DOI: 10.18372/2410-7840.22.14662
УДК 621.391:519.2

ПОРІВНЯЛЬНИЙ АНАЛІЗ СКЛАДНОСТІ МЕТОДІВ ЛІНЕАРИЗАЦІЇ ТА ПЕРЕБОРУ РОЗВ'ЯЗАННЯ СИСТЕМ НЕЛІНІЙНИХ БУЛЕВИХ РІВНЯНЬ

Владислав Лещенко, Ніна Пекарчук, Михайло Савчук

Проблема знаходження розв'язків систем нелінійних рівнянь з багатьма змінними над скінченними алгебраїчними структурами та побудови ефективних алгоритмів їх пошуку є важливою для багатьох прикладних задач у різноманітних галузях і актуальність цієї проблеми зростає з часом. Стійкість багатьох існуючих криптосистем базується на складності задачі розв'язання систем нелінійних рівнянь багатьох змінних над скінченними полями. В загальному вигляді ця задача є задачею NP-повною. Але існує багато випадків, коли до таких систем можна запропонувати методи більш швидкі ніж методи повного перебору. Оскільки вибір методу може значно зменшити час та необхідні ресурси на знаходження розв'язків системи, природньо виникають питання оцінки складності різних методів розв'язання для систем з різними наборами параметрів, а також пошуку спеціальних найбільш ефективних методів для конкретного класу систем. У статті розглядаються найбільш важливі для криптографії та криптоаналізу системи нелінійних рівнянь з багатьма змінними над скінченним полем F_2. Предметом дослідження є порівняльний аналіз складності методу лінеаризації з введенням нових змінних для розв'язання систем нелінійних рівнянь над полем F_2 з багатьма невідомими та методу повного перебору в залежності від параметрів системи. Метою роботи є отримання середніх оцінок складності методів та знаходження межі в області зміни параметрів перевизначеної сумісної системи рівнянь, яка дає можливість з двох вказаних методів вибрати більш швидкий і ефективний. Запропоновані імовірнісні моделі для отримання теоретичних, асимптотичних оцінок середньої складності методів та проведення низки статистичних експериментів з отриманням середніх оцінок методом Монте-Карло. Показано, що існує границя в області зміни параметрів, що залежить, перш за все, від співвідношення максимального степеня рівнянь системи та числа невідомих, яка визначає, коли метод лінеаризації працює краще за повний перебір. Теоретичні та експериментальні дані застосовано для побудови цієї границі. Аналітичний вираз для лінії розмежування в області зміни параметрів системи отримано з використанням методу найменших квадратів.

Ключові слова: *системи нелінійних випадкових булевих рівнянь, алгоритми розв'язку систем, методи лінеаризації та повного перебору, складність алгоритмів, статистичне моделювання, криптоаналіз.*

Вступ

Проблема розв'язання систем нелінійних рівнянь з багатьма змінними над скінченними алгебраїчними структурами та побудови ефективних алгоритмів знаходження розв'язків важлива для прикладних задач в різнних галузях. Стійкість багатьох існуючих криптосистем базується на складності знаходження рішення систем нелінійних рівнянь багатьох змінних. Одним з підходів для оцінки середньої складності алгоритмів є побудова ймовірнісних моделей: визначення ймовірнісних мір на певній множині систем рівнянь і вивчення характеристик таких випадкових систем.

У роботах [1] описано загальні підходи, класифікація, постановки задач теорії випадкових систем рівнянь та в підрозділі про аналітичні методи криптоаналізу описано коротко алгоритми розв'язання систем нелінійних рівнянь над полем F_2, а також ідея методів лінеаризації. Монографія [4] містить низку тверджень та граничних теорем відносно ймовірнісних характеристик випадкових систем рівнянь над полем F_2. Класичним способом розв'язання таких систем нелінійних рівнянь над скінченними полями є алгоритм Бухбергера для обчислення базису Грьобнера та його варіанти [14]. Алгоритм впорядковує мономи (як правило в лексикографічному порядку) і усуває верхній моном, об'єднуючи два рівняння з відповідними поліноміальними коефіцієнтами. Моном складається з числового множника (коефіцієнта) і однієї або декількох букв (змінних), узятих кожна з тим або іншим цілим додатнім показником степеня. Цей процес повторюється доки всі крім однієї змінної не виключаються, а потім розв'язується залишкове рівняння (наприклад, за допомогою алгоритму Берлекемпа над простим або розширеним полем).

Жан-Шарль Фуджер запропонував два удосконалені варіанти алгоритму Бухбергера, [10, 12] та у ряді робіт застосував модифіковані алгоритми в криптоаналізі (наприклад [11, 13]). Аналіз складності обох алгоритмів опублікували Бардт та інші тільки в 2005 році [7]. Кіпніс і Шамір розробили більш простий, але більш повільний алгоритм, що називається лінеаризація (Relinearization) для криптоаналізу їх криптосистеми відкритого ключа HFE в 1999 році [15].

Рік потому Куртуа та інші придумали аналогічний але масштабований алгоритм під назвою розширена лінеаризація (eXtended Linearization) або *XL* [9]. Аналіз складності *XL* був надзвичайно поганим і сильно відрізнявся від справжньої складності лінеаризації, проте сам алгоритм використовувався для криптоаналізу (наприклад,[8]). Г. Арс та інші [6] у 2004 році та Альбрехт [5] у 2011 році показали, що *XL* є надлишковою версією алгоритму Фуджера [12], тобто *XL* виробляє більше рівнянь, ніж це необхідно, і не є швидшим. З іншого боку Мохамед та інші ввели варіант *XL* з використанням алгоритму Вільємана [16], завдяки цьому *XL* споживає менше пам'яті.

Існує багато модифікацій алгоритмів лінеаризації систем над скінченними полями. При достатньої кількості лінійно незалежних рівнянь у системі вони працюють за однією схемою, основою якої є методи розв'язання систем лінійних рівнянь. У даній статті розглядається саме випадок перевизначених систем рівнянь. Існуючи результати, наприклад, Кіпніса і Шаміра щодо верхньої поліноміальної оцінки асимптотичної складності лінеаризації для квадратичних рівнянь[15] не дають можливості знайти чітку границю, за якою метод лінеаризації працює краще ніж метод перебору.

