1. Introducere în tehnica sistemelor de teletransmisie

1.1 Consideraţii generale

Obiectul cursului îl constituie prezentarea tehnicilor şi a procedurilor de transmisie la distanţă a informaţiei, în scopul conducerii automate a unui proces industrial.

Informaţia poate fi definită ca ştire, veste, în strânsă legătură cu conceptul de *comunicaţie* şi cu modul de propagare a energiei asociate semnalului intermediar, precum şi cu modalităţile de stocare a informaţiei.

Până în zilele noastre, cel mai important mijloc de stocare a informaţiei l-a constituit cuvântul scris, iar utilizarea tiparului a însemnat o revoluţie în sensul posibilităţilor de răspândire pe arii largi a informaţiei.

Odată cu începutul erei industriale, s-au dezvoltat tehnici de transmisie rapidă a informaţiei (folosind semnale electrice) pe distanţe mari, în timp relativ scurt: telegraf, telefon, televiziune. Dezvoltări semnificative ale comunicaţiei prin semnale electrice au avut loc în timpul celui de al doilea război mondial , nu numai tehnic – sonarul, radarul – dar şi conceptual, prin dezvoltarea teoriei generale a transmiterii informaţiei (Shannon).
Progresele tehnologice (tranzistori, circuite integrate, microprocesoare, sateliţi de comunicaţie) au făcut ca în prezent sistemele evoluate de comunicaţie să permită transportul în orice punct de pe glob a oricărui tip de informaţie: voce, text, desen.

Totodată, epoca industrială a însemnat creşterea gradului de automatizare a proceselor industriale şi a posibilităţilor de conducere cu calculatorul ale acestora.

Această evoluţie a condus la necesitatea comunicaţiei între diferite echipamente şi operatorul uman. Natura informaţiei transmise a evoluat deci spre simbolurile din tehnica discretă, care a înlocuit în mare măsură tehnica analogică de transmitere a informaţiei.

1.2 Structuri de sisteme de teletransmisie. Mesaje

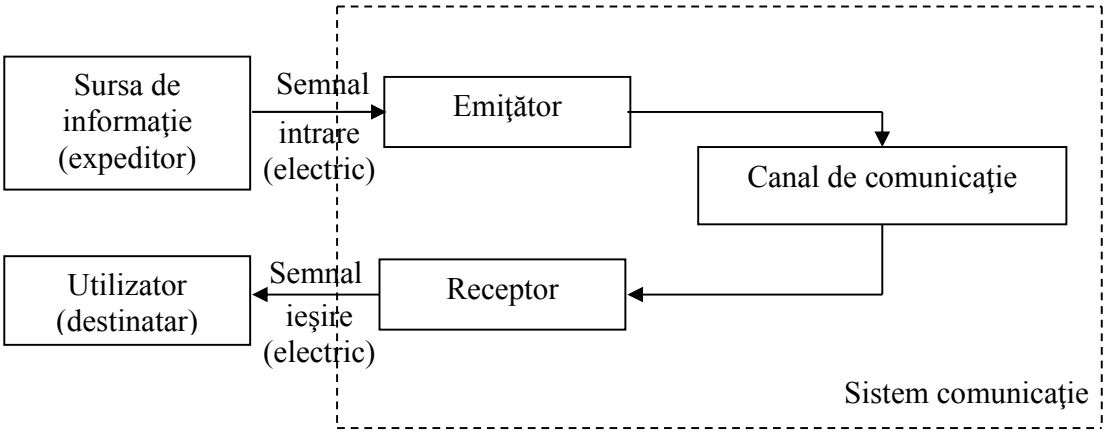

Fig. 1.1 Schema bloc funcţională a unui sistem de comunicaţie

În figura 1.1 este prezentată schema bloc funcţională a unui sistem de comunicaţie în sensul cel mai larg, având ca obiectiv transmiterea informaţiei în timp şi în spaţiu de la un punct numit *sursă* la un alt punct numit *utilizator*. În mod particular, pentru un proces industrial, sursa de informaţie poate fi un traductor, destinatarul fiind un calculator de proces.

Singura restricţie în modelul general din figura 1.1 o constituie natura electrică a semnalelor de intrare şi de ieşire, ceea ce implică necesitatea ca o sursă neelectrică de informaţie să posede un mecanism de conversie a informaţiei în semnal electric variabil în timp, care va fi denumit *semnal mesaj*. La rândul său, canalul de comunicaţie trebuie să permită transmiterea semnalului electric, dar natura sa poate fi diversă: fibră optică, canal radio.

Semnalul poate fi definit ca o mărime fizică de o anumită natură, luând valori într-un domeniu dat, utilizată într-un domeniu aplicativ.

Semnalul poate fi util sau perturbator. La transmiterea prin canalul de comunicaţie poate să apară o degradare a semnalului datorată perturbaţiilor sau distorsiunilor provocate de tehnica de transmisie.

În acest context, principalele cerinţe pentru un sistem de comunicaţie sunt:

- evitarea distorsiunilor;
- minimizarea efectelor perturbaţiilor (rejecţia perturbaţiilor).

Pentru îndeplinirea acestor cerinţe, emiţătorul va prelucra mesajul iniţial, pentru a asigura o transmisie eficientă. Principalele operaţii efectuate sunt: *amplificare, filtrare* şi *modulare*, ultima fiind esenţială în adaptarea semnalului mesaj la caracteristicile canalului. Ea oferă totodată posibilităţi de reducere a efectelor perturbaţiilor şi de transmitere simultană a mai multor mesaje.

La rândul său, receptorul va fi astfel conceput încât să permită extragerea cât mai fidelă a semnalului mesaj din forma degradată a semnalului de ieşire din canal. Acest lucru se obţine esenţial prin operaţia de *demodulare*, la care se adaugă *filtrare* şi *amplificare*.

În funcţie de metoda de modulaţie folosită şi de natura semnalului de ieşire al sursei de informaţie, sistemele de comunicaţie se clasifică astfel:

- *sisteme analogice de comunicaţie*, care se caracterizează prin faptul că transmit informaţii analogice folosind tehnică analogică de comunicaţie;
- *sisteme numerice de comunicaţie*, care transmit informaţii numerice folosind tehnică digitală de comunicaţie;
- *sisteme hibride de comunicaţie* care folosesc tehnici numerice de modulaţie pentru a transmite valori discretizate în timp / nivel ale unor mesaje analogice.

În curs, referirile se vor face exclusiv la sistemele numerice de transmitere a informaţiei sub formă de secvenţe de simboluri (date numerice), cu unele completări referitoare la alte categorii de sisteme de comunicaţie.

În figura 1.2 este prezentat modelul cu blocuri funcţionale al unui sistem numeric de comunicaţie, în care mesajele sursă şi utilizator sunt secvenţe de simboluri binare.

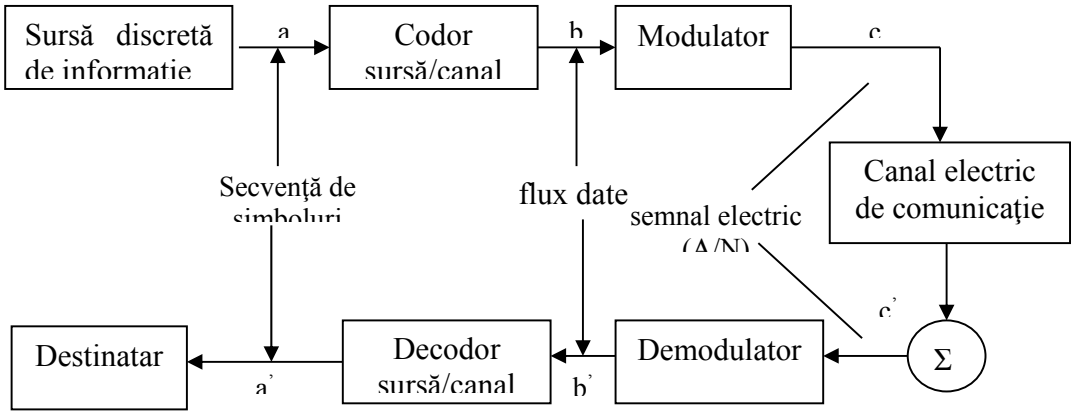

Fig.1.2 Sistem numeric de comunicaţie

În mod suplimentar faţă de schema din figura 1.1 apar blocurile de codare/decodare, specifice tratării discrete a informaţiei. Blocul de codare are în componenţă 2 subansamble:

- blocul de codare sursă (care transpune mesajul în alfabetul sursei);
- blocul de codare canal (care transpune mesajul în alfabetul canalului).

Prin tehnicile de codare, o secvenţă de simboluri capătă o anume semnificaţie, anumite reguli semantice permiţând depistarea la decodare a eventualelor erori apărute în timpul transmisiei şi, în unele cazuri, corectarea acestora. Tehnicile de codare/decodare permit, în plus, şi creşterea vitezei de transmisie în canal.

În continuare sunt prezentate câteva consideraţii generale privind specificitatea diferitelor blocuri din schema din figura 1.2.

a) sursa de informaţie

Există 2 categorii, după natura semnalului de ieşire:

- surse analogice (continue)

exemplu: semnalul oferit de un microfon la care se vorbeşte;

- surse numerice (discrete)

exemplu: ieşirea calculatorului spre imprimantă.

Sursele numerice sunt caracterizate de:

- alfabetul sursei, definit ca o mulţime finită de simboluri ireductibile care conţin informaţii;
- viteza de emisie a simbolurilor;
- probabilitatea de apariţie a unui simbol.

b) blocuri de codare/decodare

Intrarea în codor este o secvenţă de simboluri ce apar cu viteza v_s (simb/s). Codorul sursă converteşte secvenţa de simboluri într-o secvenţă de valori binare 0 sau 1, iar codorul canal grupează aceste simboluri binare în cuvinte. Cuvintele pot fi de lungime fixă sau variabilă, alegerea eficientă a lungimii făcîndu-se în funcţie de probabilitatea de apariţie a simbolurilor şi de nivelul perturbaţiilor în canal.

Problema esenţială a codării constă în găsirea unui compromis între o transmisie eficientă (caracterizată de o viteză mare) şi una cât mai sigură (cu o rată a erorii cât mai redusă). Ultima cerinţă impune folosirea unor simboluri de corecţie suplimentare, având drept consecinţă creşterea timpului de transmisie.

c) blocuri modulator/demodulator

Modulatorul asigură minimizarea efectelor perturbatoare ale canalului, prin folosirea unor semnale de putere şi bandă sporită.

Demodulatorul are drept efect extragerea mesajul din semnalul obţinut la ieşirea canalului, prin tehnici adecvate ce depind evident de tipul de modulaţie utilizat.

d) canal de comunicaţie

Este un circuit fizic de tip electric/electromagnetic, cu o bandă de trecere limitată şi un anumit efect alternator asupra semnalului. La aceasta se mai adaugă zgomotele aleatoare care degradează semnalul-mesaj iniţial. De aceea, canalul va fi caracterizat esenţial prin raportul semnal/zgomot *s/z* ce poate fi menţinut la ieşirea canalului.

e) alte blocuri funcţionale, nefigurate în schemă – blocuri de filtrare, circuite de ceas şi de sincronizare, blocuri de egalizare/adaptare pentru compensarea schimbărilor caracteristicilor canalului. Existenţa unor astfel de blocuri conduce la structuri diferite ale sistemelor de comunicaţie.

In acest punct al prezentării, trebuie subliniat faptul că figura 1.2 este o schemă pur teoretică, deoarece priveşte unilateral transmisia de date. Ca urmare, trebuie adoptată o soluţie practică prin care să fie asigurată circulaţia datelor în două sensuri, între 2 ETTD (echipamente terminale de transmisie de date, care înglobează blocurile codor/decodor şi de sincronizare), soluţie prezentată în figura 1.3.

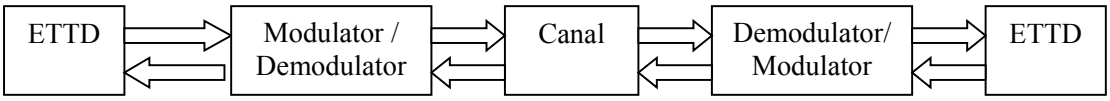

Fig. 1.3 . Sistem de comunicaţie bidirecţional

Un astfel de sistem de comunicaţie este un sistem comun de tip *punct la punct* / *port la port*, ce reprezintă doar o etapă în evoluţia acestui tip de sisteme, care s-au dezvoltat ulterior sub formă de sisteme de comunicaţie *multipunct* şi de *reţele de transmisie de date*, în care numeroase terminale pot efectua schimburi complete de informaţie prin sisteme standard de interfaţă fizică şi logică (reţele de calculatoare).

De altfel, mijloacele actuale de comunicaţie fac astăzi posibile lucruri care amintesc de science-fiction-ul de ieri:

- enciclopedii într-un disc de 4 inchi;
- cumpărături făcute de acasă prin calculator;
- teleconferinţe internaţionale;
- telefonie mobilă.

În mod remarcabil, nici unul dintre aceste sisteme şi servicii nu erau disponibile acum 20 de ani. Evident, există şi o ierarhizare a lor, dictată de anumite criterii generale cum sunt:

- *mobilitatea*, care se referă la situaţiile în care mediul este utilizat de emiţători/receptori în locaţii fixe sau mobile;
- *formatul de transmisie*: imagine, text, audio şi date;
- *capacitatea de transmisie*, caracterizată de faptul că mesajele variază în complexitate în cadrul aceleiaşi categorii de format şi între formate. În general, cu cât este

mai mare complexitatea mesajului, cu atât trebuie să fie mai mari viteza de transmisie şi capacitatea de procesare;

- *combinaţia emiţător – receptor*, care se referă la modul în care sunt conectate cele trei mari categorii de perechi emiţător – receptor: persoane (p) , grupuri de persoane (g) şi maşini (m), creîndu-se astfel *reţele informaţionale,* conform celor prezentate în tabelul 1.1.

Tabelul 1.1

p *la* p	p *la* g	p *la* m
g *la* p	g *la* g	g *la* m
m *la* p	m *la* g	m *la* m

- *game de semnale*: locale, regionale, naţionale, internaţionale, globale;
- *arii de răspândire*, după cum urmează:

- un sistem prin *cablu* poate acoperi un singur complex de locuinţe sau un întreg oraş;
- un sistem *pager* poate ajunge la limitele oraşului sau poate fi extins la o regiune;
- un sistem *teletext* poate avea o audienţă locală sau naţională prin distribuţie prin satelit;

- *interactivitate*, care se referă la cele două mari categorii:

- transmitere într-un singur sens / fără interacţiune;
- transmitere în două sensuri, cu nivel înalt de interactivitate, cel mai simplu exemplu fiind cel al sistemul telefonic, cu varianta sa modernă – tehnologia mobilă, care asigură ultimul tip de legătură şi cel mai complex din sistemul de comunicaţii: *legătura dintre oameni* şi *reţelele informaţionale,* materializată prin serviciul de telefonie mobilă, care a cunoscut o extindere notabilă începând cu anul 1949. În aceeaşi categorie trebuie menţionaţi şi *sateliţii de comunicaţie*, o prezenţă familiară în peisajul comunicaţional actual, care şi-au făcut debutul în octombrie 1957 (primul satelit de comunicaţie Sputnik -URSS), cu observaţia că în SUA, uzul acestora a fost limitat până în 1972 la domeniul militar şi guvernamental. Din 1972, domeniul s-a extins, apărând aşa-numiţii "domsats", cu utilizări şi în alte domenii, preponderent ştiinţifice.

1.3 Informaţia. Măsura cantităţii de informaţie

Materialul prezentat în această parte a capitolului se bazează pe munca de pionierat a lui Shannon, prezentată pentru prima dată în anul 1948, în "Bell Technical Journal", unde au fost expuse rezultatele cercetărilor sale care reprezintă *bazele tehnologiei comunicaţiei*.

1.3.1.Scurt istoric al conceptului de informaţie

În perioada anilor 1920–1930, Robert Fischer a stabilit criterii pentru evaluarea estimaţiilor statistice, astfel încât, plecând de la date de observaţie, să se poată estima parametrii unei distribuţii de frecvenţe, numite *probabilităţi*. El a observat că poate izola un termen care nu depinde de datele de observaţie, ci numai de probabilităţile efective. Această expresie matematică a fost denumită *informaţie conţinută în observaţie*, şi este prima menţionare ştiinţifică a noţiunii.

În anul 1927 Robert Hartley pune bazele teoriei statistice a comunicaţiei. El încearcă stabilirea unei măsuri cantitative prin care să se poată compara capacităţile diferitelor sisteme de a transmite informaţie. Hartley adoptă ca măsură practică a informaţiei *logaritmul numărului de secvenţe de simboluri posibile*, definind *capacitatea de informaţie a unui sistem* prin

$$C = \log N = \log m^n = n \log m \qquad (1.1)$$

cu:
m- numărul de stări posibile ale unei unităţi de memorie;
m^n- numărul de stări pentru *n* unităţi (secvenţe de simboluri).

Analiza acestei măsuri evidenţiază atât un avantaj, care se referă la faptul că mărimea permite comparaţii cantitative uşoare, cât şi un dezavantaj, legat de procesul de selecţie a semnalului ce trebuie transmis. În acest sens, măsurarea propriu–zisă a informaţiei este dificilă, procesul fiind cu atât mai complex cu cât setul de semnale din care se face alegerea este mai mare; drept urmare, s-a decis (Shannon, 1948) luarea în consideraţie a *probabilităţii de apariţie a unui anumit tip de semnal*.

În anul 1948 Shannon stabileşte, aşadar, unitatea de măsură a informaţiei, care să nu depindă de natura acesteia (aşa cum starea unui corp nu depinde de natura fizică a acestuia).
Shannon porneşte de la premisa că orice informaţie asupra unor evenimente contribuie la scăderea gradului de incertitudine asupra realizării evenimentelor respective. Astfel, din punctul de vedere al utilizatorului, comunicaţia este o variabilă aleatoare, continutul informaţional al unei ştiri fiind cu atât mai mare cu cât există mai puţine aşteptări referitoare la realizarea acesteia.

1.3.2.Formularea matematică a problemei

Fie A un experiment care evidenţiază n evenimente aleatoare $a_1, a_2, ..., a_n$ cu probabilităţile de apariţie aferente $p_1, p_2, ..., p_n; p_i > 0, i = \overline{1,4}$ şi $\sum_1^n p_i = 1.$

Acest *experiment* evidenţiază un anumit *câmp de probabilităţi* $\{A, a_i, p_i\}$, caracterizat de *repartiţia* $A = \begin{vmatrix} a_1 & a_2 & \dots & a_n \\ p_1 & p_2 & \dots & p_n \end{vmatrix}$.

De exemplu, experimentul A, caracterizat de câmpul de probabilităţi $\begin{vmatrix} a_1 & a_2 \\ 0.1 & 0.9 \end{vmatrix}$, are, din punct de vedere *calitativ*, un grad de incertitudine mai mic decât experimentul B, caracterizat de câmpul de probabilităţi $\begin{vmatrix} b_1 & b_2 \\ 0.5 & 0.5 \end{vmatrix}$

Dar *cantitativ*, se impune referirea la noţiunea de *probabilitate condiţionată*.
Fie A şi B două evenimente; se defineşte probabilitatea condiţionată

$$P(A/B) = \frac{P(A*B)}{P(B)} = \frac{P(A \cap B)}{P(B)} \tag{1.2}$$

ca schimbarea probabilităţii $P(A)$ *de apariţie a evenimentului A când s-a realizat evenimentul B.*
În cazul particular

$$A \subset B \Rightarrow P(A*B) = P(A) \rightarrow P(A \setminus B) = \frac{P(A)}{P(B)} \geq P(A) \tag{1.3}$$

Se observă că informaţia "*B* realizat", adică $P(B)=1$, creşte probabilitatea lui *A*, adică se micşorează incertitudinea asupra realizării evenimentului *A*.

Utilizând o funcţie logaritmică, se *apreciază numeric incertitudinea asupra realizării evenimentului A, I(A)*.

$$\log\frac{1}{P(A/B)}=\log\frac{1}{P(A)/P(B)}=\log P(B)-\log P(A)=$$
$$\log\frac{1}{P(A)}-\log\frac{1}{P(B)}\le\log\frac{1}{P(A)} \tag{1.4}$$

ceea ce se scrie

$$I(A/B)=I(A)-I(B) \tag{1.5}$$

unde s-a folosit notaţia $I(A)=\log\frac{1}{P(A)}=-\log P(A)$

Un caz particular sugestiv este $A=B\Rightarrow I(A/A)=I(A)-I(A)=0$, adică incertitudinea asupra lui A se anulează la realizarea lui *A*.

Rezultă că se poate stabili o echivalenţă între incertitudinea asupra realizării unui eveniment şi realizarea lui.

Informaţia care se obţine prin realizarea evenimentului x_i *de probabilitate* p_i va fi

$$I(x_i)=-k\log_b p(x_i),\ \text{cu}\ k=k(b) \tag{1.6}$$

A alege o unitate de informaţie (de incertitudine) revine la a-l alege pe *b*. Există următoarele cazuri:

- b=e, situaţie în care unitatea de informaţie se numeşte *nit* sau *nat* (natural unit);
- b=2, situaţie în care unitatea de informaţie se numeşte *bit* (binary unit), care nu trebuie confundat cu bit (binary digit) corespunzător cifrelor binare 0/1;
- b=10, situaţie în care unitatea de informaţie se numeşte *decit* (decimal unit).

Se defineşte astfel *bit-ul ca informaţia care se obţine prin realizarea unui eveniment din două evenimente echiprobabile* ($p(x_i)=\frac{1}{2}$) ($k=1$)

$$I=-\log_2\frac{1}{2}=1bit \tag{1.7}$$

$$1nit=-\log_2\frac{1}{e}=\frac{1}{\ln 2}=1.44bit \tag{1.8}$$

$$1decit=-\log_2\frac{1}{10}=\frac{1}{\lg 2}=3.32bit \tag{1.9}$$

În tabelul 1.2 este prezentată comparativ corespondenţa unităţilor de informaţie.

Tabelul 1.2

	1 bit	1 nit	1 decit
1 bit	1	0.693	0.301
1 nit	1.443	1	0.434
1 decit	3.322	2.303	1

În cele ce urmează, *va fi folosit exclusiv bit-ul ca unitate de măsură a informaţiei*, ceea ce subliniază importanţa utilizării tehnicii binare în codificare.

Aplicaţia 1.1.
Considerând 4 mesaje m_1,m_2,m_3,m_4, cu probabilităţile de apariţie asociate $p_1=\frac{1}{2};\ p_2=\frac{1}{4};\ p_3=\frac{1}{8};\ p_4=\frac{1}{16}$, să se determine informaţia conţinută în fiecare mesaj.

$$I(m_i)=-\log_2 p_i \tag{1.10}$$

$$I(m_2)=-\log_2\frac{1}{4}=-\log_2 1+\log_2 4=2biti \tag{1.11}$$

$$I(m_1)=1bit;\ \ I(m_3)=3bit;\ \ I(m_4)=4bit \tag{1.12}$$

1.3.3. Entropia informaţională. Definiţie. Proprietăţi

Pentru un experiment cu N rezultate echiprobabile, se presupune că fiecare rezultat în parte introduce o nedeterminare egală cu a $\frac{1}{N}$- a parte din nedeterminarea totală, deci cu

$$\frac{1}{N}\log N = -\frac{1}{N}\log\frac{1}{N} \tag{1.13}$$

Generalizând, se poate concluziona că măsura nedeterminării unui experiment cu n evenimente $a_1,...,a_n$, caracterizate de probabilităţile $p_1,...,p_n$, $p_i \geq 0, \sum_1^n p_i = 1$, este expresia

$$H(p_1, p_2,..., p_n) = -\sum_1^n p_i \log p_i \tag{1.14}$$

şi este denumită *entropie* (Shannon).

În acest context, câteva dintre proprietăţile relevante ale entropiei sunt prezentate în continuare.

1. Cu convenţia $p \log p = 0$, entropia anterior definită este o funcţie pozitivă, simetrică şi continuă.

2. $(\forall)\ p_1,\ p_2,...,\ p_N,\ H(p_1, p_2,..., p_n) \leq H\left(\frac{1}{n}, \frac{1}{n},..., \frac{1}{n}\right)$ (1.15)

Această proprietate arată că entropia este maximă atunci când evenimentele sunt echiprobabile.

3. Se consideră câmpul de evenimente

$$A = \begin{vmatrix} a_1 & a_2,... & a_{n-1} & ...a_n \\ p_1 & p_2,... & p_{n-1} & ...p_n \end{vmatrix} \tag{1.16}$$

Fie evenimentul a_n împărţit în evenimentele disjuncte $b_1,..., b_m$, $a_n = b_1 \cup b_2 \cup ... \cup b_m$, cu probabilităţile asociate $q_1, q_2,... q_m$, $\sum_1^m q_j = p_n$.

S-a format astfel un nou câmp de evenimente

$$(A, B) = \begin{vmatrix} a_1 & a_2 & ...a_{n-1} & b_1... & b_m \\ p_1 & p_2 & p_{n-1} & q_1... & q_m \end{vmatrix} \tag{1.17}$$

cu entropia

$$H(p_1,...,p_{n-1},q_1,...q_m) \geq H(p_1,...,p_n) \tag{1.18}$$

Prin împărţirea unui eveniment în cât mai multe evenimente, entropia nu poată să scadă (de regulă creşte).

În vederea caracterizării sistemului de transmisie de date, de un interes deosebit este studiul entropiei legilor compuse.

Fie două experimente *A,B,* caracterizate prin câmpurile:

$$A = \begin{vmatrix} a_1 .... a_n \\ p_1 ... p_n \end{vmatrix} \; ; \; B = \begin{vmatrix} b_1 ... b_m \\ q_1 ... q_m \end{vmatrix} \tag{1.19}$$

$$p_k > 0, k = \overline{1;n}, \sum_1^m p_k = 1$$

cu

$$q_l > 0, l = \overline{1;m}, \sum_1^m q_l = 1$$

Când evenimentele din cele două experimente nu se condiţionează reciproc, *evenimentul cumulat (A,B)* definit de apariţia simultană a unui eveniment a_k din *A* şi a unui eveniment b_l din *B* este caracterizat prin probabilitatea

$$\pi_{kl} = p_k \cdot q_l \text{ , cu } \sum_1^n \sum_1^m \pi_{kl} = 1 \tag{1.20}$$

Entropia experimentului cumulat va fi

$$H(A,B) = -\sum_1^n \sum_1^m \pi_{kl} \log \pi_{kl} = H(A) + H(B) \tag{1.21}$$

Entropia unui experiment alcătuit din mai multe experimente independente este egală cu suma entropiilor experimentelor independente.

Situaţia se modifică atunci când probabilităţile de apariţie a evenimentelor b_1,..., b_m *sunt condiţionate* de apariţia evenimentelor a_1,...a_n..

Se consideră că apariţia unui eveniment a_k din *A* implică pentru *B* o schemă de repartiţie de forma

$$a_k \rightarrow \begin{vmatrix} b_1 ..... b_m \\ q_{k1} ... q_{km} \end{vmatrix} \text{ , cu } \sum_1^m q_{kl} = 1 \tag{1.22}$$

Experimentul compus care reflectă realizarea evenimentului b_l condiţionată de apariţia evenimentului a_k este în acest caz caracterizat de proprietatea

$$\pi_{kl} = p(a_k, b_l) = p(a_k) \cdot q_{kl} \tag{1.23}$$

şi în această situaţie există un câmp complet de evenimente, deoarece

$$\sum_{k=1}^{n}\sum_{l=1}^{m}(p_k \cdot q_{kl}) = 1.$$

Entropia experimentului *B* condiţionat de apariţia evenimentului a_k este dată de relaţia:

$$H_k(B) = H(q_{k1},...,q_{km}) = -\sum_{1}^{m} q_{kl} \log q_{kl} \tag{1.24}$$

iar entropia experimentului *B* condiţionat de realizarea experimentului *A* va fi

$$H_A(B) = H(B/A) \underline{\Delta}\sum_{1}^{n} p_K H_K(B) = -\sum_{k1}^{n} p_k \sum_{l1}^{m} q_{kl} \log q_{kl} \tag{1.25}$$

Entropia experimentului compus *(A,B)* se calculează

$$H(A,B) = -\sum_{1}^{n}\sum_{1}^{m} \pi_{kl} loh\pi_{kl} = ... = H(A) + H(B/A) = H(B) + H(A/B) \tag{1.26}$$

Aşadar, în cazul evenimentelor condiţionate, entropia experimentului compus este mai mică decât în cazul evenimentelor independente.

1.4 Caracterizarea entropică a sistemelor de transmisie de date

1.4.1 Definiţii. Proprietăţi

Transmiterea de date (de informaţie) poate fi considerată un exemplu particular de experiment compus. În acest sens se pot face următoarele consideraţii:

1. *Sursa de transmitere a informaţiei este considerată experimentul X* reprezentat prin câmpul de probabilităţi $\{X, x, p(x)\}$ şi schema de repartiţie

$$X = \begin{vmatrix} x_1 & x_2 & .....x_n \\ p(x_1) & p(x_2) & ...p(x_n) \end{vmatrix} \tag{1.27}$$

unde x_i sunt simbolurile alfabetului sursă, $i = \overline{1,n}$, iar $p(x_i) > 0, p(x_i)$ este probabilitatea ca să fie emis simbolul x_i, $\sum_{1}^{n} p(x_i) = 1$

Sursa este caracterizată de entropia

$$H(X) = -\sum_{1}^{n} p(x_i) \log p(x_i) \tag{1.28}$$

2. *Receptorul sistemului de transmitere a informaţiei este considerat experimentul Y* reprezentat prin câmpul de probabilităţi $\{Y,y,p(y)\}$, cu schema de repartiţie

$$Y = \begin{vmatrix} y_1 & y_2 & .....y_n \\ p(y_1) & p(y_2) & ...p(y_n) \end{vmatrix} \tag{1.29}$$

unde y_j sunt simbolurile alfabetului recepţiei, $j = \overline{1,m}$

$p(y_j)$ este probabilitatea să fie recepţionat simbolul y_j, şi $\sum_1^m p(y_j) = 1$

Recepţia este caracterizată de entropia

$$H(Y) = -\sum_1^m p(y_j) \log p(y_j) \tag{1.30}$$

3. *Experimentul compus care caracterizează transmiterea informaţiei (X,Y)* constă în realizarea evenimentului (x_i, y_j), ceea ce înseamnă recepţia simbolului y_j atunci când a fost emis simbolul x_i şi este caracterizat de

$$\{X, Y), (x, y), p(x, y)\}, \text{ cu } \sum_{i=1}^n \sum_{j=1}^m p(x_i) p(y_j) = 1 \tag{1.31}$$

Acest experiment compus este definit de entropia

$$H(X,Y) = -\sum_1^n \sum_1^m p(x_i, y_j) \log p(x_i, y_j) \tag{1.32}$$

Se deduc relaţiile

$$\sum_{j=1}^m p(x_i, y_j) = p(x_i) \text{ şi } \sum_{i=1}^n p(x_i, y_j) = p(y_j) \tag{1.33}$$

În cazul în care transmisia se efectuează fără perturbaţii, cunoaşterea câmpului de evenimente de la recepţie permite identificarea mesajului emis. În realitate, *existenţa perturbaţiilor conduce la incertitudine asupra mesajului emis*. Valoarea medie a acestei incertitudini este dată de entropia câmpului *X* condiţionat de câmpul *Y, H(X|Y)*.

În aceste condiţii, considerând probabilitatea condiţionată $p(x_i / y_j)$ ca la intrarea în canal să fie emis simbolul x_i când la ieşire se recepţionează simbolul y_j, formula de calcul este

$$p(x_i / y_j) = \frac{p(x_i, y_j)}{p(y_j)} \equiv p(x / y) = \frac{p(x, y)}{p(y)} \qquad (1.34)$$

de unde rezultă probabilitatea condiţionată *H(X|Y)*.

Analog, considerând probabilitatea de a recepţiona semnalul y_j când se emite semnalul x_i

$$p(y_j / x_i) = \frac{p(x_i, y_j)}{p(x_i)} \qquad (1.35)$$

de unde rezultă probabilitatea condiţionată *H(Y |X)*.

Cunoaşterea probabilităţii condiţionate $p(y / x)$ *înseamnă, de fapt, cunoaşterea canalului de transmitere a informaţiei*. ***Configuraţia*** $\{X, p(y / x), Y\}$ ***reprezintă configuraţia de bază a sistemului de transmitere a informaţiei.***

Proprietăţile cele mai importante care definesc din punct de vedere entropic sistemul de transmisie de date sunt prezentate în continuare.

Echivocaţia $H(x \setminus y)$, definită ca fiind măsura echivocului care există asupra câmpului de intrare *X* când se cunoaşte câmpul de ieşire Y.

Eroarea medie de transmisie $H(y / x)$, definită ca măsura incertitudinii care există asupra câmpului de ieşire când se cunoaşte câmpul de intrare.

$$H(y,x) = H(x) + H(y/x) = H(y) + H(x/y) \qquad (1.36)$$

cu situaţiile definitorii:

a) la perturbaţie nulă

$$H(y/x) = 0 \Rightarrow H(x,y) = H(x) = H(y) = 0 \qquad (1.37)$$

b) la perturbaţii foarte puternice, câmpurile de intrare / ieşire în/din canal devin independente şi deci

$$H(x/y) = H(x);\ H(y/x) = H(y) \ \Rightarrow\ H(x,y) = H(x) + H(y) \qquad (1.38)$$

Din punct de vedere al transmisiei, cea mai relevantă caracterizare o oferă *cantitatea de informaţie medie care trece prin canal*, adică valoarea medie a informaţiei care se obţine asupra câmpului de la intrare, X, când se cunoaşte câmpul de ieşire Y. Această mărime este denumită *transinformaţia I(x,y)* şi se calculează

$$I(x,y) = H(x) - H(x/y) \tag{1.39}$$

Capacitatea canalului este definită ca fiind valoarea maximă a transinformaţiei

$$C = \max I(x, y) \tag{1.40}$$

Redundanţa canalului $R_C = C - I(x, y)$ (1.41)

Eficienţa canalului, care arată cât de mult se apropie transinformaţia de valoarea ei maximă.

$$\eta_C = \frac{I(x, y)}{C} \tag{1.42}$$

Aplicaţia 1.2.

Se consideră o sursă discretă care emite la fiecare milisecundă un simbol din cinci simboluri posibile ale căror probabilităţile asociate sunt $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16}$. Se cere entropia sursei şi viteza de transmisie a sursei, v_s.

$$H(s) = H(x) = -\sum_1^5 p_i \log p_i = 1.875 \quad \text{bit/simbol}$$

$$v_s = v \cdot H(x)$$

cu v – viteza fixă cu care sunt emise simbolurile;

v_s – viteza / rata de transmisie a sursei.

Rezultă

$v_s = 1000 \cdot 1.875 = 1875 bit/s$.

1.4.2 Modele statistice pentru sursele de informaţie

Analiza modelelor statistice pentru sursele de informaţie este realizată în condiţiile considerării următoarelor ipoteze:

- sursa este *staţionară* (probabilităţile de apariţie a diferitelor simboluri nu depind de timp);
- sursa este *regulată* (nu există posibilitatea de a nu fi emise toate mesajele posibile).

