DOI 10.15276/imms.v12.no4.337
УДК 004.056.53

Informatics and Mathematical Methods in Simulation
Vol. 12 (2022), No. 4, pp. 337-345

ВИЯВЛЕННЯ ТА ВИПРАВЛЕННЯ ПОМИЛОК У ЗАХИЩЕНИХ СИСТЕМАХ ЗБЕРІГАННЯ ДАНИХ НА ОСНОВІ ОБЧИСЛЕННЯ ПРОЕКЦІЙ ЧИСЛА

С.В. Кулина

Західноукраїнський національний університет, вул. Львівська, 11,
м. Тернопіль, 46020, Україна; e-mail: sersks@wunu.edu.ua

Система розподіленого зберігання даних включає в себе окрім самих пристроїв на яких буде зберігатися інформація, також і канали зв'язку по яких її пересилають. При обміні інформацією на значні відстані виникає ряд нових проблем, які потребують вирішення. Аналіз сучасного стану та розвитку систем передачі та зберігання даних свідчить про те, що є необхідність застосування різноманітних технічних та програмних методів для захисту від спотворення, втрати чи крадіжки інформації. Використання лише одного з них не є ефективним, тому зазвичай використовують їх поєднання. Одним із поширених методів захисту даних від спотворень є використання коригуючих кодів, серед яких важливе місце займають коди на основі системи залишкових класів (СЗК). Завдяки представленню інформації у вигляді залишків від ділення даних на кожен елемент системи взаємно простих модулів досягається зменшення розрядності чисел з якими проводяться операції. У роботі проаналізовано подання інформації у вигляді залишків для реалізації методу розподіленого зберігання даних, завдяки чому можна спростити реалізацію систем збору та передачі інформації. Запропоновано використати для захисту даних розширення початкового діапазону обчислення за рахунок введення надлишкових модулів, завдяки чому ймовірність виявити помилку становить 100%. Проаналізовано переваги та недоліки виявлення та виправлення помилок з використанням коригуючих кодів у СЗК за допомогою методів обчислення синдрому та за допомогою обчислення проекції числа. Кількість інформаційних модулів у проектованих системах може змінюватися у залежності від поставленої задачі, тому у проведених дослідження було визначено, що при збільшенні числа інформаційних модулів та комбінацій спотворених символів кількість невиявлених помилок у відсотковому поданні практично залишається без змін. Запропоновано при проектуванні систем збору інформації використовувати СЗК без додаткового виправлення помилок, оскільки є можливість підібрати набір модулів таким чином, що максимальна кількість помилок що виникатиме не впливатиме на роботу системи.

Ключові слова: Коригуючі коди, система залишкових класів, китайська теорема про залишки, розподілене зберігання даних, методи зворотного перетворення, виправлення помилок.

Вступ

Основою будь-якої системи розподіленого зберігання даних окрім пристроїв на яких буде зберігатися інформація є також канали зв'язку по яких вона пересилається. При цьому варто враховувати, що передача інформаційних пакетів між пристроями однієї мережі є відносно захищеною та менше піддається впливу завад. Проте при обміну інформацією за межами цієї мережі значно зростає ймовірність спотворення, втрати чи крадіжки інформації. В залежності від каналу обміну інформацією використовуються різноманітні технічні рішення, завдяки яким зменшується ймовірність втрати чи спотворення повідомлення.

Проте використання лише технічних рішень не є ефективним, тому поряд з цими методами захисту інформації також застосовують різноманітні алгоритмічні рішення – коригуючі коди, методи криптографічного захисту та багато іншого [1-5]. На даний час дослідження захисту від спотворень, втрати чи несанкціонованого

доступу до інформації при передачі між користувачами та в межах однієї системи є важливою галуззю наукових досліджень [6-7].

Одним із таких методів захисту даних є використання коригуючих кодів на основі системи залишкових класів (СЗК). СЗК дозволяє окрім паралельної обробки даних також і захистити інформацію від спотворення, втрати чи крадіжки шляхом введення додаткових перевірочних модулів, що є значно ефективнішим ніж використання резервного копіювання чи дзеркального збереження масивів даних [8].