В даній роботі розглядаються найбільш важливі для криптографії та криптоаналізу системи нелінійних рівнянь з багатьма змінними над полем F_2. Предметом дослідження є метод лінеаризації в порівнянні з методом повного перебору розв'язання перевизначених, надлишкових систем нелінійних рівнянь над полем F_2 з багатьма змінними та методи отримання оцінок їх складності. Метою роботи є знаходження межі в області зміни параметрів перевизначених і напевно сумісних систем рівнянь, яка дає можливість з двох вказаних методів вибрати більш швидкий і ефективний. Показано, що має існувати границя, що залежить перш за все від співвідношення максимального степеня рівнянь системи та числа невідомих, коли метод лінеаризації працює краще за повний перебір. Запропонована математична імовірнісна модель для теоретичного аналізу та проведення низки статистичних експериментів. Застосовано теоретичний та асимптотичний аналіз для побудови лінії розмежування між методами по ефективності. Для цієї границі отримано аналітичний вираз.

1. Алгоритм розв'язання нелінійних перевизначених систем над полем F_2 методом лінеаризації з введення нових змінних

Розглядається системи нелінійних рівнянь багатьох змінних над полем F_2 з елементами поля $\{0,1\}$.

$$\left\{\begin{array}{l} f_1(x_1,...,x_n)=y_1, \\ f_2(x_1,...,x_n)=y_2, \\ ..... \\ f_m(x_1,...,x_n)=y_m \end{array}\right\}, \qquad (1)$$

де n - кількість змінних (невідомих) в системі; m - кількість рівнянь системи, x_i, i=1,2,…,n, - невідомі змінні, $x_i \in \{0,1\}$; y_j, j =1,2,…,m, праві частини j-го рівняння, $y_j \in \{0,1\}$; булеві функції f_j задано алгебраїчною нормальною формою (поліномами Жегалкіна) без вільного доданку

$$f_j(x_1,...,x_n)=\sum_{k=1}^{r}\sum_{i_1<i_2<....<i_k} a^j_{i_1 i_2 ....i_k} x_{i_1} x_{i_2} ... x_{i_k},$$
$$j=1,2,...,m, \qquad (2)$$

праві частини y_j та коефіцієнти $a^j_{i_1 i_2 ....i_k} \in \{0,1\}$, $i_1 < i_2 < .... < i_k, k=1,2,...,r$, відомі величини F_2.

Припущення А. Відносно системи (1) зробимо такі припущення:

– А1. Алгебраїчний степінь кожного рівняння в (1) не перевищує деякого натурального числа $r: \deg f_j \le r \le n, j=1,2,...,m.$

– А2. Існує рівняння системи, яке має алгебраїчний степінь r.

– А3. Система сумісна, тобто існує вектор $X^0=(x_1^0, x_2^0, ..., x_n^0)$, який є розв'язком системи.

– А4. Система перевизначена (надлишкова), тобто має число рівнянь m набагато більше ніж число невідомих n.

Зауваження до припущення А4. Для багатьох застосувань це припущення не є надто жорстким. Наприклад, в криптоаналізі симетричних систем шифрування кількість рівнянь в системі (1) відносно невідомих бітів ключа може бути отримана достатньо великою.

Метод лінеаризації введенням нових змінних розв'язання системи (1) базується на введенні таких додаткових змінних, відносно яких система буде лінійною. Введемо позначення:

$$x_{i_1} x_{i_2} ... x_{i_k} = u_{i_1 i_2 ....i_k},$$
$$1 \le i_1 < i_2 < .... < i_k \le n,\ k=1,2,...,r, \qquad (3)$$

та будемо вважати $u_{i_1 i_2 \ldots i_k}$ новими невідомими системи (1). Коли $k=1$, $u_1 = x_1, u_2 = x_2, \ldots, u_n = x_n$. Відносно нових змінних система (1) буде лінійною, та матиме вигляд

$$\sum_{k=1}^{r} \sum_{i_1<i_2<\ldots<i_k} a^j_{i_1 i_2 \ldots i_k} u_{i_1 i_2 \ldots i_k} = y_j,\ j = 1,2,\ldots,m. \quad (4)$$

Число невідомих N системи (4) задовольняє нерівностям

$$N \le n + C_n^2 + C_n^3 + \ldots + C_n^r = N(n,r) \le 2^n - 1. \quad (5)$$

Розв'язок система (1) зводиться до розв'язку лінійної системи (4) та систем спеціального виду (3). Для подальшого дослідження описаного методу лінеаризації та порівняння з методом повного перебору лінійну систему (4) будемо розв'язувати методом Гаусса. Для дослідження середньої складності алгоритмів побудуємо ймовірнісні моделі систем (1), (4) та процесів експериментального і теоретичного отримання оцінок складності.

2. Ймовірнісні моделі системи нелінійних рівнянь та асимптотичні оцінки

Для отримання середніх оцінок складності методу лінеаризації введенням нових змінних, який описано в розділі 2, побудуємо імовірнісні моделі щодо розподілу коефіцієнтів систем (1) і (4), враховуючи припущення А.

Імовірнісна модель М1. При умові виконання припущень **А1-А4** для системи (1) виконується такі умови:

a) Усі $m \times N(n,r)$ коефіцієнтів системи (1) є реалізацією незалежних рівноймовірних булевих випадкових величин:

$$P\{a^j_{i_1 i_2 \ldots i_k} = 1\} = P\{a^j_{i_1 i_2 \ldots i_k} = 0\} = \frac{1}{2},$$

$$1 \le i_1 < i_2 < \ldots < i_k \le n, k = 1,2,\ldots,r, \quad j = 1,2,\ldots,m.$$

b) Справжній розв'язок системи $X^0 = (x_1^0, x_2^0, \ldots, x_n^0)$ є реалізацією випадкового n - вимірного булевого вектора з рівноймовірним розподілом на множині усіх 2^n можливих векторів.

c) Праві частини $y_j = f_j(x_1^0, \ldots, x_n^0),\ j = 1, 2, \ldots, m.$

Імовірнісна модель М2. При умові виконання припущень **А1, А3, А4** для системи (1) виконується умови a), b) та c) імовірнісної моделі М1.