Practic, aproape toate sursele de informaţie emit mesaje static dependente de succesiunea mesajelor transmise anterior (de exemplu, considerând un oarecare text, care nu e complet aleator, există frecvenţe diferite de apariţie a literelor A şi X) .

În acest context, modelul cel mai des întâlnit este *modelul Markov staţionar discret,* definit de următoarele caracteristici:

1. Sursa se află în una din cele n stări posibile $1...n$ la începutul fiecărui interval elementar de emitere a unui simbol. Ea îşi schimbă o singură dată starea pe durata unui interval, din starea iniţială *i* în starea finală *j,* cu probabilitatea p_{ij}, numită probabilitate de tranziţie. Această probabilitate rămâne constantă pe toată durata procesului.
2. Când sursa trece din starea i in starea j, este emis un simbol care depinde de starea i şi de tranziţia $i \to j$.
3. Fie $S_1,..., S_M$ – simbolurile alfabetului sursei, iar $x_1,...,x_k,...$– secvenţa de variabile aleatoare, cu x_k– simbolul evenimentului *k* din şirul simbolurilor emise de sursă. Probabilitatea ca acest simbol să fie S_q va fi condiţionată de celelalte simboluri emise anterior.

$$p(x_k = S_q | x_1, x_2, ..., x_{k-1}) \tag{1.43}$$

4. În conexiune cu 3., influenţa reziduală a simbolurilor $x_1,..., x_{k-1}$ defineşte starea sistemului la începutul intervalului *k;* fie ea S_k -atunci

$$p(x_k = S_q | x_1, x_2, ..., x_{k-1}) = p(x_k = S_q | S_k) \tag{1.44}$$

5. La începutul primului interval de emisie, sistemul se află în una din cele n stări posibile $1...n$, cu probabilităţile $p_1(1), p_2(1),..., p_n(1)$, $\sum_1^n p_i(1) = 1$.
6. Dacă probabilitatea ca sistemul să fie în starea *j* la începutul intervalului *k* este $p_j(k)$, tranziţia sistemului se reprezintă prin

$$p_j(k+1) = \sum_{i=1}^{n} p_i(k) p_{ij} \tag{1.45}$$

În acest sens, considerând $P(k)$ vector coloană cu $p_i(k)$ în poziţia *i*

$$P(k) = \begin{bmatrix} P_1(k) \\ \vdots \\ P_i(k) \\ \vdots \\ P_n(k) \end{bmatrix} \tag{1.46}$$

şi matricea $\phi = M_{m \times n}$, de forma

$$\phi = \begin{pmatrix} p_{11} & p_{12} \cdots p_{1n} \\ p_{21} & p_{22} \cdots p_{2n} \\ \cdots & \cdots \cdots \\ p_{n1} & p_{n2} \cdots p_{nn} \end{pmatrix} \quad (1.47)$$

se poate scrie relaţia matriceală

$$P(k+1) = \phi^T \cdot P(k) \quad (1.48)$$

Matricea ϕ se numeşte *matricea probabilităţilor de tranziţie a procesului Markov*, cu proprietatea că un proces Markov este staţionar dacă

$$P(k) = \phi^T P(k), k = 1. \quad (1.49)$$

Sursele Markov discrete se pot reprezenta prin *grafuri*, având în noduri stările, iar arcele reprezentând tranziţiile între stări.

Aplicaţia 1.3.
Considerând o sursă Markov reprezentată prin graful din figura 1.4.,

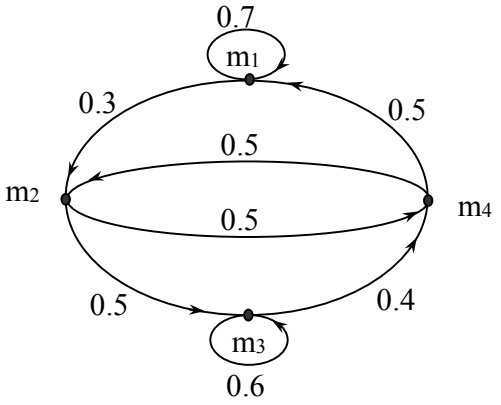

Fig.1.4. Graful unei surse Markov

matricea probabilităţilor de tranziţie asociată este de forma

$$\phi = \begin{bmatrix} 0.7 & 0.3 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0.6 & 0.4 \\ 0.5 & 0.5 & 0 & 0 \end{bmatrix} \quad (1.50)$$

1.5 Caracterizarea entropică a canalelor de comunicaţie

1.5.1 Canale discrete

Fie un canal discret de comunicaţie, caracterizat prin

- alfabetul de intrare: $X = (x_1, x_2, ..., x_n)$
- alfabetul de ieşire: $Y = (y_1, y_2, ..., y_m)$
- legea de tranziţie π_{kl}, definită prin probabilitatea condiţionată $P(y_j / x_i)$ de apariţie la ieşirea canalului a simbolului y_j când la intrare a fost simbolul x_i.

Câteva dintre proprietăţile relevante ale canalelor de comunicaţie sunt prezentate în continuare.

- Canalul este staţionar, dacă pentru fiecare pereche (x_i, y_i), $p(y_i \setminus x_i)$ nu depinde de timp;
- Canalul este fără memorie, dacă $p(y_i / x_i)$ nu depinde de natura semnalelor transmise anterior;
- Legea de tranziţie π este reprezentată de matricea

$$\pi = \begin{bmatrix} p_1(1) & p_1(2) & \dots p_1(m) \\ p_2(1) & p_2(2) & \dots p_2(m) \\ \dots & \dots & \dots\dots \\ p_m(1) & P_m(2) & \dots p_m(m) \end{bmatrix} \tag{1.51}$$

cu $\sum_j p_i(j) = 1, p_{ij} \geq 0.$

Matricea π caracterizează perturbaţia de pe canal, fiind denumită şi *matrice de zgomot*, semnificaţia ei fiind deosebit de importantă în contextul analizei transmisiei pe canal.

Cunoscând câmpul de probabilitate al sursei, deci $p(x_i)$, $i = \overline{1, n}, \sum_1^n p(x_i) = 1$, cu relaţia:

$$p(x_i, y_j) = p(y_j / x_i) \cdot p(x_i) \tag{1.52}$$

se poate calcula matricea $P(x, y)$, denumită şi *matricea probabilităţilor câmpurilor reunite*, cu proprietăţile:

- suma elementelor pe linie $\sum_{j=1}^{m} p(x_i, y_j) = p(x_i), \sum_{i=1}^{n} p(x_i) = 1$
- suma elementelor pe coloană $\sum_{i=1}^{n} p(x_i, y_j) = p(y_j), \sum_{1}^{m} p(y_j) = 1$

a) Dacă matricea de zgomot este formată numai din linii obţinute prin permutarea aceluiaşi set de probabilităţi $p_1,..., p_m$, canalul se numeşte *uniform faţă de intrare*.
b) Analog, dacă matricea de zgomot este formată numai din linii obţinute prin permutarea aceluiaşi set de probabilităţi $q_1,..., q_n$, canalul se numeşte *uniform faţă de ieşire*.

Un canal uniform atât faţă de intrare cât şi faţă de ieşire este un *canal dublu uniform*, situaţie în care $m = n$.
În cazul în care alfabetul de intrare şi cel de ieşire sunt identice şi, $(\forall) i \neq j$, se poate scrie

$$p_i(j) = p_m = \frac{1-q}{m-1} = ct \tag{1.53}$$

cu q–probabilitatea recepţionării fără eroare, canalul se numeşte *simetric.*

Capacitatea unui canal discret simetric se obţine, conform definiţiei, prin maximizarea transinformaţiei

$$C = \max[H(y) - H(x)] = H\left(\frac{1}{m}, \frac{1}{m}, ..., \frac{1}{m}\right) + \sum_{j=1}^{m} p_i(j) \log p_i(j) \tag{1.54}$$

$$C = \log m + \sum_{j=1}^{m} p_i(j) \log p_i(j) \tag{1.55}$$

Un caz particular îl constituie canalul simetric la care trecerile la acelaşi indice se fac cu aceeaşi probabilitate, iar celelalte treceri se fac cu alte probabilităţi, toate egale

$$\pi = \begin{pmatrix} 1-p & q & \ldots\ldots & q \\ q & 1-p & \ldots & q \\ \ldots & \ldots\ldots & \ldots & \\ q & q & \ldots\ldots & 1-p \end{pmatrix}, \text{ cu } q = \frac{p}{m-1} \tag{1.56}$$

Capacitatea unui astfel de canal va fi, pentru $m = n$, dată de relaţia

$$C = \log n + (1-p)\log(1-p) + (n-1)\cdot\frac{p}{n-1}\log\frac{p}{n-1} = \log n + (1-p)\log(1-p) + p\log - p\log(n-1) \quad (1.57)$$

1.5.1.1 Tipuri caracteristice de canale discrete utilizate în transmisia de date

În echipamentele de transmisie de date la care, în majoritatea cazurilor, se transmit simboluri binare, canalul cel mai des folosit este *canalul binar simetric* (CBS), reprezentat în schema din figura 1.5:

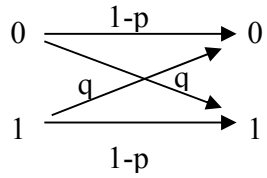

Fig.1.5 Canal binar simetric

şi caracterizat de matricea de zgomot

$$\pi = \begin{pmatrix} 1-p & q \\ q & 1-p \end{pmatrix},\ q = \frac{p}{m-1}\Big|_{m=2} = p \quad (1.58)$$

Există, deci, aceeaşi probabilitate ca un simbol binar de intrare să apară la ieşirea canalului sub forma 1 sau 0.

Capacitatea acestui canal este

$$C_{CBS} = 1 + (1-p)\log(1-p) + p\log p \quad (1.59)$$

Viteza de transmitere a informaţiei pe un canal discret V_s este inferioară vitezei medii de transmitere a informaţiei către sursă, v_s

$$V_s = H(x) + v_s \quad (1.60)$$

deoarece apar erori pe parcursul canalului.

În acest context, apare necesară definirea *debitului mediu* al transmisiei pe canal

$$D_t = [H(x) - H(x/y)]\cdot V_s = I(x,y)\cdot V_s \ \text{[bit/s]} \quad (1.61)$$

Aplicaţia 1.4

Să se calculeze capacitatea şi debitul mediu pentru un canal binar simetric care emite simboluri echiprobabile cu $v_s = 1000 simbol/s$, dacă probabilitatea de recepţie eronată este $p = 0.1$ şi $p = 0.4$.

Entropia sursei este $H(x) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2} = 1bit / simbol$

Debitul sursei rezultă $V_s = v_s \cdot H(x) = 1000bit / s$.

Informaţia medie se obţine $I(x,y) = H(x) - H(x,y) = \begin{cases} 0.531, & p = 0.1 \\ 0.029, & p = 0.4 \end{cases}$

Debitul mediu pe canal se calculează

$$D_t = I(x,y) \cdot v_s \rightarrow D_t = \begin{cases} 531bit, & p = 0.1 \\ 29bit, & p = 0.4 \end{cases}$$

Capacitatea canalului $C = \begin{cases} 0.531bit, & p = 0.1 \\ 0.029bit, & p = 0.4 \end{cases}$

Se poate observa că, în acest caz, capacitatea coincide cu transinformaţia deoarece

$$p(x=0) = p(y=1) = 1/2$$

$$\begin{pmatrix} p(y=0) = p(y=0/x=0) \cdot p(x=0) + p(y=0/x=1) \cdot p(x=1) = \\ (1-p) \cdot \frac{1}{2} + p \cdot \frac{1}{2} = \frac{1}{2} \end{pmatrix}$$

Un alt model de canal utilizat în teletransmisie este *canalul binar cu zonă de anulare*, CBZA. Acest tip de canal prezintă 2 simboluri în alfabetul de intrare: $x_1 = 0; x_2 = 1$şi 3 simboluri în alfabetul de ieşire: $y_1 = 0; y_2 = 1; y_3 = x$ (*x*– stare indiferentă distinctă), fiind descris de reprezentarea din figura 1.6.

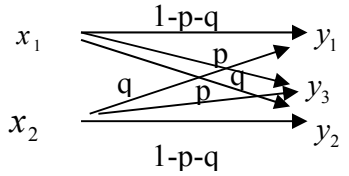

Fig.1.6 Canal binar cu zonă de anulare

şi având matricea de zgomot $\pi = \begin{pmatrix} 1-p-q & q & p \\ q & 1-p-q & q \end{pmatrix}$

Pentru CBZA, un caz interesant este cel pentru care q=0, adică y_1 nu poate proveni decât din x_1, iar y_2 nu poate proveni decât din x_2. În acest caz, $C_{CBZA} = 1 - p$.

1.5.1.2 Erori caracteristice canalelor binare

Erorile care apar în procesul transmiterii informaţiei într-un canal binar pot fi:

- singulare;
- grupate în pachete.

Pachetul de erori este o succesiune de simboluri de o anumită lungime, caracterizată printr-un număr de simboluri între prima şi ultima eroare din succesiune. Prin analogie, intervalul fără eroare este caracterizat de numărul de simboluri dintre ultima eroare a unui pachet de erori şi prima eroare din pachetul de erori următor.

Pentru a caracteriza statistic complet un canal, se iau în consideraţie următorii parametri:

- probabilitatea de eroare a unui simbol;
- repartiţia intervalelor fără erori;
- probabilitatea apariţiei pachetelor de erori de o anumită lungime;
- repartiţia erorilor multiple într-o secvenţă de o anumită lungime.

Cercetările statistice asupra perturbaţiilor ce apar în canalele de transmisie au arătat că, de regulă, erorile nu sunt independente, fiind necesară elaborarea unor modele matematice care să descrie repartiţia lor. Un astfel de model trebuie să fie:

- suficient de general pentru a putea fi adaptat la diferite tipuri de canale, oferind posibilitatea modificării parametrilor săi;
- suficient de simplu pentru a nu apela la prea mulţi parametri descriptivi.

Dintre modelele matematice care descriu repartiţia erorilor pot fi menţionate următoarele modelele:

- binomial, Salinger, Eliott, care nu iau în consideraţie decât erori singulare;
- Gilbert, care ia în consideraţie fenomenele fizice care duc la apariţia erorilor caracterizate prin lanţuri Markov;

- Benett-Froehlich, Kuhn, care iau în consideraţie fenomenele fizice care duc la apariţia erorilor caracterizate prin pachete de erori;
- Mertz, caracterizate prin lanţuri de pachete de erori.

1.5.2 Canale continue

Referirile se vor face la porţiunea $C - C'$, "porţiunea analogică", cuprinsă între modulator şi demodulator, a sistemului de comunicaţie prezentat în figura 1.2.
În această porţiune, cea a canalului electric de comunicaţie, semnalele de intrare sunt funcţii continue de timp care ar trebui să fie reproduse identic la ieşirea canalului. Acest fapt nu se întâmplă însă datorită existenţei perturbaţiilor, conform reprezentării din figura 1.7.

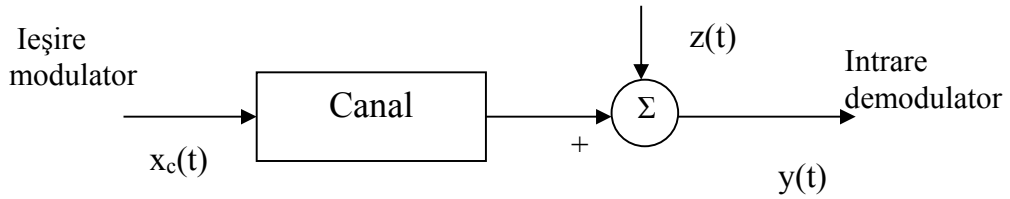

Fig.1.7 Influenţa perturbaţiilor asupra semnalului de ieşire din canal

Pentru schema din figura 1.8, $y(t) = x_c(t) + z(t)$, pentru care $z(t)$ sunt perturbaţiile considerate zgomote gaussiene în bandă limitată *B*, iar $x_c(t)$ este intrarea în canal, considerată o mărime aleatoare, canalul fiind de tip filtru trece-jos, cu banda de trecere *B*.

În continuare va fi prezentată maniera de apreciere a capacităţii de transfer a informaţiei pe această porţiune de canal.
Datorită benzii de trecere *B*, şi semnalele transmise au un spectru limitat, în gama $(-B, B)$. Conform teoriei eşantionării (Shannon), un astfel de semnal este complet determinat de un minim de eşantioane separate de intervale $\tau = \frac{1}{2B}[s]$, deci viteza de transmitere a informaţiei este *2B* simb/s.

Capacitatea canalului este $C = \max I(x, y)$, iar debitul de informaţie pe canal se obţine $D = \frac{C}{\tau} = 2BC$.

În ipotezele:

- semnalul emis este o funcţie aleatoare staţionară, cu puterea *S* definită ca un moment de ordin doi
$$S = ct = E[x^2(t)]$$
iar zgomotul gaussian $z(t)$ cu puterea $z = E[z^2(t)]$;
- puterile sunt aceleaşi cu mediile pătratice temporale $S = x^2(t)$;
- zgomotul este independent de semnal $z = z^2(t)$
$$y(t) = x(t) + z(t)$$
$$H(y/x) = H(x/x) + H(z/x) = H(z)$$

se obţine celebra *formulă Hartley-Tuller-Shannon* ce defineşte capacitatea temporală, C_τ, sau debitul de transmitere a informaţiei pe canal.

$$C_\tau = D_t = B\log\left(1+\frac{s}{z}\right)\ [bit/s], \qquad (1.62)$$

unde *B* defineşte banda de trecere, iar *s,z* puterile semnalului, respective ale zgomotului.

Formula Hartley-Tuller-Shannon are aplicaţii practice, chiar dacă se presupune sursa X gaussiană. Ea este foarte utilă pentru că subliniază corelaţia între banda de trecere şi raportul semnal–zgomot (unul dintre aceşti doi factori creşte în detrimentul celuilalt).
De asemenea, formula Hartley-Tuller-Shannon arată că pe un canal având $C < v_s$ (adică capacitatea canalului este mai mică decât viteza sursei) nu este posibilă transmisia fără eroare. Invers, impunând o anumită viteză de transmisie şi cunoscând *B*, se poate calcula raportul s/z minim.

O interpretare concretă a formulei este cea care consideră *informaţia transmisă discretizată.* Se consideră că zgomotul devine supărător dacă se depăşeşte nivelul unei cuante elementare. Numărul de niveluri discernabile este în acest caz finit şi poate fi estimat prin $q = \sqrt{\frac{s+z}{z}}$.

Mai mult, capacitatea canalului nu poate creşte oricât, numai prin creşterea benzii *B*, dacă raportul s/z rămâne acelaşi. Capacitatea temporală a unui canal are o limită.

În concluzie

- un canal fără zgomot are capacitatea infinită (concluzie amendată de practică – zgomot există întotdeauna).
- un sistem de comunicaţie ideal poate fi considerat cel care transmite informaţie cu debitul $D = B\log\left(1+\frac{s}{z}\right)$

Aplicaţia 1.5 Compunerea unui semnal sinusoidal cu semnal zgomot.

Scrieţi un program in Matlab care să reprezinte un semnal sinusoidal cu amplitudinea de 2V,frecventa f=100Hz,faza=0 esantionat cu frecvenţa fes=1000Hz peste care se suprapune un semnal uniform distribuit [-0.25;0.25].Să se reprezinte grafic şi să se calculeze raportul semnal zgomot al acestui semnal.

Semnalul sinusoidal este de forma:

$$X(t)=A*\sin(2*pi*f*t+faza)=A*\sin(2*pi*f*t)$$

deoarece faza=0, unde A-amplitudinea semnalului sinusoidal, f-frecvenţa semnalului,t-timpul, pi=3,14

Raportul semnal-zgomot - se exprimă de regula în decibeli[dB] şi este calculat cu ajutorul relaţiei:

$$Rsz[dB]=10lg(P_{semnal}/P_{zgomot})$$

unde P_{semnal} reprezintă puterea semnalului, iar P_{zgomot} reprezintă puterea zgomotului

```
%Programul in Matlab

amp=2;fes=1000;f=100;n=200;a=0.25
%definirea          variabilelor  amp-amplitudinea,fes-
frecv.esantionare,f-frecv,a-ampl.semnalului     uniform
distribuit,n=nr.eşantioane
rsz=10*log10((amp^2/2)/((2*a)^2/12))
%calculul   raportului   semnal   zgomot,log10-   functie
matlab pentru functia matem lg
t=(0:n-1)/fes
sig1=2*sin(2*pi*100*t)
```

```
%definirea semnalului sinusoidal
sig2=(2*a*rand(size(t))-a);
%definirea zgomotului
sig=sig1+sig2;
%compunerea semnalelor
plot(t,sig) %trasarea graficului
title ('Sinusoida având zgomot')
%titlul reprezentarii grafice
xlabel('t[s]')
%etichetare axa Ox
ylabel('Amplitudinea[V]')
%etichetare axa Oy
grid on
%trasare retea grafic
```

Dupa rularea programului graficul trasat cu functia *plot* este de forma reprezentata in figura 1.8.

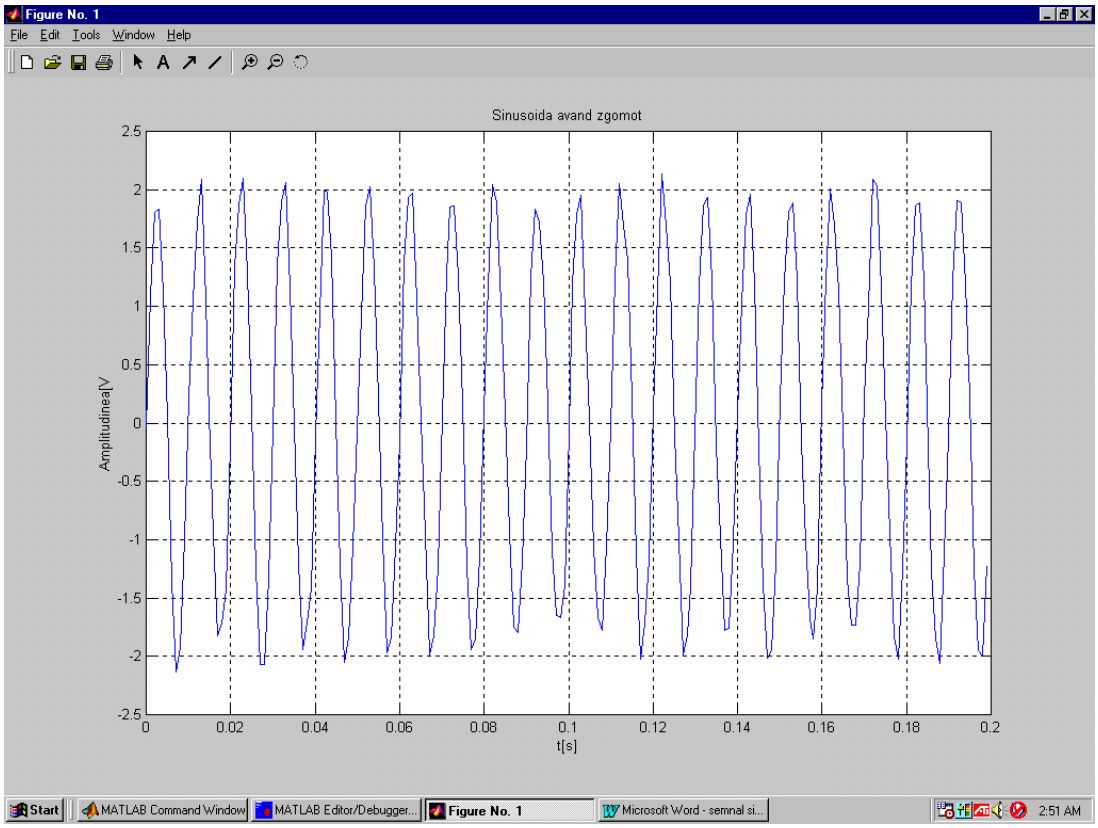

Fig.1.8. Reprezentare grafica semnal sinusoidal

a) Reprezentare grafică a semnalului sinusoidal eşantionat

```
%definirea variabilelor amp-amplitudinea,fe-
frecv.esantionare,f-frecv,a-ampl.es,n=nr es
y='Raportul semnal zgomot este egal cu:'
%afisare mesaj inainte de obtinerea valorii raportului
semnal-zgomot
rsz=10*log10(amp^2/2/(2*a)^2/12)
%calculul raportului semnal zgomot,log10 functie matlab
pentru fctia matem lg
z='Esantionarea  semnalului:'
%afisare mesaj
t=(0:n-1)/fes
sig=2*sin(2*pi*100*t)
%definirea semnalului sinusoidal
plot(t,sig)
%trasarea graficului
title ('Sinusoida esantionata')
%titlul reprezentarii grafice
xlabel('t[s]')
%etichetare axa Ox
ylabel('Amplitudinea[V]')
%etichetare axa Oy
grid on
%trasare retea grafic
```

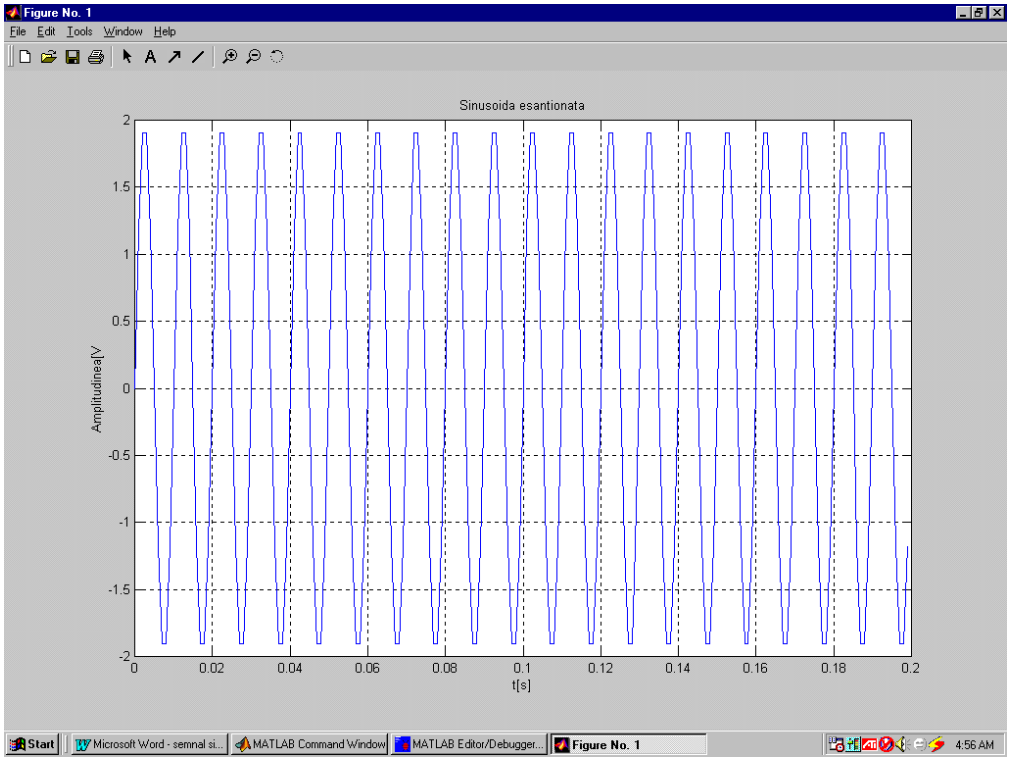

Fig.1.9. Reprezentare grafica semnal sinusoidal esantionat

b) Reprezentare grafică a zgomotului eşantionat

```
amp-amplitudinea,fe-frecv.esantionare,f-frecv,a-
ampl.es,n=nr es
y='Raportul semnal zgomot este egal cu:' %afisare mesaj
inainte de obtinerea valorii raportului semnal-zgomot
rsz=10*log10(amp^2/2/(2*a)^2/12) %calculul raportului
semnal zgomot,log10 functie matlab pentru fctia matem lg
z='Esantionarea  semnalului:' %afisare mesaj inainte de
esantionarea celor 2 semnale
t=(0:n-1)/fes
sig=2*a*rand(size(t))-a %definirea zgomotului
plot(t,sig) %trasarea graficului
title ('Esantionarea zgomotului') %titlul reprezentarii
grafice
xlabel('t[s]') %etichetare axa Ox
ylabel('Amplitudinea[V]') %etichetare axa Oy
grid on %trasare retea grafic
```

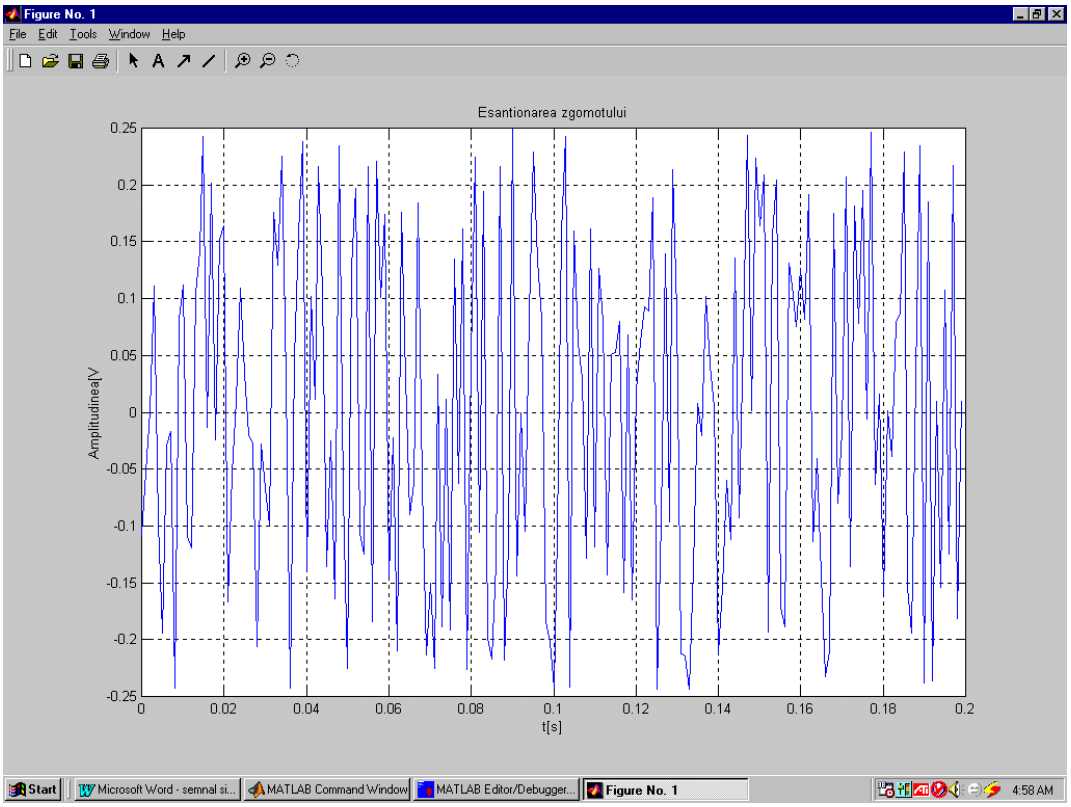

Fig.1.9. Reprezentare grafica zgomot esantionat

Aplicaţia 1.6

Se cere raportul $\left(\frac{s}{z}\right)_{\min}$ necesar pentru a transmite informaţia cu viteza 10^4 $[bit/s]$ pe un canal cu $B_1 = 3000Hz$, $B_2 = 10kHz$.

Din formula Hartley-Tuller-Shannon se calculează succesiv

$$\left(\frac{s}{z}\right)_1 = 9, \left(\frac{s}{z}\right)_2 = 1$$

Rezultă de aici că restrângerea benzii de la 10 kHz la 3 kHz necesită o creştere de nouă ori a puterii semnalului.

Caracteristici ale canalelor continue de comunicaţie

Un canal ideal din punct de vedere al transmisiei unui semnal electric, de exemplu o mărime $u_1(t)$, ar trebui să aibă o funcţie de transfer liniară, astfel încât la ieşirea canalului să se obţină

$$u_2(t) = k \cdot u_1(t)$$

$$H(\omega) = \frac{U_2(\omega)}{U_1(\omega)} = A \cdot e^{j\phi(\omega)}$$

Aceste caracteristici ideale nu se întâlnesc în practică; apar neliniarităţi, atenuări şi distorsiuni de fază care pot uneori afecta definitiv forma semnalului.

O altă problemă o constituie fenomenele de interferenţă datorate transmisiei simultane a mai multor semnale pe acelaşi suport.

Problema cea mai serioasă în transmisia datelor pe canal rămâne cea a *zgomotelor* datorate mediului fizic. În acest mediu se pot deosebi mai multe tipuri de canale de comunicaţie, dintre care cele mai importante sunt prezentate în continuare.

1. Circuitele / liniile fizice independente reprezintă categoria cea mai largă de canale. Există numeroase tipuri constructive care pot fi analizate comparativ prin capacitatea de a realiza un anumit număr de legături bidirecţionale, tip legătură telefonică:

- pereche de fire libere de cupru sau aliaje, care permit crearea a până la 24 canale telefonice;

- pereche torsadată de fire (fire împletite şi izolate pentru reducerea interferenţelor);
- cablu telefonic, conţinând mai multe perechi de fire torsadate, întregul grup fiind îmbrăcat într-un înveliş protector, câteodată cu ecran protector / masă de protecţie (frecvenţa uzuală la care se ajunge la o astfel de transmisie fiind în gama 268 kHz→1Mhz);
- cablu coaxial, alcătuit dintr-un miez cilindric de cupru şi un înveliş conductor cilindric între care se află un material dielectric sau aer. Mai multe cabluri coaxiale pot fi grupate într-un trunchi, permiţând crearea a 3600-10800 căi.
- ghiduri de undă, sub forma unor tuburi metalice traversate de unde radio de foarte înaltă frecvenţă, până la 100MHz. Se pot astfel asigura simultan peste 200000 legături telefonice.

Caracteristicile unor astfel de linii sunt exprimabile sub forma unor constante primare, şi anume rezistenţă, inductanţă, conductanţă şi capacitanţă pe unitate de lungime de linie şi sub forma unor parametri secundari – coeficient de atenuare, impedanţă caracteristică, capacitate.

Dintre parametrii primari, rezistenţa este cea mai puternic influenţată de temperatură, conform relaţiei:

$$R_\theta = R_0 \cdot [1 + \alpha \cdot (\theta - \theta_0)]$$

unde R_θ, R_0 sunt, respectiv, rezistenţele la temperaturile θ şi $0^\circ C$, iar α este coeficientul de variaţie al rezistenţei cu temperatura.

Pentru cabluri şi linii aeriene, caracteristicile primare (pe unitate de lungime tur/retur) la frecvenţă şi rezistenţă $R_0 = 20^\circ C$ sunt prezentate în tabelul 1.4.

2. Canale radio

Sunt mai puţin utilizate în transmisia de date cu caracter industrial, fiind însă deosebit de importante în tehnica telecomunicaţiilor. Există mai multe categorii, în funcţie de tipul de antenă utilizat, frecvenţă şi mod de propagare:

- cu propagare în linie dreaptă, situaţie în care antena de emisie şi cea de recepţie sunt reciproc vizibili, cu frecvenţe relative joase 3…30MHz (specifice telegrafiei fără fir sau radiofoniei pe mare);

- microunde radio, folosite în transmisia radio şi TV, care ocupă gama de până la 10 GHz. Sunt afectate de perturbaţii atmosferice, variaţii de temperatură şi umiditate;
- canale cu disipare troposferică, ce folosesc antene de mari dimensiuni ($\phi : 18 - 30mm$), bazate pe reflecţii în troposferă;
- transmisii prin satelit, care asigură transmisii multiple în bandă largă.