Подання чисел у вигляді залишків також дає змогу спростити реалізацію систем збору та передачі інформації, а також дозволяє вирішувати клас задач, що складніше реалізуються у позиційних системах числення [9-10].

Значний теоретичний внесок у розвиток системи залишкових класів та її застосування в системах збору та обробки інформації зробили такі вчені: І. Я. Акушський, Д. І. Юдицький, В. М. Амербаєв, В. А. Торгашев, А. О. Коляда, В. А. Краснобаєв, Я. М. Николайчук, М. І. Червяков, О. А. Фінько, А. Омонді (A. Omondi), Б. Премкумар (B. Premkumar), Дж. Кардаріллі (G. Cardarilli).

Дослідження, що сприяли розвитку мережного кодування даних належать іншим відомим науковцям, серед яких: С. Г.Бунін, В. О. Романов, В. А. Романюк, I. Акиілдіз (I. Akyildiz), Р. Ахлсведе (R. Ahlswede), К. Фрагоулі (C. Fragouli) та ін.

Мета і задачі дослідження

Метою роботи є підвищення надійності систем зберігання даних шляхом використання коригуючих кодів системи залишкових класів.

Для досягнення поставленої мети необхідно вирішити наступні задачі: - розглянути спосіб прямого перетворення СЗК на основі китайської теореми про залишки; - проаналізувати існуючі шляхи виправлення помилок в СЗК за допомогою методу обчислення синдрому та методу проекцій; - провести експериментальні дослідження виявлення помилок у інформаційному повідомленні при одному перевірочному модулі різної розрядності; - провести експериментальні дослідження виявлення помилок у інформаційних повідомленнях при зміні кількості інформаційних модулів.

Надлишкова система залишкових класів

В СЗК інформацію представляють у вигляді залишків $(x_1, x_2, ..., x_i, ...x_k)$ від ділення даних на кожен елемент системи взаємно простих модулів $(p_1, p_2, ..., p_i, ...p_k)$, де k кількість модулів системи [11].

Числове значення даних X, подане в позиційній системі числення, повинно лежати в діапазоні $0 < X < P$, де

$$P = \prod_{i=1}^{k} p_i \,. \tag{1}$$

Пошук залишків відбувається за формулою:

$$x_i = X(\mathrm{mod}\, p_i). \tag{2}$$

Згідно виразу (2) всі числа з діапазону значень P можна представити у вигляді матриці:

$$B=\begin{vmatrix} p_1 & p_2 & ... & p_k \\ 0 & 0 & ... & 0 \\ 1 & 1 & ... & 1 \\ ... & ... & ... & ... \\ p_1-1 & p_1-1 & ... & p_1-1 \\ 0 & p_1 & ... & p_1 \\ ... & ... & ... & ... \\ x_{i,1} & x_{i,2} & ... & x_{i,k} \\ ... & ... & ... & ... \\ p_1-1 & p_2-1 & ... & p_k-1 \end{vmatrix}$$

Розглянемо систему з 3 модулів: $p_i=[3,5,7]$. Загальний діапазон даної системи модулів становитиме: $P=3\cdot5\cdot7=105$, а будь-яке значення X можна знайти згідно формули обчислення залишків (2). Для прикладу, нехай X=10, тоді представлення цього числа у СЗК матиме вигляд (1,0,3).

СЗК дозволяє захистити інформацію при розширенні початкового діапазону обчислення за рахунок введення надлишкових модулів. Основною перевагою цього є можливість виявлення помилки в одному символі з ймовірністю 100% при використанні одного додаткового модуля [12]. При роботі з надлишковими модулями їх називають перевірочними та вводять поняття робочого та загального діапазонів.

Перевірочні символи в коригуючих кодах СЗК обчислюються за формулою [11]:

$$x_{k+a}=X\left(\mathrm{mod}\,p_{k+a}\right),$$

де k – кількість інформаційних модулів, а a кількість перевірочних модулів.

Для виявлення помилок використовують поняття загального $H=\prod_{i=1}^{k+a}p_i$ та робочого (формула 1) діапазонів.