В моделі М2 ймовірність того, що в системі (4) будуть присутні всі $N(n,r)$ змінних, дорівнює

$$P\{N = N(n,r)\} = (1 - 2^{-m})^{N(n,r)}.$$

Логарифм правої частини

$$\ln(1-2^{-m})^{N(n,r)} = -N(n,r)$$
$$(2^{-m} + \frac{1}{2}2^{-2m} + \frac{1}{3}2^{-3m} + \frac{1}{4}2^{-4m} + \ldots). \quad (6)$$

Якщо в (6) $N(n,r)2^{-2m} \to 0$, коли $m \to \infty$ (при цьому n, r також можуть прямувати до нескінченності), то

$$P\{N = N(n,r)\} \sim e^{-N(n,r)2^{-m}}, \quad (7)$$

а якщо $N(n,r)2^{-m} \to 0$, то

$$P\{N = N(n,r)\} \to 1. \quad (8)$$

Тобто якщо виконується умови (8), то можна вважати, що асимптотично число невідомих в системі (4) $N = N(n,r)$. Оцінимо точність останньої рівності.

Нехай виконується умова а) в моделі М2, тоді число невідомих системи (4) має біноміальній розподіл $B(1-2^{-m}, N(n,r))$, оскільки присутність кожної змінної можна інтерпретувати як успіх у схемі Бернуллі з ймовірністю $1-2^{-m}$ і числом іспитів $N(n,r)$. Коли $N(n,r) \to \infty, N(n,r)2^{-m} \to \lambda > 0,$ то число «відсутніх» змінних $\bar{N} = N(n,r) - N$ має розподіл Пуассона з параметром λ і математичне сподіванням $\lambda \sim N(n,r)2^{-m}$ [2]. Тоді ймовірність того, що

$$P\{\bar{N} = 0\} = P\{N = N(n,r)\} \sim e^{-\lambda}. \quad (9)$$

Асимптотичне співвідношення (9) підтверджує (7), а математичне сподівання числа невідомих систми (4) N в цьому випадку дорівнює $E(N) = N(n,r) - \lambda$. Оскільки далі для розв'язання системи (1) методом лінеаризації за припущенням А4 буде вибиратися система з числом рівнянь m рівним, або одного порядку з числом нових невідомих $N(n,r)$, то виконується (8) і співвідношення

$$\lambda = N(n,r)2^{-m} \to 0,$$
$$E(N) = N(n,r) + o(1). \quad (10)$$

Головним предметом дослідження в роботі є середні оцінки складності методу розв'язку лінійної системи (4) в моделях М1, М2 та порівняння з оцінками знаходження розв'язку повним перебором. Зі співвідношень (7)-(10) випливає, що асимптотично (для великих $N(n,r)$) оцінки складності розв'язання лінійної системи (4) методом Гаусса в моделях М1 та М2 будуть співпадать. Враховуючи припущення А та ймовірнісні моделі М1 і М2, можна вважати, що додатковою складністю систем

спеціального виду (3) можна знехтувати. В порівнянні методів лінеаризації та повного перебору і знаходженні лінії розподілу $n=\varphi(r)$ або $r=\varphi^{-1}(n)$ за складністю між цими методами будемо спиратися на середні оцінки трудомісткості метода Гаусса з $N(n,r)$ невідомими та середні оцінки методу перебору.

Коли число незалежних рівнянь $m=N(n,r)$, $N(n,r)\to\infty$ максимальна складність розв'язку системи (4) алгоритмом Гаусса $L_{\max}=\frac{N(n,r)^3}{3}$ в бітових операціях, а середня

$$L_{сер}\sim\frac{N(n,r)^3}{6}. \qquad (11)$$

Якщо число рівнянь менше $N(n,r)$, то до рішень системи лінійних рівнянь додаються рівняння спеціального виду:

$$x_{i_1}\ldots x_{i_k}=u_{i_1i_2\ldots i_k}, \qquad (12)$$

де $(i_1,\ldots,i_k)$ усі невпорядковані набори попарно нерівних індексів, $k=2,\ldots,r$. Рівняння (12) також значно прискорять обернений хід алгоритму Гаусса.

Таким чином, далі для теоретичних розрахунків вважаємо, що $N=N(n,r)$ і знайдемо більш просту, ніж за формулою (5), оцінку число невідомих системи (4).

1) Отримаємо оцінку зверху для $N(n,r)$, яка буде оцінкою зверху і для числа невідомих. Нехай $r=[\alpha n],\alpha\in(0,\frac{1}{2})$, де $[z]$ – ціла частина дійсного числа z. Тоді

$$C_n^{r-1}/C_n^r=\frac{n(n-1)\ldots(n-r+2)}{(r-1)!n(n-1)\ldots(n-r+1)}=$$
$$\frac{r}{n-r+1}\sim\frac{\alpha n}{n(1-\alpha)}=\frac{\alpha}{1-\alpha}<1,$$
$$N(n,r)<C_n^r\sum_{k=0}^{\infty}\left(\frac{\alpha}{1-\alpha}\right)^k=$$
$$C_n^r\frac{1-\alpha}{1-2\alpha}=C_n^{[\alpha n]}\frac{1-\alpha}{1-2\alpha}. \qquad (13)$$

Так, наприклад, для $\alpha=1/3$ таким чином отримаємо розклад $N(n,r)=C_n^r(2-\frac{12}{n}+O(\frac{1}{n^2}))$, $n\to\infty$. Оцінка (13) для $N(n,r)$ тим точніша, чим менше α. Враховуючи (5), (8), отримаємо

$$N=N(n,r)\sim C_n^{[\alpha n]}\frac{1-\alpha}{1-2\alpha},\ n\to\infty. \qquad (14)$$

