

Разликата се вижда още при малък полезен товар. При 200 символа латинският текст генерира QR версия 12 и използва 5.87% от пикселите, докато кирилският текст достига версия 16 и използва 8.72%. При 1000 символа съответните стойности са версия 27 и 19.47% за латиница срещу версия 37 и 32.93% за кирилица. Следователно еднакъв брой знаци не означава еднакво стеганографско натоварване.

В разглежданата конфигурация последният успешно генериран QR код за латински текст е при 2100 символа - версия 40 и 37.66% използвани пиксели; при 2200 символа капацитетът на QR представянето е превишен. За кирилица последният успешен тест е при 1100 символа - версия 39 и 36.05%, а при 1200 символа QR код вече не може да бъде генериран в допустимия диапазон. Тези прагове са специфични за използваното криптиране, QR ниво на корекция, кодиране и изображение носител и не трябва да се разглеждат като универсални ограничения.

При общите тестови дължини 200, 400, 800 и 1000 символа броят битове за вграждане при кирилица е съответно с приблизително 48.6%, 58.2%, 68.3% и 69.2% по-голям спрямо латиница. Нарастването на тази разлика не е напълно плавно, тъй като QR кодът преминава дискретно между отделни версии. Практически това означава, че оценката на капацитета трябва да се извършва след реалното кодиране и криптиране на текста, а не чрез предварително фиксиран брой символи.

4 Влияние върху качеството на стего-изображението

Увеличаването на полезния товар води до модифициране на повече LSB позиции и закономерно увеличава числовата разлика спрямо оригиналното изображение. Таблица 2 обобщава количествените показатели за всички успешно обработени дължини в двете групи.

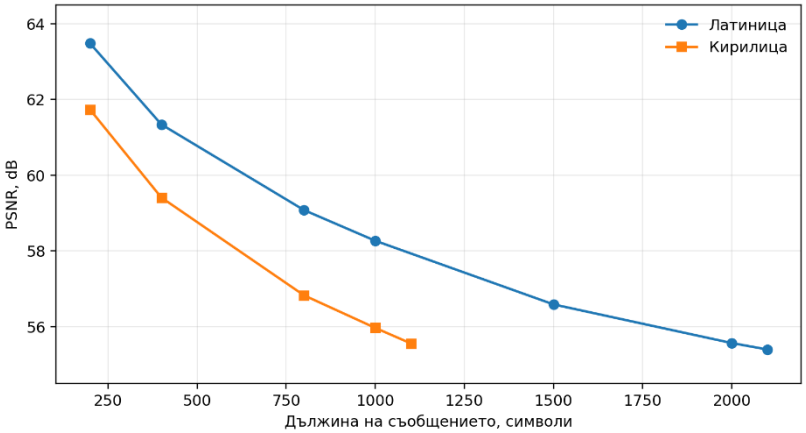
Таблица 2.

Количествена оценка на качеството при различен размер на текста

Кодиране	Символи	MSE	PSNR, dB	SSIM	BER
Латиница	200	0.029183	63.48	0.999831	0%
Латиница	400	0.047802	61.34	0.999724	0%
Латиница	800	0.080361	59.08	0.999537	0%
Латиница	1000	0.096815	58.27	0.999439	0%
Латиница	1500	0.142639	56.59	0.999175	0%
Латиница	2000	0.180235	55.57	0.998959	0%
Латиница	2100	0.187441	55.40	0.998918	0%
Кирилица	200	0.043671	61.73	0.999745	0%
Кирилица	400	0.074485	59.41	0.999572	0%
Кирилица	800	0.134940	56.83	0.999222	0%
Кирилица	1000	0.164481	55.97	0.999048	0%
Кирилица	1100	0.180668	55.56	0.998952	0%

При латиница MSE нараства от 0.029183 при 200 символа до 0.187441 при 2100 символа, докато PSNR намалява от 63.48 dB до 55.40 dB. SSIM се изменя от 0.999831 до 0.998918. При кирилица аналогичната тенденция се достига при по-малък брой символи: MSE нараства от 0.043671 при 200 символа до 0.180668 при 1100 символа, а PSNR намалява от 61.73 dB до 55.56 dB. Във всички успешно изпълнени тестове BER остава 0%, т.е. в контролираните

условия вградената битова последователност е възстановена без отчетени грешки.



Фигура 2. Изменение на PSNR при увеличаване на броя символи

Сравнението при еднакъв брой символи показва ефекта на байтовото представяне особено ясно. При 200 символа MSE е 0.029183 за латиница и 0.043671 за кирилица, а PSNR е съответно 63.48 и 61.73 dB. При 1000 символа MSE достига 0.096815 за латиница и 0.164481 за кирилица, докато PSNR е 58.27 срещу 55.97 dB. Следователно по-ниските стойности при кирилица не са следствие от специфично разпределение на самите букви в изображението, а от по-големия брой байтове, преминали през криптирането и QR представянето, и съответно от по-големия брой модифицирани позиции.