Для відновлення позиційного представлення числа X використовується формула зворотного перетворення на основі китайської теореми про залишки [12]:

$$X=\left(\sum_{i=1}^{k+a}x_i\cdot B_i\right)\mathrm{mod}M, \qquad (3)$$

де B_i – базисні числа СЗК, які обчислюються згідно рівняння:

$$B_i=\frac{M}{p_i}\cdot t_i\equiv1(\mathrm{mod}\,p_i),$$

де t_i – набір коефіцієнтів, які забезпечують ортогональність перетворень та задовольняють умову: $0<t_i<p_i$.

Виправлення помилок методом обчислення синдрому

На даний час основними методами виявлення та виправлення помилок з використанням коригуючих кодів на основі СЗК є виправлення помилок за допомогою обчислення синдрому та за допомогою обчислення проекції числа [13].

Згідно із алгоритмом виявлення помилок за допомогою обчислення синдрому відновлення позиційного числа відбувається на основі формули 3 окремо за кожним модулем робочого та перевірочного діапазонів [14]:

$$Y=\left(\sum_{i=1}^{k}x_i\cdot B_i\right)\mathrm{mod}\,P,$$

$$E = \left(\sum_{i=k+1}^{k+a} x_i \cdot B_i \right) \bmod P_c ,$$

де $P_c = \prod_{i=k+1}^{k+a} p_i$ і задовольняється умова $P_c > P$,

Саме обчислення синдрому відбувається за формулою:

$$s = (Y - E) \bmod P_c .$$

Приклад. Розглянемо використання системи з трьома інформаційних та двома перевірочними модулями:

– система модулів: $p_i = [3,5,7,11,13]$;

– робочий діапазон даної системи модулів буде становити: P=105;

– перевірочний діапазон: $P_c = 143$.

Іншим способом реалізації системи з таким самим робочим діапазоном може бути використання набору модулів з трьома інформаційними і одним перевірочним $p_i = [3,5,7,107]$, що дозволить задовільнити умову $P_c > P$. Проте, як ми можемо бачити розрядність перевірочного модуля значно зростає, що ускладнює реалізацію відновлення.

Проте основою виявлення помилок за допомогою обчислення синдрому є створення таблиці синдрому, у якій кожному значенні помилки відповідає лише одне значення синдрому. Кількість елементів у такій таблиці буде рівне P_c та дозволяє зменшити кількість необхідних обчислень, проте залишається проблема відновлення позиційного числа при зростанні кількості символів у повідомленні [14].

Також варто враховувати те, що при роботі з багаторозрядними модулями таблиця буде збільшуватися у геометричній прогресії, що значно збільшить час та ускладнить вибірку необхідного значення синдрому.

Виправлення помилок методом обчислення проекції числа

Проаналізувавши існуючий метод обчислення проекцій числа, який базується на визначенні, якщо з числа $X = (x_1, x_2, \ldots, x_k)$, вилучити модуль p_i, то саме значення X не зміниться а також проекції цього числа за всіма модулями будуть рівні.

Для виявлення помилки необхідно обчислити позиційне представлення числа X, при цьому, якщо отримане число у діапазоні $[0, \mathrm{P_k})$, то помилки немає, або пошкоджень зазнали залишки за двома і більше модулями.

У випадку якщо значення $X > P$, то сталася помилка у одному з залишків. Для її виправлення необхідно обчислити проекцій числа X за всіма модулями p_i . Якщо значення проекції $X_i < P$, то помилка відбулася за модулем p_i .

У загальному випадку при обчисленні проекцій X_i необхідно розраховувати базисні числа B_i для кожної з проекцій, тобто при використанні системи з шести модулів необхідно розрахувати та зберегти крім базового набору для всіх шести модулів ще і додатково шість наборів по шість модулів для систем проекцій.

У запропонованому спрощеному методі для знаходження проекцій числа X_i використовуються базисні числа B_i, які були розраховані для числа X, а формула зворотного перетворення матиме вигляд:

$$X_j = \left(\sum_{i=1}^{n} x_i \cdot B_i \right) \bmod \frac{P_n}{p_j} \qquad (4)$$

де *j* – номер проекції.