Tabelul 1.4

Caract / Tip circuit	Dist. între linii [cm]	Diam. sârmă [mm]	Rezist. [Ω/km]	Induct. [mH/km]	Capacitanţă [μF/Km]	Rezistenţă izolaţie minim [MΩ/km]	de între fire maxim [MΩ/km]
oţel 25-125	60	3	39.1	12.64	0.0049	2	25-125
	20	3	39.1	11.21	0.006	2	25-125
	60	4	22	9.4	0.0051	2	25-125
	20	4	22	0.96	0.0063	2	25-125
cupru	60	4	2.84	2.38	0.0051	2	25-125
	20	4	2.84	1.94	0.0063	2	25-125
aliaj oţel-cupru	60	4	6.44	2.39	0.0051	2	25-125
	20	4	6.44	1.94	0.0063	2	25-125

3. Canale cu fibră optică

Sunt des utilizate în aplicaţii industriale, datorită certelor avantaje comparativ cu alte tipuri de canale: viteze foarte mari de transmisie 1Mbit/sec-1Gbit/sec, lărgime mare de bandă, dimensiuni şi greutăţi mici, izolaţie electrică foarte bună, posibilitatea de a lucra în medii puternic perturbate.

Principiul de realizare a transmiterii de date pe fibră optică se bazează pe modularea (în amplitudine, frecvenţă sau polarizare) a fasciculului luminos cu ajutorul semnalului informaţional. Ca surse de lumină se utilizează de regulă laserul, mail ales cel cu injecţie, care permite modularea la viteze de peste 1 bit/sec prin simpla modulare a curentului de injecţie.

Tot ca ghiduri optice se folosesc şi fibrele optice cilindrice, atât pentru distanţe mici, cât mai ales pentru distanţe mari.
De exemplu, cu un ghid de undă de 70 mm diametru, având banda de frecvenţă cuprinsă în gama 30…110 GHz, se pot realiza până la 10^5 canale, pe o distanţă de până la 50 km.

2. Modulaţia semnalelor informaţionale

În vederea transmiterii semnalului purtător de informaţie pe un canal de comunicaţie, este necesar să fie efectuate operaţii de prelucrare a acestuia care să asigure compatibilitatea cu caracteristicile canalului şi combaterea într-o măsură cât mai mare a perturbaţiilor de pe canal.

Principala operaţie care are loc în acest sens este ***modulaţia***, care realizează modificarea parametrilor semnalului purtător (numit „*purtătoare*") sub acţiunea semnalului care deţine informaţia, adică semnalul mesaj (numit „*modulator*"). Se obţine astfel un „*semnal modulat"*.

Modulaţia are ca scop deplasarea semnalelor purtătoare de informaţie din aşa-numita bandă de bază în benzi de frecvenţe superioare. *Banda de bază* este constituită din frecvenţele obişnuite prezente în spectrul unui semnal, spectru situat uzual în zona frecvenţelor joase. Transmiterea tuturor semnalelor în banda de bază, adică în forma lor originară, în plus faţă de faptul că ar aglomera peste măsură zona frecvenţelor inferioare, s-ar realiza şi cu o eficienţă de cele mai multe ori modestă. Evitarea supraaglomerării benzii de bază se mai practică şi din cauza consumurilor energetice inacceptabile, dar şi datorită concurenţei perturbaţiilor.

Clasificarea tehnicilor de modulaţie, în funcţie de criteriile specifice, este următoarea:

1. după tipul purtătoarei:

    - *modulaţie armonică,* pentru care purtătoarea este o sinusoidă;

- *modulaţie de impulsuri*, pentru care purtătoarea este un tren de impulsuri.

2. după tipul semnalului modulator:

- *modulaţie analogică,* definită de un semnal modulator analogic şi tehnici de prelucrare analogice;
- *modulaţie numerică,* definită de un semnal modulator numeric şi tehnici de prelucrare numerice.

2.1. Modulaţia liniară

Fie un mesaj $m(t)$ şi transformata Fourier a acelui semnal, $M(\omega)$, reprezentată în figura 2.1.

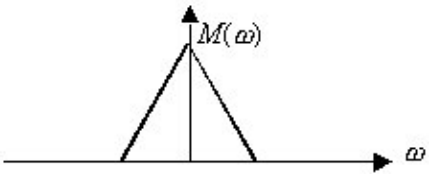

Fig. 2.1. Transformata Fourier a unui semnal mesaj

Multiplicarea semnalului în forma lui temporală cu o sinusoidă

$$A_1 \cos(\omega_1 t + \varphi_1) = \frac{A_1}{2}\left(e^{j(\omega_1 t+\varphi_1)} + e^{-j(\omega_1 t+\varphi_1)}\right) \tag{2.1}$$

produce în domeniul frecvenţelor semnalul

$$S(\omega) = \frac{A_1}{2}\left[e^{j\varphi_1} M(\omega-\omega_1) + e^{-j\varphi_1} M(\omega+\omega_1)\right] \tag{2.2}$$

conform unei *teoreme a modulaţiei* cunoscută sub denumirea de convoluţie în domeniul frecvenţelor. Figura 5.2 ilustrează efectul de translaţie a spectrului semnalului din banda de bază într-o bandă de aceeaşi lărgime, centrată pe frecvenţa sinusoidei purtătoare.

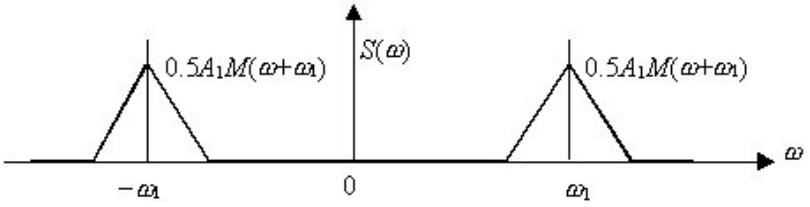

Fig. 2.2. Translaţia spectrului semnalului mesaj din banda de bază

Dacă semnalul mesaj $m(t)$ se combină cu purtătoarea conform relaţiei

$$s(t) = A_1\left[1 + \frac{a}{A_1}m(t)\right]\cos(\omega_1 t + \varphi_1) \tag{2.3}$$

atunci, în domeniul frecvenţelor se obţine semnalul

$$\begin{aligned} S(\omega) = \pi A_1\left[e^{j\varphi_1}\delta(\omega-\omega_1) + e^{-j\varphi_1}\delta(\omega+\omega_1)\right] + \\ + \frac{a}{2}\left[e^{j\varphi_1}M(\omega-\omega_1) + e^{-j\varphi_1}M(\omega+\omega_1)\right] \end{aligned} \tag{2.4}$$

care are graficul prezentat în figura 5.3, asemănător cu precedentul, dar care pune în evidenţă şi existenţa purtătoarei în integritatea ei.

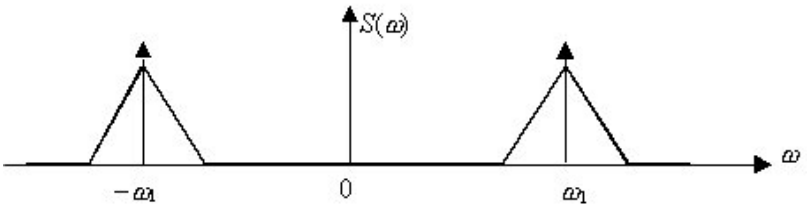

Fig. 2.3. Graficul semnalului modulat

Se observă simetria celor două benzi laterale faţă de purtătoare.

În cazul unui *semnal modulator sinusoida*l, $m(t) = \cos\Omega t$, *semnalul modulat în domeniul timp* are expresia

$$s(t) = A_1\left[1 + \frac{a}{A_1}\cos\Omega t\right]\cos(\omega_1 t + \varphi_1) \tag{2.5}$$

şi prin relaţii cunoscute din trigonometrie se poate rescrie sub forma

$$s(t) = A_1\cos(\omega_1 t + \varphi_1) + \frac{a}{2}\cos[(\omega_1 + \Omega)t + \varphi_1] + \frac{a}{2}\cos[(\omega_1 - \Omega)t + \varphi_1] \tag{2.6}$$

Spectrul semnalului este un spectru de linii şi benzile laterale sunt reduse la liniile de frecvenţe $\omega_1 \pm \Omega$. Puterile celor trei componente ale spectrului sunt, respectiv, $A_1^2/2$, $a^2/8$, $a^2/8$.

Se defineşte în continuare un aşa-numit *grad de modulaţie* $\alpha = a/A_1$. Numărul α trebuie să fie între 0 şi 1 sau, în procente, între 0 şi 100%. Depăşirea gradului de modulaţie de 100% duce la obţinerea unui semnal supramodulat şi semnalul modulator nu mai poate fi recuperat la recepţie.

Modulaţia în versiunea ultimă este o *modulaţie de anvelopă.* Anvelopa, înfăşurătoarea semnalului modulat, urmăreşte forma semnalului modulator. Pentru a nu avea o supramodulaţie cu consecinţe grave asupra recuperabilităţii semnalului modulator la utilizator, factorul care multiplică purtătoarea trebuie să păstreze permanent un semn constant, de exemplu $\left[1+\frac{a}{A_1}m(t)\right] \geq 0$.

Liniaritatea modulaţiei de amplitudine, fie în prima variantă, prin multiplicare, fie în a doua variantă, cea pe anvelopă, se poate verifica simplu ţinând seama de liniaritatea transformării Fourier. Semnalul modulator $m(t)$ multiplicat cu un scalar α se regăseşte în semnalul modulat multiplicat cu acelaşi scalar α. Suma a două semnale modulatoare $m(t) = m_1(t) + m_2(t)$ se regăseşte în semnalul modulat ca sumă a două semnale modulate de semnalele $m_1(t)$ şi $m_2(t)$. Pentru modulaţia produs, de exemplu

$$\begin{aligned} S(\omega) &= \frac{A_1}{2}\left[e^{j\varphi_1}\alpha M(\omega-\omega_1) + e^{-j\varphi_1}\alpha M(\omega+\omega_1)\right] = \\ &= \alpha\frac{A_1}{2}\left[e^{j\varphi_1}M(\omega-\omega_1) + e^{-j\varphi_1}M(\omega+\omega_1)\right] = \alpha S(\omega) \end{aligned} \tag{2.7}$$

şi cu o indexare uşor de înţeles

$$\begin{aligned} S(\omega) &= \\ &= \frac{A_1}{2}\left\{e^{j\varphi_1}\left[M_1(\omega-\omega_1)+M_2(\omega-\omega_1)\right] + e^{-j\varphi_1}\left[M_1(\omega+\omega_1)+M_2(\omega+\omega_1)\right]\right\} = \\ &= \frac{A_1}{2}\left[e^{j\varphi_1}M_1(\omega-\omega_1) + e^{-j\varphi_1}M_1(\omega+\omega_1)\right] + \\ &+ \frac{A_1}{2}\left[e^{j\varphi_1}M_2(\omega-\omega_1) + e^{-j\varphi_1}M_2(\omega+\omega_1)\right] = S_1(\omega) + S_2(\omega) \end{aligned} \tag{2.8}$$

Perturbaţiile care afectează procesul de modulaţie liniară pot fi, în general, coerente sau necoerente.

Perturbaţiile coerente provin din suprapunerea unui alt canal (învecinat) peste canalul de interes. Fie în acest caz semnalul util

$$s_1(t) = A_1 e^{j\omega_1 t} \tag{2.9}$$

şi semnalul perturbator

$$s_2(t) = A_2 e^{j\omega_2 t} \tag{2.10}$$

La intrarea în demodulator cele două semnale se aplică aditiv ca un singur semnal, deoarece circuitele premergătoare demodulatorului trebuie să fie şi sunt liniare, conform relaţiei

$$s(t) = s_1(t) + s_2(t) = A_1 e^{j\omega_1 t}\left(1 + \frac{A_2}{A_1} e^{j\omega_\Delta t}\right) \tag{2.11}$$

Se consideră că raportul $a = A_2/A_1$ este în valoare absolută mult inferior unităţii. Paranteza din expresia de mai sus se poate reprezenta grafic prin diagrama cu fazori din figura 5.4.

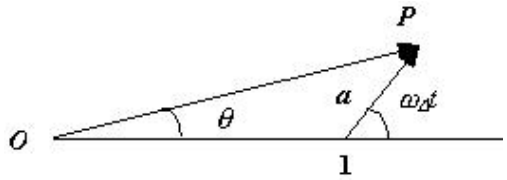

Fig. 5.4. Diagrama fazorială corespunzătoare aplicării aditive a semnalelor

în care $OP = \sqrt{1 + a^2 + 2a\cos\omega_\Delta t}$ şi $\theta = \arctan\dfrac{a\sin\omega_\Delta t}{1 + a\cos\omega_\Delta t}$. Semnalul rezultant se exprimă ca

$$s(t) = A_1\sqrt{1 + a^2 + 2a\cos\omega_\Delta t}\, e^{j(\omega_1 t + \theta)} \tag{2.12}$$

cu o amplitudine variabilă periodic cu frecvenţa diferenţă $\omega_\Delta = |\omega_2 - \omega_1|$ şi cu faza şi ea variabilă.

Descompunerea Fourier a amplitudinii, care este o funcţie periodică, se prezintă aproximativ conform relaţiei (5.13)

$$A(t) = A_1\left(1+\frac{a^2}{4}+..\right)+A_1 a\left(1-\frac{a^2}{8}+...\right)\cos\omega_\Delta t+... \qquad (2.13)$$

cu partea de frecvenţă nulă dependentă de A_1, cu fundamentala dependentă de $A_2 = A_1 a$. În plus, apar şi eventuale armonice, atâtea câte încap în banda de trecere a demodulatorului, limitată natural şi uzual de o frecvenţă □. Dacă $|\omega_2 - \omega_1| > \Omega$, atunci nu trece nici măcar fundamentala.

Raportul dintre semnal şi perturbaţie se evaluează în condiţia $a << 1$, pe baza aproximării

$$A(t) = A_1\sqrt{1+a^2+2a\cos\omega_\Delta t} \approx A_1\sqrt{1+2a\cos\omega_\Delta t} \approx A_1(1+a\cos\omega_\Delta t) \qquad (2.14)$$

$$A(t) \approx A_1 + A_2\cos\omega_\Delta t \qquad (2.15)$$

Dacă se admite că ambele semnale sunt modulate în amplitudine, $A_1 + m(t)$, $A_2 + n(t)$, şi raportul amplitudinilor lor este permanent subunitar, atunci

$$A(t) \approx A_1 + m(t) + A_2\cos\omega_\Delta t + n(t)\cos\omega_\Delta t \qquad (2.16)$$

şi raportul puterilor medii semnal util / semnal perturbator este

$$\frac{S}{P} = \frac{\overline{m^2(t)}}{\overline{[A_2+n(t)]^2\cos^2\omega_\Delta t}} \qquad (2.17)$$

Dată fiind independenţa celor doi factori, media temporală a produsului de la numitor este produsul mediilor, astfel încât raportul semnal/perturbaţie devine

$$\frac{S}{P} = \frac{\overline{m^2(t)}}{\frac{1}{2}\left[A_2^2+\overline{n^2(t)}\right]} \qquad (2.18)$$

Dacă cele două purtătoare au aceeaşi frecvenţă, atunci

$$A(t) = A_1 + A_2 \cos\varphi \tag{2.19}$$

luând în consideraţie numai defazajul □□ între cele două purtătoare.

Cu modulaţii, relaţia (5.19) devine

$$A(t) = A_1 + m(t) + A_2 \cos\varphi + n(t)\cos\varphi \tag{2.20}$$

şi raportul semnal/perturbaţie se scrie

$$\frac{S}{P} = \frac{\overline{m^2(t)}}{\overline{n^2(t)}\cos^2\varphi} \tag{2.21}$$

La o diferenţă între faze de $\pm\pi/2$, raportul este foarte favorabil semnalului util deoarece numitorul din relaţia (5.21) este foarte mic.

La modulaţia de produs, cele două semnale modulate au expresiile binecunoscute $s_1(t) = m_1(t)\cos\omega_1 t$ şi $s_2(t) = m_2(t)\cos\omega_2 t$. La demodulare, care se realizează prin multiplicarea din nou cu o sinusoidă de frecvenţă □ 0 şi fază □ 0 sunt recuperate semnalele

$$n_1(t) = m_1(t)\cos[(\omega_0 - \omega_1)t + \varphi_0 - \varphi_1] \tag{2.22}$$
$$n_2(t) = m_2(t)\cos[(\omega_0 - \omega_2)t + \varphi_0 - \varphi_2] \tag{2.23}$$

O reglare de frecvenţă şi de fază a oscilatorului de la recepţie poate aduce situaţii avantajoase, respectiv egalizarea frecvenţei la recepţie cu aceea a semnalului parazit, □ 0 = □ 2 şi aranjarea fazelor, astfel încât diferenţa de fază $\varphi_0 - \varphi_2 = (2k+1)\frac{\pi}{2}$ face să dispară complet semnalul parazit.

Transmiterea unei singure benzi laterale

După cum se poate observa, cele două benzi laterale ale unui semnal modulat în amplitudine sunt identice, exceptând o simetrie faţă de frecvenţa purtătoare. Ideea suprimării uneia din benzile laterale este cât se poate de firească: banda candidată la eliminare conţine aceeaşi informaţie ca şi cealaltă bandă laterală, consumă o energie uneori prea importantă pentru a repeta transmiterea (în consecinţă, redundantă) a aceleiaşi informaţii şi ocupă un spaţiu în banda de frecvenţe disponibilă care poate fi alocat unei alte căi de transmisie.

Expresiile în domeniul frecvenţă şi în domeniul timp ale semnalelor transmise în sistemul cu bandă laterală unică (BLU) sunt prezentate în continuare. Dacă se elimină banda inferioară se obţine

$$S_{1+}(\omega) = \frac{1}{2} e^{j\varphi_1} M_+(\omega - \omega_1) + \frac{1}{2} e^{-j\varphi_1} M_-(\omega + \omega_1) \tag{2.24}$$

iar dacă se elimină banda superioară rezultă

$$S_{1-}(\omega) = \frac{1}{2} e^{j\varphi_1} M_-(\omega - \omega_1) + \frac{1}{2} e^{-j\varphi_1} M_+(\omega + \omega_1) \tag{2.25}$$

cu indicii + sau – asociaţi cu poziţiile benzilor laterale, la dreapta, respectiv la stânga purtătoarei, în reprezentarea simetrică pe axa reală a frecvenţelor.

Prin transformarea Fourier inversă, pentru cazul transmiterii benzii superioare se obţine succesiv

$$s_1(t) = \frac{e^{j\varphi_1}}{4\pi} \int_{\omega_1}^{\omega_1+\Omega_M} M(\omega - \omega_1) e^{j\omega t} d\omega + \frac{e^{-j\varphi_1}}{4\pi} \int_{-\omega_1-\Omega_M}^{-\omega_1} M(\omega + \omega_1) e^{j\omega t} d\omega$$

$$s_1(t) = \frac{e^{j(\omega_1+\varphi_1)}}{4\pi} \int_{0}^{\Omega_M} M(\omega) e^{j\omega t} d\omega + \frac{e^{-j(\omega_1+\varphi_1)}}{4\pi} \int_{-\Omega_M}^{0} M(\omega) e^{j\omega t} d\omega \tag{2.26}$$

$$\begin{aligned} s_1(t) = {} & \frac{1}{2}\cos(\omega t + \varphi_1)\left[\frac{1}{2\pi}\int_0^{\infty} M(\omega) e^{j\omega t} d\omega + \frac{1}{2\pi}\int_{-\infty}^{0} M(\omega) e^{j\omega t} d\omega\right] + \\ & + \frac{1}{2}\sin(\omega t + \varphi_1)\left[\frac{1}{2\pi}\int_0^{\infty} jM(\omega) e^{j\omega t} d\omega + \frac{1}{2\pi}\int_{-\infty}^{0} -jM(\omega) e^{j\omega t} d\omega\right] \end{aligned} \tag{2.27}$$

Cu notaţia $N(\omega) = jM(\omega)$ pentru $\omega > 0$ şi $N(\omega) = -jM(\omega)$ pentru $\omega < 0$ şi cu $n(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} N(\omega)e^{j\omega t}d\omega$, se obţine ca factor în expresia semnalului în domeniul timp, *transformata Hilbert* a semnalului *m*(*t*), $H\{m(t)\}$, aşa încât

$$s_1(t) = \frac{1}{2}m(t)\cos(\omega_1 t + \varphi_1) + \frac{1}{2}H\{m(t)\}\sin(\omega_1 t + \varphi_1) \tag{2.28}$$

şi, analog, în cazul transmiterii benzii inferioare

$$s_1(t) = \frac{1}{2}m(t)\cos(\omega_1 t + \varphi_1) - \frac{1}{2}H\{m(t)\}\sin(\omega_1 t + \varphi_1) \tag{2.29}$$

5.2. Modulaţia exponenţială

Fie purtătoarea $E_0 e^{j\omega_0 t}$ multiplicată cu $e^{jk_\varphi m(t)}$ sau cu $e^{jk_f\int_0^t m(\tau)d\tau}$. În ambele cazuri exponentul este funcţie de mesajul *m*(*t*). Numerele k□□□□si k_f sunt constante ale modulatorului care nu au o importanţă deosebită pentru dezvoltarea teoretică următoare, astfel încât atribuirea valorii 1 acestor constante nu impietează asupra adevărului demostraţiilor de mai jos.

Cu precizarea anterioară, se pot scrie expresiile semnalelor cu exponentul modulat, respectiv

$$s(t) = E_0 e^{j\omega_0 t}e^{jm(t)} = E_0 e^{j[\omega_0 t + m(t)]} \tag{2.30}$$

$$s(t) = E_0 e^{j\omega_0 t}e^{j\int_0^t m(\tau)d\tau} = E_0 e^{j[\omega_0 t + \int_0^t m(\tau)d\tau]} \tag{2.31}$$

şi, în general,

$$s(t) = E_0 e^{j\varphi(t)} \tag{2.32}$$

Pentru tratarea şi înţelegerea modulaţiei exponenţiale este necesară introducerea noţiunii de *frecvenţă instantanee*.

Fie semnalul descris de relaţia (5.32) şi semnalul

$$e(t) = Ee^{j\omega t} \tag{5.33}$$

Pentru ca cele două semnale să fie aproximativ egale pe un interval de timp foarte scurt Δt în jurul unui punct (ceea ce înseamnă să fie local egale), este necesar ca funcţiile să fie egale şi un număr de derivate ale lor să fie de asemenea egale în acel punct. Dezvoltările Taylor ale celor două semnale, de forma

$$s(t+\Delta t) = s(t) + \frac{\Delta t}{1!} s'(t) + \ldots \tag{2.34}$$

$$e(t+\Delta t) = e(t) + \frac{\Delta t}{1!} e'(t) + \ldots \tag{2.35}$$

produc egalităţile

$$E = E_0\,,\quad s(t) = e(t)\,,\quad s'(t) = e'(t) \tag{2.36}$$

Relaţiile (5.34) şi (5.35) exprimă necesitatea ca cele două semnale să fie egale ca amplitudine şi să aibă instantaneu aceeaşi fază. Din relaţia (5.36) rezultă expresia frecvenţei instantanee $\omega = \varphi'(t)$.

La modulaţia de fază M□□□□faza se modifică în conformitate cu mesajul

$$\varphi(t) = \omega_0 t + m(t) \tag{2.37}$$

La modulaţia de frecvenţă, *MF*, frecvenţa este aceea care urmăreşte mesajul

$$\varphi(t) = \omega_0 t + \int_0^t m(\tau)d\tau \quad \text{şi} \quad \varphi'(t) = \omega_i = \omega_0 + m(t) \tag{2.38}$$

Spectrele semnalelor modulate cu mesaje bogate în frecvenţe sunt extrem de complicate. Ele pot fi analizate prin studiul cazurilor mai simple când mesajele sunt sinusoidale.

Fie, aşadar, mesajul $m(t) = M\cos\Omega t$ cu □ << □ . Se definesc deviaţiile maxime de fază □□ = M şi de frecvenţă □□ = M. Semnalele modulate se scriu

$$s_\varphi(t) = E_0 e^{j[\omega_0 t + \Delta\varphi\cos\Omega t]} \tag{2.39}$$

$$s_f(t) = E_0 e^{j[\omega_0 t + \frac{\Delta\omega}{\Omega}\sin\Omega t]} \tag{2.40}$$

Numerele □□ şi □□□□ se mai numesc *indici de modulaţie* şi sunt notaţi cu □.

Cheia aprecierii spectrului unui semnal modulat în fază sau în frecvenţă o constituie următoarea dezvoltare matematică

$$\exp\left[\frac{1}{2}\left(t-\frac{1}{t}\right)z\right]=\sum_{k=-\infty}^{\infty}J_k(z)t^k \tag{2.41}$$

care generează funcţiile Bessel $J_k(z)$, de specia I, de indice întreg k. O înlocuire a variabilei t cu $e^{j\theta}$ produce dezvoltarea

$$e^{jz\sin\theta}=\sum_{k=-\infty}^{\infty}J_k(z)e^{jk\theta} \tag{2.42}$$

Acum, expresia semnalului pentru cazul modulaţiei de frecvenţă, de exemplu, se poate rescrie sub forma

$$s_f(t)=E_0e^{j\omega_0t}e^{j\beta\sin\Omega t}=E_0e^{j\omega_0t}\left[\sum_{k=-\infty}^{\infty}J_k(\beta)e^{jk\Omega t}\right] \tag{2.43}$$

şi se pot pune în evidenţă componente de spectru cu frecvenţe de forma □ = □ 0 + k□ □ cu k întreg.

Figura 5.5. prezintă graficele câtorva funcţii Bessel de specia I, pentru indici întregi de la 0 la 7 inclusiv.

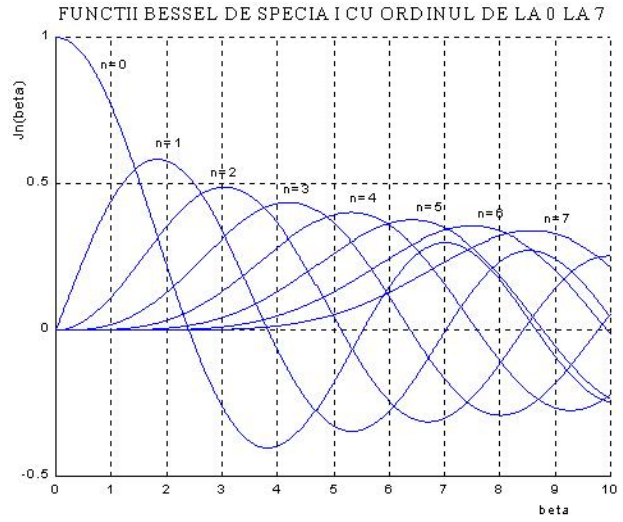

Fig. 2.5. Funcţii Bessel de specia I cu ordinul de la 0 la 7

Pe diagrama din figura 5.5. pot fi determinate, pentru un indice □ dat, amplitudinile componentelor spectrale pentru k = 0, 1, …, 7, precum şi pentru k

= –1, –2, …, –7. De exemplu, pentru = 7, componenta cu $k = 1$ se anulează, iar componenta cu $k = 5$ apare ca fiind cea mai importantă.

Este de reţinut că puterea semnalului modulat este constantă, oricare ar fi indicele de modulaţie. Puterea este distribuită în spectrul de frecvenţe astfel încât suma puterilor componentelor este mereu aceeaşi. Spectrul este teoretic infinit. Sub aspect practic, un număr finit de componente de ordine k inferioare cumulează cvasitotalitatea puterii semnalului. Într-o situaţie diferită, modulaţiile exponenţiale ar fi fost inutilizabile: o singură purtătoare modulată în această manieră ar fi ocupat tot spectrul de frecvenţe disponibil.

Interferenţe

Semnalele

$$s_1(t) = A_1 \exp\left[j\int_0^t \omega_1(\tau)d\tau \right] \tag{2.44}$$

$$s_2(t) = A_2 \exp\left[j\int_0^t \omega_2(\tau)d\tau \right] \tag{2.45}$$

în condiţiile $A_1 > A_2$ sau $a = A_2/A_1 < 1$ şi perturbare mutuală aditivă se constituie în semnalul

$$s(t) = s_1(t) + s_2(t) = A_1 \exp\left[j\int_0^t \omega(\tau)d\tau \right]\left\{1 + a\exp\left[j\int_0^t \omega_\Delta(\tau)d\tau \right]\right\} \tag{2.46}$$

care este modulat concomitent în amplitudine şi în fază

$$s(t) = A_1(1 + a^2 + 2a\cos\alpha)\exp\left[j\int_0^t \omega(\tau)d\tau + j\varphi \right] \tag{2.47}$$

cu $\alpha = \int_0^t \omega(\tau)d\tau$ şi cu $\tan\varphi = (a\sin\alpha)/(1 + a\cos\alpha)$.

Faza semnalului interesant se modifică în conformitate cu relaţia (5.48)

$$\Psi(t) = \int_0^t \omega(\tau)d\tau + \varphi = \omega_0 t + \Phi(t) + \varphi \tag{2.48}$$

şi frecvenţa instantanee, succesiv

$$\omega_i(t) = \omega_1(t) + \frac{d\varphi}{dt} = \omega_0 + \Omega_1(t) + \frac{d\varphi}{dt} \tag{2.49}$$

$$\omega_i(t) = \omega_1(t) + \frac{d\varphi}{dt} = \omega_0 + \Omega_1(t) + a\omega_\Delta(t)\frac{a+\cos\alpha}{1+2a\cos\alpha+a^2} \qquad (2.50)$$

Pentru semnale nemodulate frecvenţa este de forma

$$\omega_i(t) = \omega_{10} + \frac{d\varphi}{dt} \qquad (2.51)$$

şi media temporală a acestei frecvenţe este

$$\overleftrightarrow{\omega_i(t)} = \lim_{T\to\infty}\frac{1}{T}\int_0^T \omega_i(\tau)d\tau = \omega_{10} + \lim_{T\to\infty}\frac{\varphi(T)-\varphi(0)}{T} = \omega_{10} \qquad (2.52)$$

aşadar frecvenţa instantanee este cea a semnalului mai puternic.
Dacă $a << 1$, atunci

$$\Psi(t) = \omega_{10}t + \Phi_1(t) + a\sin\alpha \qquad (2.53)$$

şi

$$\omega_i(t) = \omega_{10} + \Omega_1(t) + a\omega_\Delta(t)\cos\alpha \qquad (2.54)$$

Separat, pentru modulaţia de fază (M□) şi pentru modulaţia de frecvenţă (*MF*), semnalul recepţionat este, respectiv

$$n_\Phi(t) = \Phi_1(t) + a\sin\int_o^t \omega_\Delta(\tau)d\tau \qquad (2.55)$$

$$n_F(t) = \Omega_1(t) + a\omega_\Delta(t)\cos\int_o^t \omega_\Delta(\tau)d\tau \qquad (2.56)$$

Dacă deviaţiile de frecvenţă sunt mici, atunci $\omega_\Delta = \omega_{10} - \omega_{20}$ şi

$$n_\Phi(t) = \Phi_1(t) + a\sin(\omega_{10}-\omega_{20})t \qquad (2.57)$$

$$n_F(t) = \Omega_1(t) + a\omega_\Delta(t)\cos(\omega_{10}-\omega_{20})t \qquad (2.58)$$

Diferenţa de frecvenţă $\omega_\Delta = \omega_{10} - \omega_{20}$ trebuie comparată cu frecvenţa maximă □$_M$ din spectrul semnalului. Dacă □□ > □$_M$, atunci efectul perturbator poate fi ignorat.

Rapoartele semnal/perturbaţie sunt:

$$\left(\frac{S}{P}\right)_{\Phi} = \frac{\overleftrightarrow{m^2(t)}}{\frac{1}{2}a^2} = 2\frac{A_1^2}{A_2^2}\overleftrightarrow{m^2(t)} \tag{2.59}$$

$$\left(\frac{S}{P}\right)_{F} = \frac{\overleftrightarrow{m^2(t)}}{\frac{1}{2}(\omega_{10}-\omega_{20})^2 a^2} = 2\frac{A_1^2}{A_2^2}\frac{\overleftrightarrow{m^2(t)}}{\omega_{\Delta}^2} \tag{2.60}$$

Zgomotele perturbatoare pot fi de impulsuri sau de fluctuaţii. Cele două tipuri de perturbaţii necoerente se tratează întrucâtva diferit.

Zgomotele de impulsuri pot fi eliminate prin limitare, deoarece ele afectează mai curând amplitudinea.

Zgomotele de fluctuaţii care interferă cu semnale modulate exponenţial se tratează pornind de la forma

$$z(t) = V(t)\cos[\omega_0 t + \Theta(t)] \tag{2.61}$$

cu amplitudinea $V(t)$ aleatoare, de exemplu gaussiană, cu faza $\Theta(t)$ aleatoare şi uniform repartizată pe întervalul $[0, 2\pi]$.

Tratarea este în bună măsură analogă celei din cazul perturbaţiilor coerente.

5.3. Modulaţia secvenţelor de impulsuri periodice

Secvenţele de impulsuri periodice, conform reprezentării din figura 5.6, sunt caracterizate printr-o amplitudine (A), prin perioada (T)/frecvenţa lor, printr-o durată (τ) şi prin poziţia/faza lor. Fiecare dintre aceste caracteristici parametrice este susceptibilă ca prin modulare să transforme secvenţa de impulsuri într-o secvenţă purtătoare de informaţie.

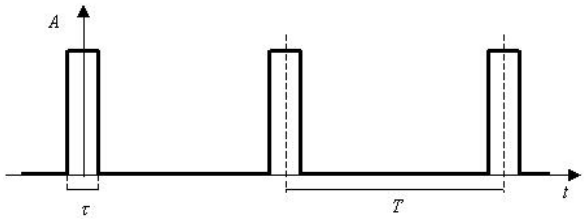

Fig. 2.6. Secvenţă de impulsuri periodice

Modulaţia impulsurilor în amplitudine (MIA) este similară modulaţiei liniare a purtătoarelor sinusoidale: amplitudinea se modifică în ritmul semnalului modulator.

Modulaţia de poziţie a impulsurilor (MIP) este analogă într-un fel modulaţiei de fază aplicată purtătoarelor sinusoidale.

Modulaţia în durată (MID) nu are un echivalent între modulaţiile purtătoarelor sinusoidale.

Modulaţia în frecvenţă a impulsurilor este posibilă, dar este complicată sub aspect ingineresc şi nu aduce avantaje noi faţă de modulaţia în poziţie.

Fiecare dintre cele trei modalităţi de modulare a impulsurilor se poate implementa în varianta *uniformă* sau în varianta *naturală*.

În cazul *uniform* eşantionarea semnalului modulator se efectuează la intervale regulate şi parametrul modulat urmează variaţia acelor eşantioane.

În cazul *natural* eşantionarea aceasta nu este supusă regulii uniformităţii. Diferenţele apar mai clar în prezentările specifice date în continuare.