2) При виконанні умови а) в імовірнісних моделях М1, М2 для асимптотичної оцінки $N(n,r)$ можна використати формулу для надвеликих ухилень з роботи [3, гл. II, п. 6]:

$$P(\nu\le\alpha n)\sim(2\pi n\alpha\beta)^{-1/2}\frac{q\alpha}{q\alpha-p\beta}\left(\frac{p^\alpha q^\beta}{\alpha^\alpha\beta^\beta}\right)^n. \qquad (15)$$

де ν – число успіхів в схемі Бернуллі с параметрами (p,n), $q=1-p$, $\beta=1-\alpha$, $\alpha<p$. Використовуючи (15) при $p=1/2$ з рівності $N(n,r)+1=2^nP(\nu\le\alpha n)$, отримаємо при $n\to\infty$ асимптотичну досить якісну оцінку

$$N(n,r)\sim(2\pi n\alpha\beta)^{-1/2}\frac{\alpha}{2\alpha-1}\left(\frac{1}{\alpha^\alpha\beta^\beta}\right)^n. \qquad (16)$$

Формули (5), (14), (16) будемо використовувати в різних областях зміни параметрів n та r для теоретичного підрахунку оцінок середньої складності розв'язання лінійної системи рівнянь (4) методом Гаусса. Імовірнісні моделі М1 і М2 будемо також використовувати для отримання теоретичних оцінок середньої складності розв'язання системи (1) методом повного перебору. Для $n=\overline{10,18}$ та $r=\overline{2,6}$, тобто $n\in\{10,11,\ldots,18\}$ та $r\in\{2,3,4,5,6\}$ у наступному розділі отримано оцінки середньої складності алгоритму лінеаризації методом статистичного моделювання.

3. Експериментальне дослідження складності алгоритму лінеаризації

Задача і умови експерименту. Створено програми, які при заданій кількості змінних, степені систем та кількості систем, що генеруються випадковим чином, виводить середню кількість арифметичних булевих операцій на розв'язання однієї системи. Експеримент для знаходження складності алгоритму лінеаризації розбивається на етапи.

Етап 1. Генерація системнелінійних рівнянь.

Моделювання випадковим чином згідно з **припущеннями А** та ймовірнісною **моделлю М1** масиву напевно сумісних систем рівнянь над F_2 виду (1) з $m=N(n,r)$ рівняннями в кожній системі з n невідомими та максимальним степенем r.

Системи отримуємо випадковим чином, всі коефіцієнти в системі задовольняють умові а) моделі М1 – незалежні і рівноймовірні. Згідно з припущеннями А будуємо системи з n невідомими зі

степенем систем r, системи сумісні та усі різні. Розглядаємо випадок, коли рівнянь достатньо і непотрібно до систем додавати нові рівняння з мономів виду (3) для відкидання зайвих розв'язків. При цьому справжні розв'язки кожної системи, які використовується для побудови рівнянь на етапі 1 при обчисленні правих частин (4) та при розв'язанні системи на етапі 2, або генерується як незалежні рівноймовірні вектори довжини n, або перебираються усі 2^n векторів. Середня кількість операцій бітового множення в розрахунку монома дорівнює 1. Вважаємо, що обчислення одного моному та операцію $\oplus$ мають однакову складність - одну булеву операцію .

Вхід: n – кількість змінних, r – степінь системи, v – кількість систем.

Вихід: v матриць – $N(n,r)\times(N(n,r)+1)$.

Схематичний опис алгоритму:

1. Випадково генеруємо булевий масив довжини 89 – початковий стан генератора псевдовипадкових двійкових послідовностей.

2. За допомогою генератора псевдовипадкових двійкових послідовностей, що задається формулою $x_t = x_{t-38} \oplus x_{t-89}$, за початковим вектором з кроку 1 отримуємо булеву послідовність M довжини 2^{50}. Лінійна рекурентна послідовність генератора має період $2^{89}-1$ і якісні ймовірнісні характеристики.

3. Фіксуємо справжній розв'язок $X^0 = (x_1^0, x_2^0,...,x_n^0) \in \{0,1\}^n$ з випадковими, або новими послідовними значеннями (якщо робимо повний перебір по всім n-векторам).

4. З послідовності M отримуємо коефіцієнти при мономах в рівняннях системи (1).

Перші $N(n,r)$ біт масиву M – це перший рядок матриці $N(n,r)\times N(n,r)$. Другі $N(n,r)$ біт M – це другий рядок матриці $N(n,r)\times N(n,r)$ і так до $N(n,r)$ рядку матриці. Тобто i-ий рядок матриці складається з біт послідовності M від $(i-1)N(n,r)+1$ до $i\times N(n,r)$, $i=1,2,...,N(n,r)$.

5. Перевіряємо рядки матриці на повторення. Замість рядків, що повторюються, беремо нові рядки з M.

6. Розраховуємо праві частини рівнянь (1), підставляючи розв'язок X^0 в отримані функції (2).

7. Порівнюємо матрицю з попередніми для цього розв'язку, якщо така матрицю вже існує повторюємо кроки 4-6.

8. Зберігаємо матрицю.

9. Повторюємо кроки 4-8 $[v/2^n]+1$ разів, де $[x]$ – ціла частина числа x, якщо робимо повний перебір по всім $X^0 = (x_1^0, x_2^0,...,x_n^0) \in \{0,1\}^n$. Або повторюємо кроки 4-8 $[v/u]+1$ разів, якщо випадковим чином вибираємо u справжніх розв'язків X^0.

10. Повторюємо кроки 3-9 для всіх 2^n можливих розв'язків X^0, або відповідно повторюємо u разів для випадкових розв'язків X^0.

В результаті отримаємо v або більше сумісних систем з лінійно незалежними рівняннями. Кожна система матиме один розв'язок, коли кількістю рівнянь дорівнює кількості змінних $N(n,r)$. Системи модельовані за допомогою різних початкових розв'язків будуть різні і їх непотрібно перевіряти на повторення.