Як видно з формули (4) для обчислення проекцій використовуються ті самі базисні числа, що і для відновлення початкового значення числа X. Це забезпечує спрощення алгоритму виправлення помилок на основі коригуючих кодів СЗК за рахунок зменшення в n раз кількості операцій множення при зворотному перетворенні.

Приклад. Розглянемо приклад використання системи з чотирма інформаційних та двома перевірочними модулями.

Система модулів: $p_i = [5,7,11,13,17]$.

Загальний діапазон для даної системи модулів буде становити: $P_n = 765765$.

Робочий діапазон: $P_k = 3465$.

Базисні числа $M_i = [306306, 656370, 680680, 556920, 412335, 450450]$.

Нехай число X=73, яке у СЗК з обраними модулями матиме вигляд (3, 3, 1, 7, 8, 5) отримало спотворення в третьому розряді, тоді спотворене повідомлення буде мати вигляд (3, 3, 0, 7, 8, 5). Обчислені згідно формули 4 значення Хпоч наведені в таблиці 1.

Таблиця 1

Обчислення значень Хпоч згідно номера проекції

Хпоч	p_1	p_2	p_3	p_4	p_5	p_6	j
85158		3	0	7	8	5	1
85158	3		0	7	8	5	2
73	3	3		7	8	5	3
15543	3	3	0		8	5	4
26253	3	3	0	7		5	5
40113	3	3	0	7	8		6

Як видно з таблиці 1 при обчисленні значення Хпоч, у всіх випадках окрім вірного значення Хпоч $> P_k$. Це наочно показує, що при втраті або пошкодженні значення за одним із модулів, ми можемо відновити втрату маючи два перевірочних символи. На основі проведених обрахунків запропонований метод обчислення проекцій забезпечує спрощення обчислень за рахунок зменшення в шість разів кількості операцій множення при зворотному перетворенні. Це досягається завдяки використанню початкових базисних чисел.

Експериментальні дослідження виявлення помилок у послідовностях інформаційних символів

Умову перевірки на спотворення залишку можна використати для виявлення помилок у двох і більше символах. Для цього були проведені експериментальні дослідження, результатом яких було визначення відсотки виявлення помилок в послідовностях інформаційних символів. При проведених дослідженнях був використаний набір модулів необхідний для передачі 24 бітного інформаційного повідомлення та перевірочний символ різної розрядності, а саме: набір інформаційних модулів p_1=257, p_2=263, p_3=269 та перевірочний модуль починаючи з p_4=271. Оскільки незначне збільшення перевірочного модуля суттєво не впливало на обчислення, то кожне наступне значення перевірочного модуля було більше на один розряд.

У таблиці 2 розрахований загальний діапазон, робочий діапазон та кількість комбінацій з помилками в трьох символах для даних наборів модулів. Не залежно від значення перевірочного модуля робочий діапазон не змінювався і становив 18181979, а кількість комбінацій залишалася рівною 17975296.

Таблиця 2

Обчислення кількісного значення невиявлених помилок

Значення p_4	Загальний діапазон	Робочий діапазон	Кількість комбінацій	Кількість невиявлених помилок	Кількість невиявлених помилок, %
271	4927316309	18181979	17975296	66327	0, 36899
509	9254627311	18181979	17975296	35315	0, 19646
1021	18563800559	18181979	17975296	17606	0, 09795
2039	37073055181	18181979	17975296	8817	0, 04905
4093	74418840047	18181979	17975296	4393	0, 02444
8191	148928589989	18181979	17975296	2195	0, 01221
16273	295875344267	18181979	17975296	1105	0, 00615
32742	595441630271	18181979	17975296	549	0, 00305
65521	1191301446059	18181979	17975296	274	0, 00152

Графічне відображення залежності відсотка невиявлених помилок від перевірочного модуля наведено на рисунку 1.