5 Анализ

Получените резултати показват, че при оценяване на текстов полезен товар в стеганографска система е по-коректно да се използва байтовият обем след кодиране, а не само броят на символите. Тази разлика е особено важна при многоезични приложения. В разглеждания пример 1000 символа на кирилица създават стеганографско натоварване, което е значително по-

близо до горната граница на QR капацитета, отколкото 1000 символа на латиница. Така влиянието на UTF-8 кодирането се проявява не директно върху LSB алгоритъма, а чрез цялата последователност кодиране - криптиране - Base64 - QR.

Резултатите за PSNR и SSIM показват, че визуалното качество остава високо при всички успешно изпълнени конфигурации, но с увеличаването на товара тенденцията е еднозначна: MSE се увеличава, а PSNR и SSIM намаляват. Това е в съответствие с общия компромис между капацитет и незабележимост, характерен за стеганографските системи [1, 3]. Нулевата BER при проведените опити показва функционална коректност на вграждането и извличането при съхраняване на носителя без загуба на данни.

Изследването има ясни ограничения. Основната серия за размера на полезния товар използва едно 24-битово BMP изображение с фиксирана разделителна способност, една схема за криптиране и едно ниво за корекция на QR кода. Следователно установените прагове 2100 и 1100 символа са характеристики на конкретната експериментална конфигурация, а не общи свойства на латиницата и кирилицата. Освен това LSB вграждането е чувствително към операции, които променят пикселните стойности; в отделни проверки на системата преобразуване през JPG, преоразмеряване и изрязване водят до неуспешно възстановяване. Това ограничение трябва да се отчита при практическо използване.

Следващо изследване може да разшири експеримента с независим набор от изображения с различна текстура и размер, с различни нива на QR корекция и с други кодировки и писмени системи. Особено полезно би било сравнението между директно вграждане на криптиран полезен товар и вграждане след QR представяне, за да се измери количествено цената на допълнителния QR слой спрямо възможните му практически предимства.

6 Заключение

Представеното експериментално изследване установява пряка зависимост между байтовото представяне на текстовото съобщение и стеганографското натоварване при последователност Fernet - QR - LSB. При еднакъв брой символи кирилският UTF-8 текст генерира по-голям криптиран и QR представен полезен товар, което води до по-висока QR версия, използване на по-голям дял от пикселите, по-високо MSE и по-нисък PSNR.

В конкретната тестова конфигурация латинският текст достига QR версия 40 при 2100 символа, докато при кирилица последният успешен тест е при 1100 символа. Независимо от увеличаването на натоварването, при всички валидни тестове BER е 0%, PSNR остава над 55 dB, а SSIM - над 0.9989. Основният практически значим извод е, че капацитетът на подобна система трябва да се планира по реалния брой байтове след кодиране и криптографска обработка, а не единствено по броя на текстовите символи.

ЛИТЕРАТУРА:

[1] Cheddad, A., Condell, J., Curran, K., Mc Kevitt, P., Digital image steganography: Survey and analysis of current methods. Signal Processing, 90 (3) (2010), 727-752. DOI: 10.1016/j.sigpro.2009.08.010.

[2] Provos, N., Honeyman, P., Hide and seek: An introduction to steganography. IEEE Security & Privacy, 1 (3) (2003), 32-44. DOI: 10.1109/MSECP.2003.1203220.

[3] Sahu, A.K., Sahu, M., Digital image steganography and steganalysis: A journey of the past three decades. Open Computer Science, 10 (1) (2020), 296-342. DOI: 10.1515/comp-2020-0136.

[4] Rani, M.M., Euphrasia, K.R., Data security through QR code encryption and steganography. Advanced Computing: An International Journal, 7 (1) (2016), 1-7.

[5] Alajmi, M., Elashry, I., El-Sayed, H.S., Faragallah, O.S., Steganography of encrypted messages inside valid QR codes. IEEE Access, 8 (2020), 27861-27873. DOI: 10.1109/ACCESS.2020.2971984.

[6] Koptyra, K., Ogiela, M.R., Steganography in QR Codes - Information Hiding with Suboptimal Segmentation. Electronics, 13 (13) (2024), 2658. DOI: 10.3390/electronics13132658.
[7] Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing, 13 (4) (2004), 600-612. DOI: 10.1109/TIP.2003.819861.
[8] Matsumoto, M., Nishimura, T., Mersenne Twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. ACM Transactions on Modeling and Computer Simulation, 8 (1) (1998), 3-30.

Павел Е. Йорданов
студент IV курс, специалност „Софтуерно инженерство“
Месторабота: Шуменски университет „Епископ Константин Преславски“, Факултет по математика и информатика
E-mail: 2230150004@shu.bg

Христо Параскевов
ORCID ID: 0000-0001-8097-6131
Месторабота: Шуменски университет „Епископ Константин Преславски“, Факултет по математика и информатика
E-mail: h.paraskevov@shu.bg