Етап 2. Розв'язання систем лінеаризацією.

Перетворення кожної системи виду (1) масиву в систему виду (4). Розв'язання кожної системи виду (4) методом Гаусса, підрахунок числа операцій та часу для кожної системи та усереднення характеристик складності.

Складність алгоритму лінеаризації на етапі 2 залежить від двох кроків. Перший: кількість операцій для розв'язання лінійної системи (4). Другий: складність розв'язання додаткових рівнянь (3). Якщо кількість рівнянь системи (4) дорівнює або більше ніж число усіх невідомих N, то складністю другого кроку можна знехтувати.

Створено програму, яка розв'язує системи (4) з використанням метода Гаусса та рахує кількість операцій. Системи з випадковими коефіцієнтами отримуємо за допомогою попередньої програми. Проведено експерименти та отримано значення складності розв'язання систем лінеаризацією для систем з кількістю змінних $n \in \{10,11,...,18\}$ та $r \in \{2,3,4,5,6\}$. Для кожного значення пари (n,r) порахована середня складність за 10 000 змодельованих випадково нелінійних систем. Результати експериментів наведено в розділах 4 та 5.

4. Порівняння експериментальних результатів середньої складності для методів лінеаризації та повного перебору

Математичне сподівання складності методу перебору для систем вигляду (1) нелінійних рівнянь багатьох змінних над полем F_2 оцінюємо за ймовірнісною моделлю М2 формулою

$$L_{пер} = 2^{n+1} N(n,r). \qquad (17)$$

У формулі враховується кількість усіх можливих розв'язків 2^n, середнє число мономів в одному рівнянні $\frac{N(n,r)}{2}$, середнє число додавань мономів і порівняння з правою частиною $\frac{N(n,r)}{2}$, середнє число рівнянь, на яких відбраковується хибні розв'язки, а саме 2. Також, як і раніше, і розрахунок одного моному, і кожну операцію $\oplus$ будемо вважати за одну операцію.

Зробимо порівняння отриманих значень складності лінеаризації і повного перебору для кількості змінних $n = \overline{10,18}$ і степеня систем $r = \overline{2,6}$.

Для систем степеня 2 лінеаризація працює завжди швидше ніж перебір для всіх досліджуваних значень n. Причому зі збільшенням n відношення складності методу повного перебору до складності методу лінеаризації зростає дуже швидко.

Для систем степеня 3 метод лінеаризація працює гірше (з більшою складністю) ніж перебір для систем з $n < 15$ та починає працювати краще ніж перебір (з більшою ефективністю) коли $n \geq 15$. Відношення складності методу повного перебору до складності методу лінеаризації також зростає, але повільніше ніж у попередньому випадку. Отже, для систем степеня 3 експериментально знайдено граничне значення n, після якого метод лінеаризації починає працювати краще за перебір.

Приклад порівняння складності методів зображено на рисунку 1 для степеня $r = 3$. На осі абсцис відкладена кількість початкових змінних в системі (1). На осі ординат відкладена складність - середня кількість операцій при рішенні системи.

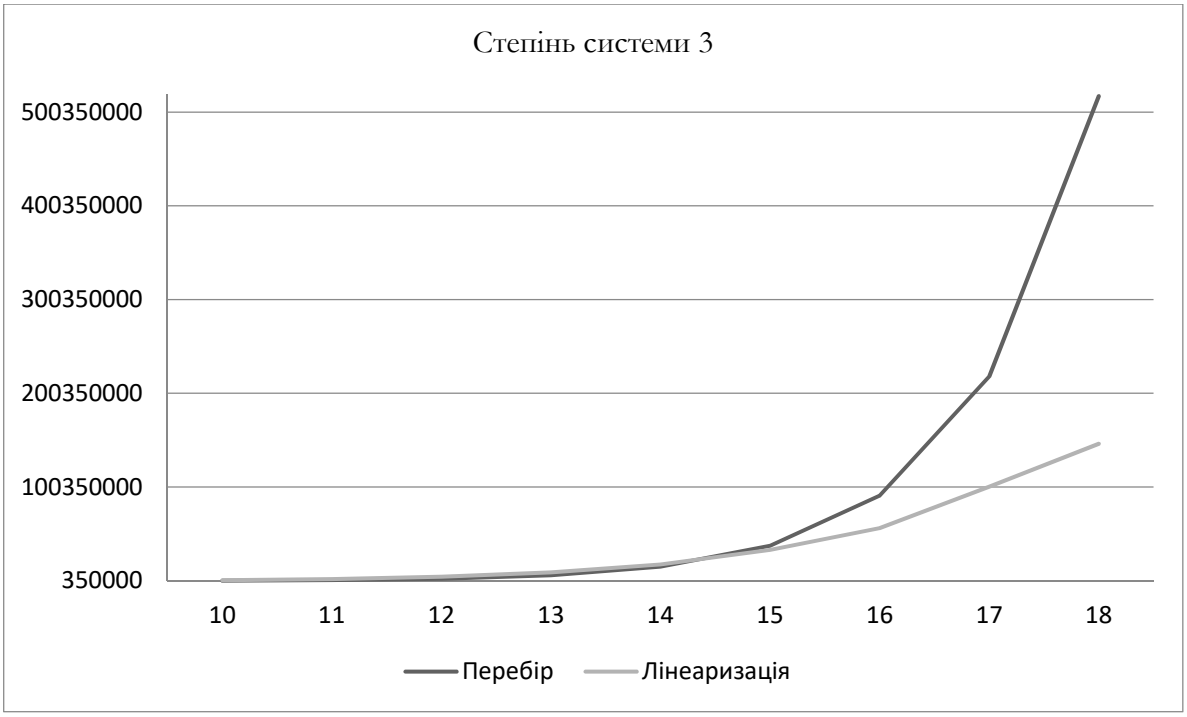

Рис. 1. Складність розв'язання систем степеня 3

Для систем степеня 4 та 5 лінеаризація працює гірше для всіх експериментально досліджених значень n.