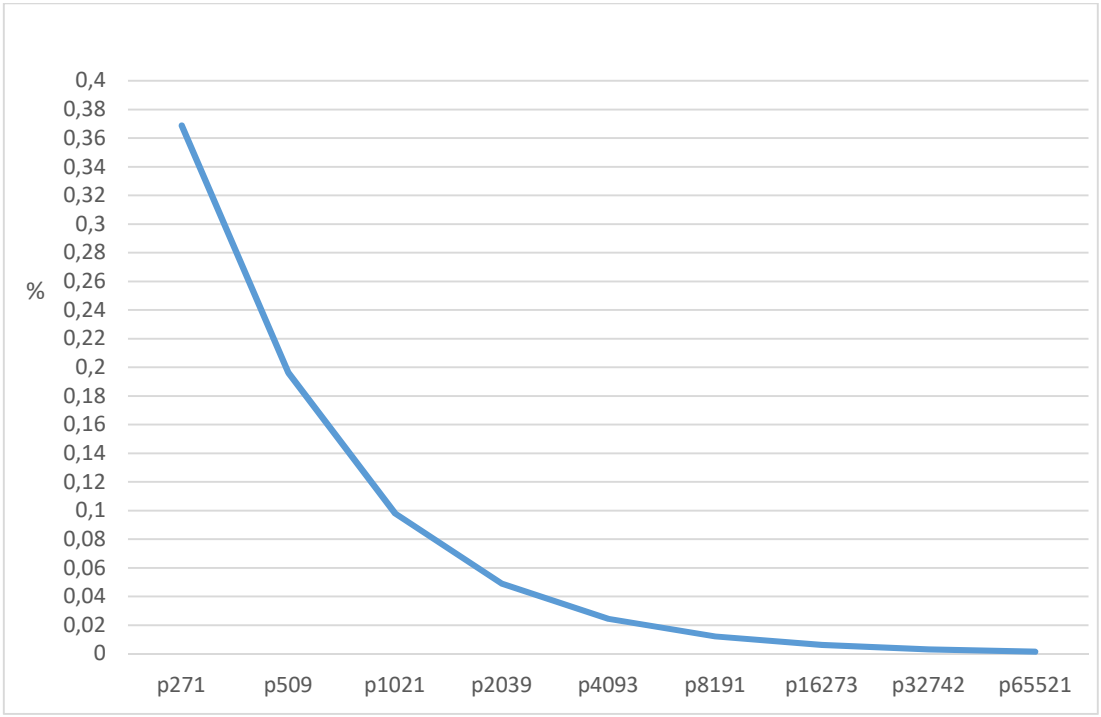

Рис. 1. Залежність кількості (%) невиявлених помилок від перевірочного модуля

За результатами проведеного експериментального дослідження можна зробити висновок, що при збільшенні значення перевірочного модуля на один розряд, кількість невиявлених помилок в трьох символах у заданому діапазоні зменшується в 1,88-2,02 рази. Також з таблиці 2 видно, що збільшення перевірочного модуля з 271 до 65521 дозволяє зменшити кількість невиявлених помилок приблизно у 240 разів.

Зазвичай при зростанні розрядності повідомлення зростає ймовірність спотворення більшої кількості символів. Для обчислення цього значення було проведено експериментальне дослідження залежності кількості невиявлених помилок. Результати розрахунків при спотворенні усіх інформаційних символів та при збільшенні їх кількості наведені в таблиці 3.

Таблиця 3

Обчислення кількісного значення виявлених помилок у всіх модулях окрім перевірочного

Кількість модулів, (n, k)*	Загальна кількість комбінацій	Кількість невиявлених помилок	Кількість невиявлених помилок, %
(3, 2)	67072	251	0,374
(4, 3)	17975296	66327	0,369
(5, 4)	4853329920	17521051	0,361
(6, 5)	1339519057920	4755290656	0,355

* - де n – загальна кількість модулів; k – кількість інформаційних модулів.

Графічне відображення залежності % невиявлених помилок від кількості модулів наведено на рисунку 2.