Modulaţia în amplitudine cu eşantionare uniformă (MIA/U)

Mesajul eşantionat are în domeniul frecvenţelor expresia

$$M^*(\omega) = a\sum_{n=-\infty}^{\infty} M(\omega - n\omega_0) \tag{2.62}$$

Din eşantioane trebuie să fie formate impulsuri rectangulare. Se utilizează un aşa-numit *filtru de formare* cu funcţia de transfer

$$H_f(\omega) = \frac{\sin\omega\frac{\tau}{2}}{\omega\frac{\tau}{2}} e^{-j\omega\frac{\tau}{2}} \tag{2.63}$$

Semnalul obţinut este

$$S(\omega) = M^*(\omega)H_f(\omega) = a\frac{\sin\omega\frac{\tau}{2}}{\omega\frac{\tau}{2}}e^{-j\omega\frac{\tau}{2}}\sum_{n=-\infty}^{\infty}M(\omega - n\omega_0) \tag{2.64}$$

cu $M(\square) = 0$ pentru $|\omega| > (1/2)\omega_0$, conform teoremei de eşantionare.

Pe această expresie se pot face studii detaliate asupra spectrului de frecvenţe.

De altfel, la fiecare tip de modulaţie a secvenţelor de impulsuri, scopul scrierii expresiilor în domeniul frecvenţelor pentru semnalele modulate are ca scop ultim crearea posibilităţii de a studia pe spectrul semnalului modalităţile de recuperare a informaţiei.

Modulaţia în amplitudine cu eşantionare naturală (MIA/N)

Cu notatiile $e_T(t)$ pentru secvenţa periodică de impulsuri rectangulare de valoare medie a şi cu $m(t)$ pentru semnalul modulator, semnalul modulat în amplitudine cu eşantionare naturală are expresia

$$s(t) = m(t)e_T(t) = a\sum_{n=-\infty}^{\infty}\frac{\sin n\omega_0\frac{\tau}{2}}{n\omega_0\frac{\tau}{2}}m(t)\cos n\omega_0 t \tag{2.65}$$

În domeniul frecvenţelor acelaşi semnal se exprimă ca

$$S(\omega) = \frac{1}{2}a\sum_{n=-\infty}^{\infty}\frac{\sin n\omega_0\frac{\tau}{2}}{n\omega_0\frac{\tau}{2}}[M(\omega - n\omega_0) + M(\omega + n\omega_0)] \tag{2.66}$$

Pe această expresie, care este diferită de aceea obţinută pentru eşantionarea uniformă, se poate studia importanţa diferitelor componente ale spectrului.

Modulaţia în poziţie cu eşantionare uniformă (MIP/U)

Ca şi la modulaţia de fază/frecvenţă a purtătoarelor sinusoidale, pentru simplificarea şi esenţializarea discuţiei, se consideră că semnalul mesaj este sinusoidal

$$m(t) = \square\cos\square t \tag{2.67}$$

Poziţia impulsurilor se modifică în conformitate cu relaţia

$$\Box p \sin \Box nT, \ \Box p = k\Box \tag{2.68}$$

Suprapunerea a două impulsuri succesive este evitată dacă $\Box p < T / 2$.

Se consideră următoarea secvenţă de impulsuri $\Box_T$ periodică, cu poziţiile modificate conform eşantioanelor semnalului mesaj prelevate uniform

$$d(t) = aT \sum_{n=-\infty}^{\infty} \delta(t - nT + \Delta p \sin \Omega nT) \tag{2.69}$$

În domeniul frecvenţelor acest semnal este

$$D(\omega) = aT \sum_{n=-\infty}^{\infty} \int_{-\infty}^{+\infty} \delta(t - nT + \Delta p \sin \Omega nT) e^{-j\omega t} dt = aT \sum_{n=-\infty}^{\infty} e^{-j\omega(nT - \Delta p \sin \Omega nT)} \tag{2.70}$$

Dar, conform unei discuţii purtate la modulaţia de fază/frecvenţă a purtătoarelor sinusoidale, $e^{j\omega \Delta p \sin \Omega nT} = \sum_{m=-\infty}^{\infty} J_m(\omega \Delta p) e^{jmn\Omega T}$, deci

$$D(\omega) = aT \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} J_m(\omega \Delta p) e^{-j\omega nT} e^{jmn\Omega T} = aT \sum_{m=-\infty}^{\infty} J_m(\omega \Delta p) \sum_{n=-\infty}^{\infty} e^{j(m\Omega - \omega)nT} \tag{2.71}$$

Fie funcţia auxiliară periodică de perioada neprecizată, $\Box_{\Box\Box}$

$$F(u) = \sum_{n=-\infty}^{\infty} \delta(u - n\omega_0) = \sum_{n=-\infty}^{\infty} C_n e^{jn\frac{2\pi}{\omega_0}u} \tag{2.72}$$

Coeficienţii Fourier ai acestei funcţii sunt

$$C_n = \frac{1}{\omega_0} \int_{-\frac{\omega_0}{2}}^{\frac{\omega_0}{2}} F(u) e^{-jn\frac{2\pi}{\omega_0}u} du = \frac{1}{\omega_0} \int_{-\frac{\omega_0}{2}}^{\frac{\omega_0}{2}} \delta(u) e^{-jn\frac{2\pi}{\omega_0}u} du = \frac{1}{\omega_0} \tag{2.73}$$

aşadar

$$F(u) = \sum_{n=-\infty}^{\infty} \delta(u - n\omega_0) = \sum_{n=-\infty}^{\infty} \frac{1}{\omega_0} e^{jn\frac{2\pi}{\omega_0}u} \tag{2.74}$$

Dacă $u = m\Box\Box\Box\Box\Box$, atunci

$$F(m\Omega-\omega)=\sum_{n=-\infty}^{\infty}\delta(m\Omega-\omega-n\omega_0)=\frac{T}{2\pi}\sum_{n=-\infty}^{\infty}e^{j(m\Omega-\omega)nT} \tag{2.75}$$

şi printr-o substituţie se obţine

$$D(\omega)=2\pi a\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}J_m(\omega\Delta p)\delta(m\Omega-\omega-n\omega_0) \tag{2.76}$$

După trecerea prin filtrul de formare se obţine

$$S(\omega)=D(\omega)H_f(\omega)=2\pi a\frac{\sin\omega\frac{\tau}{2}}{\omega\frac{\tau}{2}}e^{-j\omega\frac{\tau}{2}}\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}J_m(\omega\Delta p)\delta(m\Omega-\omega-n\omega_0) \tag{2.77}$$

şi din nou se crează posibilitatea studiului comparativ al componentelor din spectrul semnalului modulat.

Modulaţia în poziţie cu eşantionare naturală (MIP/N)

Fie semnalul *i(t),* cu *u*(□)) semnalul treaptă unitate

$$i(t)=\frac{d}{dt}\sum_{n=-\infty}^{\infty}u(\theta_n) \tag{2.78}$$

şi fie mesajul

$$m(t)=\square\ \sin\ \square t \tag{2.79}$$

şi $\square p = k\square$, $\theta_n = t - nT + \Delta p\sin\Omega t$. Salturile treaptă apar la $\square_n$=0, adică la momentele definite de relaţia $t_n = nT - \Delta p\sin\Omega t_n$. Eşantionarea este naturală, deplasarea impulsului depinde de eşantionul la momentul t_n.

Se mai poate scrie

$$i(t)=aT\sum_{n=-\infty}^{\infty}\frac{du(\theta_n)}{d\theta_n}\frac{d\theta_n}{dt}=aT\sum_{n=-\infty}^{\infty}\delta(\theta_n)\frac{d\theta_n}{dt} \tag{2.80}$$

$$i(t) = aT\sum_{n=-\infty}^{\infty}\delta(t - nT + \Delta p\sin\Omega t)(1 + \Delta p\Omega\cos\Omega t) =$$

$$= a(1 + \Delta p\Omega\cos\Omega t)\sum_{n=-\infty}^{\infty}e^{jn\omega_0(t+\Delta p\sin\Omega t)} = a(1 + \Delta p\Omega\cos\Omega t)\sum_{n=-\infty}^{\infty}e^{jn\omega_0 t}\sum_{m=-\infty}^{\infty}J_m(n\omega_0\Delta p)e^{jm\Omega t} =$$

$$= a\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}J_m(n\omega_0\Delta p)\left(e^{j(n\omega_0+m\Omega)t} + \frac{\Omega\Delta p}{2}e^{j[n\omega_0+(m-1)\Omega]t} + \frac{\Omega\Delta p}{2}e^{j[n\omega_0+(m+1)\Omega]t}\right)$$

(2.81)

Semnalul $i(t)$ în domeniul frecvenţelor este

$$I(\omega) = a\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}J_m(n\omega_0\Delta p).\{\delta(\omega - n\omega_0 - m\Omega) +$$

$$+\frac{\Omega\Delta p}{2}\delta[\omega - n\omega_0 - (m-1)\Omega] + \frac{\Omega\Delta p}{2}\delta[\omega - n\omega_0 - (m+1)\Omega]\} \qquad (2.82)$$

Prin multiplicarea cu funcţia de transfer a filtrului de formare rezultă semnalul $S(\square)$, spectrul semnalului modulat. Studiul componentelor spectrului este acum deplin posibil.

Modulaţia în durată cu eşantionare uniformă şi cu eşantionare naturală (MID/U şi MID/N)

Fie semnalul

$$g(t) = i(t) - \square_T(t - \square_0) \qquad (2.83)$$

Atunci, semnalul *s(t)* din relaţia (5.84) este modulat în durată.

$$s(t) = \int_{-\infty}^{t} g(t)dt \qquad (2.84)$$

Dacă se uzează de perechile Fourier $i(t) \Leftrightarrow I(\omega)$ şi $\delta_T(t - \tau_0) \Leftrightarrow \Delta(\omega)$, atunci

$$S(\omega) = \frac{1}{j\omega}G(\omega) = \frac{1}{j\omega}[I(\omega) - \Delta(\omega)] \qquad (2.85)$$

$$\Delta(\omega) = \int_{-\infty}^{+\infty} \delta_T(t-\tau_0) e^{-j\omega t} dt = 2\pi a e^{-j\omega\tau_0} \sum_{n=-\infty}^{\infty} \delta(\omega - n\omega_0) \tag{2.86}$$

Acum, pentru eşantionarea *uniformă*, adică pentru $\tau = \tau_0 - \Delta\tau \sin\Omega nT$, rezultă

$$I(\omega) = 2\pi a \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} J_m(\omega\Delta\tau)\delta(m\Omega - \omega - n\omega_0) \tag{2.87}$$

$$S(\omega) = 2\pi \frac{a}{j\omega} \sum_{n=-\infty}^{\infty} \left[\sum_{m=-\infty}^{\infty} J_m(\omega\Delta\tau)\delta(m\Omega - \omega - n\omega_0) - e^{-j\omega\tau_0}\delta(\omega - n\omega_0) \right] \tag{2.88}$$

Pentru eşantionarea *naturală*, $\tau = \tau_0 - \Delta\tau \sin\Omega t$ şi

$$S(\omega) = 2\pi \frac{a}{j\omega} \frac{1}{T} \sum_{n=-\infty}^{\infty} \left\{ \sum_{m=-\infty}^{\infty} J_m(n\omega_0\Delta\tau) \left[\delta(\omega - n\omega_0 - m\Omega) + \right.\right.$$
$$\left. + \frac{\Omega\Delta\tau}{2}\delta[\omega - n\omega_0 - (m-1)\Omega] + \frac{\Omega\Delta\tau}{2}\delta[\omega - n\omega_0 - (m+1)\Omega] \right] -$$
$$\left. - e^{-j\omega\tau_0}\delta(\omega - n\omega_0) \right\} \tag{2.89}$$

Tabloul distribuţiilor spectrale este acum complet. Pentru toate tipurile uzuale de modulaţie a secvenţelor periodice de impulsuri rectangulare, pe expresiile obţinute, se pot observa elemente cu semnificaţie inginerească. De exemplu, în unele cazuri semnalele modulatoare pot fi recuperate prin simpla filtrare *trece-jos*. Pe relaţiile stabilite se pot observa de asemenea distorsiunile semnalelor mesaj la demodulare.

5.4. Modulaţia diferenţială

Modulaţia diferenţială (delta) este gândită ca un mijloc de transmitere nu a semnalului însuşi ci a sensului în care semnalul (analogic) se modifică într-un interval finit de timp. Modificările pot fi de creştere sau de scădere şi transmiterea este binară, un exemplu putând fi unitatea pentru creştere, zero pentru scăderi, ceea ce este o alegere cât se poate de naturală.

Pentru implementarea modulaţiei diferenţiale se crează un semnal $g(t)$ care variază în trepte egale şi care urmăreşte semnalul $m(t)$. În acest context, se spune că primul semnal este *aservit* celui din urmă. Semnalul $g(t)$ are forma

$$g(t) = g_0 + \sum_{n=0}^{\infty} \Delta_n u(t - n\tau) \tag{2.90}$$

în care $u(t)$ este funcţia treaptă unitară, g_0 este o constantă şi $\Delta_n = \pm\Delta$, adică poate lua valori negative sau pozitive, dar de amplitudine fixă □. La intervale regulate de timp, la momente discrete n□, cele două funcţii/semnale se compară. Dacă $m(n$□$) > g(n$□$)$, atunci $\Delta_n = +$□, dacă $m(n$□$) < g(n$□$)$, $\Delta_n = -$□.

Decodarea se realizează prin integrarea semnalului, de exemplu cu un circuit *RC* simplu sau multiplu.
Mărimea cuantelor □ este legată de intervalul de comparare □ prin relaţia

$$\left|\frac{dm(t)}{dt}\right|_{\max} \le \frac{\Delta}{\tau} \tag{2.91}$$

Dacă semnalul $m(t)$ variază mai rapid decât permite relaţia de mai sus, atunci semnalul $g(t)$ nu mai poate urmări semnalul căruia îi este aservit.

Variante posibile de implementare a modulaţiei diferenţiale sunt prezentate în figura 2.7.

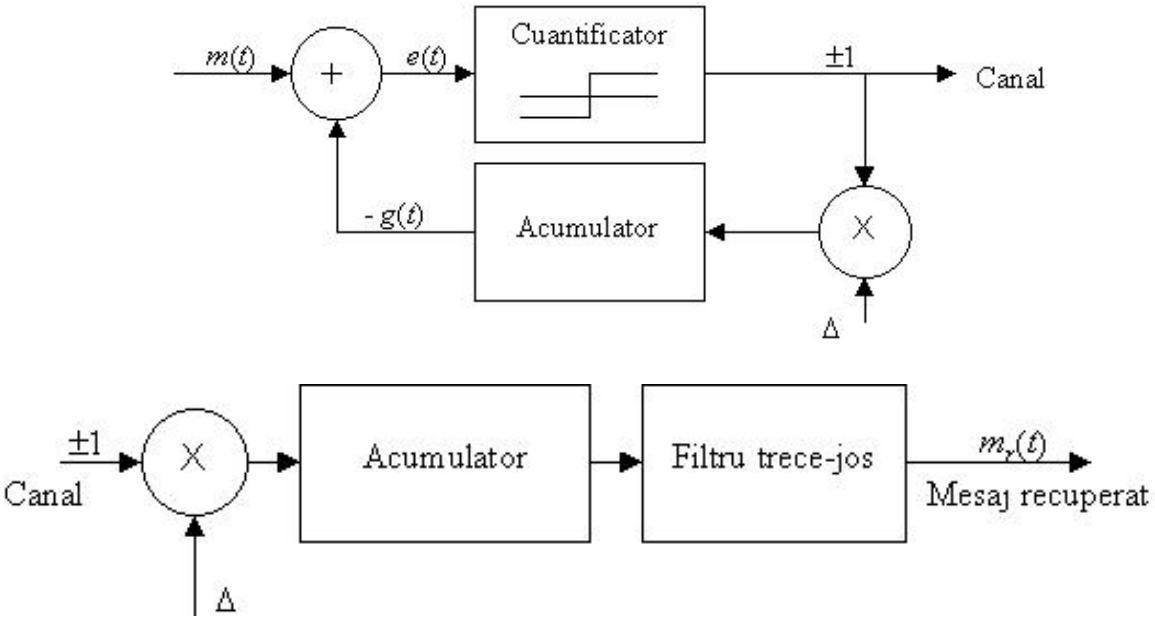

Fig.2.7. Variante de implementare a modulaţiei diferenţiale

Efectul cuantificării

Eşantionarea mesajului se face conform relaţiei

$$m(t) = \sum_{k=-\infty}^{\infty} m\left(\frac{k}{2w}\right) \frac{\sin 2\pi w\left(t - \frac{k}{2w}\right)}{2\pi w\left(t - \frac{k}{2w}\right)} \tag{2.92}$$

cu regula cunoscută de eşantionare care ia în consideraţie lărgimea spectrului semnalului.

Dacă are loc o cuantificare se transmite de fapt semnalul

$$m_q(t) = \sum_{k=-\infty}^{\infty} m_q\left(\frac{k}{2w}\right) \frac{\sin 2\pi w\left(t - \frac{k}{2w}\right)}{2\pi w\left(t - \frac{k}{2w}\right)} \tag{2.93}$$

care diferă de eşantioanele adevărate conform relaţiei

$$m_q\left(\frac{k}{2w}\right) = m\left(\frac{k}{2w}\right) + \Theta_k q \tag{2.94}$$

cu q mărimea cuantei şi Θ_k o variabilă aleatoare uniform repartizată pe intervalul (– 0,5, 0,5). Aşadar, există o eroare de cuantizare care este $e_k = \Theta_k q$. Cu notaţia

$$s_k(t) = \frac{\sin 2\pi w\left(t - \frac{k}{2w}\right)}{2\pi w\left(t - \frac{k}{2w}\right)} \tag{2.95}$$

semnalul eroare se scrie

$$e(t) = q \sum_{k=-\infty}^{\infty} \Theta_k s_k(t) \tag{2.96}$$

Puterea medie a zgomotului de cuantificare este

$$\overleftrightarrow{e^2(t)} = q^2 \overleftrightarrow{\sum_{k=-\infty}^{\infty} \Theta_k s_k(t) \sum_{j=-\infty}^{\infty} \Theta_j s_j(t)} =$$

$$= q^2 \lim_{T \to \infty} \frac{1}{2T} \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \Theta_k \Theta_j \int_{-T}^{T} s(t)s(t)dt \tag{2.97}$$

Dar funcţiile eşantion sunt ortogonale, iar variabilele Θ_k , Θ_j sunt independente, de medie nulă şi de dispersie relativ uşor de calculat. Aşadar, energia semnalului eroare este

$$q^2 \lim_{n \to \infty} \frac{2w}{2n} \sum_{k=-n}^{n} \Theta_k^2 \frac{1}{2w} = \lim_{n \to \infty} \frac{1}{2n} \sum_{k=-n}^{n} \Theta_k^2 \tag{2.98}$$

ultimul fiind un moment centrat de ordinul al doilea egal cu 1/12.

Aşadar, în final

$$\overleftrightarrow{e^2(t)} = \frac{1}{12} q^2 \tag{2.99}$$

5.5. Transmisiuni multiple

Utilizarea multiplă a canalelor de transmitere a informaţiei este un capitol de mare interes în teoria comunicaţiilor. Curent, în practică sunt folosite trei tipuri de multiplicare a transmisiunilor:

- Cu diviziunea căilor în fază;
- Cu diviziunea căilor în frecvenţă;
- Cu diviziunea căilor în timp.

Pentru ca separarea canalelor să fie posibilă, este necesar ca $\int_{-\infty}^{\infty} s_k(t)s_l(t)dt = 0$ pentru $k \neq l$ sau, echivalent, $\int_{-\infty}^{\infty} S_k(\omega)S_l^*(\omega)dt = 0$ pentru $k \neq l$, unde $s_k(t)$ şi $S_k(\Theta)$ sunt semnale pereche Fourier transmise pe calea marcată de indicele *k*. Notaţia $S_l^*(\omega)$ este pentru conjugatul semnalului $S_l(\Theta)$).

Diviziunea căilor în fază

Semnalele

$$s_1(t) = m_1(t)\cos(\omega_0 t + \varphi_0) \tag{2.100}$$

$$s_2(t) = m_2(t)\cos\left(\omega_0 t + \varphi_0 + \frac{\pi}{2}\right) \tag{2.101}$$

sunt ortogonale dacă frecvenţa maximă din spectrul mesajelor este inferioară frecvenţei ω_0, adică dacă $M_1(\omega) = 0$ şi $M_2(\omega) = 0$ pentru $|\omega| \geq \omega_0$.

Într-adevăr

$$\int_{-\infty}^{\infty} s_1(t)s_2(t)dt = \int_{-\infty}^{\infty} m_1(t)m_2(t)\cos(\omega_0 t + \varphi_0)\sin(\omega_0 t + \varphi_0)dt = \\ = \frac{1}{2}\int_{-\infty}^{\infty} m_1(t)m_2(t)\sin(2\omega_0 t + 2\varphi_0)dt \tag{2.102}$$

şi cu notaţia $m(t) = m_1(t)\sin(2\omega_0 t + 2\varphi_0)$, se obţine

$$M(\omega) = \frac{1}{2} j e^{j2\varphi_0} M_2(\omega - 2\omega_0) + \frac{1}{2} j e^{-j2\varphi_0} M_2(\omega + 2\omega_0) \tag{2.103}$$

Teorema de convoluţie în domeniul frecvenţelor conduce la

$$\int_{-\infty}^{\infty} m_1(t)m(t)e^{-j\omega t}dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} M_1(\Omega)M(\omega - \Omega)d\Omega \tag{2.104}$$

ceea ce pentru $\omega = 0$ produce un rezultat util şi utilizabil

$$\int_{-\infty}^{\infty} s_1(t)s_2(t)dt = \frac{1}{2}\int_{-\infty}^{\infty} m_1(t)m(t)dt = \frac{1}{4\pi}\int_{-\infty}^{\infty} M_1(\Omega)M(-\Omega)d\Omega = \\ = \frac{1}{8\pi} j \int_{-\infty}^{\infty}\left[e^{j2\varphi_0} M_1(\Omega)M_2(-\Omega - 2\omega_0) + e^{-j2\varphi_0} M_1(\Omega)M_2(-\Omega + 2\omega_0)\right]d\Omega \tag{2.105}$$

Dar spectrul semnalului $M_1(\omega)$ şi spectrele oricăruia dintre semnalele $M_2(-\omega - 2\omega_0)$ şi $M_2(-\omega + 2\omega_0)$ nu au puncte în care să fie concomitent nenule, de unde verificarea condiţiei de ortogonalitate.

Ca şi în alte cazuri în care faza trebuie să fie riguros controlată, şi aici este necesară o sincronizare a oscilatorului local prin transmiterea unei mici reminiscenţe a purtătoarei, conform reprezentării din figura 2.8.

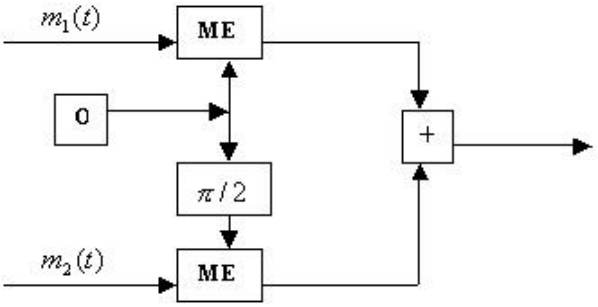

Fig.2.8. Sincronizarea oscilatorului local la diviziunea căilor în fază

În cazul separării căilor în frecveţă sau în timp, condiţiile de ortogonalitate sunt evident îndeplinite.

Diviziunea căilor în frecvenţă

Semnalele au spectru limitat. Ele sunt separabile dacă spectrele lor nu se suprapun. Matematic, aceasta înseamnă că

$$\int_{-\infty}^{\infty} S_i(\omega) S_j(\omega) d\omega = 0 \qquad (2.106)$$

adică cele două semnale de indici *i* şi *j*, din mai multe posibile, sunt ortogonale în domeniul frecvenţelor. Oricare două canale care îndeplinesc condiţia de ortogonalitate de mai sus sunt separabile.
Separarea efectivă la utilizare se realizează cu filtre trece-bandă potrivite.

Diviziunea căilor în timp

Factorul de umplere al secvenţelor periodice de impulsuri rectangulare este mult sub unitate. Există, aşadar, un timp de absenţă a impulsurilor în care se poate intercala o altă secvenţă de impulsuri rectangulare care poate purta alt mesaj. Dacă impulsurile uneia dintre căi nu se suprapun cu cele ale altei căi, atunci cele două căi sunt separabile printr-un sistem de porţi adecvat. Într-o exprimare matematică, lipsa suprapunerii este exprimată ca

$$\int_{-\infty}^{\infty} s_k(t) s_l(t) dt = 0 \quad \text{pentru } k \neq l \qquad (2.107)$$

Separarea necesită transmiterea unor semnale de sincronizare de la emiţător la receptor, la intervale de timp cu atenţie selectate.

3. Codificarea şi decodificarea informaţiei

În capitolul anterior au fost descrise procedurile de prelucrare a semnalelor purtătoare de informaţie în vederea transmiterii pe canale perturbate, cu predilecţie, a datelor binare. S-a constatat că posibilitatea de eroare reziduală este dependentă de raportul semnal-zgomot de la intrarea în receptor şi de viteza de transmisie a biţilor. În anumite situaţii, această probabilitate de eroare nu scade sub limite acceptabile, ceea ce impune recurgerea la utilizarea de coduri detectoare de erori şi la tehnici de corecţie a acestora.

Detectarea şi corectarea erorilor sunt în strânsă legătură cu noţiunea de redundanţă care se referă la adăugarea unor biţi de control, pe lângă cei purtători de informaţie, care permit depistarea unor secvenţe eronate de biţi. Procedurile de codificare/decodificare nu acţionează deci la nivel de bit, ci la nivel de mesaj (secvenţe de biţi, cuvinte, blocuri).

3.1 Codificarea şi decodificarea pe canale fără perturbaţii

3.1.1 Definirea unui cod

Fie o sursă discretă, fără memorie, având alfabetul $S=\{S_1,S_2,...,S_N\}$, cu probabilităţile de apariţie asociate $p(S_i)=p_i, P=\{p_1,p_2,...,p_N\}$ şi fie ansamblul finit de semne (caractere, litere) al alfabetului canalului $X=\{x_1,x_2,...x_q\}$, care, în particular pentru cazul binar, este $X=\{0,1\}$.

Ansamblul de secvenţe finite de litere $X_{a_1},X_{a_2},...,X_{a_n}$ este reuniunea extensiilor lui X

$$X^*=\bigcup_{n\geq 1}X^n, \text{ cu } X^*=\{x_{a_1},...,x_{a_n}\} \tag{3.1}$$

Orice aplicaţie $S \to X^*$ se numeşte *codificarea* ansamblului S prin alfabetul X. Elementul lui X^*, S_i^*, ce corespunde lui S_i, este un *cuvânt de cod*. Lungimea cuvântului de cod este egală cu numărul de litere care îl formează.

Totalitatea cuvintelor de cod constituie *codul* lui S, cu menţiunea că X^* poate conţine şi combinaţii care nu aparţin codului, numite *cuvinte fără sens*. Altfel spus, *un cod este o corespondenţă biunivocă între mulţimea mesajelor sursă şi o mulţime de cuvinte de cod*, astfel încât un text constituit dintr-o secvenţă de mesaje $m_j = \{S_{i_1}, S_{i_2}, ..., S_{i_k}\}$ este codificat printr-o secvenţă de cuvinte de cod, cu sens $m_j^* = \{S_{i_1}^*, S_{i_2}^*, ..., S_{i_k}^*\}$

În replică, operaţia de *decodificare* (decodare) implică posibilitatea de a separa cuvintele de cod în mod unic, ceea ce se poate scrie $(\forall) S_i \neq S_j \Rightarrow S_i^* \neq S_j^*$, funcţia $S \to X^*$ să fie injectivă. În aceste condiţii, codul este *regulat* sau *nesingular*.

Dar regularitatea nu este suficientă pentru înlăturarea ambiguităţii. De exemplu, fie secvenţele $S_1 = 0; S_2 = 10; S_3 = 01$. Un text codificat 010 poate fi interpretat atât ca $S_1 S_2$, cît şi ca $S_3 S_1$. Pentru a distinge fără ambiguitate un text trebuie ca fiecărei succesiuni de cuvinte să-i corespundă şi o succesiune unică de litere. Codurile de acest tip se numesc *unic decodabile/descifrabile*. Printre condiţiile suficiente care asigură descifrabilitatea, cele mai importante sunt:

- utilizarea cuvintelor de cod de aceeaşi lungime (bloc);
- utilizarea unui semn distinct, de separare, între cuvinte.

Există însă şi coduri care nu necesită asemenea artificii suplimentare, numite *coduri separabile*. Un exemplu este cel prezentat în tabelul 3.1

Tabelul 3.1

Mesaje	A	B	C	D
S_0	00	0	0	0
S_1	01	10	01	10
S_2	10	110	011	110
S_3	11	1110	0111	111

Codurile exemplificate în tabelul 3.1 sunt definite astfel:

A. cod ponderat binar natural;
B. cod care are întotdeauna ultima literă "0";
C. cod care are întotdeauna prima literă "0";
D. o variantă a codului B.

În aceste condiţii, considerând succesiunea $S_3S_1S_0S_2$ codificată în variantele B şi C, aceasta se prezintă sub formele:

B: 111010.0110
C: 011101.0011

În cazul în care se propune descifrarea secvenţei până la punct, pentru codul B, succesiunea S_3S_1 este *descifrabilă,* în timp ce pentru codul C, există *ambiguitate*: după S_3 ar putea fi S_2, S_3, S_1.

Rezultă de aici condiţia necesară şi suficientă ca un cod să fie *ireductibil* (instantaneu): nici un cuvânt de cod să nu fie prefix al altui cuvânt de cod. Această condiţie este cunoscută sub numele de *condiţia de prefix.*
Codurile A şi B sunt ireductibile, în timp ce codul C nu este ireductibil (nu satisface condiţia de prefix).

3.1.2 Criterii de apreciere a unui cod

Deoarece la transmisia mesajelor, costul exploatării unui sistem de transmisie creşte liniar cu timpul, un criteriu convenabil de apreciere a unui cod este *lungimea medie a unui cuvânt*

$$\pi = \sum_{i=1}^{n} p_i \cdot n_i \tag{3.2}$$

cu: p_i – probabilităţile de apariţie asociate, $\sum_{i=1}^{n} p_i = 1$;

n_i – numărul de litere din cuvântul de cod cu indicele i;

π – este un parametru ce precizează compactitatea codului şi este evident că trebuie să fie cât mai mic (pentru $\pi_{\min} \rightarrow$ coduri compacte/cvasioptimale).

Un al doilea criteriu de apreciere a unui cod este prin calculul entropiei sursei, care conduce la determinarea *eficienţei codului*

$$\eta = \frac{H}{\bar{n}\log q}\Bigg|_{q=2} = \frac{H}{\bar{n}} \tag{3.3}$$

Exemplul 3.1
Pentru sursa prezentată în tabelul 3.1 se consideră următoarele probabilităţi de apariţie a mesajelor: $p_1 = \frac{1}{2}; p_2 = \frac{1}{4}; p_3 = p_4 = \frac{1}{8}$. Să se determine eficienţa fiecărui cod.

Rezolvare

Entropia sursei $H = -\sum_1^n p_i \log p_i = \frac{7}{4} bit$

$\bar{n}_A = 2; \quad \bar{n}_B = \frac{15}{8}; \quad \bar{n}_C = \frac{15}{8}; \quad \bar{n}_D = \frac{7}{4}$

$\eta_A = \frac{7}{8}\left(\frac{7}{4} \cdot \frac{1}{2}\right); \quad \eta_B = \frac{14}{15}; \quad \eta_C = \frac{14}{15}; \quad \eta_D = 1$

Prima teoremă a lui Shannon

Pentru orice sursă omogenă, există un cod ireductibil pentru care lungimea medie a cuvintelor este oricât de apropiată de marginea inferioară.

Interesul practic al acestei teoreme în transmiterea informaţiilor se limitează la sistemele la care se doreşte să se codifice un număr cât mai mare de texte cu un număr dat de caractere.

3.1.3 Metode de elaborare a codurilor compacte

A. Metoda Shannon

Se consideră dată o listă a mesajelor m_i emise de sursă, cu probabilităţile aferente p_i. Principiul metodei constă în aranjarea mesajelor în ordinea descrescătoare a probabilităţilor lor de apariţie $p_1 \geq p_2 \geq ... \geq p_N$ şi determinarea celor mai mici întregi n_i astfel încât

$$n_i \geq \frac{\log\frac{1}{p_i}}{\log q}\Bigg|_{q=2(0,1)} \Rightarrow n_i \geq \log\frac{1}{p_i} \tag{3.4}$$

ceea ce conduce la $n_1 \leq n_2 \leq ... \leq n_N$.

B. Metoda Shannon-Fano

Pentru codificarea binară, în aceleaşi condiţii iniţiale ca şi în cazul A, metoda constă în gruparea mesajelor în două grupe, cu probabilităţile cumulate cât mai apropiate. Se codifică fiecare grupă cu 0, respectiv 1, apoi se repetă procedura în cadrul fiecărei grupe, până când în fiecare grupă rămân doar două mesaje.

Exemplul 3.2

Pentru setul de mesaje (s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8) cu probabilităţile aferente (0.4, 0.18, 0.1, 0.1, 0.07, 0.06, 0.05, 0.04), maniera de lucru este cea prezentată în figura 3.1.

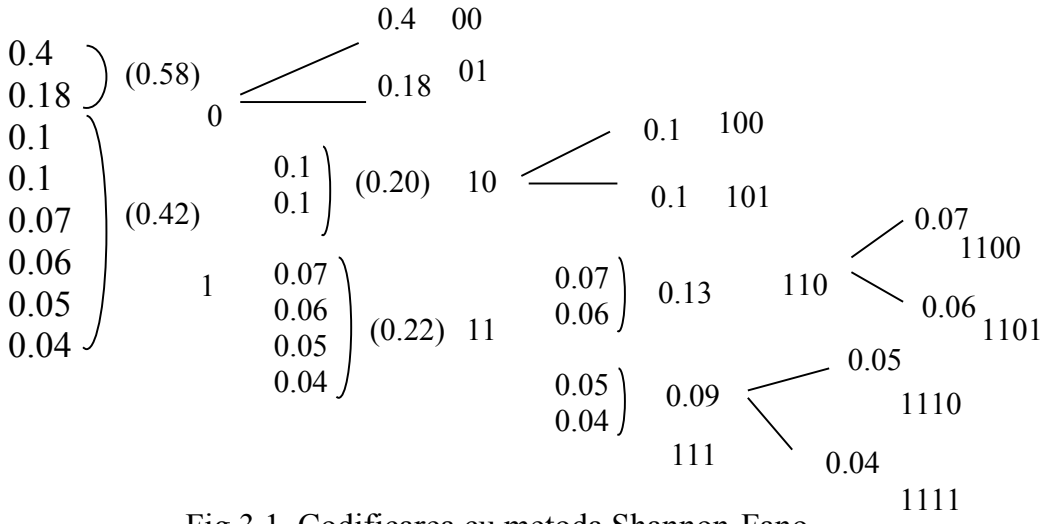

Fig.3.1. Codificarea cu metoda Shannon-Fano

C. Metoda Huffman – varianta Schwartz

Metoda se bazează pe ideea alfabetului Morse şi anume, caracterele mai frecvente sunt reprezentate printr-o codificare binară mai scurtă. În plus, nu există separatori, fiind necesar să se respecte condiţia de prefix.