5. Порівняння теоретичної і експериментальної складності алгоритму лінеаризації

Теоретичну складність алгоритму лінеаризації розраховуємо як середню складність алгоритму Гаусса в умовах ймовірнісної моделі М2 і в прийнятих операціях. Якщо враховувати прямий і обернений хід алгоритму Гаусса, то формула середньої складності реалізації алгоритму для системи (4) приймає вигляд

$$L_{гаусс} = \frac{1}{6}N(n,r)^3 + \frac{3}{4}N(n,r)^2 - \frac{5}{6}N(n,r). \qquad (18)$$

Для теоретичного обчислення $N(n,r)$ застосовувались формули (5), (14), (16).

Порівняння теоретичної та експериментальної складності алгоритму лінеаризації для з оцінкою відхилень для значень $n = \overline{10,18}$ та $r = \overline{2,5}$ представлено у таблицях 1-4.

Як бачимо для систем ступеня 2 експериментальні значення майже не відрізняються від теоретичних. Похибка коливається від [-4; 8] відсотків. Середнє значення модуля похибки 4,7%.

Таблиця 1

Середня кількість операцій методу лінеаризації

для систем ступеня 2

n	Експериментальна складність	Теоретична складність	Відхилення %
10	30845	29947	-3
11	52656	51122	-3
12	76896	83583	8
13	123819	131722	6,5
14	197088	201110	2
15	310637	298690	-4
16	407011	432990	6
17	583629	614346	5
18	803833	855142	6

Таблиця 2

Середня кількість операцій для систем ступеня 3

n	Експериментальна складність	Теоретична складність	Відхилення %
10	906877	916037	1
11	1989497	2094207	5,5
12	4118774	4476928	8
13	9307791	9036690	-3
14	15795925	17358159	9
15	29697073	31932337	7
16	61079324	56554930	-8
17	89106210	96854576	8
18	161785425	160980522	-0,5

Похибка коливається в межах [-8;9] відсотків. Середнє значення модуля похибки 5,5%.

Таблиця 3

Середня кількість операцій для систем ступеню 4

n	Експериментальна складність	Теоретична складність	Відхилення %
10	9573810	9621920	0,5
11	27882224	29661940	6
12	78568759	83583786	6
13	198308833	217921795	9
14	562902217	531039827	-6
15	1109943612	1219718255	9
16	2579455753	2659232735	3
17	5868057902	5535903681	-6
18	10395819002	11059381917	6

Похибка коливається від [-6; 9] відсотків. Середнє значення модуля похибки 5,7%. Для систем ступеня 4 відносна різниця між значеннями майже не змінилась порівняно з системами ступеню 3.

Похибка за таблицею 5коливається від [-10; 0,5] відсотків. Середнє значення модуля похибки 4,2%. Для систем ступеню 6 відносна різниця між значеннями зменшилась порівняно з системами ступеня 5.

Таким чином, теоретичні та експериментальні значення складності лінеаризації співпадають з відносною похибкою меншою 10%. Теоретичну оцінку значення середньої складності лінеаризації для систем (4) використаємо для прогнозування складності систем з більшою кількістю змінних та степеню систем.

Таблиця 4

Середня кількість операцій для систем ступеня 5

n	Експериментальна складність	Теоретична складність	Відхилення %
10	101302459	101811517	0,5
11	552920202	547445745	-1
12	2848081013	2637112049	-8
13	11686581229	11457432577	-2
14	47087222272	45276175262	-4
15	164975148022	164154376141	-0,5
16	594651871957	550603585145	-8
17	1893761645314	1721601495740	-10
18	5254498940086	5052402827006	-4

6. Рівняння для лінії розмежування

Побудова функціцї для лінії розмежування виконувалася за допомогою спеціальної програми. Для значень r, починаючи з $r=3$ з кроком 3, розраховуємо, використовуючи формули (5), (14)-(18), значення n, при яких середня складність $L_{лін}$ лінеаризації менше середньої складності $L_{пер}$ перебору

$$L_{лін} < L_{пер}. \qquad (19)$$

Програма дозволяє розрахувати до значення $r=114$ та $n=1018$. За цими значеннями побудовано графік – рисунок 2. На осі абсцис знаходяться значення r, на осі ординат – значення n. В області під лінією ефективніше метод перебору, а на лінії та вище – метод лінеаризації має меншу складність.

Аналітичний вираз для лінія розділення, побудовано за допомогою методу найменших квадратів за поліноміальною моделлю степеню 2. Рівняння лінії розмежування має вигляд

$$n = 0,00051 \times r^2 + 8,982 \times r - 12,683. \qquad (20)$$

Для перевірки (20) розроблена незалежна програма для n до 1753 та r до 195. Результати роботи програми підтвердили попередні дані та висновки. Основні результати нової програми наведено в таблиці 5.

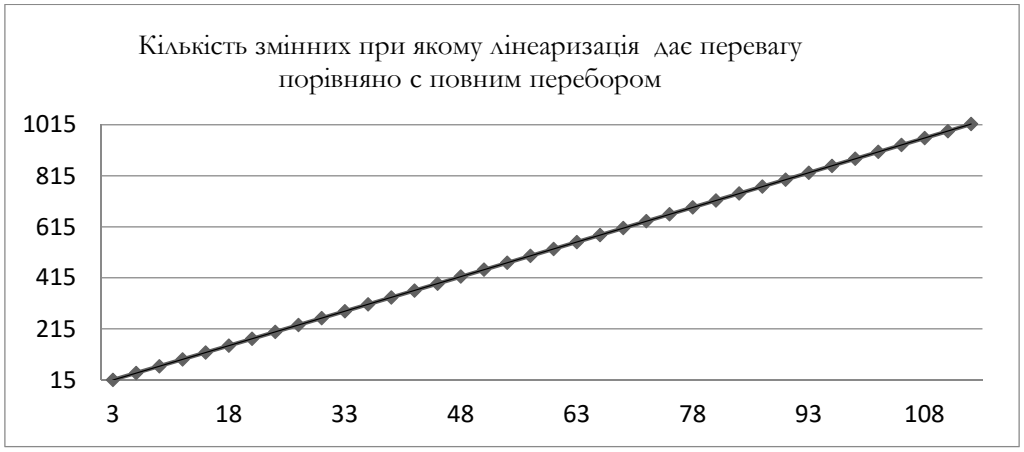

Рис. 2. Лінія розмежування – мінімальна кількість змінних, при якому лінеаризація дає перевагу порівняно з повним перебором для степеня r