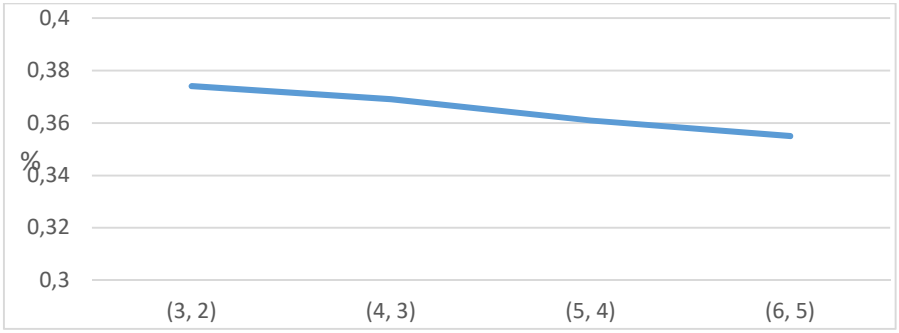

Рис.2. Залежність % невиявлених помилок від кількості модулів

Як видно з рисунку 2 зі збільшенням числа інформаційних модулів та комбінацій спотворених символів кількість невиявлених помилок у відсотковому поданні практично залишається без змін.

На основі результатів дослідження можна зробити висновок, що збільшення кількості помилок при зміні кількості інформаційних модулів прямо пропорційне збільшенню робочого діапазону.

Висновки

Поєднання технічних та програмних методів дозволяє більш ефективно захистити інформацію при передачі по каналам зв'язку. Використання коригуючих кодів на основі СЗК, як одного із методів програмного захисту, дозволяє нам завдяки рівноцінності інформаційних та перевірочних символів відновити втрачену інформацію. Використання методу обчислення синдрому вимагає створення та зберігання таблиці синдромів і тоді виправлення однієї помилки здійснюється шляхом зчитування відповідного значення з таблиці. Зберігання великого масиву даних може негативно позначитись на системах обробки інформації. Іншим недоліком є те, що необхідно задовільнити умову $P_c > P$, проте перевагою даного методу є швидкість виправлення помилок. На відміну від методу пошуку синдрому метод проекцій не потребує збереження масивів даних, проте після виявлення спотворення необхідно провести додаткові обчислення, кількість яких залежать від кількості модулів.

Для обчислення ймовірності виникнення помилок при передачі інформації були проведені експериментальні дослідження, які показали, що при збільшенні значення перевірочного модуля на один розряд, кількість невиявлених помилок в інформаційних символах зменшується в 1,88-2,02 рази.

Кількість інформаційних модулів у залежності від поставленої задачі може змінюватись. У проведених дослідження було визначено, що при збільшенні числа інформаційних модулів та комбінацій спотворених символів кількість невиявлених

помилок у відсотковому поданні практично залишається без змін. При ймовірності помилки менше 0,4 % можливе використання приведених комбінацій модулів для проектування систем захищеного зберігання інформації.