În cazul codificării, proprietatea se referă la faptul că într-un cod optimal, la $p_i > p_j$ corespunde $n_i < n_j$ şi este, în plus, îndeplinită cerinţa ca cele mai puţin probabile două mesaje să aibă aceeaşi lungime. Tehnica de codificare constă în rescrierea tabelei de probabilităţi, intercalând în ordine descrescătoare suma ultimelor două mesaje – cele mai puţin probabile –, iteraţia oprindu-se când în tabel rămân numai două mesaje. Combinaţia de cod se scrie urmărind fiecare grupare în parte şi formând codul de la dreapta la stânga.

Codurile obţinute pentru fiecare dintre cele 8 mesaje sunt cele prezentate in continuare.

$0.4 \rightarrow 1$
$0.18 \rightarrow 001$
$0.1 \rightarrow 011$
$0.1 \rightarrow 0000$
$0.07 \rightarrow 0100$
$0.06 \rightarrow 0101$
$0.05 \rightarrow 00010$
$0.04 \rightarrow 00011$

În tabelul 3.2 este prezentată comparativ codificarea setului de 8 mesaje, obţinută prin aplicarea metodelor Shannon, Shannon-Fano şi Huffman.

Tabelul 3.2

Mesaj	p_i	$\log 1/p_i$	n_i	Shannon	Shannon-Fano	Huffman
S_1	0.4	1.32	2	00	00	1
S_2	0.18	2.47	3	010	01	001
S_3	0.10	3.32	4	0110	100	011
S_4	0.10	3.32	4	0111	101	0000
S_5	0.07	3.83	4	1000	1100	0100
S_6	0.06	4.06	5	10010	1101	0101
S_7	0.05	4.32	5	10011	1110	00010
S_8	0.04	4.64	5	10100	1111	00011

D. Codificarea aritmetică

Dezavantajul algoritmilor pentru compresia datelor care folosesc *arbori Huffman* este acela că fiecare simbol generat de o sursă de informaţie S este codificat folosind un număr întreg de biţi, fapt care duce la apariţia unei diferenţe mari între lungimea codificării unui şir de simboluri şi entropia acestuia, dacă nu se consideră lungimea, în biţi, a dicţionarului.

Nu există algoritmi care să elimine complet redundanţa unei surse de informaţie deoarece, în primul rând, entropia unui şir de simboluri generat de o sursă este un număr real şi bitul este o unitate atomică, şi, în al doilea rând, trebuie transmis dicţionarul pentru ca informaţia să poată fi reconstituită.

Cea mai eficientă metodă entropică de compresie a datelor care elimină aproape complet redundanţa unei surse de informaţie este *compresia aritmetică.*

Compresia aritmetică a fost descoperită de către cercetătorii *Peter Elias, Jorma J. Rissanen* şi *Richard C. Pasco.* Ideea care stă la baza acestei metode de compresie este aceea de a codifica un şir de simboluri folosind un număr real cuprins în intervalul $[0; 1)$. Datorită faptului că această metodă de compresie este entropică sunt necesare probabilităţile p_i ($0 \le i < m$, unde m este numărul de simboluri pe care le poate genera o sursă de informaţie S) de apariţie ale simbolurilor. Pot fi construite variantele *statică, semi-statică* şi *dinamică* ale algoritmului de compresie aritmetică.

Este prezentat în continuare modul în care se realizează codificarea unui şir de simboluri generat de o sursă de informaţie S pentru varianta *statică*. În cazul variantei *semi-statice* probabilităţile se calculează pe baza simbolurilor generate de sursa de informaţie S, fiind nevoie de două parcurgeri ale şirului de simboluri la fel ca în cazul algoritmului *Huffman.*

Se consideră un interval $[a, b) \subseteq [0; 1)$, $b > a$. Fiecărui simbol care poate fi generat de o sursă de informaţie S i se ataşează un subinterval al lui $[a, b)$ cu proprietatea că lungimea subintervalului corespunzător unui simbol este direct proporţională cu probabilitatea de apariţie a simbolului respectiv şi oricare două subintervale corespunzătoare a două simboluri distincte nu au puncte de intersecţie.

Fie P_i probabilitatea cumulată a simbolului $A_i, P_0 = 0$ şi $P_i = p_0 + \ldots + p_{i-1}, 0 \le i < m$, şi fie $l = b - a$ lungimea intervalului $[a, b)$. Din faptul că suma tuturor probabilităţilor de apariţie ale simbolurilor A_i i se va ataşa intervalul $[a + l \cdot P_i, a + l \cdot (P_i + p_i))$. Lungimea subintervalului corespunzător unui simbol A_i este egală cu $l \cdot p_i$.

În cazul în care o sursă de informaţie nu generează toate simbolurile pe care le poate genera, atunci probabilitatea de apariţie a unor simboluri poate fi 0 deci, lungimea subintervalelor ataşate simbolurilor care nu sunt generate este 0, caz în care avem intervale degenerate, şi, dacă p_i este 0 şi p_{i+1} este diferit de 0, atunci intersecţia dintre subintervalele corespunzătoare celor două simboluri este diferită de mulţimea vidă şi

nu mai sunt respectate condiţiile enunţate mai sus. Dacă se elimină subintervalele de lungime 0, atunci condiţiile sunt respectate.

Algoritmul de compresie aritmetică se bazează pe această împărţire a unui interval. Faptul că există intervale degenerate nu va influenţa funcţionalitatea algoritmului, deoarece o sursă de informaţie nu va genera niciodată simboluri care au probabilitatea de apariţie 0.

Algoritmul de codificare constă în alegerea unui subinterval al intervalului $[0; 1)$ corespunzător primului simbol generat de o sursă S şi apoi, ca nou interval se consideră subintervalul ales şi se alege subintervalul corespunzător celui de-al doilea simbol.

Algoritmul de compresie aritmetică folosit pentru a codifica un şir de simboluri generat de o sursă de informaţie S este următorul:

- se consideră intervalul care are extremitatea din stânga a şi lungimea l;
- fie p_i probabilităţile de apariţie ale simbolurilor care pot fi generate de o sursă de informaţie S;
- fie P_i probabilităţile cumulate ale simbolurilor;
- pentru fiecare simbol A_i generat de sursa de informaţie S execută:
  - $a \leftarrow a + l \cdot P_i$;
  - $l \leftarrow l \cdot p_i$;
- $rezultat \leftarrow a + l/2$.

În ciclul repetitiv prezentat anterior se schimbă intervalul iniţial cu subintervalul corespunzător simbolului generat de sursa de informaţia S. Datorită faptului că intervalele ataşate simbolurilor care au probabilitatea de apariţie diferită de 0 sunt disjuncte, fiecare simbol este unic determinat de orice număr care aparţine subintervalului corespunzător. În teorie se foloseşte ca interval iniţial intervalul $[0; 1)$, deci cu capătul din stânga 0 şi lungime 1.

După parcurgerea şirului de simboluri generate de o sursă de informaţie se transmite un număr real din intervalul $[0; 1)$ care reprezintă codificarea şirului de simboluri. Numărul real trebuie transmis cu o precizie foarte mare. Acest număr este dat, de obicei, de mijlocul ultimului interval calculat cu ajutorul algoritmului prezentat anterior: $a + l/2$. Algoritmul de codificare mai poate fi construit folosind limitele inferioară şi superioară ale unui interval în locul limitei inferioare a lungimii intervalului. În acest caz, după linia în care se calculează lungimea intervalului se adaugă linia $b \leftarrow a + l$. Dacă intervalul iniţial este $[0; 1)$, atunci precizia cu care trebuie calculată

limita din stânga pentru un simbol cu probabilitatea p_i, $0 \le i < m$, este de $\lceil -\log_2 p_i \rceil$ biţi.

Dacă analizăm modul de construire al numărului real plecând de la intervalul $[0; 1)$, atunci precizia cu care trebuie transmis numărul care reprezintă mijlocul ultimului interval calculat este de $\left\lceil n \cdot \left(\sum_{i=0}^{m-1} p_i \cdot \log_2 p_i \right) \right\rceil + 1$ biţi, unde n reprezintă numărul de simboluri generate de sursa S. Se poate observa foarte uşor că diferenţa dintre entropia şirului generat de sursa de informaţie S şi numărul mediu de biţi necesari transmiterii unui simbol este foarte mică (mai mică de 1 bit).

De exemplu, fie o sursă de informaţie S care poate genera simbolurile ‘a’, ‘b’ şi ‘c’ cu probabilităţile $1/2$, $1/4$ şi $1/4$. În figura 3.2. se poate observa modul de codificare a şirului de simboluri ‘*abac*’. Precizia cu care trebuie transmis rezultatul este de 7 biţi (rezultatul este 0.3046875 = 0.0100111_2 ceea ce înseamnă că se transmit biţii 0100111, deoarece numărul este cuprins între 0 şi 1 şi ne interesează dar porţiunea care se află după virgulă).

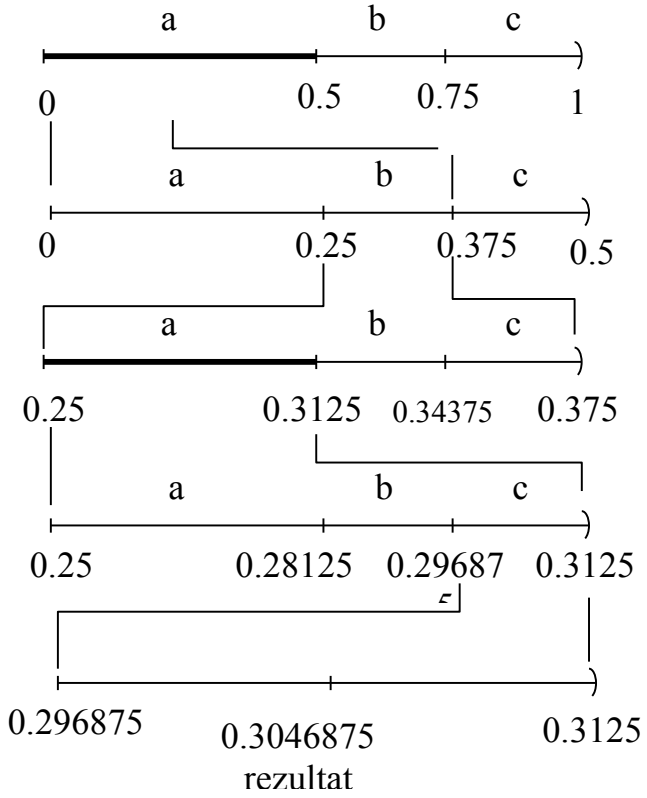

Fig. 3.2.

Pe lângă rezultatul obţinut în urma codificării mai trebuie transmis *dicţionarul* care este format din probabilităţile de apariţie ale simbolurilor care pot fi generate de sursa de informaţie S.

Motivul pentru care rezultatul îl reprezintă mijlocul ultimului interval calculat, şi deci o creştere a preciziei cu un bit, este acela că în practică există situaţii în care reprezentarea în baza 2 ale celor două limite ale intervalului, folosind aceeaşi precizie, sunt egale pentru că microprocesoarele existente realizează operaţii cu numere reale cu o precizie finită de până la 80 de biţi.

Din punct de vedere teoretic este suficient ca rezultatul să fie constituit de limita inferioară a ultimului interval găsit, calculată cu precizia de $\left\lceil n \cdot \left(-\sum_{i=0}^{m-1} p_i \cdot \log_2 p_i \right) \right\rceil$ biţi, deoarece, bazându-ne pe cele enunţate anterior, un simbol căruia i s-a ataşat intervalul $[a; b)$ este unic determinat de orice număr real care aparţine intervalului.

Se poate pune întrebarea "*De ce nu se poate lua ca rezultat limita superioară a intervalului corespunzător ultimului simbol codificat?*".

Răspunsul este acela că limita superioară a intervalului corespunzător ultimului simbol codificat nu aparţine intervalului corespunzător simbolului care urmează în ordine lexografică.

Algoritmul de decompresie

Pentru a decodifica un număr real, cuprins între 0 şi 1, a cărui lungime în biţi se cunoaşte, trebuie să avem probabilităţile de apariţie a simbolurilor care au fost folosite în procesul de codificare şi numărul total n al simbolurilor care au fost codificate.

La începutul procesului de decodificare se consideră intervalul $[a; a+l)$, unde $a = 0$ şi $l = 1$. Fiecărui simbol îi corespunde un subinterval al acestui interval.

În continuare se caută subintervalul căruia îi aparţine numărul care trebuie decodificat. După ce s-a găsit acest subinterval se transmite simbolul corespunzător acestuia şi noul interval devine cel găsit. Acest pas se execută până în momentul în care am decodificat n simboluri.

Dacă nu se transmite numărul n de simboluri care au fost codificate, în momentul compresiei alfabetului sursei S de informaţie se poate extinde cu un simbol suplimentar care are semnificaţia de *sfârşitul*

codificării care se va codifica după ce sursa S nu mai generează simboluri. În concluzie, în momentul în care numărul se va afla în subintervalul corespunzător simbolului de *sfârşit de codificare*, decodificarea se va încheia.

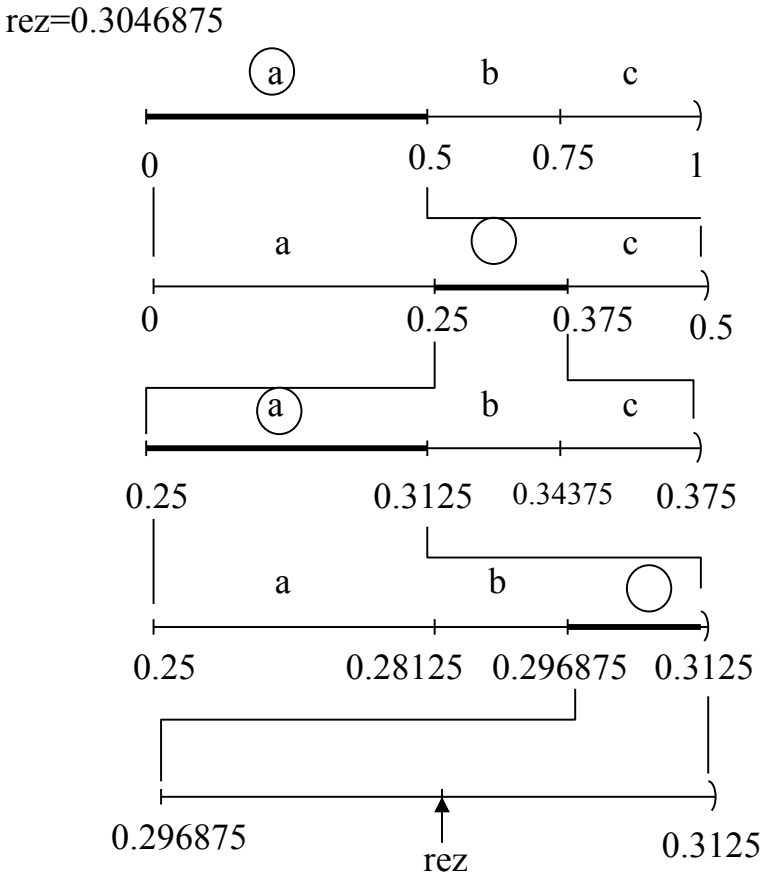

Fig. 3.3.

La începutul procesului de decodificare se consideră intervalul $[a; a+l)$, unde $a=0$ şi $l=1$. Fiecărui simbol îi corespunde un subinterval al acestui interval.

În continuare se caută subintervalul căruia îi aparţine numărul care trebuie decodificat. După ce s-a găsit acest subinterval se transmite simbolul corespunzător acestuia şi noul interval devine cel găsit. Acest pas se execută până în momentul în care am decodificat n simboluri.

Dacă nu se transmite numărul n de simboluri care au fost codificate, în momentul compresiei alfabetului sursei S de informaţie se poate extinde cu un simbol suplimentar care are semnificaţia de *sfârşitul codificării* care se va codifica după ce sursa S nu mai generează simboluri. În concluzie, în momentul în care numărul se va afla în

subintervalul corespunzător simbolului de *sfârşit de codificare*, decodificarea se va încheia.

În figura 3.3 se poate observa modul de decodificare a numărului 0.3046875 pentru alfabetul format din simbolurile ‘a’, ‘b’ şi ‘c’, care au probabilităţile $1/2$, $1/4$ şi $1/4$, şi numărul de simboluri care au fost codificate $n = 4$.

La primul pas se observă că numărul *rez* aparţine subintervalului $[0,0.5)$, subinterval corespunzător simbolului ‘a’, deci acest simbol se va transmite şi noul interval va deveni $[0,0.5)$.

La al doilea pas numărul *rez* aparţine subintervalului $[0.25,0.375)$ care corespunde simbolului ‘b’, acest simbol se va transmite şi pe noul interval $[0.25,0.375)$.

La al treilea pas numărul *rez* aparţine subintervalului $[0.25,0.3125)$ care corespunde simbolului ‘a’, acest simbol se va transmite şi noul interval este $[0.25,0.3125)$.

La al patrulea pas numărul *rez* aparţine subintervalului $[0.296875,0.3125)$ care corespunde simbolului ‘c’, acest simbol se va transmite şi noul interval este $[0.296875,0.3125)$. În acest moment se încheie procesul de decodificare deoarece au fost decodificate $n = 4$ simboluri. În timpul decodificării a fost transmis şirul de simboluri ‘abac’.

3.1.4 Concluzii privind compresia datelor

Metodele de compresie a datelor se încadrează în două mari categorii :

- statice;
- dinamice.

O metodă este *statică* dacă fixează corespondenţa dintre mesaje şi cuvintele de cod înainte de începerea codificării şi o păstrează pe toată durata ei. Exemplul clasic este cel al codificării prin metoda Huffman, la care corespondenţa se bazează pe probabilitatea de apariţie a mesajelor în secvenţe de mesaje (mesajelor mai frecvente le sunt asociate cuvinte de cod mai scurte).

O metodă este *dinamică* (*adaptivă*) atunci când corespondenţa dintre mesaje şi cuvintele de cod se modifică în timp. Astfel, codificarea Huffman adaptivă actualizează această corespondenţă pe baza frecvenţelor relative de apariţie a mesajelor, calculate pe măsura transformării lor. În acest mod, un acelaşi mesaj poate fi reprezentat prin cuvinte de cod diferite, după poziţia la care respectivul mesaj apare: la începutul şirului de mesaje sau la sfârşitul lui, dacă frecvenţa sa relativă se modifică pe parcurs.

O măsură a compresiei este *redundanţa codului*. Fie o sursă fără memorie

$$S = \{S_1, S_2, \ldots, S_n\}$$

$$p(S_i) = p_i, \sum_1^n p_i = 1$$

Alfabetul codului este $\{0,1\}$, cele două simboluri având costuri egale, de transmisie sau de memorare. În aceste condiţii, se pot determina:

- măsura informaţiei conţinută de un mesaj $(-\log p_i)$;
- entropia sursei $H = -\sum_1^n p_i \log p_i$, care reprezintă conţinutul informaţional mediu al mesajelor sursei.

Un cod este cu atât mai eficient/bun cu cât diferenţa dintre media ponderată a lungimii cuvântului de cod n_i şi entropie este mai mică. Această diferenţă se numeşte *redundanţa codului*.

$$\sum_{i=1}^n p_i \cdot n_i - \sum_1^n p_i \log p_i$$

Deoarece, în general, valoarea entropiei nu este un număr întreg, utilizarea unor cuvinte de cod de lungime variabilă devine o condiţie necesară pentru realizarea unei redundanţe cât mai mici.

O altă măsură a eficienţei codului este *rata de compresie*, definită ca raportul dintre lungimea medie a mesajelor şi lungimea medie a cuvântului de cod corespunzător. Evident, un cod este cu atât mai bun cu cât rata de compresie este mai mare.

O problemă deosebit de importantă legată de compresia datelor se referă la *stabilitatea la perturbaţii*. Deşi aceste metode sunt elaborate în contextul unui mediu de transmisie fără perturbaţii, este interesantă analiza susceptibilităţii la erori a codului rezultat din compresie.

O primă observaţie este vulnerabilitatea mare pe care o prezintă metodele adaptive, datorită redefinirii dinamice a codului, pe măsura efectuării transmisiei. Afectarea unui cod în această manieră poate determina desincronizarea completă a transmiţătorului şi a receptorului, cu efecte deosebite asupra celui din urmă.

În schimb, în cazul codurilor statice se poate vorbi de resincronizare după producerea unor *erori de fază* (pierderea sau adăugarea unui simbol de cod, altfel spus, modificarea cu un simbol de cod) sau *erori de amplitudine* (înlocuirea unui simbol de cod cu altul).

Recuperarea acestor erori este ilustrată în exemplul 3.4 pentru un cod Huffman aplicat unui mesaj *BCDAEB*.

Exemplul 3.4.

Recuperarea erorilor de fază

011.010.001.1.000.011 — *BCDAEB*
1.1.010.001.1.000.011 — bitul 1 pierdut $\Rightarrow$ *AACDAEB*
010.1.000.1.1.000.011 — bitul 2 pierdut $\Rightarrow$ *CAEAAEB*
011.1.000.1.1.000.011 — bitul 4 pierdut $\Rightarrow$ *BAEAAEB*

Recuperarea erorilor de amplitudine

111.010.001.1.000.011 — bitul 1 inversat $\Rightarrow$ *DCDAEB*
001.010.001.1.000.011 — bitul 2 inversat $\Rightarrow$ *DCDAEB*
011.110.00.1.1.000.011 — bitul 4 inversat $\Rightarrow$ *BAAEAAEB*

3.2 Codificarea/decodificarea pe canale perturbate

3.2.1 Eroarea în transmisia datelor

Se vor face referiri numai la coduri bloc, adică la coduri care au aceeaşi lungime *n*.

Conform celor prezentate în paragraful 3.1, codificarea unei surse $S=\{S_1,S_2,...,S_n\}$ înseamnă definirea unei funcţii f pe mulţimea simbolurilor sursei cu valori în mulţimea q^n de succesiuni de litere. (q=2).

Transmisia este considerată corectă dacă cuvântul de cod recepţionat v_i conduce, după decodificare, la mesajul S_i emis.

Astfel, considerând $S=\{u_1,u_2,...,u_n\}$ codul sursei şi $R=\{v_1,v_2,...y_m\}$ ansamblul cuvintelor de lungime n care pot fi recepţionate, regula de decizie atribuie fiecărui cuvânt recepţionat v_i un cuvânt de cod u_i. Decodificarea este unică, adică $u_i=f(v_i)$, dacă funcţia f este bijectivă.

În aceste condiţii, funcţia inversă determină un ansamblu T_i de cuvinte v_j astfel încât

$$T_i=f^{-1}(u_i)=\{v_j \mid f(v_j)=u_i\} \tag{3.6}$$

Schema unei astfel de decodificări este prezentată în figura 3.4, cu menţiunea că grupele T_i se numesc ansambluri decodificatoare.

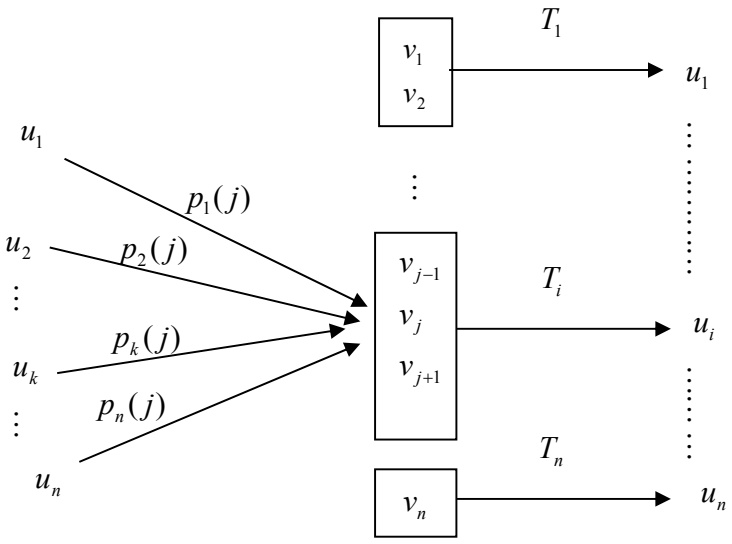

Fig. 3.4. Ansambluri decodificatoare

Apariţia unei erori e_i constă în aceea că un cuvânt recepţionat nu aparţine unui subansamblu T_i atunci când cuvântul emis este cel căruia îi este asociat subansamblul, adică $e_i=v_j$. Probabilitatea de apariţie a erorii va fi

$$p(e_i)=p(T_i^C/u_i)=1-p(T_i\setminus u_i) \tag{3.7}$$

cu T_i^C – ansamblul complementar al lui T_i. Asociind acestei probabilităţi probabilitatea de emisie a mesajului u_i, egală cu p_i, se poate obţine probbilitatea medie a erorii de codificare

$$p(e) = \sum_1^N p_i \cdot p\left(\frac{T_i^C}{u_i}\right) \tag{3.8}$$

relaţie care se poate scrie şi sub forma

$$p(e) = \sum_{i=1}^n p_i - \sum_{i=1}^N p(u_i) \cdot p(T_i \setminus u_i) \tag{3.9}$$

Considerând pentru fiecare cuvânt recepţionat v_i diferitele probabilităţi condiţionate $p(u_k / v_i)$, o regulă naturală de decizie este aceea care consideră drept cel mai verosimil cuvânt u_i – dacă este unic – pe cel care maximizează probabilitatea

$$p(u_i / v_i) \geq p(u_k / v_k), \; k \neq i \tag{3.10}$$

iar decodificarea constă în partiţia

$$T_i = \{v_j \,|\, p(u_i / v_j) = \max p(u_k / v_j)\} \tag{3.11}$$

Exemplul 3.5.
Fie o sursă cu 4 mesaje, codificate astfel încât fiecare combinaţie să difere de oricare alta prin cel puţin 3 poziţii.

$u_1 = 00000$

$u_2 = 01101$

$u_3 = 10110$

$u_4 = 11011$

Teoretic, se pot recepţiona $2^5 = 32$ mesaje. Dacă se consideră $p =$probabilitatea de eroare a unui bit, rezultă următoarele situaţii:

- probabilitatea de a obţine un cuvânt fără eroare
$$p(k=0) = q^5 = (1-p)^5 \sim 1-5p$$
- probabilitatea de a obţine un cuvânt cu 1 eroare
$$p(k=1) = C_5^1 q^4 p \sim 5p$$
- probabilitatea de a obţine un cuvânt cu *r* erori
$$p(k=r) = 0 \;\; (r > 1)$$

Se constată că în situaţia în care v_i este identic cu un cuvânt de cod u_i, probabilitatea de transmisie corectă este foarte mare (1-5p), în timp ce probabilitatea de a avea o eroare (cel puţin 3 caractere modificate) este neglijabilă şi, deci, este posibilă o decizie de tip $u_i^* = u_i$. Această decizie este posibilă şi dacă există eroare la un singur caracter, deoarece eroarea $p(v_i \setminus u_i) = pq^4 \sim p$ este foarte mică.

Se poate concluziona că s-a obţinut o protecţie satisfăcătoare la perturbaţii datorită diferenţei de minim 3 caractere între combinaţiile de cod.

3.2.2. Distanţa Hamming

Distanţa dintre două cuvinte binare de lungime n

$$u = x_1,...,x_n$$

$$v = y_1,...,y_n$$

este dată de numărul poziţiilor de acelaşi rang în care cele două cuvinte diferă.

$$d(u,v) = \sum_{i=1}^{n} x_i \oplus y_i \tag{3.12}$$

Proprietăţile distanţei Hamming sunt cele cunoscute ale unei distanţe, şi anume:

$$\begin{aligned} &1. \quad d(u,v) = d(v,u) \geq 0 \\ &2. \quad d(u,v) = 0 \Leftrightarrow u = v \\ &3. \quad d(u,v) \leq d(u,w) + d(w,v) \end{aligned} \tag{3.13}$$

Ansamblul cuvintelor de cod *n* a căror distanţă la un cuvânt de cod *u* este cel mult egală cu *r*, se numeşte *sfera de centru* u_0 *şi rază r* şi se notează

$$\varphi_r(u_0) = \{u \mid d(u_0,u) \leq r\} \tag{3.14}$$

O reprezentare geometrică a lui *u* poate fi un punct de coordonate $x_1,...,x_n$ în $\Re^n$. Cele 2^n combinaţii de succesiuni de *n* simboluri 0 sau 1 (posibile cuvinte de cod) au ca imagine vârfurile unui hipercub de latură 1, conform reprezentării din figura 3.5..

Se defineşte *distanţa Hamming* între 2 vârfuri ca fiind cel mai mic număr de laturi care le uneşte.

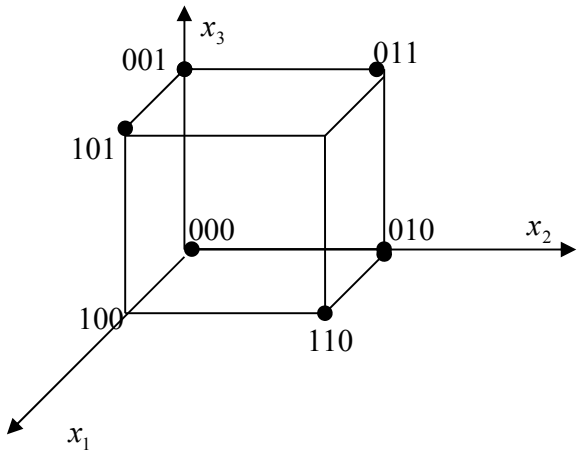

Fig.3.5. Hipercubul cuvintelor de cod din R^3

Dacă toate cuvintele de cod ar avea sens, atunci orice eroare – modificare de caracter – ar conduce la un alt cuvânt de cod, neputând fi depistată. Dacă însă sunt separate din cele 2^n cuvinte de cod numai 2^k, atunci e posibil să fie depistate unele *erori singulare* – care modifică un singur bit – deoarece $2^m = 2^{n-k}$ combinaţii nu au sens.

Astfel, dacă pe cubul din figura 3.3., combinaţiile cu sens sunt 000, 101, 110, 011, cuvinte separate prin câte 2 biţi, se poate *detecta* orice eroare singulară. Mai mult, dacă combinaţiile cu sens sunt numai 000, 111, care diferă între ele prin trei poziţii, se poate aprecia că un cuvânt din subansamblul (100,001,010) provine din 000, iar unul din subansamblul (110,101,011) provine din111, deci se poate *corecta* orice eroare singulară.

3.3. Coduri detectoare şi corectoare de erori

3.3.1 Teorema fundamentală a teoriei informaţiei

Formulată de Shannon în anul 1949, teorema fundamentală a teoriei informaţiei se referă la posibilitatea ca o sursă de entropie $H<C$, cu C- capacitatea sursei, să fie codată astfel încât rata de emisie R să fie oricât de apropiată de C.

Fondul teoremei constă din următoarele două idei de bază:

a) Dacă $H \le C$, există codificări care asigură transmiterea mesajelor cu eroare arbitrar de mică (a cărei probabilitate de apariţie e finită) de decodificare.
b) Dacă $H > C$, nici o metodă nu poate asigura transmisia fără eroare a cărei probabilităţi de apariţie e finită.

Concluzia teoremei este aceea că pentru a micşora eroarea trebuie să fie crescută lungimea cuvântului de cod. Pe de altă parte, în practică trebuie să se utilizeze cuvinte de cod cât mai scurte. Este deci necesar să se ajungă la un compromis între eliminarea perturbaţiilor şi creşterea eficienţei. Acest lucru este realizabil fie prin creşterea debitelor, fie prin utilizarea unor algoritmi rapizi de decodificare.

Cel mai bun exemplu în acest sens este cel al *codurilor sistematice* care conţin cuvinte de *n* caractere destinate codificării unor surse echiprobabile de 2^m mesaje. Astfel, considerând *k* numărul de caractere suplimentare, destinate asigurării corecţiei, fiecare cuvânt eronat trebuie să fie acoperit de un cuvânt de control astfel încât

$$2^k > \sum_{i=0}^{r} C_n^i \tag{3.15}$$

cu *r*– numărul poziţiilor în care pot să apară erori.
(în cazul particular în care se urmăreşte doar corecţia erorilor unitare ($i = 1$) $2^k > n$ sau $n_{\max} = 2^k - 1$, $n_{\max}$ = margine Hamming).

Dacă *n* este foarte mare, $2^k = n \Rightarrow k = \log n$ şi $m = n - \log n$ şi deci *eficienţa* codului va fi

$$\eta = \frac{m}{n} = 1 - \frac{\log n}{n} \tag{3.16}$$

Se observă că η creşte când *n* creşte, ceea ce corespunde şi concluziilor teoremei fundamentale. Esenţială este totuşi asigurarea unor posibilităţi de detecţie şi eventual corecţie a erorilor, pentru că nu se poate admite o creştere exagerată a lungimii cuvintelor de cod.

3.3.2 Coduri detectoare şi corectoare de erori cu controlul parităţii

De regulă, în echipamentele de transmisie de date se folosesc coduri bloc cu cuvinte de lungime constantă *n*. Un cuvânt de cod va fi notat $u = a_1 a_2 .. a_n$, constituind unul din ansamblurile $B^n = \{0,1\}^n$ ale succesiunii de *n* simboluri binare.

Un cod de N cuvinte din B^n se numeşte cod de lungime n şi dimensiune (talie) N, cu n – numărul de simboluri din cuvânt şi N – numărul de cuvinte.

Se defineşte *ponderea cuvântului* $\pi(n)$ ca suma obişnuită a cifrelor "1" dintr-un cuvânt.

$$\pi(n)=\sum_{i=1}^{n} a_i \qquad (3.17)$$

În funcţie de ponderea cuvântului se apreciază *paritatea unui cod*, testele de paritate la care se vor face referiri în continuare reprezentând calculul parităţii modulo 2. În acest sens, un cod are paritate pară dacă $\pi(n)=0$.

Se defineşte *adunarea a două cuvinte de cod* ca suma modulo 2 dintre elementele de acelaşi rang.

Cu proprietăţile prezentate mai sus, se poate aprecia că B^n are o structură de *grup abelian*, orice subgrup din acest grup fiind numit *cod de grup* (sunt îndeplinite axiomele grupului abelian, cu menţiunea că elementul neutru este cuvântul de cod $u_n=0..0$ şi fiecare cuvânt de cod u are şi simetric u' astfel încât $u+u'=u_n$)

Un cod de grup binar poate fi identificat cu un vector

$$\pi=(a_1,a_2,...a_n) \text{ cu } a_i=0 \text{ sau } 1 \qquad (3.18)$$

În aceste condiţii, generarea unui cuvânt de cod de grup poate fi făcută plecând de la baza canonică formată din cele n cuvinte de pondere 1.

$e_1=100..0$
$e_2=010..0$
$e_3=001..0$
……
$e_n=000..1$

astfel încât

$$u=a_1e_1+a_2e_2+...+a_ne_n \qquad (3.19)$$

Controlul *simplu de paritate* constă în completarea unui cuvânt de *n-1* simboluri cu un simbol 0 sau 1 care face ca ponderea totală să fie pară (paritate pară) sau impară (paritate impară).

De exemplu, alegând un cod cu paritate pară, pentru care $\pi(n)=0$, dacă la recepţie se primeşte cuvântul u', pot exista următoarele situaţii

- $\pi(u')=0$: fie nu a existat eroare, fie au fost eronate un număr par de simboluri;
- $\pi(u')=1$: au fost eronate un număr impar de simboluri.