Таблиця 5

Мінімальних значень n при заданих значеннях r, для яких $L_{лін} < L_{пер}$ (лінеаризація дає перевагу).

r	Min n	n	Min n
3	15	102	909
6	42	105	936
9	68	108	963
12	95	111	990
15	122	114	1018
18	149	117	1045
21	176	120	1072
24	203	123	1099
27	230	126	1127
30	257	129	1154
33	284	132	1181
36	311	135	1208
39	338	138	1235
42	365	141	1263
45	392	144	1290
48	420	147	1317
51	447	150	1344
54	474	153	1371
57	501	156	1399
60	528	159	1426
63	555	162	1453
66	583	165	1480
69	610	168	1508
72	637	171	1535
75	664	174	1562
78	691	177	1589
81	718	180	1617
84	746	183	1644
87	773	186	1671
90	800	189	1698
93	827	192	1725
96	854	195	1753
99	882		

Висновки

Запропоновано ймовірнісні моделі для теоретичного та експериментального аналізу середньої складності методів лінеаризації та повного перебору розв'язання систем нелінійних рівнянь над полем F_2 при певних умовах на ймовірнісний розподіл коефіцієнтів системи. Проведено статистичне моделювання, експериментально отримані середні складності лінеаризації, перебору та здійснено їх порівняння в різних випадках. Для цього створено програми для генерації випадкових систем рівнянь, та розв'язування їх експериментально методом лінеаризації та підрахунку кількості операцій. Під час експериментального дослідження пораховано середня складність лінеаризації для систем з числом змінних $n = \overline{10,18}$ та $r = \overline{2,6}$. Виконано порівняння з теоретичними та асимптотичними результатами. Для числа змінних n від 3 до 1753 на основі експериментальних даних та теоретичного аналізу знайдено аналітичний вираз для лінії розмежування, яка визначає коли метод лінеаризації розв'язання систем нелінійних рівнянь над скінченним полем більш ефективний ніж метод повного перебору.

ЛІТЕРАТУРА

[1]. А. Бабаш, Г. Шанкин, *Криптография*, М.: СОЛОН-Р, 2002, 512 с.

[2]. А. Кобзарь, *Прикладная математическая статистика. Для инженеров и научных работников,* М.: ФИЗМАТЛИТ, 2006, 816 с.

[3]. И. Коваленко, А. Филиппова, *Теория вероятностей и математическая статистика, 2-е изд., перераб. и доп.* – М.:Высш.школа, 1982, 256 с.

[4]. В. Колчин, *Случайные графы*, М.: ФИЗМАТЛИТ, 2000, 256с.

[5]. M. Albrecht, C. Cid, J.-C. Faugère, L. Perret, *On the relation between the mutant strategy and the normal selection strategy in gröbner basis algorithms*, Eprint Report 2011/164, 2011.

[6]. G. Ars, J.-C. Faugère, H. Imai, M. Kawazoe, M. Sugita, "Comparison between xl and gröbner basis algorithms", *In ASIACRYPT 2004, Lecture*, pp. 338-353, 2004.

[7]. M. Bardet, J.-C. Faugère, B. Salvy, B.-Y. Yang, "Asymptotic behavior of the degree of regularity of semi-regular polynomial systems", In P. Gianni, editor, *MEGA 2005,* Sardinia (Italy), 2005.

[8]. N. Courtois, J. Pieprzyk, "Cryptanalysis of block ciphers with overdefined systems of equations", *In Advances in Cryptology _ ASIA-CRYPT 2002*, vol. 2501 of Lecture Notes in Computer Science, pp. 267-287, 2002.

[9]. N. Courtois, A. Klimov, J. Patarin, A. Shamirv, "Efficient algorithms for solving overdefined systems of multivariate polynomial equations", *In Advances in Cryptology _ EUROCRYPT 2000*, vol. 1807, pp. 392-407, 2000.

[10]. J.-C. Faugère, "A new e-cient algorithm for computing Gröbner bases without reduction to zero (F5)", *In International Symposium on Symbolic and Algebraic Computation _ ISSAC 2002*, pp. 75-83, 2002.

[11]. J.-C. Faugère, A. Joux, "Algebraic cryptanalysis of Hidden Field Equations (HFE) using Gröbner bases, *In Advances in Cryptology _ CRYPTO 2003*, vol. 2729, pp. 44-60, 2003.

[12]. J.-C. Faugère, "A new e-cient algorithm for computing Gröbner bases (F4)", *Journal of Pure anNational Institute of Standards and Technologyd Applied Algebra*, vol. 139, pp. 61-88, 1999.

[13]. J.-C. Faugère, A. Otmani, L. Perret, J.-P. Tillich, "Algebraic cryptanalysis of McEliece variants with compact keys", *In EUROCRYPT*, vol. 6110, pp. 279-298, 2010.

[14]. I.A. Ajwa, Z. Liu, P. S. Wang, "Grobner Bases Algorithm", *ICM Technical Reports*, 1995.

[15]. A. Kipnis, A. Shamir, "Cryptanalysis of the HFE public key cryptosystem", *In Advances in Cryptology _ CRYPTO 1999*, vol. 1666, pp. 19-30, 1999.

[16]. W.S.A. Mohamed, J. Ding, T. Kleinjung, S. Bulygin, J. Buchmann, "Pwxl: A parallel wiedemann-xl algorithm for solving polynomial equations over gf(2)", *Proceedings of the 2nd International Conference on Symbolic Computation and Cryptography (SCC 2010)*, pp. 89-100, 2010.