Список літератури

1. Ляшук О. М. Mhed – високоефективний метод захисту даних на основі багатошарового гібридного шифрування. *Вісник НТУУ "КПІ"*. Серія Радіотехніка, Радіоапаратобудування. 2014. №56. С. 144–151.
2. Polotai O., Kukharska N., Lagun A. The steganographic approach to data protection using arnold algorithm and the pixel-value differencing method. *7 Proceedings of the 2020 IEEE 3rd International Conference on Data Stream Mining and Processing, DSMP 2020*. 2020. P. 174–177.
3. Клименко С.В., Малайчук В.П., Селіванов Ю.М., Петренко О.М., Астахов Д.С. Система передачі інформації із застосуванням інтерактивного блокового криптографічного алгоритму TWOFISH. *Актуальні проблеми автоматизації та інформаційних технологій*. 2021. С. 62–71.
4. Barannik V., Havrilov D., Shulgin S., Parkhomenko M., Yroshenko V. The possibility of using the arithmetic coding method in the systems of cryptographic information protection. *Ukrainian Scientific Journal of Information Security*. 2020. Vol. 26. issue 3. P. 156-167.
5. Цаволик Т.Г. Коригуючі коди в системі залишкових класів зі спеціальними модулями. *Вимірювальна та обчислювальна техніка в технологічних процесах*. Технічні наук и. 2016. № 3. С. 100–104.
6. Kaganyuk, O., Chernyashchuk N., Burchak I. Аналіз апаратних та програмних методів захисту інформації від несанкційного доступу до інформаційного каналу передачі даних. *COMPUTER-INTEGRATED TECHNOLOGIES: EDUCATION, SCIENCE, PRODUCTION 47*. 2022. С. 76–82.
7. Горбенко І.Д., Гріненко Т.О. Захист інформації в інформаційно-телекомунікаційних системах. *Харківський національний ун-т радіоелектроніки*. - Х.: ХНУРЕ, 2004. С. 364-368.
8. Яцків, В.В., Кулина С.В. Метод надійного зберігання даних на основі надлишкової системи залишкових класів. *Вісник Хмельницького національного університету*. 2019. С. 98-104.
9. Волинський О. Систематизація характеристик теоретико-числових базисів та їх застосування для побудови високопродуктивних спецпроцесорів. *Вісник Тернопільського національного технічного університету*. 2011. Том 16. №3. С. 183–189.
10. Alrajeh, N.A., Marwat, U., Shams, B., Shah, S.S.H. Error correcting codes in wireless sensor networks: an energy perspective. *Applied Mathematics & Information Scienc-es.* 2015. P. 809-818.
11. Omondi A., Premkumar B. Residue Number System: Theory and Implementation. Imperial College Press. 2007. Vol. 2. 296 P.
12. Yuke Wang. Residue-to-Binary Converters Based On New Chinese Remainder Theorems. *IEEE Transactions on Circuits and Systems – II: Analog and Digital Signal Processing*. Vol. 47. No. 3. March 2000. P. 197–205.
13. Яцків В.В. Виявлення та виправлення багатократних помилок на основі модулярних коректуючих кодів. *Інформаційні технології та комп'ютерна інженерія*. 2015. Том 33. № 2. С. 77–82.
14. Яцків В.В. Теоретичні основи створення і структурна організація компонентів безпровідних сенсорних мереж підвищеної ефективності : дис.... д-ра техн. наук : 05.13.05 / Тернопільський нац.екон. ун-т. Тернопіль, 2016. 280 с.

С.В. Кулина

ERROR DETECTION AND CORRECTION IN PROTECTED DATA STORAGE SYSTEMS BASED ON THE CALCULATION OF NUMBER PROJECTIONS

S.V. Kulyna

West Ukrainian National University, 11 11 Lvivska st.,
Ternopil, 46020, Ukraine; e-mail: sersks@wunu.edu.ua

The system of distributed data storage includes, in addition to the devices themselves, on which information will be stored, as well as communication channels through which it is forwarded. When transmitting information to considerable distances, there are a number of new problems that need to be solved. An analysis of the current state and development of data transmission and storage systems shows that it is necessary to use various technical and software methods of information protection from distortion, loss or theft. The use of only one of them is not effective, so they usually use their combination. The most widely used data protection methods are the methods based on the use of correction codes, particularly the residue number system (RNS) codes. Due to the presentation of information in the form of remainders from the division of data into each element of the system of mutually simple modules, a reduction in the number of digits with which operations are carried out is achieved. The simplification of information collection and transmission system due to the presentation of information in the form of residuals is investigated in this paper. The initial range of calculations expansion implemented by the introduction of additional backup modules is suggested for data protection, that allows to increase the probability of error detection to 100%. The advantages and disadvantages of detecting and correcting errors with the use of corrective codes in RNS using the methods of calculating the syndrome and calculating the number projection is analyzed. The number of information modules in the designed systems can vary depending on the task, therefore, in the conducted studies, it was determined that with an increase in the number of information modules and combinations of distorted symbols, the number of undetected errors in percentage compliance remains practically unchanged. It is suggested to use RNS without additional error correction for designing information collection systems, as it is possible to choose a set of modules in such a way that the maximum number of errors that will not affect the system functionality.

Keywords: Corrective codes, residual number system, Chinese remainder theorem, distributed data storage, reverse transformation methods, error correction.