A. Coduri sistematice de tip Hamming

Aşa cum a fost deja subliniat, deoarece informaţia transmisă printr-un canal de comunicaţie este supusă perturbaţiilor, sunt necesare măsuri de detecţie şi eventual de corecţie a erorilor. O metodă este folosirea unui codificator/decodificator la emisie/receptie care să permită implementarea acestor măsuri. Structura unui astfel de montaj este prezentată în figura 3.4.

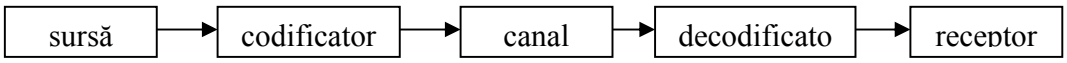

Fig. 3.4 Sistem de detectie a erorilor

Sursa emite o secvenţă de cuvinte binare $S=\{a_1,a_2,...a_i,...\}$ codificată în $\{c_1,c_2,...c_i,...\}$ şi apoi transmisă prin canalul de comunicaţie pentru a ajunge la recepţie într-o formă, foarte posibil, perturbată.

Decodificarea generează secvenţa $\{a_1',a_2',...a_i',...\}$ care poate fi:

- coincidentă cu cea a sursei, dacă se foloseşte un cod corector de erori;
- diferită de cea a sursei şi însoţită de un indicator de eroare, dacă se foloseşte un cod detector de erori. În acest caz, corecţia se face prin retransmisia mesajului, folosind sistemul *ARQ* (Automatic Repetition Request).

La codificare, secvenţele $\{a_1,...a_i,...\}$ sunt, de regulă, decupate în blocuri, fiecărui bloc asociindu-i-se un număr de biţi de control. O astfel de codificare, numită *sistematică*, prezintă următoarele caracteristici:

- cuvintele de cod sunt formate dintr-un număr total de *n* biţi;
- biţii purtători de informaţie sunt în număr de *m* şi formează *partea semnificativă* a cuvântului de cod;
- există *n-m=k* biţi de control care formează *partea de test* a cuvântului de cod.

În ceea ce priveşte maniera în care se generează biţii de control, pot fi evidenţiate două situaţii:

- când biţii de control se deduc din biţii informaţionali ai blocului curent, situaţie în care este vorba despre despre un *cod bloc*, folosit frevent în transmiterea şi prelucrarea datelor;
- când biţii de control depind şi de blocurile anterioare, situaţie în care codul se numeşte *convoluţional* sau *recurent*.

Un cod sistematic se notează *(n,m),* exemplul cel mai elocvent fiind *codul Hamming*, cu variantele (3,1), (7,4) , (15,11), (31,26).

Codurile Hamming sunt coduri de grup pentru care biţii de control sunt determinaţi în funcţie de biţii informaţionali prin relaţii de condiţie care asigură paritate prin suma modulo 2.

Codurile Hamming pot fi:

- sistematice, caz în care primii *m* biţi sunt informaţionali;
- ponderate, atunci când biţii de control apar pe poziţii corespunzători puterilor lui 2 (poziţiile 1, 2, 4, 8, …).

Se propune în continuare studiul detaliat al *codului sistematic Hamming de tip (7,4)* (n=7, m=4).

Se consideră cuvântul de cod de 7 biţi $u = a_1 a_2 a_3 a_4 a_5 a_6 a_7$ şi cuvântul recepţionat $u' = a_1' a_2' a_3' a_4' a_5' a_6' a_7'$

Erorile singulare care pot să apară la recepţie, în care e_1, e_2, e_3 sunt simbolurile pentru cei trei biţi de test, corespunzători celor opt erori posibile, sunt prezentate în tabelul 3.3.

Tabelul 3.3

Eroare asupra	$e_3 e_2 e_1$
nici unei cifre	000
$a_1^{'}$	001
$a_2^{'}$	010
$a_3^{'}$	011
$a_4^{'}$	100
$a_5^{'}$	101
$a_6^{'}$	110
$a_7^{'}$	111

Din examinarea tabelului, rezultă condiţiile ca e_1, e_2, e_3 să aibă valoarea 1, şi anume:

$$\begin{aligned} e_1 &= a_1^{'} + a_3^{'} + a_5^{'} + a_7^{'} \\ e_2 &= a_2^{'} + a_3^{'} + a_6^{'} + a_7^{'} \\ e_3 &= a_4^{'} + a_5^{'} + a_6^{'} + a_7^{'} \end{aligned} \tag{3.20}$$

Pentru a determina biţii de control a_5, a_6, a_7, în funcţie de biţii informaţionali a_1, a_2, a_3, a_4, este suficient să punem condiţia de nonexistenţă a erorii, şi anume:

$$a_i^{'} = a_i, \; i = \overline{1,7}$$

şi deci $e_1 = e_2 = e_3$, astfel încât se obţin următoarele relaţii:

$$\begin{aligned} a_1 + a_3 + a_5 + a_7 &= 0 \\ a_2 + a_3 + a_6 + a_7 &= 0 \\ a_4 + a_5 + a_6 + a_7 &= 0 \end{aligned} \tag{3.21}$$

care generează condiţiile:

$$\begin{aligned} a_5 &= a_2 + a_3 + a_4 \\ a_6 &= a_1 + a_3 + a_4 \\ a_7 &= a_1 + a_2 + a_4 \end{aligned} \tag{3.22}$$

În cazul în care codurile Hamming se scriu în formă ponderată, adică în forma în care biţii de control ocupă poziţiile corespunzătoare puterilor crescătoare ale lui 2 $(a_1, a_s, a_4, ...)$, regulile de control devin:

$$\begin{aligned} a_1 &= a_3 + a_5 + a_7 \\ a_2 &= a_3 + a_6 + a_7 \\ a_4 &= a_5 + a_6 + a_7 \end{aligned} \tag{3.23}$$

Condiţiile de control (3.22) pot fi scrise şi sub formă matriceală, şi anume:

$$[a_5 \ a_6 \ a_7] = [a_1 \ a_2 \ a_3 \ a_4] \begin{bmatrix} 011 \\ 101 \\ 110 \\ 111 \end{bmatrix} \tag{3.24}$$

sau, altfel, $<t> = <s>. \ \Gamma_{4\times3}$ (3.25)

şi $[a_1 \ a_2 \ a_3 \ a_4 \ a_5 \ a_6 \ a_7] = [a_1 \ a_2 \ a_3 \ a_4] \begin{bmatrix} 1000011 \\ 0100101 \\ 0001111 \end{bmatrix}$ sau

$<u> = <s> G_{4\times7} = <s> [I_4 / \Gamma_{4\times3}]$

cu $<t>$– matricea de test

$<s>$– matricea biţilor informationali

$\Gamma_{4\times3}$– sau $H_{3\times7} <u>_{7\times1} = <0>_{3\times1}$; H– matrice de control

Generalizând, toate cele 2^m combinaţii semnificative ale unui cod (n,m) sunt generate prin combinaţii liniare plecând de la o bază de n cuvinte liniar independente:

$$u = \sum_{i=1}^{n} \lambda_i \cdot u_i \tag{3.26}$$

Matricea generatoare va fi matricea $m \times n$, de rang m, în care liniile sunt cuvintele de bază u_i .

$$G = \begin{bmatrix} <u_1> \\ <u_2> \\ \ldots\ldots\ldots \\ <u_m> \end{bmatrix} \tag{3.27}$$

sau, sub formă redusă

$$G=[I_m/\Gamma_{m\times k}],\ G=[\Gamma_{m\times k}/I_m] \tag{3.28}$$

după cum primii sau ultimii m biţi din n sunt alocaţi pentru partea semnificativă.

Orice matrice dedusă din G prin permutări de coloane generează un cod echivalent.

Codul poate fi definit şi plecând de la o matrice de test $k \times n$,

$$H=\begin{bmatrix} < v_1 > \\ < v_2 > \\ \ldots\ldots\ldots \\ < v_k > \end{bmatrix} \tag{3.29}$$

ale cărei linii sunt cuvintele de test v_j.

Spaţiul generat de H este ortogonal cu Γ, deci $H(u)=0$ (condiţia necesară şi suficientă ca un cuvânt u să aparţină codului Γ), unde:

(u) – vector coloană asociat cuvântului de cod u

(0) – vector coloană cu k elemente nule.

Explicit, codul Γ este situaţia sistemului omogen de rang *m*

$$\sum_{j=1}^{m} h_{ij} \cdot a_i = 0;\quad i=\overline{1,k}$$

cele k ecuaţii permiţând calculul simbolurilor de control în funcţie de simbolurile semnificative.

Codurile Hamming permit corectarea erorilor singulare. Procedura se bazează pe faptul că $H(u')=H(u)+H(e)=H(c)=(c)$, unde *e* este un vector eroare astfel încât $u'=u+e$, iar *c* este un vector coloană cu *k* elemente, numit vector corector al lui u'.

Explicit: $c_i=\sum_{j=1}^{n} h_{ij} \cdot e_i$.

De exemplu, când există erori singulare – un singur 1 pe poziţia r a cuvântului $e \to c_i = h_{ir}$, ceea ce arată că eroarea se află în poziţia în care se află coloana h_{ir} în H.

Codurile H au distanţa 3, dar pot exista coduri liniare şi cu $d > 3$; condiţia necesară şi suficientă ca distanţa minimă dintre cuvintele unui cod liniar să fie d este să nu existe combinaţii liniar independente de mai puţin de d coloane în matricea de control a codului.

B. Coduri ciclice

Definiţii

Codurile ciclice sunt o subclasă a codurilor liniare, frecvent utilizate în practică. Ele prezintă interes pentru echipamentele de transmisie de date din următoarele considerente:

- pot fi generate simplu cu scheme secvenţiale folosind registre de deplasare;
- permit detecţia şi corecţia pachetelor de erori;
- pot fi studiate riguros folosind teoria polinoamelor algebrice.

Un cod liniar este ***ciclic*** dacă orice permutare ciclică a unui cuvânt de cod este, de asemenea, un cuvânt de cod. Astfel, dacă se consideră cuvântul de cod $u = a_1 a_2 a_3 ... a_{n-1} a_n$, prin deplasarea în inel a simbolurilor ce alcătuiesc cuvântul de cod se obţine tot un cuvânt de cod:

$$\begin{aligned} u^{(1)} &= a_2 a_3 ... a_n a_1 \\ u^{(2)} &= a_3 a_4 ... a_1 a_2 \\ &..... \\ u^{(n-1)} &= a_n a_1 ... a_{n-1} \\ u^{(n)} &= a_1 a_2 ... a_{n-1} a_n = u \end{aligned} \qquad (3.33)$$

Uzual, considerând componentele unui cuvânt de cod din $C(B^n)$ $u = u(n-1),...,u(0)$ drept coeficienţii unui polinom $u(x)$

$$u(x) = u(n-1)x^{n-1} + ... + u(1)x + u(0) \qquad (3.34)$$

condiţia precedentă se exprimă sub forma

$$u(x) \in C(B^n) \Leftrightarrow x^i u(x) \bmod(x^n + 1) \in B^n \qquad (3.35)$$

Se demonstrează că orice cuvânt al unui cod ciclic *(n,k)* este un multiplu al unui polinom generator *g(x)* de grad *n-k*, asociat ultimei linii a matricii generatoare *G*. De asemenea, *g(x)* este divizor al lui x^n+1.

Un cod ciclic de lungime *n* şi talie *N* este definit de un ansamblu Γ de *N* cuvinte extrase din ansamblul de 2^n cuvinte cu *n* poziţii aparţinând lui B^n şi stabil în raport cu adunarea şi permutarea circulară.

El este echivalent cu ansamblul $\Gamma(x)$ al polinoamelor de cod *u(x),* $\Gamma(x)$ fiind o parte a inelului claselor reziduale modulo x^n-1.

Teoremă
Condiţia necesară şi suficientă ca Γ să fie un cod ciclic este ca $\Gamma(x)$ să fie un ideal al claselor de resturi modulo x^n-1.

Demonstraţia constă în verificarea condiţiilor de definiţie ale unui ideal:
a) $\Gamma(x)$ conţine $u(x)+v(x)$ dacă conţine $u(x)$ şi $v(x)$;
b) $\Gamma(x)$ conţine $x^k u(x)$ şi deci există relaţia $\sum_k a_k x^k u(x) = u(x)p(x)$

Metode de construcţie a codurilor ciclice

De regulă, există două modalităţi de exprimare a polinomului de cod *u(x):*

- cu evidenţierea părţii semnificative (informaţionale) *s(x)* şi a părţii de test *t(x),* sub forma $u(x)=s(x)+t(x)$;
- cu evidenţierea proprietăţii oricărui cuvânt de cod ciclic de a se divide cu un polinom generator *g(x)*, sub forma $u(x)=g(x)q(x)$

Pentru generarea codurilor ciclice, sunt utilizate următoarele patru metode:

- metoda directă;
- metoda matricii generatoare;
- metoda matricii de control;
- metoda rădăcinilor polinomului generator.

Metoda directă generează un cod ciclic în două moduri, după cum urmează:

- *prin înmulţire*, situaţie în care se consideră partea semnificativă de forma $s(x)=a_1x^{m-1}+...+a_mx^0$, *m* fiind numărul de biţi semnificativi (informaţionali), şi cuvântul de cod *u(x)* se obţine cu relaţia $u(x)=s(x)g(x)$;
- *prin împărţire*, care foloseşte exprimarea cuvântului de cod cu evidenţierea separată a părţii semnificative *s(x)* şi a părţii de test *t(x)* : $u(x)=s(x)+t(x)$. Partea de test *t(x)* se calculează ca fiind restul împărţirii părţii semnificative *s(x)* la polinomul generator *g(x)*.

Exemplul 3.6.
Se consideră codul ciclic (7,4) cu polinomul generator $g(x) =x^3+x+1$, divizor al lui x^7+1.
Considerând partea semnificativă *s* de forma *0111*, acesteia îi corespunde un polinom de cod ataşat $s(x)=x^5+x^4+x^3$, ceea ce conduce la obţinerea părţii de test de forma $t(x))=x$. Aşadar, cei trei biţi de control vor fi *010*, astfel încât cuvântul de cod complet va fi *u=0111010*.

În tabelul 3.4. este prezentat codul (7,4) generat prin împărţire.

Tabelul 3.4.

s	s(x)	t(x)	u
0000	0	0	0000000
0001	x^3	x+1	0001011
0010	x^4	x^2+x	0010110
0011	x^4+ x^3	x^2+1	0011101
0100	x^5	x^2+x+1	0100111
0101	x^5+x^3	x^2	0101100
0110	x^5+x^4	1	0110001
0111	x^5+x^4+x^3	x	0111010
1000	x^6	x^2+1	1000101
1001	x^6+x^3	x^2+x	1001110
1010	x^6+x^4	x+1	1010011
1011	x^6+x^4+x^3	0	1011000
1100	x^6+x^5	x	1100010
1101	x^6+x^5+x^3	1	1101001
1110	x^6+x^5+x^4	x^2	1110100
1111	x^6+x^5+x^4+x^3	x^2+x+1	1111111

Metoda matricii generatoare

Codul Γ fiind de dimensiune *n*, poate fi generat de la o bază de *m* cuvinte, fiind complet definit de polinomul generator *g(x)* format de coeficienţii $g=0000g_0g_1...g_k$. Deoarece orice polinom de cod este un multiplu al lui *g(x)*, se poate scrie:

$$u(x)=g(x)q(x)=g(x)(b_0+...+b_{m-1}x^{m-1}) \tag{3.36}$$

$$u=b_0g^0+b_1g^1+...b_{m-1}g^{m-1} \tag{3.37}$$

ceea ce evidenţiază faptul că orice cuvânt de cod admite ca bază cuvântul *g* şi cele *m-1* permutări circulare ale sale.

Matricial,

$$<u>=<b>G_{mxn} \tag{3.38}$$

unde $<b>=[b_0\ b_1...b_{m-1}]$

$$G_{mxn}=\begin{bmatrix} 00...0g_{0.}...g_k \\ 00...g_0g_1...0. \\ . \\ . \\ g_0g_1...g_ko...0 \end{bmatrix}$$

Metoda matricii de control

Această metodă de construcţie a codurilor ciclice pleacă de la premisa că polinomul generator divide x^n-1. Astfel, se poate scrie $x^n-1=g(x)h(x)$, câtul $h(x)=h_mx^m+h_{m-1}x^{m-1}`+...+h_1x_+h_0$ fiind de grad $m=n-k$..

În inelul claselor de resturi $mod(x^n-1)$, produsul claselor *g(x)* şi *h(x)* este clasa *0*, astfel încât *h* se numeşte polinom ortogonal codului Γ.

Teoremă
Condiţia necesară şi suficientă ca un polinom să fie polinom de cod este ca produsul său prin polinomul ortogonal să fie divizibil cu x^n-1.

$$H_{kxn}(n) = 0 \tag{3.39}$$

unde s-au folosit notaţiile:

u: matrice coloană a cuvântului de cod, *0*: coloană de zerouri, *H*: matrice de control, de forma

$$H_{kxn} = \begin{bmatrix} h_0 h_1 ... h_m 0 ... 0 \\ 0 h_0 ... h_{m-1} h_m ... 0 \\ ..... \\ 00 ... h_0 h_1 ... h_m \end{bmatrix} \qquad (3.40)$$

Exemplul 3.7.
Considerând codul Γ (7,4), relaţiile de calcul sunt, succesiv:

$h(x)=(x^7+1)/(x^3+x+1)=x^4+x^2+x+1$
$h_4=1;\ h_3=0;\ h_2=1;\ h_1=1;\ h_0=1$

$$H_{3x7} = \begin{bmatrix} 1110100 \\ 0111010 \\ 0011101 \end{bmatrix}$$

În aceste condiţii, considerând cuvântul de cod $u = \begin{bmatrix} u_1 \\ u_2 \\ . \\ u_7 \end{bmatrix}$, se obţin relaţiile de control

$$\begin{aligned} &a_1+a_2+a_3+a_5=0 \\ &a_2+a_3+a_4+a_6=0 \\ &a_3+a_4+a_5+a_7=0 \end{aligned} \qquad (3.41)$$

Dacă $a_1,\ a_2,\ a_3,\ a_4$ sunt biţii informaţionali, biţii de control pot fi obţinuţi cu relaţiile:

$$\begin{aligned} a_5&=a_1+a_2+a_3 \\ a_6&=a_2+a_3+a_4 \\ a_7&=a_1+a_2+a_4 \end{aligned} \qquad (3.42)$$

Metoda rădăcinilor polinomului generator

Condiţia necesară şi suficientă ca un cuvânt de cod *u* să aparţină codului Γ este ca rădăcinile polinomului generator să fie rădăcini ale polinomului asociat *u(x)*.

Dacă rădăcinile polinomului generator *g(x)* sunt $\alpha_1,\alpha_2,...,\alpha_s$, atunci condiţia $u(\alpha_i)=a_1+a_2\alpha_i+...+a_n\alpha_i^{n-1}=0$ se poate scrie sub formă matricială

$$\begin{bmatrix} 1\alpha_1\alpha_1^2..\alpha_1^{n-1} \\ 1\alpha_2\alpha_2^2..\alpha_2^{n-1} \\ ..... \\ 1\alpha_s\alpha_s^2..\alpha_s^{n-1} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ . \\ a_n \end{bmatrix} = 0 \qquad (3.43)$$

sau

$H'_{sxn}<a>=0$, cu H'_{sxn} – matrice de test.

Erori simple – orice eroare simplă este o permutare a cuvântului *e=0 0 ...1,* deci *e(x)=1* şi, prin urmare, nu poate fi divizibil cu nici un polinom diferit de o constantă..

Erori duble – o eroare dublă este de forma *e=0...010...01,* având « 1 » în poziţiile *1* şi *i*, deci $e_i(x)=x^i+1$. Pentru a fi depistată, trebuie ca *g(x)* să nu dividă nici unul dintre polinoamele $e_i(x)$, deci să fie primitiv.

Pachete de erori – se defineşte un pachet de erori de lungime cel mult *r* o permutare circulară a cuvântului $e=0...0\ e_1...e_r$, unde nu toţi e_i sunt nuli. De exemplu, *e= 00101001* conţine un pachet de erori de lungime 6.

După cum se poate observa, se pleacă de la premisa că *problema detectabilităţii* se pune în funcţie de divizibilitatea lui *e(x)* la *g(x).* În general, există 2^p-1 cuvinte ale căror polinoame asociate sunt de grad inferior lui *p*. Deci, pentru ca aceste polinoame să nu fie divizibile cu *g(x),* trebuie ca *g(x)* să fie de grad cel puţin p; aşadar, codurile *(n,m)* detectează pachete de erori de lungime *k=n-m*, *k>=1*.

Corecţia erorilor se bazează pe proceduri asemănătoare. Fie o eroare pe poziţia *n-i* a unui cuvânt de cod *u*. Aceasta echivalează cu recepţia cuvântului eronat al cărui polinom asociat este $u'(x)=u(x)+x^i$. La analiza divizibilităţii lui *u'(x)* la *g(x)* se va constata că numai o parte e divizibilă, ceea ce permite localizarea erorii.

Tehnica utilizată pentru corectarea erorilor simple este cea folosită pentru detectarea erorilor duble. *A corecta p erori este echivalent cu a detecta 2p erori.*

Astfel, se consideră F o familie anume de erori (simple, duble, pachete de o anumită lungime). Se asociază fiecărui cuvânt u_i din Γ ansamblul F_i al tuturor cuvintelor $u_i'=u_i+e$, $e \in F$, deci $F_i=u_i+F$. Subansamblele F_i pot să intersecteze Γ, situaţie în care unele erori scapă detecţiei, sau să se interinfluenţeze (un cuvânt recepţionat poate proveni din două cuvinte distincte emise).

Se acceptă următoarele afirmaţii:

- pentru ca un *cod* Γ *să detecteze familia F* este necesar şi suficient ca F_i să fie disjuncte cu Γ
- pentru ca un *cod* Γ *sa corecteze familia F* este necesar şi suficient ca F_i să fie două câte două disjuncte şi disjuncte de Γ

În acest caz, F_i se numesc *ansambluri decodificatoare*.

De fapt, această a doua condiţie înseamnă că, dacă $u_i \neq u_j \Rightarrow u_i^{'} \neq u_j^{'}$, unde $u_i^{'} = u_i + e, u_j^{i} = u_j + e^{'}, e, e^{'} \in F.$

Utilizarea *codurilor ciclice redundante CRC* (Cyclic Redundancy Check) este o măsură puternică de detecţie a erorilor apărute în transmiterea informaţiei. De exemplu, dintre performanţele utilizării unui cod de control de 16 biţi pot fi enumerate:

- erori de 1,2 sau un număr impar de biţi şi erori în rafale de mai puţin de 16 biţi: detecţie în proporţie de 100%;
- erori în rafală de 17 biţi: detecţie în proporţie de 99,9969%;
- erori în rafală de >17 biţi: detecţie în proporţie de 99,9948%.

Deoarece valoarea de control asociată unui şir de biţi se obţine utilizând teoremele aritmeticii modulo 2, acest lucru conduce la simplificarea implementării hardware, care nu necesită decât registre de deplasare şi porţi de tip SAU exclusiv, încorporate în dispozitivul de control al comunicaţiei.

4. Criptarea datelor

7.1. Introducere în criptografie

Criptografia descrie câmpul larg al comunicaţiilor secrete, fiind definită prin totalitatea mijloacelor şi metodelor utilizate pentru protecţia interceptării pasive (înregistrarea mesajului transmis) sau/şi active (modificarea informaţiei sau introducerea de mesaje false pe canalul transmisiei, între emiţătorul şi receptorul legali).

Criptografia are o lungă şi fascinantă istorie. A fost folosită pentru prima dată de către egipteni acum patru mii de ani, iar în secolul XX a jucat un rol hotărâtor în cele două războaie mondiale. Cei care practicau această adevărată artă aveau strânse legături cu domeniile militar, diplomatic şi de guvernământ. Criptografia a fost folosită iniţial pentru a transmite şi proteja strategiile şi secretele naţionale. Începutul utilizării pe scară tot mai largă, în deceniul 7 al secolului trecut, a calculatoarelor şi a sistemelor de comunicaţii a adus cu sine cererea de mijloace de protejare a informaţiilor în format digital şi oferirea de servicii de securitate.

În sens clasic, operaţia de ***criptare*** constă în aplicarea unei transformări E_k asupra *mesajului* (text clar) *M* care aparţine mulţimii (spaţiului) mesajelor notată cu {M}. Se urmăreşte, astfel, obţinerea *criptogramei* (text criptat) *C* care aparţine mulţimii (spaţiului) criptogramelor notată cu {C}, atunci când se utilizează cheia secretă *k*, aparţinând mulţimii (spaţiului) cheilor {K}. *Cheia criptografică* *k* este o secvenţă secretă, relativ scurtă, de caractere, care identifică transformarea E_k utilizată. Această cheie este selectată din spaţiul cheilor {K}, dintr-un număr mare de chei posibile, şi este cunoscută numai de

către corespondenţii legitimi. Formalizarea matematică este cea prezentată în relaţia

$$E_k(M)=C \tag{4.1}$$

Determinarea mesajului M presupune aplicarea la recepţie a unei transformări D_k, corespunzătoare cheii k, printr-un proces denumit ***decriptare***, conform relaţiei

$$D_k(C)=M \tag{4.2}$$

Un sistem de criptare şi de decriptare se numeşte *criptosistem* şi este proiectat de către un *criptograf*. Acesta îşi propune să găsească metode pentru a asigura secretul şi/sau autentificarea mesajelor.

Criptanaliza se referă la tehnicile folosite pentru a intra în posesia informţiei originale prin alte mijloace decât cele disponibile receptorului legal (fără cunoaşterea cheii k). În acest context, *criptanalistul* este un receptor ilegal care foloseşte metode de lucru specifice criptanalizei.

Sistemele criptografice oferă patru tipuri de servicii principale:

- *Secretizare*, prin intermediul căruia persoanelor neautorizate nu li se permite accesul la informaţia corespunzătoare textului clar;
- *Autentificare*, prin care, de regulă, este validată sursa mesajului iniţial. Autentificarea se adresează atât entităţilor cât şi informaţiei. Două părţi care comunică ar trebui să se identifice una pe cealaltă. O informaţie transmisă pe un canal trebuie să-şi autentifice originea, datele despre origine, conţinutul datelor, timpul transmisiunii. Din aceste motive acest aspect a criptografiei se împarte în două clase majore: autentificarea entităţilor şi autentificarea originii datelor. Autentificarea originii datelor oferă implicit şi integritatea datelor, deoarece dacă mesajul a fost modificat şi sursa a fost schimbată.
- *Integritate*, care permite luarea în considerare numai a anumitor mesaje transmise către persoana identificată şi stabileşte că aceste mesaje nu fac parte din categoria celor transmise anterior. Serviciul de integritate asigură faptul că mesajul nu a fost transmis accidental, în timpul operaţiilor de transmitere, inserţie sau ştergere;

- *Nerepudierea originii*, care oferă protecţie împotriva unui transmiţător de mesaje care neagă, ulterior, transmiterea acestora. Acest serviciu previne neacceptarea unei identităţi de a îndeplini angajamente sau acţiuni asumate anterior.

Protocolul criptografic constă dintr-un algoritm care urmăreşte realizarea în secret a comunicaţiilor între diferiţi parteneri. Acesta poate fi reprezentat printr-un număr relativ mare de proceduri matematice, numite *transformări*, care definesc modul în care o secvenţă de date inteligibile, care reprezintă mesajul *M*, este schimbată într-o secvenţă aparent aleatoare, care constituie criptograma *C*.

Dezvoltarea metodelor de criptare pentru sistemele de transmisii de date se bucură în prezent de un mare interes. Există două direcţii de lucru, care de cele mai multe ori se întrepătrund, şi anume:

- Elaborarea de algoritmi de criptare cât mai puternici, cu implementare hardware şi/sau software;
- Elaborarea şi proiectarea de reguli şi protocoale specifice pentru utilizarea unui anumit algoritm.

În funcţie de existenţa sau nonexistenţa unor proceduri de criptare şi autentificare, canalul de comunicaţie al unei reţele de transmisii de date poate fi:
- *Public*, atunci când nu există proceduri de criptare şi autentificare;
- *Privat*, definit de existenţa exclusivă a procedurilor de criptare;
- *Cu semnătură*, caracterizat in exclusivitate de proceduri de autentificare;
- *Sigur*, atunci când există atât proceduri de criptare, cât şi de autentificare.

O primă clasificare a sistemelor criptografice, realizată în funcţie de maniera în care este prelucrat mesajul, este următoarea:

- *Sisteme criptografice cu criptare secvenţială*, caz în care mesajul este tratat pe porţiuni mici (biţi sau caractere), generându-se o seccesiune pseudoaleatoare de simboluri;
- *Sisteme criptografice cu criptare bloc*, situaţie în care sistemul devine pur combinaţional, la nivel de blocuri mari din mesaj. La aceste sisteme, o modificare apărută în blocul de intrare determină o modificare majoră în blocul de ieşire, printr-un proces numit de propagare a erorii.

În funcţie de relaţia dintre cheile de la emisie şi de la recepţie, notate cu k şi k', sistemele criptografice pot fi:

- *Sisteme criptografice cu chei secrete*, cunoscute şi sub numele de *sisteme criptografice simetrice*, pentru care cheile k şi k' sau cu uşurinţă deduse prin calcul una din cealaltă. Aceste sisteme necesită dezvoltarea unor servicii suplimentare de management al cheilor secrete;
- *Sisteme criptografice cu chei publice*, cunoscute şi sub numele de *sisteme criptografice asimetrice,* pentru care cheile k şi k' sunt diferite, imposibil de dedus una din cealaltă în anumite condiţii. În plus, este posibil ca aceste sisteme să fie:
  - *bidirecţional asimetrice*, atunci când cele două chei k şi k' nu pot fi deduse una din cealaltă;
  - *înainte asimetrice*, atunci când cheia k' nu poate fi dedusă din k;
  - *înapoi asimetrice*, atunci când cheia k nu poate fi dedusă din k'.

Plecând de la probabilitatea cunoaşterii de către criptanalist a criptosistemului folosit, există câteva posibilităţi prin care acesta poate intra în posesia mesajului (a textului clar), şi anume:

- *Atacul cu text criptat cunoscut*, atunci când criptanalistul are la dispoziţie eşantioane de criptotext, suficient de lungi, astfel încât el se poate servi de cunoaşterea caracteristicilor specifice ale limbajului utilizat;
- *Atacul cu text clar cunoscut*, situaţie în care criptanalistul cunoaşte perechi text clar-text criptat corespunzător, de tipul $(M, E_k(M))$;
- *Atacul cu text clar ales*, când pentru anumite mesaje M, alese de către criptanalist, sunt cunoscute criptogramele $E_k(M)$. Criptanalistul poate emite ipoteze asupra cheii k sau poate pretinde că este un utilizator autorizat al sistemului în discuţie;
- *Atacul cu metoda de criptare E_k cunoscută*, caz în care criptanalistul are la dispoziţie un timp suficient de lung pentru a putea determina metoda de decriptare D_k, înainte chiar de a primi la recepţie un eşantion de criptotext.

Un sistem criptografic este considerat *sigur* dacă criptanaliza nu îşi atinge obiectivul de determinare a mesajului.

Un sistem secret care rezistă la orice atac criptanalitic, indiferent de volumul calculelor care se cer, se numeşte *sigur necondiţionat*.

Sistemul secret este *sigur computaţional* atunci când se recunoaşte posibilitatea criptanalistului de a intra în posesia mesajului după o cantitate finită de calcule care ocupă un volum de calcul foarte mare, nejustificat din punct de vedere economic.

În acest context, în literatura de specialitate sunt citate frecvent *dezideratele lui Kerckhoff,* care sunt următoarele :

1. sistemul trebuie să fie, dacă nu teoretic, cel puţin practic, de nepătruns;
2. compromiterea amănuntelor despre sistem nu ar trebui să producă inconveniente corespondenţilor;
3. cheia trebuie să poată fi memorată fără a necesita transcrierea ei şi să poată fi schimbată uşor;
4. criptograma trebuie să poată fi transmisă prin telegraf;
5. aparatul de criptare trebuie să fie portabil şi să poată fi operat de o singură persoană;
6. sistemul trebuie să fie uşor de înţeles şi să nu necesite nici cunoaşterea unor liste lungi de reguli, nici o capacitate intelectuală dezvoltată.

Aceste enunţuri au fost elaborate în 1883 şi majoritatea lor rămân folositoare şi astăzi. Punctul al doilea permite clasei funcţiilor de criptare folosită să fie cunoscută public şi securitatea sistemului să fie bazată doar pe cheia aleasă [26].

În orice criptosistem sunt implicate două alegeri statistice: alegerea mesajului şi alegerea cheii. Pentru evaluarea cantităţii de informaţie, atunci când se alege un mesaj *M*, se utilizează noţiunea de *entropie a spaţiului mesajelor*

$$H(M) = -\sum_{i=1}^{n} p(M_i) \log p(M_i) \tag{4.3}$$

cu notaţia $p(M_i)$ – probabilitatea de a fi emis mesajul M_i .

Gradul de nesiguranţă asociat alegerii cheii *k* este dat de

$$H(k) = -\sum_{j=1}^{m} p(k_j) \log p(k_j) \tag{4.4}$$

Cantitatea maximă de informaţie se obţine atunci când mesajele sunt echiprobabile, fiind egală cu *log(n).* Informaţia este complet ascunsă atunci când nedeterminarea cheii este maximă şi este verificată condiţia

$$H(k) = \log(n) \tag{7.5}$$

Această valoare reprezintă şi cantitatea de nedeterminare maximă care poate fi introdusă într–un sistem secret. Se poate deduce de aici un principiu general, şi anume acela că incertitudinea care poate fi introdusă într–un criptosistem nu poate fi mai mare decât incertitudinea cheii. Teoretic, dacă numărul de mesaje este infinit, ncio cheie finită nu asigură secretul perfect. Rezultă aşadar, că din punct de vedere al criptanalistului, un sistem secret este identic cu un sistem de comunicaţii influenţat de perturbaţii, cu următoarele diferenţe:

- Transformarea prin criptare este mai complexă decât perturbaţiile din canalul de comunicaţie;
- Cheia unui sistem secret este aleasă dintr–o mulţimefinită, în timp ce zgomotul din canal este continuu şi face parte dintr–o mulţime infinită.

Considerând *n* criptograme, *q* chei şi *l* mesaje, reprezentând toate criptogramele, cheile, respectiv mesajele de o anumită lungime *N*, ca indici ai secretului teoretic se folosesc două tipuri de echivoc, şi anume:

- Echivocul asociat cheii, notat *H(k/C);*
- Echivocul asociat mesajului, notat *H(M/C)*, de forma:

$$H(k/C) = -\sum_{i=1}^{q}\sum_{j=1}^{n} p(k_i \cap C_j)\log p(k_i \cap C_j) \tag{4.6}$$

$$H(M/C) = -\sum_{i=1}^{l}\sum_{j=1}^{n} p(M_i \cap C_j)\log p(M_i \cap C_j) \tag{4.7}$$

O clasificare riguroasă a metodelor criptografice este destul de dificil de realizat, datorită numărului mare de criterii posibile şi a numărului foarte mare de metode aflate în uz. O clasificare foarte generală, care ia în consideraţie evoluţia acestor metode, este următoarea:

- metode criptografice clasice
  – prin substituţie (monoalfabetică, poligrafică, polialfabetică);
  – prin transpoziţie
- metode criptografice computaţionale
  – simetrice;
  – cu chei publice
  - metode criptografice cu coduri redundante.

4.2. Criptosisteme clasice

4.2.1. Cifrul lui Caesar

Cifrul lui Caesar este un cifru cu substituţie în care fiecare literă din grup este înlocuită pentru deghizare cu o altă literă. Acest algoritm este unul dintre cele mai vechi cifruri cunoscute şi este atribuit lui Julius Caesar. În această metodă, *A* devine *D*, *B* devine *E*, *C* devine *F*, …, *X* devine *A*, *Y* devine *B*, *Z* devine *C*, conform reprezentării din tabelul 7.1.

Tabelul 4.1.

Alfabet mesaj	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Alfabet criptogramă	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C

De exemplu, mesajul

ACESTA ESTE UN TEXT CODIFICAT

devine

DFHVWD HVWH XQ WHAW FRGLILFDW.

O mică *generalizare a cifrului lui Caesar* permite alfabetului textului cifrat să fie deplasat cu *k* litere, în loc de a fi deplasat întotdeauna cu trei litere. În acest caz, *k* devine o cheie pentru metoda generală a alfabetelor deplasate circular.

Matematic, cifrul lui Caesar generalizat se exprimă conform relaţiei (4.1), iar printr-o transformare liniară a funcţiei, se obţine

$$C_i = (m_i + k) \bmod p \tag{4.8}$$

unde $i = \overline{1,n}$, n este lungimea mesajului, p este lungimea alfabetului şi k este cheia, $k \in [1, p-1]$.

Pentru o cheie k=3, se obţine cifrul lui Caesar.

Pentru o cheie k=9, alfabetul de 26 de litere (p=26) şi criptograma corespunzătoare lui sunt cele prezentate în tabelul 7.2.

Tabelul 4.2.

Alfabet mesaj	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Alfabet criptogramă	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I

Dacă mesajului

ACESTA ESTE UN TEXT CODIFICAT

i se aplică corespondenţa din tabelul 7.2, rezultă următorul mesaj criptat

JLNBCJ NBCN DW CNGC LXMRORLJC

7.2.2.Criptarea prin substituţie monoalfabetică

Principala îmbunătăţire a acestui tip de criptare o reprezintă stabilirea pentru fiecare simbol din textul clar (pentru simplitate cele 26 de litere de mai sus), o corespondenţă cu o altă literă.

Matematic, dacă există o singură lege de corespondenţă notată cu *f* (între elementele alfabetului mesajului şi elementele alfabetului criptogramei), substituţia este monoalfabetică. Astfel, pentru mesajul $M=m_1, m_2,...,m_n$, se obţine criptograma $C=c_1, c_2,...,c_n$

$$C=E_k(M)=f(m_1), f(m_2),..., f(m_n) \tag{4.9}$$

printr-o transformare liniară de forma

$$C_i=(a\ m_i+b)\ mod\ p \tag{4.10}$$

unde, suplimentar faţă de notaţiile din relaţia (7.8), *a* şi *b* sunt două numere de tip întreg, iar cheia *k* este dată de ansamblul *(a,b)*.

Criptarea care foloseşte substituţia monoalfabetică este slabă la atacuri criptanalitice (în principal cu text criptat), dat fiind faptul că identificarea cheii conduce la obţinerea întregului mesaj.

Ca un caz particular, este prezentat **cifrul aleator de substituţie**. Cheia este constituită din 26 de perechi de numere echivalente de forma *(a,b)*, cu *(a,b)* ϵ (1,2,3,...,26) . Într-un mod pseudoaleator, fiecărei litere a alfabetului primar îi corespunde o literă a alfabetului secundar. Literele alfabetului de substituire sunt static independente, dar există dezavantaje legate de generarea, transmiterea şi păstrarea cheii. Un exemplu este cel prezentat în tabelul 7.3.

Tabelul 4.3.

Alfabet mesaj	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Alfabet criptogramă	Q	W	E	R	T	Y	U	I	O	P	A	S	D	F	G	H	J	K	L	Z	X	C	V	B	N	M

Cheia acestui tip de codificare o reprezintă şirul de 26 de litere corespunzând întregului alfabet. În aceste condiţii, mesajul

ACESTA ESTE UN TEXT CODIFICAT

devine

QETLZQ TLZT XF ZTBZ EGROYOEQZ.

4.2.3. Criptarea prin transpoziţie pe coloane

Spre deosebire de cifrurile cu substituţie, cifrurile cu transpoziţie *reordonează caracterele, dar nu le deghizează.*

În exemplul următor este prezentată maniera de realizare a criptogramelor folosind un *cifru cu transpoziţie pe coloane* (tabelul 7.4). Cifrul are drept cheie un cuvânt sau o expresie care nu trebuie să conţină caractere repetate (cheia este cuvântul „universal”). Textul clar (necodificat) este scris orizontal, pe rânduri.

Scopul cheii este să stabilească numărul de coloane şi să ordoneze caracterele, coloana 1 fiind sub litera din cheie cea mai apropiată de începutul alfabetului.

Text clar: *ACEST ALGORITM DE CODIFICARE ESTE FOARTE BUN.*

Tabelul 4.4.

u	**n**	**i**	**v**	**e**	**r**	**s**	**a**	**l**
8	**5**	**3**	**9**	**2**	**6**	**7**	**1**	**4**
A	C	E	S	T		A	L	G
O	R	I	T	M		D	E	
C	O	D	I	F	I	C	A	R
E		E	S	T	E		F	O
A	R	T	E		B	U	N	.

Text criptat : *LEAFNTMFT EIDETG RO.CRO R IEBADC UAOCEASTISE*

7.2.4. Metoda cheilor acoperitoare

Construirea unui cifru imposibil de spart este actualmente destul de simplă. Tehnica este cunoscută de decenii, având următoarele etape:

- se alege un şir aleatoriu de biţi pe post de cheie;
- se converteşte textul clar într-un şir de biţi;
- pentru a obţine codificarea textului clar, se calculează *xor* între cheie şi textul clar, bit cu bit;
- pentru a realiza decodificarea se calculează tot *xor* între aceeaşi cheie şi textul codificat, tot bit cu bit.

Această metodă, cunoscută sub numele de *metoda cheilor acoperitoare (one-time pad),* are următoarele avantaje:

- textul codificat nu oferă criptanalistului nici o informaţie;
- se poate codifica orice tip fişier (text, imagine, sunet, video, bază de date, executabil) ;
- chiar şi atunci când află metoda de codificare utilizată, criptanalistul nu are şanse să deducă cheia, deoarece şirul de biţi al cheii poate avea orice lungime în raport cu şirul de biţi corespunzător textului clar.

Cu toată siguranţa pe care o oferă această metodă, ea prezintă şi dezavantaje practice importante:

- cheia nu poate fi memorată, necesitând un suport, de preferinţă electronic, o copie a ei putând ajunge oricând în posesia unei persoane neautorizate;
- indiferent că este consultată parţial sau în întregime, informaţia codificată trebuie decodificată în întregime.
  Acesta este un foarte mare dezavantaj în cazul bazelor de date, utilizatorul fiind obligat să decodifice şi să codifice respectiva bază de date în întregime, indiferent de numărul de înregistrări pe care le consultă/modifică;
- cu cât fişierul codificat este mai mare, cu atât manipularea lui (consultare/modificare) este mai greoaie, încetinind operarea cu produse software în care este implementată acestă metodă de codificare.

Un exemplu de criptare prin metoda cheilor acoperitoare este cel prezentat în continuare.

Mesaj: 00101010 10101000 10101010 01111100 01011101 11101111
Cheie: 10010111 00111011 00001111
Criptogramă: 10111101 10010011 10100101 11101011 01100110 11100000

4.3. Criptosisteme computaţionale

4.3.1. Noţiuni generale

Dacă la algoritmii clasici de criptare secretul este asigurat, în principal, de folosirea unor chei de lungimi mari, la cele computaţionale accentul este pus pe *complexitatea algoritmilor de criptare, respectiv de decriptare*. Criptarea utilizând tehnica de calcul generează cifruri greu de spart chiar şi de către un criptanalist care dispune de cantităţi mari de text cifrat.

Primul sistem de criptare bloc, *criptosistemul Lucifer*, a fost creat de IBM în 1970 pentru a mări siguranţa transferurilor bancare. În variantele acestui criptosistem apar pentru prima dată elementele unei reţele de substituţie-permutare. Sunt criptate şi decriptate mesaje de orice lungime, în grupuri de 128 biţi, sub controlul unei chei de 128 biţi aleşi aleator. Cheia se putea obţine dintr-

o cartelă magnetică sau o memorie de tip ROM [ ]. Acest tip de criptare prezintă următoarele avantaje:

- pentru aceeaşi secvenţă de mesaj, sunt generate criptograme diferite, la momente de timp diferite;
- sistemul sesizează situaţiile în care există o eroare de transmisie sau când este utilizată o cheie inadecvată.

Cu toate aceste avantaje certe, niciuna dintre variantele acestui criptosistem nu s-a bucurat de prea multă încredere din partea utilizatorilor, fiind considerate nesigure.

Schema generală a criptării folosind tehnica cheilor publice este prezentată în figura 4.1.

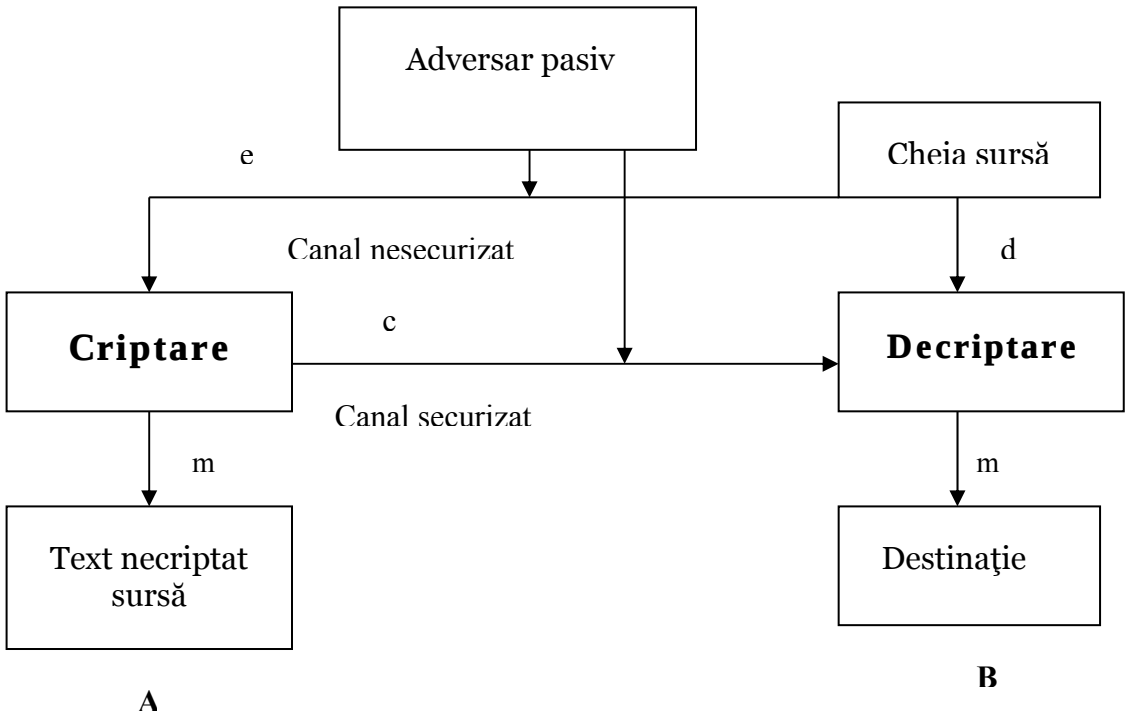

Fig. 4.1. Criptare folosind tehnica cheilor publice

Se consideră o comunicare între două entităţi, A şi B. B selectează perechea de chei (*e,d*) şi trimite cheia de criptare *e*, numită *cheie publică*, lui A, prin oricare dintre cele două canale, dar păstrează cheia secretă de decriptare *d*, numită *cheie privată*. A îi poate astfel trimite un mesaj lui B aplicând funcţia de criptare determinată de cheia publică a lui B pentru a obţine criptograma *c*, $c=E_e(m)$. B decriptează textul cifrat *c*, aplicând inversa transformării, D_d, determinată în mod unic de către *d*. Deoarece nu este necesar ca cheia de criptare *e* să fie ţinută

secretă, ea poate fi făcută publică, astfel încât orice entitate care doreşte să îi transmită un mesaj criptat lui B să poată face acest lucru, iar B să poată realiza decriptarea. Figura 7.2. ilustrează această posibilitate, unde A_1, A_2, A_3 sunt entităţi distincte. Este de subliniat faptul că, dacă A_1 distruge mesajul m_1 după ce îl criptează în c_1, atunci nici măcar A_1 nu mai poate recupera m_1 din c_1.

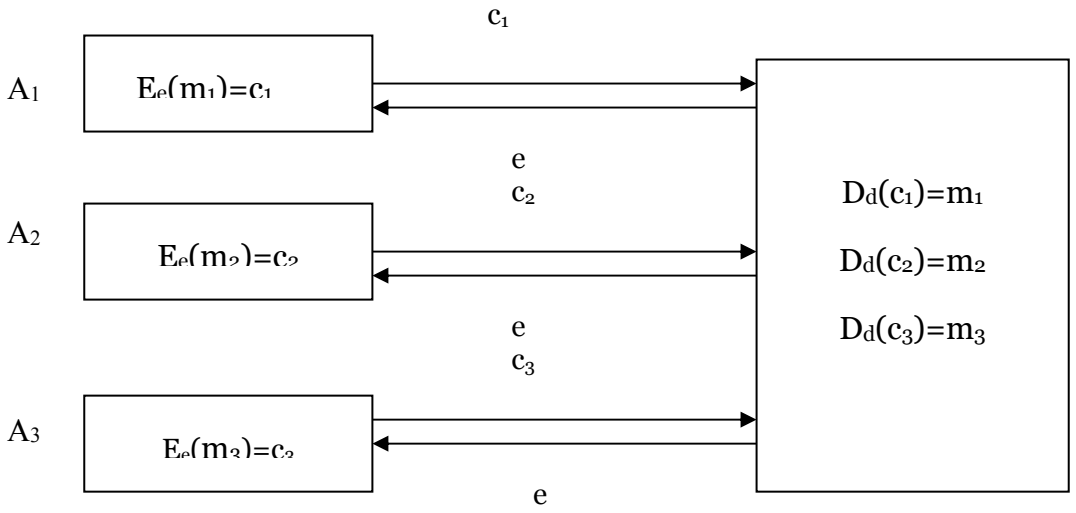

Fig. 4.2. Criptarea cu chei publice cu trei entităţi distincte

Criptarea cu chei publice ar putea fi considerată, în sensul celor prezentate până în acest punct, un sistem ideal, care nu are neapărată nevoie de un canal securizat pentru transmiterea cheii de criptare. Acest lucru ar implica faptul că două entităţi ar putea comunica peun canal nesecurizat, fără a se fi întâlnit vreodată pentru a face schimb de chei. În realitate, însă, un adversar poate "sparge" sistemul, decriptând mesajele celei de a doua entităţi, aşa cum este prezentat în figura 7.3.

În acest scenariu, al atacului asupra unei comunicări în doi cu asumarea unei identităţi false, adversarul îşi asumă identitatea entităţii B, trimiţându-i entităţii A o cheie publică e' pe care A o consideră, în mod incorect, ca fiind cheia publică a lui B.

Adversarul interceptează mesajele criptate ale lui A către B, le decriptează folosind propria cheie privată d', recriptează mesajul cu cheia publică e a lui B şi îl trimite lui B. Acest fapt evidenţiază necesitatea autentificării cheilor publice pentru a obţine autentificarea originii datelor şi a cheilor publice (A trebuie să fie sigur că B este deţinătorul legitim al cheii publice sub care a efectuat criptarea).

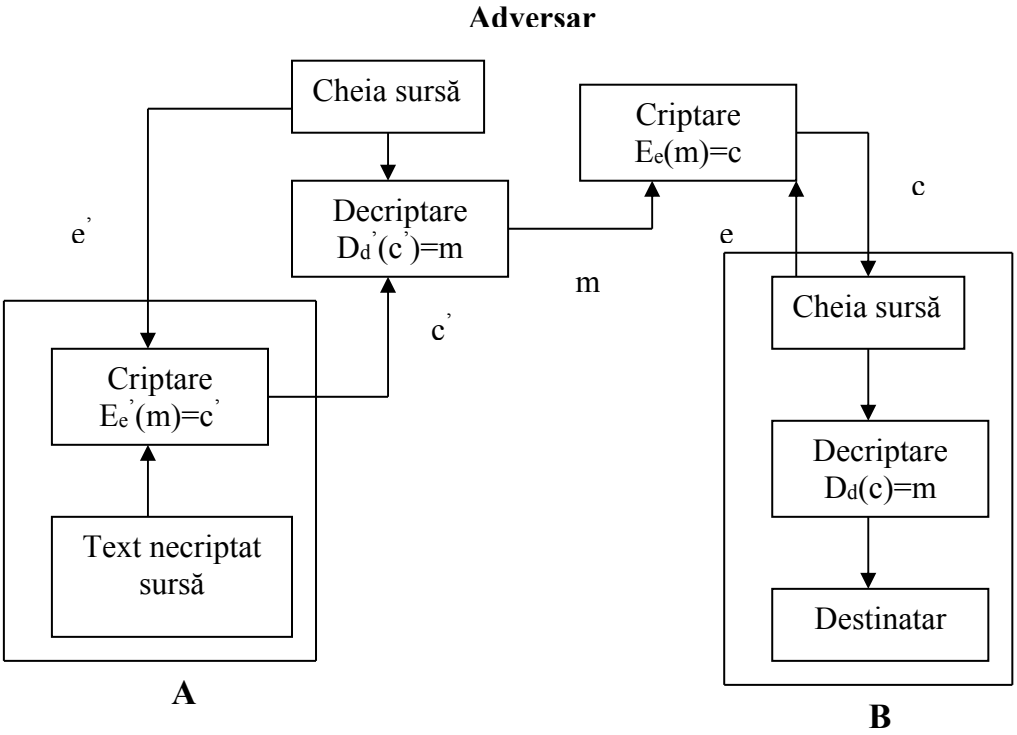

Fig. 4.3. Atacul asupra unei comunicări în doi cu asumarea unei identităţi false

4.3.2. Algoritmul R.S.A

Deoarece toţi criptologii au considerat întotdeauna ca de la sine înţeles faptul că atât pentru criptare cât şi pentru decriptare se foloseşte aceeaşi cheie şi că aceasta trebuie distribuită tuturor utilizatorilor sistemului, părea a exista întotdeauna aceeaşi problemă inerentă: cheile trebuiau protejate împotriva furtului dar, în acelaşi timp, ele trebuiau să fie distribuite, astfel încât nu puteau fi sechestrate într-un seif de bancă.

În 1976, doi cercetători, Diffie şi Hellman, au propus un tip radical nou de criptosistem în care cheile de criptare şi decriptare sunt diferite, iar cheia de decriptare nu poate fi dedusă din cheia de criptare. În propunerea lor, algoritmul (cheia) de criptare *E* şi algoritmul (cheia) de decriptare *D*, trebuiau să satisfacă trei cerinţe. Aceste trei cerinţe pot fi exprimate simplificat după cum urmează:

- *D(E(M))=M*;

- Este foarte dificil să se deducă *D* din *E*;
- *E* nu poate fi spart printr-un atac cu text clar ales.

Respectându-se aceste trei condiţii, nu există nici un motiv pentru ca *E* , respectiv cheia de criptare, să nu poată fi făcută publică; dimpotrivă, toţi utilizatorii care au adoptat acest model de criptosistem trebuie să-şi facă cunoscute cheile publice.

Plecând de la aceste trei condiţii, în anul 1978 a fost inventat criptosistemul **R.S.A**. Denumirea lui provine de la numele celor trei inventatori ai acestui mod de criptare a informaţiei: Ron **R**ivest, Adi **S**hamir şi Leonard **A**delman.

Acest criptosistem stă şi astăzi, în diverse variante, la baza sistemelor de protecţie a datelor şi transmisiilor de informaţii. El se bazează pe o problemă matematică dificilă şi anume găsirea unor numere prime foarte mari, fapt care a impulsionat elaborarea unor metode specifice mai eficiente.

Pentru obţinerea cheilor (*cheia privată* şi *cheia publică*), se procedează astfel:

- Se aleg două numere prime *p* şi *q*;
- Se calculează $n = p \times q$ şi $z = (p-1) \times (q-1)$;
- Se alege un număr *e* relativ prim cu *z*, astfel încât $1 < e < z$;
- Se găseşte un număr*d*, astfel încât $(e \times d) \bmod z = 1$ şi $1 < d < z$.

Numărul *e* se numeşte *exponent public*, iar numărul *d* *exponent privat*.

În urma operaţiilor anterioare se obţin două perechi de numere (n,e) şi (n,d) care reprezintă cheia publică, respectiv cheia privată.

Pentru a obţine mesajul criptat c, mesajul clar m (privit ca şir de biţi), se împarte în k blocuri de text clar. Fiecărui bloc m_i, $(i = \overline{0, k-1})$ i se aplică funcţia:

$$c_i(n, e) = m^e \bmod n, \text{ unde } i = \overline{0,k-1} \qquad (4.11)$$

Astfel, şirul *c* obţinut reprezintă mesajul criptat.

Pentru decriptare (obţinerea mesajului clar *m*), criptogramei *c* i se aplică funcţia:

$$m_i(n, d) = c_i^d \bmod n, \text{ unde } i = \overline{0,k-1} \qquad (4.12)$$

Din motive de securitate numerele *p* şi *q* se şterg după generarea cheilor publice şi private.

Securitatea metodei se bazează pe dificultatea factorizării numerelor mari. Dacă un criptanalist ar putea factoriza numărul `n` (public cunoscut), atunci el ar putea obţine `p` şi `q`, iar din acestea pe `z`. Cu acesta din urmă aflat, se restrâng şi variantele pentru `e`, respectiv `d`. Din fericire, matematicienii încearcă de peste 300 de ani să factorizeze numere mari şi experienţa acumulată sugerează că aceasta este o problemă mai mult decât dificilă.

În figura 4.4. este ilustrat modul de funcţionare a algoritmului R.S.A.

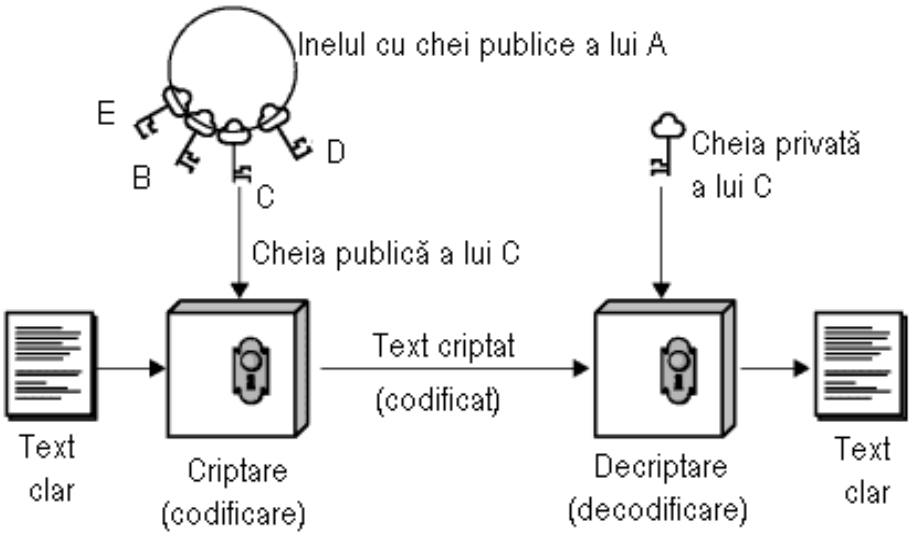

Fig. 4.4. Schema bloc a funcţionării criptosistemului R.S.A.

Persoana A deţine un grup (inel) de chei publice ale persoanelor B, C, D şi E. Pentru a transmite un mesaj criptat persoanei C, criptează mesajul (textul clar) cu cheia publică a lui C. Persoana C primeşte mesajul criptat de la A şi îl decodifică cu cheia sa privată, obţinând astfel textul clar original.

În cadrul grupului de persoane A, B, C, D, E, fiecare deţine cheile publice ale celuilalt şi le utilizează pentru transmiterea mesajelor. De asemenea, fiecare persoană îşi utilizează cheia privată (personală) pentru a decripta mesajele primite, astfel încât numai destinatarul mesajului poate citi mesajul respectiv.

Această modalitate de criptare este utilizată atunci când expeditorul este interesat ca nimeni (nici măcar cei din grup) să nu poată citi mesajul clar. Dezavantajul acestei metode este că oricine din grup poate trimite mesaje, iar destinatarul nu poate fi 100% sigur de identitatea expeditorului.

Un exemplu de calcul este prezentat în continuare.

Se aleg două numere prime (pentru o criptare eficientă p şi q se aleg mai mari de 10^{100}):

$p = 61$
$q = 53$

Se calculează:

$n = p \cdot q = 61 \cdot 53 = 3233$
$z = (p-1) \cdot (q-1) = 60 \cdot 52 = 3120$

Conform algorimului se alege $e = 17$ şi $d = 2753$

Cheia publică va fi $(n,e)=(3233,17)$
Cheia privată rezultă $(n,d)=(3233,2753)$

Se alege mesajul clar (de criptat) $m=123$.

Codificarea este

$c = m^e \mod n = 123^{17} \mod 3233 =$
$= 337587917446653715596592958817679803 \mod 3233 = 855$

Decodificarea este

$m = c^d \mod n = 855^{2753} \mod 3233 = 123$

Semnătura digitală R.S.A.

Avantajul algoritmului R.S.A. este că poate fi utilizat şi pentru semnarea mesajelor expediate. Acest tip de semnătură este cunoscut sub numele de *semnătura digitală R.S.A*.

Semnătura digitală este folosită pentru a identifica autorul unui mesaj şi reprezintă una dintre cele mai importante contribuţii ale criptării cu chei publice. Primul standard internaţional pentru semnăturile digitale , ISO/IEC 9796, bazat pe schema cheilor publice RSA, a fost adoptat în 1991. Schema bloc a sistemului de transmitere a mesajelor semnate digital este reprezentată în figura 4.5.

După cum este cunoscut, fiecare membru al unui grup deţine cheia publică a celorlalţi membri ai grupului şi cheia sa privată. Pentru a transmite persoanei C un mesaj criptat şi semnat, persoana A criptează mesajul (textul clar) cu cheia sa privată (personală). Persoana C primeşte mesajul criptat de la A şi îl decodifică cu cheia publică a acesteia (cheia publică a lui A), obţinând astfel textul clar original.

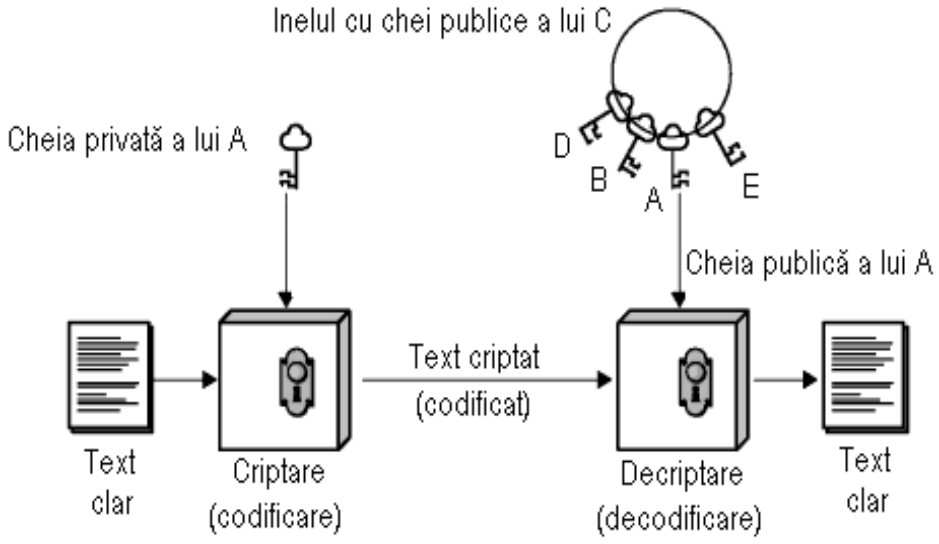

Fig. 4.5. Schema bloc a sistemului de transmitere a mesajelor semnate digital

Spre deosebire de schema de criptare din figura 7.4, în acest caz, toate persoanele din grup pot decodifica mesajul dar nu există nici un dubiu în privinţa identităţii expeditorului.

A, B, C, D şi E pot fi atât persoane cât şi programe, ceea ce înseamnă că acest sistem de criptare poate fi folosit :

- de persoane în vederea transmiterii de maseje (de exemplu transmiterea de e-mail-uri, fişiere în orice format),
- de programe (pachete de programe client/server) în vederea transmiterii de informaţii de la aplicaţia server la aplicaţia client şi/sau invers, în cadrul reţelelor de tip LAN (Local Area Network) sau WAN (World Area Network).

Folosind notaţiile de la metoda de criptare, pentru a obţine mesajul criptat c (criptogramă semnată digital), mesajul clar m (privit ca şir de biţi), se împarte în k blocuri de text clar. Fiecărui bloc m_i, $(i = \overline{0, k-1})$ i se aplică funcţia

$$c_i(n, d) = m^d \bmod n, \text{ unde } i = \overline{0,k-1} \tag{4.13}$$

Astfel şirul **c** obţinut reprezintă mesajul criptat semnat.

Pentru decriptare (obţinerea mesajului clar **m**), criptogramei **c** i se aplică funcţia

$$m_i(n, e) = c_i^e \bmod n, \text{ unde } i = \overline{0,k-1} \tag{4.14}$$

4.3.3. Algoritmul El-Gamal

Schema de criptare El-Gamal se bazează pe rezolvarea problemei logaritmilor discreţi care consideră drept date iniţiale un grup G de ordinul n şi $\beta \in G$, logaritmul discret al lui β în bază α, notat cu $\log_\alpha \beta$ este numărul întreg unic x, $0 \le x \le n-1$, astfel încât $\beta = \alpha^x$.

Cel mai important aspect al criptării cu chei publice este acela care se referă la faptul că autorul unui mesaj nu are posibilitatea să îl decripteze, după ce l-a trimis destinatarului, deoarece nu deţine cheia privată necesară decriptării. Astfel, dacă este transmis, din greşeală, un mesaj unui alt destinatar decât cel

dorit, acesta nu îl va putea decripta, păstrându-se în acest fel secretul respectivei informaţii.

În acest context, două entităţi A şi B care doresc să comunice trebuie să parcurgă următoarele etape:

- generarea cheilor pentru criptarea cu chei publice El-Gamal (algoritm 1);
- realizarea schimbului de chei publice, care poate fi făcută pe orice canal de comunicaţii, indiferent de gradul de securitate conferit de acesta (de exemplu prin e-mail sau telefonie);
- criptarea de către B a mesajului pe care doreşte să i-l transmită lui A (algoritm 2);
- decriptarea de către A a mesajului cifrat *c* primit de la B (algoritm 3).

Algoritm 1. Generarea cheilor pentru criptarea cu chei publice El-Gamal

Fiecare entitate creează o cheie publică şi o cheie privată corespunzătoare. Entitatea A trebuie să aplice următorii paşi:

- Generarea unui număr *p* prim, cu o valoare mare, şi a unui număr α din Z^*_p;
- Selectarea unui număr întreg aleator a, $1 \le a \le p\text{-}2$, şi calcularea $\alpha^a \bmod p$;
- A deţine cheia publică (p, α, α^a) şi cheia privată a.

Algoritm 2. Criptarea cu chei publice El-Gamal

Entitatea B criptează mesajul *m,* prin aplicarea următorilor paşi:

- Obţinerea cheii publice aparţinând lui A (p, α, α^a) ;
- Reprezentarea mesajului ca un număr întreg aparţinând domeniului $\{0,1,...,p\text{-}1\}$;
- Selectarea unui număr întreg aleator k, $1 \le k \le p\text{-}2$;
- Calcularea $\gamma = \alpha^k \bmod p$ şi a lui $\delta = m(\alpha^a)^k \bmod p$.
- Trimiterea către A a textului cifrat $c=(\gamma, \delta)$.

Mesajul este considerat a fi un şir de caractere. Pentru a putea fi criptat, aceste caractere sunt transformate în codul ASCII (*American Standard Code for Information Interchange*) echivalent lor. Deoarece reprezentarea mesajului se

face ca un număr întreg aparţinând domeniului {0,1,...,p-1}, p nu poate fi ales mai mic decât 128.

Algoritm 3. Decriptarea cu chei publice El-Gamal

Pentru a recupera textul m, care a fost cifrat, din c, A trebuie să aplice paşii de mai jos:

- Folosind cheia privată α, calculează $\gamma^{p-1-a} \bmod p$, cu $\gamma^{p-1-a} = \gamma^{-a} = \alpha^{-ak}$.
- Recuperarea lui m calculând $(\gamma^{-a})\delta \bmod p$.

Verificarea decriptării. Decriptarea permite recuperarea textului mesajului original, deoarece $\gamma^{-a} \cdot \delta \bmod p = \alpha^{-ak} m \alpha^{ak} \bmod p = m$.

În continuare este prezentat un exemplu de criptare folosind algoritmul El-Gamal pentru parametri mici.

Generarea cheii. Entitatea A selectează numărul prim p=2357 şi un α=2 din Z^*_{2357}. În continuare, A îşi alege cheia privată a=1751 şi calculează
$\alpha^a \bmod p = 2^{1751} \bmod 2357 = 1185$.
Cheia publică a lui A este (p=2357, α=2, α^a=1185).

Criptarea. Pentru a cripta mesajul m=2035, B selectează un număr întreg aleator k=1520 şi calculează:
$\gamma = 2^{1520} \bmod 2357 = 1430$
$\delta = 2035 \cdot 1185^{1520} \bmod 2357 = 697$.
B trimite lui A γ = 1430 şi δ = 697.

Decriptarea. Pentru a decripta, A calculează
$\gamma^{p-1-a} = 1430^{605} \bmod 2357 = 872$
şi recuperează mesajul m, calculând
$m = 872 \cdot 697 \bmod 2357 = 607\,784 \bmod 2357 = 2035$.

4.3.4. Algoritmul Merkle-Hellman

Schema de criptare Merkle-Hellman se bazează pe problema sumei subşirurilor. Ideea de bază este selectarea unei instanţe a problemei sumei subşirurilor care este uşor de rezolvat, deghizată apoi ca o instanţă a problemei generale a problemei sumei subşirurilor care este mai dificil de soluţionat. *Şirul original poate servi drept cheia privată, iar şirul transformat serveşte drept cheie publică*.

Schema de criptare Merkle-Hellman este importantă din motive istorice, deoarece a fost prima realizare concretă a unei scheme de criptare cu chei publice. Ulterior au fost realizate mai multe variante, dar toate, inclusiv cea originală, s-au dovedit nesigure.