СРАВНИТЕЛЬНЫЙ АНАЛИЗ СЛОЖНОСТИ МЕТОДОВ ЛИНЕАРИЗАЦИИ И ПЕРЕБОРА РЕШЕНИЯ СИСТЕМ НЕЛИНЕЙНЫХ БУЛЕВЫХ УРАВНЕНИЙ

Проблема нахождения решений систем нелинейных уравнений со многими переменными над конечными алгебраическими структурами и построения эффективных алгоритмов их поиска важна для многих прикладных задач в различных областях и актуальность этой проблемы возрастает со временем. Стойкость многих существующих криптосистем базируется на сложности задачи решения систем нелинейных уравнений многих переменных над конечными полями. В общем виде эта задача является задачей *NP*-полной. Но существует много случаев, когда к таким системам можно предложить методы более быстрые чем методы полного перебора. Поскольку выбор метода может значительно уменьшить время и необходимые ресурсы на нахождение решений системы, естественно возникают вопросы оценки сложности различных методов решения для систем с разными наборами параметров, а также поиска специальных наиболее эффективных методов для конкретного класса систем. В статье рассматриваются наиболее важные для криптографии и криптоанализа системы нелинейных уравнений со многими переменными над конечным полем F_2. Предметом исследования является сравнительный анализ сложности метода линеаризации с введением новых переменных для решения систем нелинейных уравнений над полем F_2 со многими неизвестными и метода полного перебора в зависимости от параметров системы. Целью работы является получение средних оценок сложности методов и нахождения границы в области изменения параметров переопределенной совместной системы уравнений, которая дает возможность из двух указанных методов выбрать более быстрый и эффективный. Предложенны вероятностные модели для получения теоретических, асимптотических оценок средней сложности методов и проведения ряда статистических экспериментов с получением средних оценок методом Монте-Карло. Показано, что существует граница в области изменения параметров, зависисящая, прежде всего, от соотношения максимальной степени уравнений системы и числа неизвестных, которая определяет, когда метод линеаризации работает лучше чем полный перебор. Теоретические и экспериментальные данные применены для построения этой границы. Аналитическое выражение для линии разграничения в области изменения параметров системы получено с использованием метода наименьших квадратов.

Ключевые слова: системы нелинейных случайных булевых уравнений, алгоритмы решения систем, методы линеаризации и полного перебора, сложность алгоритмов, статистическое моделирование, криптоанализ.

COMPARATIVE ANALYSIS OF THE COMPLEXITY OF THE LINEARIZATION AND ENUMERATION METHODS FOR SOLVING SYSTEMS OF NONLINEAR BOOLEAN EQUATIONS

The problem of finding solutions to systems of multivariate nonlinear equations over finite algebraic structures and constructing efficient algorithms for their search is important for many applied problems in various fields, and the relevance of this problem increases over time. The security of many existing cryptosystems is based on the complexity of the problem of solving systems of multivariable nonlinear equations over finite fields. In general, this task is an NP-complete problem. But there are many cases where such systems can be solved by methods faster than the brute-force methods. Since choosing a method can significantly reduce the time and resources required to find system solutions, the questions of assessing the complexity of different methods of solutions for systems with different sets of parameters, as well as finding the special best performing methods for a particular classes of systems, naturally arise. The paper deals with the most important for cryptography and cryptanalysis systems of multivariate nonlinear equations over the finite field F_2. The subject of the study is a comparative analysis of the complexity of the method of linearization with the introduction of new variables for solving systems of nonlinear equations over the field F_2 with many unknowns, and the method of complete enumeration, depending on the system parameters. The purpose of the work is obtaining average estimates of the complexity of the methods, and determining the boundary in the parameter space of the overdefined consistent system of equations, which makes it possible to choose the faster and more efficient method of the two considered. Probabilistic models for obtaining theoretical, asymptotic estimates of the average complexity of the methods, and for conducting a number of statistical experiments with obtaining the average estimates by the Monte Carlo method, are proposed. It is shown that there exists a boundary in the parameter space, depending on, first of all, the relation between the maximum degree of equations of the system and the number of unknowns, which determines when the linearization method works better than the complete enumeration. Theoretical and experimental data have been used to construct this boundary. The analytical expression for the separation line in the parameter space of the system was obtained by the least squares method.

Keywords: systems of nonlinear Boolean equations, algorithms for solving systems, linearization and enumeration methods, complexity of algorithms, statistical modeling, cryptanalysis.

Лещенко Владислав Вадимович, студент Фізико-технічного інституту Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського».
E-mail: vladon4eg1@gmail.com.
Orcid ID: 0000-0001-8601-1365.
Лещенко Владислав Вадимович, студент Физико-технического института Национального технического университета Украины «Киевский политехнический институт имени Игоря Сикорского».
Leshchenko Vladyslav, student of Institute of Physics and Technology, National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute».

Пекарчук Ніна Андріївна, аспірант Фізико-технічного інституту Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського».
E-mail: nina.pekarchuk@gmail.com.
Orcid ID: 0000-0002-9091-4329.
Пекарчук Нина Андреевна, аспирант Физико-технического института Национального технического университета Украины «Киевский политехнический институт имени Игоря Сикорского»
Pekarchuk Nina, PhD student of Institute of Physics and Technology, National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute».

Савчук Михайло Миколайович, доктор фізико-математичних наук, доцент, в.о. завідувача кафедри математичних методов захисту інформації Фізико-технічного інституту Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського».
E-mail: mikhail.savchuk@gmail.com.
Orcid ID: 0000-0001-6580-2694.
Савчук Михаил Николаевич, доктор физико-математических наук, доцент, и.о. заведующего кафедрой математических методов защиты информации Физико-технического института Национального технического университета Украины «Киевский политехнический институт имени Игоря Сикорского».
Savchuk Mykhailo, Doctor of Physical and Mathematical Sciences, docent, Acting Head of the Department of mathematical methods of information security, Institute of Physics and Technology, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute".