Článok poukazuje na možnosť použitia šifrovacieho algoritmu s použitím štandardných kódových techník z množiny lineárnych samoopravných kódov. Kvalita dešifrovacieho algoritmu je určená pre Hammingove (n, k) kódy. Zároveň je stanovené pre aké dimenzie Hammingových (n, k) kódov je uvedený algoritmus prakticky použiteľný. Uvedený princíp sa dá zosobniti aj pre iné typy samoopravných kódov.

1 ÚVOD

Komunikačný systém, ako subsystém informačného systému, predstavuje potenciálny cieľ aktívnych a pasívnych útokov na informácie prenášané medzi jednotlivými časťami informačných systémov. Otázka zaistenia bezpečnosti prenášaných informácií je zvlášť významná vo verejných sieťoch, ktoré sú v tomto smere považované za nedôveryhodné. Ochrana dát počas prenosu je preto veľmi dôležitým problémom, ktorým sa treba zaobernut.

Bezpečné služby podľa americkeho štandardu „Trusted Network Interpretation“ sú klasifikované do troch skupín [5]:

- komunikačná integrita,
- odmietnutie služby,
- ochrana dát pred uníkom.

Pracovná verzia pripravovaného telekomunikačného zákona pre Slovenské telekomunikácie sa tiež zaobera problematikou ochrany informácií, sietí a prostredia (časť IV, § 26 „Systém zväčšenej miery utajenia a ochrany prenášanej informácie“). Cieľom novej telekomunikačnej legislatívy na Slovensku je dosiahnuť úroveň telekomunikačných služieb poskytovaných štátmi Európskej únie.

Dominantnými bezpečnostnými mechanismami pri zabezpečení prenosu dát sú kryptografické algoritmy. Moderné šifrovacie systémy využívajú obvykle kombináciu symetrických a asymetrických algoritmov doplnených o certifikáty verejných kľúčov [6]. Zažívanou možnosťou je použitie blokových samoopravných kódov na účely šifrovania. Výhodou tohto riešenia je existencia dostupných kóduových zariadení a zachovanie prenosovej rýchlosti aj napriek použitiu šifrovania, ako je tomu pri použití niektorých asymetrických šifier.

This paper remarks on the possibility of the ciphering algorithm use based on the standard encoding techniques from the linear error-correcting coding area. The quality of deciphering algorithms is determined for (n, k) Hamming codes and the valid code word lengths are recommended for practical use. The presented principle can be generalised for another type of the algebraic codes.

1 INTRODUCTION

The communication subsystem as an important part of an information system is a neglected area for passive and active attacks against transferred information. Specifically, the public networks are regarded as non-trusted networks. This is why a solution for data security problems during transmission plays a very important role.

The network security services according to USA standard “Trusted Network Interpretation” are classified into three groups [5]:
- Communications Integrity,
- Denial of Service,
- Compromise Protection.

The draft of paragraph version of New Telecommunication Law solves the problem of information, networks and intermediate protection in Slovak Telecommunication, too (the part IV, §26 „System of increased rate of secrete and transmitted information protection“). The aim of the new telecommunication legislation in Slovakia is to achieve a level of telecommunication services provided in selected European Union (EU) members.

The dominant security mechanisms for data transmission are cryptographic algorithms that provide the security service as confidentiality, authentication and communications integrity. The modern cryptographic systems use the hybrid combination of symmetric and asymmetric algorithms with certification of public key [6]. The use of a block error - correcting coding in ciphering applications is an interesting possibility. The advantage of this coding is an availability of commercial ciphering equipment. The next advantage is a high code rate with ciphering use.
Suľbu dôvernosti dát pri prenose podľa odporúčania ISO 7492-2 Security Architecture možno poskytnúť v druhej, tretjej, štvrtej, šiestej alebo sedmjej vrstve referenčného modelu OSI [2]. Šifrovanie dát v druhej vrstve referenčného modelu OSI je použiteľné len pre ochranu spojenia typu bod-bod. Výhodou tejto realizácie je transparentnosť dát pre všetky siete a aplikácie. Príklad komunikácie medzi dvoma koncovými stanicami v sieťach s rozhraním typu X.25 je znázornený schematickým modelom na obr. 1. V dolnej časti obrázku je znázornené, ktoré vrstvy modelu OSI sa podieľajú na komunikácii v závislosti na sledovaného prvku siete.

If the entities A and B require to keep privacy of information, the communications system must be expanded by a ciphering encoder before the error-correcting encoder at the transmitter side and the ciphering decoder after error-control coding at the receiver side. For the increasing of data rate (mainly by modem data transmission) it is necessary to use the data compression. The advantage of the compressed cipher text is its resistantancy against some cryptoanalytic’s attack. In the paper the authors do not solve problems of data compression. Further, the described algorithm is supposed to be applied only for the noiseless channel. For the noise channel for the elimination of noise must be channel code included.

2 VLASTNOSTI KRYPTOGRAFICKÉHO SYSTÉMU NA BÁZE HAMMINGOVÝCH (n, k) KÓDOV

Odborné verejnosti je dobre známy princíp kodovania, dekóduvania, detekcie a korekcie chýb u lineárnych systematických (n, k) kódov [1], [4]. Základné principy týchto kódov možno využiť aj na šifrovanie.

2 PROPERTIES OF THE CRYPTOSYSTEM ON THE BASE OF HAMMING (n,k) CODES

Generally it is well known that principles of encoding, decoding, detection and correction of errors with linear systematic (n, k) codes use [1], [4]. The basic principles of these codes can be used also for ciphering.
Kódovanie zdrojovej kóti $z = (z_1, z_2, ..., z_k)$ lineárneho systématického kódu $k \times n$ sa realizuje prostrednictvom rovnosti:

$$u = z \cdot G,$$  \hspace{1cm} (1)

kde $G$ je generujúca matica $k \times n$ lineárneho systematického kódu, ktorá ho jednoznačne určuje.

Hammingové $(n, k)$ kódy patria do množiny lineárnych kódov a pokiaľ ich použijeme na elimináciu šumu v kanáli, majú nasledujúce vlastnosti [4]:
- dĺžka kódu $n = 2^m - 1$.
- počet informačných prvkov $k = 2^m - m - 1$.
- počet zabezpečovacích prvkov $m = n - k$.
- minimalná Hammingova vzdialenosť medzi kóduvými zložkami $d_{\text{min}} = 3$ (v prípade perfektných kódov) a korekčná schopnosť $t = 1$.

Základná myšlienka použitia lineárnych samoopravných kódov $(n, k)$ pre potreby šifrovania spočíva v „utajení“ alebo „zamaskovaní“ generujúcej maticy $G$ po vynásobení maticami $S$ a $P$. Takto získame maticu $K$, ktorá predstavuje klúč takéhoto kryptosystému:

$$K = S \cdot G \cdot P,$$  \hspace{1cm} (2)

kde: $S$ je lubovoľná invertovateľná binárna matica typu $k \times k$,
$P$ je permutačná matica typu $n \times n$, ktorá vznikne z jednotkovej mätice zámenou riadkov a stĺpcov.

Tento systém možno zaradiť medzi systémy s verejným klúčom. Súkromný (tajný) klúč pozostáva z troch matic $S$, $G$, a $P$ a verejný klúč z matic $K$. Verejný klúč je spolu s algoritmom zveřejnený.

Hammingove $(n, k)$ kódy možno použiť na šifrovanie dát, ak sú dodržané nasledujúce podmienky:
- vysielacia strana pozná maticu $K$ (verejný klúč),
- prijímač strana pozná typ Hammingovo $(n, k)$ kódu, matice $S$, $G$, a $P$ (súkromný - tajný klúč) a kontrolnú maticu $H$ na korekciu náhodne generovaného chybového vektoru.

Príslušné šifrovacie zobrazenie $T_k(z)$ takto definovaného kryptosystému je:

$$T_k(z) = z \cdot K + c,$$  \hspace{1cm} (3)

kde $c$ reprezentuje vektor dĺžky $n$, náhodne generovaný vysieľacím správy pre každý blok správy. Prijímač strana prijme signál, ktorý je reprezentovaný vektorom $y = T_k(z)$.

Desšífranie prebieha podľa nasledujúcich krokov [3]:
- nájdenie inverznej permutačnej maticy $P^{-1}$ a výpočet $y \cdot P^{-1}$,
- eliminácia chybového vektoru $c$ pomocou kontrolnej maticy $H$, t.j. výpočet $(y \cdot P^{-1}) \cdot HT$,
- nájdenie kódu $z$ $S$ pomocou generáčnej maticy $G$,
- nájdenie inverznej maticy $S^{-1}$ a výpočet originalného vektoru $z$.

The encoding of a plain text word $z = (z_1, z_2, ..., z_k)$ is as follows:

$$u = z \cdot G,$$  \hspace{1cm} (1)

where $G$ is generating matrix of linear systematic code of the size $k \times n$ and $u$ are the code words.

The Hamming $(n, k)$ codes are the linear block codes with following properties:
- code word length is $n = 2^m - 1$.
- message length is $k = 2^m - m - 1$.
- check parity is $m = n - k$.
- minimal Hamming distance $d_{\text{min}} = 3$ (for perfect codes).
- error-correcting capability $t = 1$ in each code word.

The main idea of the use Hamming $(n, k)$ codes for ciphering a plain text is based on the masking of the generating matrix $G$. Generating matrix is transformed by binary matrices $S$ and $P$ to the matrix $K$ according to:

$$K = S \cdot G \cdot P,$$  \hspace{1cm} (2)

where: $S$ is the binary convertible matrix of the side $k \times k$,
$P$ is the permutation matrix of the side $n \times n$, which it is created from the eye matrix by changing its rows and columns.

This system can be classified as the public key cryptosystem.

The private key consists of three matrices $S$, $G$ and $P$ and the public key of the matrix $K$ only, which is publicly known with the algorithm, too.

The Hamming $(n, k)$ codes can be used as the cipher codes when the following conditions are kept:
- transmitting side knows the matrix $K$ (public key),
- receiving side knows the type of Hamming $(n, k)$ code, the matrices $G$, $S$, $P$ (private key) and the check matrix $H$ for correction of random error vector.

Transformation of this cryptosystem $T_k(z)$ is given by

$$T_k(z) = z \cdot K + c,$$  \hspace{1cm} (3)

where $c$ is the $n$-bits error vector of weight $\leq t$, that is at random generated from the transmitting side for every code word. The receiving side receives the signal, which can be represented by vector $y = T_k(z)$.

The deciphering process is realised according to the following steps [3]:
- determination of the inverse permutation matrix $P^{-1}$ and calculation $y \cdot P^{-1}$,
- elimination of error vector $c$ by check matrix $H$ by calculation $(y \cdot P^{-1}) \cdot H$,
- determination of the code $z \cdot S$ by means of the $G$,
- calculation the original vector $z$ by the binary inverse matrix $S^{-1}$ use.
3 ANALÝZA VÝPOČTOVEJ ZLOŽITOSTI ŠIFRY

Kvalita šifry je daná zložitostou šifrovacieho a dešifrovacieho algoritmu. Zložitosť šifrovacieho a dešifrovacieho algoritmu možno určiť zo počtu cyklov priemerne potrebných na dešifrovanie kryptogramu. Čím je šifra zložitejšia, tým viac cyklov bude potrebných na jej prelomienie a to samozrejme zaberie viac času. Dnes je známy mnoho kryptoanalytických útokov [3]. Algoritmus kvalitnej šifry predpokladá len útok hrubou silou, t. j. vyskúšanie všetkých možných kombinácií klúča.

Kvalita analyzovanej šifry spočíva v tom, že určenie spätnej šifrovacej transformácie \( T(c)^{-1} \) nie je možné prostredníctvom výpočtu inverznej matice \( K^{-1} \), pretože každý odosielaný blok správy je po zašifrovaní znáhodňovaný súčasne s vektorom chyby.

Vzhladom na to je dešifrovanie bez znalosti tajnej časti klúča aj pre malé dimenzie \((n, k)\) ošteplé, čo súvisí s problémom reprezentácie čísla väčšieho ako \(10^{308} \) v pamäti počítača. Hodnoty označené pomlčkou sú možné výpočtené ďalej a to je dôsledkom kriptografických obmedzení, ako napríklad limita rozloženia šifrovacej transformácie \( T(c)^{-1} \) ktorá je možné implementovať iba v praxi.

| m  | k  | n  | r [%] | \( n! \) | \( v_k \) | \( P_{RM} \) |
|----|----|----|-------|--------|--------|--------|
| 3  | 4  | 7  | 42.86 | 5040   | 1181   | 20160  |
| 4  | 11 | 15 | 26.6  | 1.30767 \times 10^{12} | 5.71623 \times 10^{17} | 7.6805 \times 10^{35} |
| 5  | 26 | 31 | 16    | 8.22283 \times 10^{33} | 4.56733 \times 10^{39} | 9.0546 \times 10^{100} |
| 6  | 57 | 63 | 9.52  | 1.98260 \times 10^{47} | 3.05394 \times 10^{52} | 3.21397 \times 10^{77} |
| 7  | 120| 127| 5.51  | 3.01286 \times 10^{53} | 2.51922 \times 10^{58} | 1.96120 \times 10^{83} |
| 8  | 247| 255| 3.17  | 3.35085 \times 10^{64} | 2.36007 \times 10^{71} | 8.04651 \times 10^{164} |
| 9  | 502| 511| 1.76  | 6.79158 \times 10^{86} | 3.66854 \times 10^{94} | 0.0180495 \times 2.2 \times 10^{10} |
| 10 | 1013| 1023| 0.97 | 5.29153 \times 10^{25} | 1.13471 \times 10^{31} | 0.0180495 \times 10^{32} |
| 11 | 2036| 2047| 0.54 | 8.16744 \times 10^{50} | 8.08367 \times 10^{57} | 0.0180495 \times 10^{59} |
| 12 | 4083| 4095| 0.29 | - | 3.20101 \times 10^{87} | - |
| 13 | 8178| 8191| 0.16 | - | - | - |

Zoznamka: Všeobecné uvedené v tab. 1 boli získané v súvislosti s vyhodnotením kriptografického nástroja DERIVE, bežne dostupné programy (napr. MATLAB, EXCEL) dokážu vypočítať max. \(170!\) a obdobné obmedzenia majú aj pre iný výpočet ďalších hodnôt, čo súvisí s problémom reprezentácie čísla väčšieho ako \(10^{308}\) v pamäti počítača. Hodnoty označené pomlčkou sú možné výpočtené ďalej.

Note: The parameters shown in the table were calculated via the programme DERIVE. The commercial programmes (E.g. MATLAB, EXCEL) are able to compute values max. \(170!\). Similar limits are valid also for computation of further values, what causes the problem to represent a number larger than \(10^{308}\) in computer memory. The values marked by symbol "-" are impossible to compute by DERIVE programme.
Výpočtovú zložitosť ovplyvňujú nasledujúce faktory:

A. **výpočet inverznej permutačnej matice** $P^{-1}$
   
   Permutačná matica je veľkosťí $n \times n$ bitov. Nájdenie všetkých permutačných matíc je zložitý problém najmä pre väčšie dimenzie $n$, lebo počet kombinácií odpovedá hodnote $n!$. V tab. 1 je tento parameter vypočítaný maximálne pre kod (2047, 2036).
   
   Po nájdení všetkých inverzných permutačných matíc treba pre každý prijatý vektor $y$ vyriešiť výpočet $y \cdot P^{-1}$.

B. **eliminácia chybového vektora** $c$
   
   Chybový vektor $n$-bitov s vahou $w(c) \leq t$, kde $t$ je počet korisťovaných chýb. Celkový počet rôznych chybových vektov $c \cdot P^{-1}$ pre slová dĺžky $n$ potom je:
   
   $$p = \binom{n}{0} + \binom{n}{1} + \ldots + \binom{n}{t}.$$  \hspace{1cm} (4)
   
   Chybový vektor sa deteguje pomocou techniky známej z teórie korekčných kódov, na základe znalosti kontrolnej matice $H$ veľkosťí $n \times (n-k)$, ktorá sa určí z generujúcej matice $G$ veľkosť SPR c. Počet rôznych chybových vektov $c \cdot P^{-1}$ pre slová dĺžky $n$ potom je:
   
   $$p = \binom{n}{0} + \binom{n}{1} + \ldots + \binom{n}{t}.$$  \hspace{1cm} (4)
   
   The error vector can be detected with the help error-detecting techniques. This algorithm is based on knowledge of the check matrix $H$ of the side $n \times (n-k)$ which can be determined by generating matrix of the side $n \times k$ and from the eye matrix $I$ of the side $(n-k) \times (n-k)$. The column of matrix with error is determined according to expression $y \cdot P^{-1}$, $H^T$. In the next step the error is eliminated. (Note: $H^T$ is a transposed matrix of $H$ matrix.).

C. **hladanie povodného vektora** $z \cdot S$
   
   Ak $G$ je generujúca matica systématického kódu, je tento proces hľadania zjednodušený, pretože z priestoru možných kódových kombinácií $V_n$ stačí sledovať zdrojový podpriestor $V_c$, ktorom sa nachádza kombinácia informačnej časti. Počet takýchto možností je [3]:
   
   $$v_k = \prod_{j=0}^{k-1} \left(2^{(n-j)} - 1\right) \cdot 2^{(n-k)}.$$  \hspace{1cm} (5)
   
   $D. **hladanie inverznej maticy** S^{-1}$
   
   Matica $S$ je veľkosťí $k \times k$ bitov a musí byť invertovateľná. Všetky kombinácie regularných matíc $S$ sa môžu pre nesystématický vypočítať podľa [3]:
   
   $$P_{RM} = 2^k \prod_{j=0}^{k-1} \left(1 - 2^{-j}\right) \approx 2^k \cdot 0.29.$$  \hspace{1cm} (6)
   
   kde $k$ je dĺžka správy.

**ZÁVER**

Z tab. 1 vidieť, že už pre malé dimenzie Hammingových $(n, k)$ kódov je počet kombinácií pri výpočte čiastkových časti klúča značný. Výpočet všetkých kombinácií jednotlivých časti klúča pre dimenzie od $(n, k) \rightarrow (511, 502)$ je obtiažné realizovať v reálnom čase. Pre nájdenie originálneho klúča je potrebné všetky vypočítané kombinácie správne skombinovať, čo je tiež druhým faktorom, ktorý ovplyvňuje zložitosť.

**CONCLUSION**

Table 1 shows that the number of key pieces combination is high already for relatively small code dimensions Hamming $(n, k)$ codes. Computing combinations of all key pieces from dimension $(n, k) \rightarrow (511, 502)$ is realised problematically in real time operation. To find the original key, it is necessary combine the determined combination correctly, which it is time demanding problem, too.
časovo veľmi náročný problém. Možno konštatovať, že analyzo-
vaný šifrovací systém (za predpokladu útoku hrubou silu) je výpo-
cťového složitý.

Uvedený šifrovací systém je vhodný na použitie pre špeciali-
zované prenosy v úrovni linkovej vrstvy, kde sa vyžaduje rýchly
prenos dát pri zaručenej dôvernosti prenášaných informácií.
Výhodné by bolo, keby kóder-dekóder plnil okrem šifrovacej
funkcie aj funkciu korekčnú (myslí sa korekcia chýb spôsobených
šumom). Pre takúto aplikáciu predpokladáme možnosť využitia
algoritmov na báze samooprávnených kódov pre viacnásobné chyby
akými sú napr. algoritmy BCH kódov. U týchto kódov sa zaroveň
zvyšuje odolnosť voči prelomeniu šifry aj pri použití iných kryp-
toanalytických útokov.

Táto práca bola riešená v rámci grantových projektov číslo:
- 1/5255/98 s názvom: „Teoretický aparat pre analyzu
a syntézu systému protokolov komunikačného systému s osobit-
ým sortimentom služieb“. 
- 1/5230/98 s názvom: „Teoretický aparat pre analyzu
a syntézu s definovanou úrovňou bezpečnosti“. 

Recenzenti: D. Levický, P. Tomašov

We can say that the analysed cipher system is complex of
computation (assuming only brute force attack). We would
recommend its use in special applications for fast and confidential
transmission in the line layer of the OSI.

If cipher-decipher algorithms had also the error-correcting
coding function (i.e. elimination influences of channel), it would
be a big advantage. For this application we assume using of multi-
error correcting coding algorithms, e.g. algorithm of BCH codes,
whose ciphering algorithm is more resistant against the other
cryptoanalytic attacks.

This work is a part of grant research projects:
1/5255/98 with the title “Theoretical apparatus for analysis
and synthesis of communication system protocol with special
service set”,
1/5230/98 with the title: “Theoretical apparatus for analysis
and synthesis of system with defined level of safety”.

Recenzed by: D. Levický, P. Tomašov

LITERATÚRA - REFERENCES

[1] CLARC, G. C., CAIN, J. B.: Error Correcting Coding for Digital Communications, Plenum Press, New York, 1988
[2] DOBDA, L.: Ochrana dat v informačních systémech, GRADA, Praha, 1998
[3] GROŠEK, O., PORUBSKÝ, Š.: ŠIFROVANIE, GRADA, Praha, 1992
[4] KONVIT, M.: Teória oznamovania, ALFA, Bratislava, 1989
[5] NOVÁK, L.: Dušová série (Rainbow Series), Hodnocení informační bezpečnosti, Seminář AFOI, Praha, 1995, str. III-1 - III-15
[6] STAÚDEK, J.: Kryptografie a bezpečnost, LanCom, február 1998, str. 14 - str. 22