Definiţia problemei sumei subşirurilor poate fi formulată astfel: fiind dat un şir de întregi pozitivi $\{a_1, a_2,\ldots,a_n\}$ şi un întreg pozitiv s, să se determine dacă există sau nu un subşir a_j a cărui sumă este s, şi dacă există $x_i \in \{0,1\}$, $1 \le i \le n$ astfel încât $\sum_{i=1}^{n} a_i x_i = s$.

Algoritmul pentru generarea cheilor pentru criptarea cu chei publice Merkle-Hellman cuprinde în mod necesar următoarele etape:

Fiecare identitate îşi creează o cheie publică şi una privată corespunzătoare

1. Un număr întreg n este stabilit ca parametru comun de sistem.
2. Fiecare identitate urmează paşii 3-7.
3. Se alege un şir supercrescător ($b_1, b_2, \ldots, b_n$) şi un modul M astfel încât $M > b_1+b_2+\ldots+b_n$.
4. Se selectează un întreg aleator W, $1 \le W \le M-1$, astfel încât cel mai mare divizor comun al lui W şi M să fie 1.
5. Se alege o permutare aleatoare □□de întregi $\{1, 2, \ldots, n\}$.
6. Se calculează $a_i = W \cdot b_{□□i□} \,□\bmod M$ pentru $i=1, 2, \ldots, n$.
7. Cheia publică a lui A este ($a_1, a_2, \ldots, a_n$), iar cheia privată este (□,, M, W, ($b_1, b_2, \ldots, b_n$)).

Criptarea cu chei publice Merkle-Hellman. B criptează un mesaj m pentru A.

a. Se obţine cheia publică autentică a lui A ($a_1, a_2, \ldots, a_n$)
b. Mesajul m este reprezentat ca un şir binar de lungime n, $m=m_1m_2\ldots m_n$
c. Este calculat întregul $c=m_1a_1+m_2a_2+\ldots+m_na_n$
d. Textul cifrat c este trimis către A.

Decriptarea cu chei publice Merkle-Hellman

e. Se calculează $d = W^{-1}c \bmod M$.
f. Rezolvând problema sumei subşirurilor, sunt găsiţi întregii $r_1, r_2, \ldots, r_n$, $r_i \in \{0,1\}$, astfel încât $d=r_1b_1+r_2b_2+\ldots+r_nb_n$.
g. Biţii mesajului sunt: $m_i=r_{(i)}$, $i \in 1,2,\ldots,n$.

Algoritmul binar extins al celui mai mare divizor comun, folosit pentru a determina inversul faţă de înmulţire al x mod y

Date de intrare: două numere întregi x şi z.
Date de ieşire: întregii a şi b, astfel încât $ax + bz = 1$, unde 1 este cel mai mare divizor comun al lui x şi y.

1. cât timp x şi y sunt pare, execută: $x \leftarrow x/2$, $y \leftarrow y/2$
2. $u \leftarrow x$, $v \leftarrow y$, $A \leftarrow 1$, $B \leftarrow 0$, $C \leftarrow 0$, $D \leftarrow 1$.
3. cât timp u este par, execută:
   3.1. $u \leftarrow u/2$
   3.2. dacă $A \equiv B \equiv 0 \pmod 2$, atunci $A \leftarrow A/2$, $B \leftarrow B/2$;
   altfel, $A \leftarrow (A+y)/2$, $B \leftarrow (B-x)/2$.
4. cât timp v este par, execută:
   4.1. $v \leftarrow v/2$
   4.2. dacă $C \equiv D \equiv 0 \pmod 2$, atunci $C \leftarrow C/2$, $D \leftarrow D/2$;
   altfel, $C \leftarrow (C+y)/2$, $D \leftarrow (D-x)/2$.
5. dacă $u \geq v$ atunci $u \leftarrow u - v$, $A \leftarrow A - C$, $B \leftarrow B - D$
   altfel: $v \leftarrow v - u$, $C \leftarrow C - A$, $D \leftarrow D - B$.
6. Dacă $u = 0$, atunci $a \leftarrow C$, $b \leftarrow D$ şi returnează a, altfel reia de la pasul 3.

Definiţie: un şir ($b_1, b_2, \ldots, b_n$) de numere întregi pozitive este supercrescător dacă are proprietatea $b_i > \sum_{j=1}^{i-1} b_j$ pentru fiecare i, $2 \leq i \leq n$.

Algoritm pentru rezolvarea problemei sumei unui subşir supercrescător

Date de intrare: un şir supercrescător ($b_1, b_2, \ldots, b_n$) şi un număr întreg s, care este suma subşirului b_i.
Date de ieşire: ($x_1, x_2, \ldots, x_n$) unde $x_i \in \{0,1\}$, astfel încât $\sum_{i=1}^{n} x_i b_i = s$.

1. $i \leftarrow n$
2. cât timp i ≥ 1 execută:

dacă $s \geq b_i$, atunci $x_i \leftarrow 1$ şi $s \leftarrow s - b_i$; altfel: $x_i \leftarrow 0$.
$i \leftarrow i - 1$

3. returnează ($x_1, x_2, \ldots, x_n$)

Verificarea decriptării. Decriptarea din algoritmul de mai sus permite recuperarea textului mesajului original, deoarece $d \equiv W^{-1}c \equiv W^{-1}\sum_{i=1}^{n} m_i a_i \equiv \sum_{i=1}^{n} m_i b_{\pi(i)} \pmod{M}$. Deoarece $0 \leq d \leq M$, $d = \sum_{i=1}^{n} m_i b_{\pi i} \bmod M$, deci soluţia oferită la pasul g al decriptării indică în mod corect biţii mesajului după aplicarea permutării π.

În continuare este prezentat un exemplu de criptare folosind algoritmul Merkle-Hellman pentru parametri mici

Generarea cheilor

Fie n=6. Entitatea A alege şirul supercrescător (12, 17, 33, 74, 157, 316), M=737, W=635, şi permutarea π = {1, 2, 3, 4, 5, 6} definită prin π(1) = 3, π(2) = 6, π(3) = 1, π(4) =2 , π(5) = 5 şi π(6) =4 .
Cheia publică a lui A în urma calculelor:

$a_1 = W b_{\pi(1)} \bmod M = 635 \cdot 33 \bmod 737 = 20955 \bmod 737 = 319$
$a_2 = W b_{\pi(2)} \bmod M = 635 \cdot 316 \bmod 737 = 200660 \bmod 737 = 196$
$a_3 = W b_{\pi(3)} \bmod M = 635 \cdot 12 \bmod 737 = 70620 \bmod 737 = 250$
$a_4 = W b_{\pi(4)} \bmod M = 635 \cdot 17 \bmod 737 = 10795 \bmod 737 = 477$
$a_5 = W b_{\pi(5)} \bmod M = 635 \cdot 157 \bmod 737 = 99695 \bmod 737 = 200$
$a_6 = W b_{\pi(6)} \bmod M = 635 \cdot 74 \bmod 737 = 46990 \bmod 737 = 559$

este şirul (319, 196, 250, 477, 200, 559), iar cheia privată este (□□□*M, W,*(12, 17, 33, 74, 157, 316)).

Criptarea. Pentru a cripta mesajul m = 101101, B calculează

$$c = 1\cdot319 + 0\cdot196 + 1\cdot250 + 1\cdot477 + 0\cdot200 + 1\cdot559 = 1605$$

şi îl trimite lui A.

Decriptarea. Pentru a decripta c , A calculează $d = W^{-1}c$ mod M cu ajutorul algoritmului binar extins al celui mai mare divizor comun, sub forma $d = W^{-1}c$ mod M = 513·1605 mod 737 = 136.
Pentru a decripta mesajul trebuie rezolvată ecuaţia:
$136 = 12r_1 + 17r_2 + 33r_3 + 74r_4 + 157r_5 + 316r_6$, obţinându-se $r_1 = 1, r_2 = 1, r_3 = 1, r_4 = 1, r_5 = 0, r_6 = 0$, apoi se aplică permutarea □□ rezultând biţii mesajului: $m_1 = r_3 = 1, m_2 = r_6 = 0, m_3 = r_1 = 1, m_4 = r_2 = 1, m_5 = r_2 = 0, m_6 = r_4 = 1,$ deci mesajul decriptat este m = 101101 , acelaşi cu cel care a fost criptat.

4.3.4. Algoritmul D.E.S

Sistemul **D.E.S. (Data Encryption Standard)** este unul dintre cele mai cunoscute exemple de cifruri bloc, adoptat în 1977 în SUA de către National Bureau of Standards drept standardul federal de procesare a informaţiei pentru criptarea informaţiilor neclasificate, pornind de la un cifru elaborat de IBM. D.E.S. este un cifru bloc cu lungimea de 64 biţi prelucraţi în conjuncţie cu o cheie, compusă din 56 biţi generaţi pseudo-aleator şi 8 biţi folosiţi pentru detectarea erorilor de transmisie; fiecare din aceşti biţi reprezintă paritatea impară a celor 8 octeţi ai cheii. Aceasta este expandată la lungimea blocului şi păstrată de către toţi membrii unui grup de utilizatori.

Construcţia fundamentală a unui bloc D.E.S. este o combinaţie a două tehnici elementare de criptare, şi anume substituţie urmată de permutare, efectuate asupra textului, pe baza unei chei. Această construcţie este cunoscută sub denumirea de *rundă*. Criptosistemul D.E.S. este compus din 16 runde. Algoritmul este bazat pe un set de permutări, substituţii şi sumă modulo 2, aplicate iterativ de 16 ori, pe un bloc de 64 biţi, prin folosirea de fiecare dată a unei chei diferite de 48 biţi, extrasă dintr-o cheie de 56 biţi.

Criptarea D.E.S. constă din următoarele categorii de prelucrare efectuate asupra blocului ce conţine textul de cifrat, conform prezentării din figura 4.6 [2]:

- blocul de date de 64 biţi este supus unei permutări iniţiale "IP";
- blocul permutat trece printr-un calcul complex care depinde de cheie şi care constă din 16 iteraţii funcţional identice, parametrizate de chei diferite;
- un interschimb al celor două jumătăţi ale blocului, fiecare având 32 biţi;
- o permutare (transpoziţie) finală, care este inversa celei iniţiale.

Prelucrarea la fiecare iteraţie "*i*" constă din următoarele operaţii [2]:

- se notează cu *L(i-1) şi R(i-1)* cele două jumătăţi de 32 de biţi, stânga şi dreapta, care compun blocul supus iteraţiei respective;
- considerând *k(i)* cheia şi un bloc de 48 de biţi aleşi din cei 64 de biţi ai cheii, sunt valabile relaţiile pentru calculul ieşirilor *L(i)* şi *R(i)* :

$$\begin{aligned} L(i) &= R(i-1); \\ R(i) &= L(i-1) \oplus f(R(i-1), K(i)) \end{aligned} \tag{4.15}$$

Ultima iteraţie este diferită de celelalte, fiind definită de relaţiile:

$$\begin{aligned} L(16) &= R(15); \\ R(16) &= L(15) \oplus f(R(15), K(16)) \end{aligned} \tag{4.16}$$

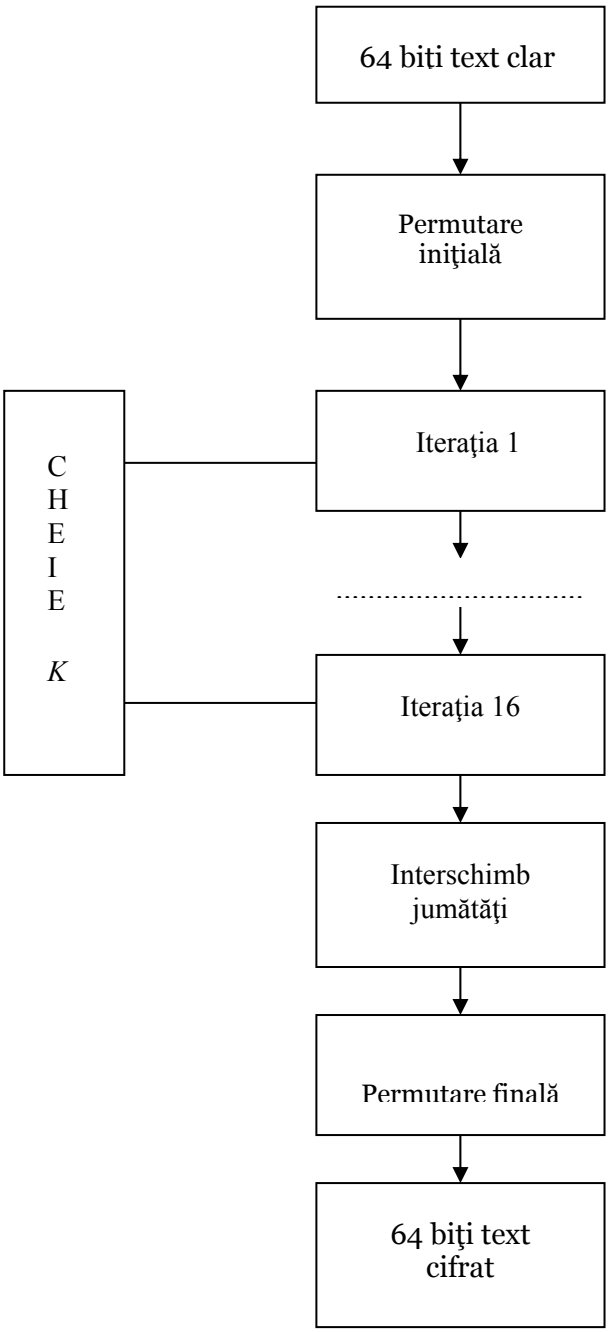

Fig. 4.6. Schema generală a criptării cu algoritmul D.E.S.

Funcţia de criptare *f* folosită realizează o substituţie neliniară, astfel încât asupra blocului iniţial de 32 de biţi se aplică o funcţie de expandare *E,* care generează 48 de biţi la ieşire. În continuarea prelucrărilor făcute de funcţia *f*, *E(R(i-1))* se însumează modulo 2 cu cei 48 de biţi ai cheii *K(i)*. Rezultatul este partiţionat în 8 blocuri de 6 biţi care constituie intrările a 8 cutii *S(j), j=1,..,8*, care realizează o substituţie neliniară cu 6 intrări şi 4 ieşiri.

Fie $S_1, S_2,...,S_8$ cele 8 cutii S, P funcţia de permutare şi E funcţia de expandare prezentate mai sus. Pentru a defini funcţia $f(R(i-1), K(i))$ se realizează blocurile $B_1,...,B_8$ de 6 biţi fiecare

$$B_1,...,B_8 = K(i) \oplus E(R(i-1)) \tag{4.17}$$

În acest caz, blocul $f(R(i-1), K(i))$ poate fi definit ca

$$f(R(i-1), K(i)) = P(S_1(B_1)\ S_2(B_2)\ ...\ S_8(B_8)) \tag{4.18}$$

După calculul format din cele 16 iteraţii descrise anterior, blocul de 32 de biţi este supus unei permutări IP^{-1}, *inversa permutării iniţiale*.

Decriptarea constă în folosirea aceluiaşi algoritm, dar cu cheile $K(i)$ aplicate în sens invers, de la K_{16} la K_1. Astfel, primul pas în decriptare este aplicarea permutării IP, care dezleagă ultimul pas IP^{-1} din operaţia de criptare. Apoi se generează în sens invers:

$$\begin{aligned} R(i-1) &= L(i) \\ L(i-1) &= R(i) \oplus f(L(i), K(i)) \end{aligned} \tag{4.19}$$

Relaţiile (4.19) urmează a fi aplicate de la $R(16)$ şi $L(16)$, generându-se în final $R(0)$ şi $L(0)$. În final, blocul de 64 de biţi obţinut este supus unei permutări inverse IP^{-1}, ceea ce conduce la obţinerea mesajului de tip text clar.

Atacuri asupra algoritmului D.E.S.

Pentru orice sistem de cifrare, cea mai populară metodă de atac este ***atacul forţă–brută***, ceea ce înseamnă *încercarea tuturor cheilor posibile pentru spargere*. Lungimea cheii este cea care determină numărul de chei posibile şi, prin urmare, posibilitatea de reuşită a acestui atac.

În mediile academice, au fost avansate mai multe propuneri pentru spargerea criptosistemului D.E.S. În 1977, Diffie şi Hellman au propus o maşină de calcul,

al cărei cost era estimat la 20 milioane USD, care putea găsi cheia D.E.S. în 24 ore. În 1993, Wiener a propus şi el o maşină de căutare a cheii care costa 1 milion USD şi care găsea cheia D.E.S. în 7 ore.

Fezabilitatea spargerii criptosistemului D.E.S. a fost demonstrată încă din 1988, când a fost înfiinţată o comunitate de spargere a D.E.S. de către Electronic Frontier Foundation (EFF). Motivaţia acestui grup era să demonstreze că D.E.S. era la fel de uşor de spart în teorie, ca şi în practică. Maşina foloseşte atacul forţat pentru a sparge cheia în mai puţin de două zile de căutare.

Există şi o a doua categorie de atacuri, numite ***atacuri mai rapide decât forţă-brută*** . În această categorie sunt incluse trei tipuri de atacuri cunoscute care pot sparge toate cele 16 faze ale D.E.S. cu o complexitate mai mică decât căutarea de tip forţă-brută: *Criptanaliza Diferenţială (DC), Criptanaliza Lineară (LC)* şi *atacul Davie*. Cu toate acestea, atacurile sunt teoretice şi sunt imposibil de pus în practică, fiind denumite şi vulnerabilităţi certificate.

Criptanaliza diferenţială a fost descoperită la sfârşitul anilor ’80 de Eli Biham şi Adi Shamir. Pentru a sparge toate cele 16 faze, criptanaliza diferenţială necesită 2^{47} secvenţe de text prestabilite. D.E.S. a fost proiectat să reziste atacului DC.

Criptanaliza lineară a fost descoperită de Mitsuru Matsui în 1994 [20], şi necesită 2^{43} secvenţe de text prestabilite; metoda a fost implementată şi a fost primul experiment criptanalitic raportat de către D.E.S. Nu există nici o probă să ateste faptul că DES a fost proiectat să reziste la acest tip de atac.

O generalizare a LC, şi anume *criptanaliza lineară multiplă,* a fost sugerată în tot în 1994 de către (Kaliski şi Robshaw [21] şi a fost rafinată mai departe de Biryukov în 2004 [22] ; analiza lor sugerează că aproximările lineare multiple pot fi folosite pentru a reduce datele necesare atacului cu cel puţin un multiplu de 4, fiind necesare astfel 2^{41} secvenţe in loc de 2^{43}.

În timp ce criptanaliza diferenţială şi cea lineară sunt tehnici generale şi pot fi aplicate unui număr mare de categorii de criptosisteme, *atacul Davie* este o tehnică specializată pentru D.E.S., propusă la început de Davie in anii ’80, şi îmbunătăţită de Biham şi Biryukov în 1997. Cea mai puternică formă de atac necesită 2^{50} secvenţe de text prestabilite, are o complexitate de ordinul 2^{50} şi are 51% rată de success.

4.3.5. Algoritmul A.E.S.

Standardul de codare avansat **A.E.S.- Advanced Encryption Standard** a fost dezvoltat în anul 1977 de doi criptografi belgieni, Joon Daemen şi Vincent Rijmen, sub denumirea de "Rijandel", un hibrid alcătuit din numele celor doi inventatori. Acest algoritm a reprezentat o îmbunătăţire a proiectului iniţial al celor doi, numit Square, fiind o reţea de tip substituţie-permutare. A.E.S. prezintă avantajul rapidităţii, este relativ uşor de implementat şi necesită puţină memorie.

Descrierea algoritmului A.E.S.

A.E.S. are o *mărime bloc* fixă de 128 biţi şi *mărimea cheii* de 128, 192 sau 256 biţi şi operează pe un tablou de biţi de dimensiune 4x4, numit "the state", conform reprezentării din figura 4.7.

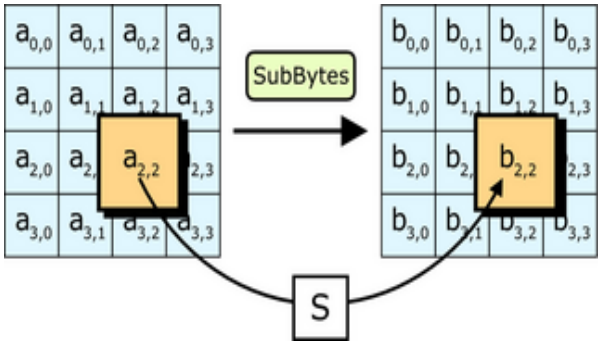

Fig. 4.7. Descrierea algoritmului A.E.S.

Pentru *codare*, fiecare *ciclu* a lui A.E.S. (exceptând ultimul) constă din *patru stadii,* şi anume:

- *Substituire*, prin efectuarea unui pas de substituţie neliniar, prin care fiecare bit este înlocuit cu un altul, conform descrierii din figura 7.7.;
- *Schimbarea rândurilor* – un pas de transpunere, unde fiecare rând al stării este schimbat ciclic la un anumit număr de pasi;
- *Amestecarea coloanelor* - o operaţie de amestecare care operează asupra coloanelor de stare, combinând cei 4 biţi din fiecare coloană folosind o transfomare liniară;

- *Cheia adăugare ciclică*, în care fiecare bit al stării este combinat cu cheia ciclului, astfel încât fiecare ciclu este obţinut din cheia zero folosind o cheie program.

Ciclul final omite stadiul amestecării coloanelor.

Pasul de substituţie

În *pasul de substituţie*, fiecare bit din tablou este actualizat folosind un S-box de 8 biţi. Această operaţie furnizează o neliniaritate în cifru. S-box-ul folosit este obţinut din inversarea funcţiei GF (2^8) cunoscută, cu bune proprietăţi neliniare. Pentru a evita atacurile bazate pe proprietăţi simple algebrice, S-box-ul este construit prin combinarea funcţiei inverse cu o transformare inversabilă. S-box-ul este de asemenea ales pentru a evita orice puncte fixe şi orice puncte fixe opuse.

Pasul de schimbare a rândurilor (Shift Rows)

Pasul de schimbare a rândurilor operează asupra rândurilor stadiului, conform prezentării din figura 7.8., schimbând ciclic biţii din fiecare linie după un anumit tipar. Pentru A.E.S., primul rând este lăsat neschimbat. Fiecare bit al celui de al doilea rând este schimbat câte unul către stânga. În mod similar, rândurile trei şipatru sunt schimbate după un tipar de doi, respectiv trei.

În acest fel, fiecare coloană a stării de ieşire a pasului 'Schimbarea rândurilor' este alcătuită din biţii fiecărei coloane a stării de intrare.

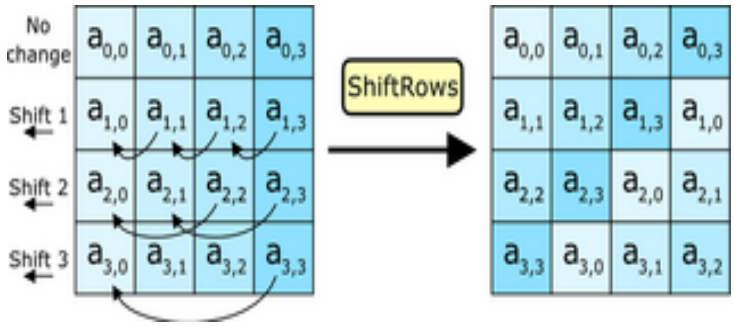

Fig. 4.8. Pasul de schimbare a rândurilor

Pasul de combinare a coloanelor (Mix Columns)

În pasul de combinare a coloanelor, cei patru biţi ai fiecărei coloane a stării sunt combinaţi folosind o transformare liniară inversabilă împreună cu schimbarea rândurilor, combinarea coloanelor furnizând difuzia în cifru. Fiecare coloană este tratată ca un polinom definit în GF(2^8) şi este atunci înmulţită modulo X^4 +1 cu un polinom fix C(X), aşa cum este ilustrat în figura 4.9.

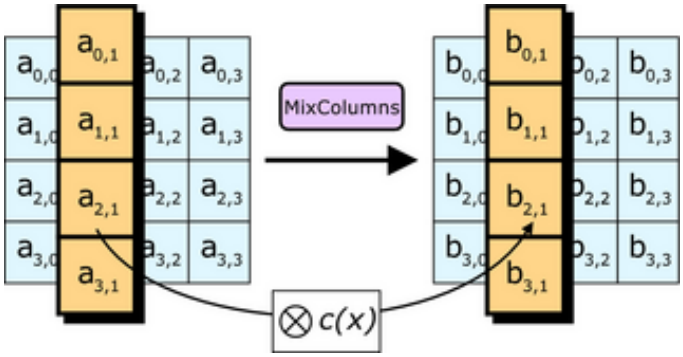

Fig. 4.9. Pasul de combinare a coloanelor

Pasul cheie adăugare ciclică (Add Round Key)

În acest pas, subcheia este combinată cu starea. Prin efectuarea fiecărui ciclu, o subcheie este obţinută din cheia principală folosind programul cheii, fiecare subcheie având aceeaşi mărime ca şi starea. Subcheia este adăugată prin combinarea fiecărui bit al stării cu bitul corespunzător al subcheii bitwise (bitul inteligent) XOR. Realizarea acestui pas poate fi observată în figura 4.10.

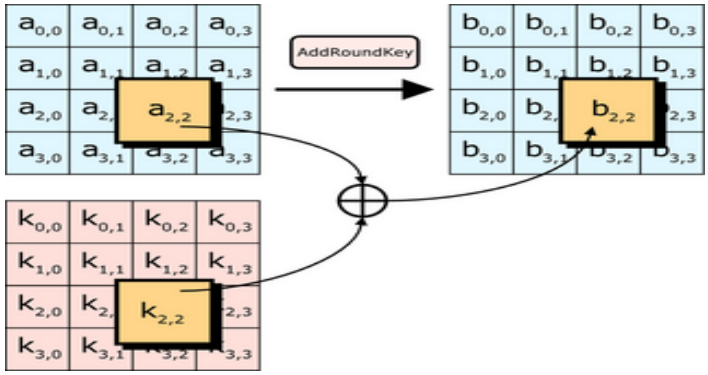

Fig. 4.10. Pasul cheie adăugare ciclică

Implementarea A.E.S. în produse informatice având misiunea de a proteja sistemele de securitate naţională şi/sau informaţiile trebuie să fie certificată de Agenţia de Securitate Naţională (NSA) înainte de achiziţie şi folosire. Aceasta marchează, în premieră, accesul publicului la un cifru aprobat de NSA pentru informaţii de tip strict secret. Este interesant de observat că multe produse publice folosesc implicit chei secrete de 128 biţi; sunt posibile, aşadar, atacuri de 14 cicluri pentru cheile de 192 biţi şi tot de 14 cicluri pentru cheile de 256 biţi. Până în prezent, cele mai cunoscute atacuri sunt cele de 7 cicluri pentru cheile de 128 biţi, 8 cicluri pentru cheile de 192 biţi şi 9 cicluri pentru cheile de 256 biţi.

Criptografii care realizează aceste chei se tem, însă, de securitatea A.E.S. Ei consideră că marginea dintre numărul de cicluri specificate în cifru şi cele mai cunoscute atacuri este prea mică pentru confortul unei siguranţe absolute. Riscul este că unele modalităţi de a îmbunătăţi aceste atacuri vor fi, evident, găsite şi, în consecinţă, cifrul poate fi spart. În acest context, o"spargere" de cifru este oricând mai rapidă decât o căutare completă; astfel un atac împotriva unei chei A.E.S. de 128 biţi, necesită "doar" 2^{120} operaţii pentru a se considera o spargere, chiar dacă acest lucru este, teoretic, complet imposibil. În consecinţă, din fericire, pentru moment, asemenea preocupări pot fi ignorate.

Cel mai mare *atac public* cunoscut ca forţă brută, a fost asupra unei chei RC5 de 64 biţi, prin *reţea distribuită.* Luând în considerare faptul că, spre deosebire de alte cifruri bloc, A.E.S. are o descriere matematică foarte riguroasă, el nu a constituit încă obiectul unui atac, dar unii cercetători sunt îngrijoraţi că viitoarele atacuri pot găsi modalitaţi de a exploata această stuctură.

4.3.6. Curbe eliptice

Studiul curbelor eliptice este o ramură importantă a matematicii. Curbele eliptice sunt simple funcţii, definite peste orice câmp de numere şi constau din elemente numite *puncte (x, y)*, cu reprezentările grafice asociate cunoscute, de tipul celor din figura 4.11. Interesul pentru aceste construcţii matematice devine, însă, cu totul special atunci când se studiază punctele în care curbele respective intersectează coordonatele întregi (*x, y*).

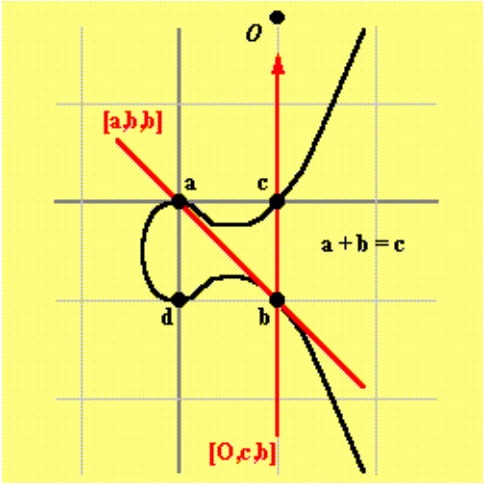

Fig. 4.11. Reprezentarea grafică a unei curbe eliptice

Domeniul curbelor eliptice a fost studiat destul de intens de către matematicieni, dar în ultima jumătate a secolului 20 au fost obţinute rezultate foarte semnificative. Una dintre aplicaţiile teoriei curbelor eliptice este *utilizarea curbelor eliptice în criptografie* [24].

Curbele eliptice pot furniza versiuni ale metodelor de criptare cu chei publice, dovedindu-se, în unele cazuri, mai rapide şi folosind chei mai scurte, cu avantajul asigurării unui nivel echivalent de securitate. Secretul constă în utilizarea unui tip aparte de grup matematic pentru aritmetica cheilor publice.

Pentru descrierea unei curbe eliptice este util un exerciţiu de imaginaţie a unei reprezentări grafice pe o coală de hârtie. Fiecare linie reprezintă un întreg şi fiecare intersecţie de linii reprezintă o pereche de intregi de coordonate (*x,y*). Desenând pe această coală de hârtie o curbă care se întinde la infinit şi care traversează un număr finit (*x,y*), la fiecare asemenea intersecţie se poate marca un punct. Aceste puncte speciale de pe curbă pot fi numărate şi poate fi definit un "operator de adunare" care combină oricare două puncte pentru a localiza astfel un al treilea punct. Acest aşa-numit operator de adunare care acţionează asupra punctelor formează un grup finit. Curba prezentată în figura 7.11. este definită de ecuaţia

$$y^2 + y = x^3 - x^2 \qquad (4.20)$$

Adăugarea de puncte pe curba eliptică

În cazul curbelor eliptice, regula referitoare la adăugarea de puncte pleacă de la premisa că fiecare linie dreaptă care trece prin curbă o intersectează în exact trei puncte. În acest context, regula de adăugare a punctelor, fie ele *u* şi *v* este următoarea : se trasează o linie dreaptă prin *u* şi *v* pentru a determina cel de al treilea punct de intersecţie *w;,* apoi se trasează o linie verticală prin *w* pentru a se determina un alt punct de intersecţie, notat cu *z*.

Astfel, suma $u + v = z$.

Dar pentru a putea utiliza această regulă în context criptografic, trebuie să fie definite unele situaţii speciale care implică existenţa unui punct imaginar suplimentar *O,* numit *origine* sau *punct la infinit*. Se admite că punctul *O* este localizat foarte sus, unde se presupune că toate liniile verticale converg. În plus, *O* este pe curbă, chiar dacă nu este caracterizat de coordonatele specifice (*x,y*). O altă presupunere este aceea că o linie tangentă la un punct se spune că intersectează punctul de două ori.

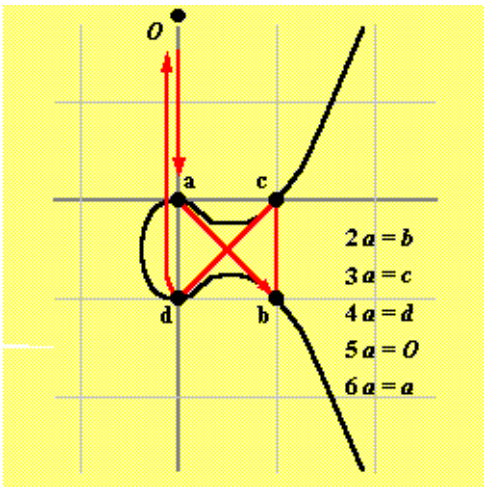

Fig. 4.12. Adăugarea de puncte pe curba eliptică

În exemplul considerat, linia care trece prin *a* şi *b* intersectează un al "treilea" punct *b,* conform reprezentării din figura 7.12**.** Se notează această linie **[***a,b,b***]**.

Această regulă de adunare creează un grup matematic de puncte. Curba reprezentată în figura 7.12. intersectează (0,0), (1,-1), (1,0), (0,-1), şi *O*, sau {*a, b, c, d, O*}. Se calculează *a*+*b* prin trasarea liniei tangente [*a, b, b*] pentru a determina *b*, şi apoi se utilizează linia [*O, c, b*] pentru a determina *c*.

Multiplicarea scalară a unui punct este o adunare repetată a punctului cu el însuşi. Ştiind că $a+a=b$ şi $a+b=c$, se poate calcula $3a=c$. De asemenea, poate fi demonstrat faptul că O este elementul neutru al grupului, deoarece pentru orice v, $v + O = v$. Punctul suplimentar considerat, O, este analog multiplicării prin 1 într-un grup de întregi modulo p.

Din considerente criptografice, vor fi considerate câmpuri de numere finite, iar aritmetica acestor curbe este operată modulo p, unde p este fie un număr prim foarte mare, fie un număr putere foarte mare a lui 2. Un grup eliptic conţine, de regulă N puncte, unde N este aproape egal cu p, $N = k * q$, q este prim, iar k este un număr mic. *Operaţia de adunare pe o curbă eliptică este corespondentă operaţiei de înmulţire în sisteme cu chei publice obişnuite, iar multiplicarea este corespondenta exponenţierii numerelor în* Z_p^*.

Considerând numărul prim p, $p =$

6.277.101.735.386.680.763.835.789.423.207.
666.416.083.908.700.390.324.961.279.

şi în spaţiul astfel definit, fie curba eliptică de forma

$$y^2 = x^3 + Ax^2 + B \ (mod\ p) \tag{4.21}$$

cu A şi B alte două numere mari, atent alese, din Z_p^*. Această curbă conţine exact N puncte, unde $N =$

6.277.101.735.386.680.763.835.789.423.337.
720.473.986.773.608.255.189.015.329.

Aceste N puncte formează un grup, conform regulii prezentate anterior. Considerând p ca un număr binar, se observă că are o formă specială, $p = 2^{192} - 2^{64} - 1$, ceea ce uşurează considerabil calculul. Oricum, calculele cu numere mari nu constituie o problemă pentru sistemele moderne, cu atât mai mult cu cât dimensiunile acestor numere sunt mult mai mici decât cele utilizate de metodele tradiţionale, de tip R.S.A.