1. ALAPOK

A távközlő hálózatok és informatikai szolgáltatások tervezésének, létesítésének, működésének, működtetésének megértéséhez és a várható fejlődési irányok érzékeléséhez nélkülözhetetlenek bizonyos alapismeretek. Jelen fejezet ezen alapismeretek közül a szerkesztőbizottság és a fejezet-szerkesztő által legfontosabbnak ítélteket igyekszik áttekinteni. Bár a szóba kerülő alapismeretek valamennyien elméleti megalapozottságúak, a fejezet mégsem az „elméleti alapok" címet viseli. Ennek oka kettős. Egyrészt a terjedelmi korlátok nem teszik lehetővé, hogy a távközlő hálózatok és informatikai szolgáltatások fő elvi törvényszerűségeit a matematikai, fizikai és gazdasági alaptudományokból indulva – bár ezt azért több helyen érzékeltetve – mutassuk be, másrészt a fő törvényszerűségek önmagukban is hordoznak közvetlen útmutatást a gyakorlati alkalmazók számára.

Az „Alapok" fejezet megszületésének egy további oka az, hogy a távközlő hálózatok és informatikai szolgáltatások hatalmas témaköre nagyon nehézzé – valószínűleg lehetetlenné- teszi a teljes könyvön végigvonuló szigorúan didaktikus tárgyalást. Ha azonban az alapismereteket a könyv elején összefoglaljuk, akkor azokra a további fejezetek akkor is hivatkozhatnak, ha egy adott ismeret részletesebb kifejtésére csak egy későbbi fejezetben kerül sor.

Az „Alapok" fejezet témáinak kiválasztása néhány koncepcionális kérdést is kifejezésre kíván juttatni. Első helyen említendő talán a távközlés, a számítástechnika és a tartalom-kezelés (ipar?) konvergenciája, egyre szorosabb összefonódása. Egy másik koncepció a gazdasági, gazdaságossági szempontokkal, valamint a minőséggel és annak biztosításával bővíti a korábban „klasszikusan" műszakinak tekintett alapok tárházát. Az egyes alfejezeteken belül pedig a ma elterjedten használt eljárások mellett a jövőben nagy valószínűséggel nagyobb jelentőségre szert tevő résztémák kapták a hangsúlyt – a terjedelmi korlátok miatt – a múlt mára már elsőrendű fontosságát elveszítő témái kárára.

Fentieknek megfelelően az első hat alfejezet nagyon hasonlít egy hírközléselméleti összefoglalóra. A jelelmélet, a természetes híranyagok jellemzői, a tömörítést is szolgáló forráskódolás, a hibakorlátozást lehetővé tevő csatornakódolás, a modulációs eljárások és a hullámtan minden ilyen

összefoglalóban szerepel. E kört azonban éppen a sávszélességgel való takarékoskodás nagy gyakorlati jelentősége miatt kibővítettük a tömörítés külön alfejezetbe foglalásával. A forgalomelmélet (azzal, hogy túllép az Erlang-i világon) az adat alanyára irányuló adatvédelem és nyilvánosság, az adat alanyától és tárgyától független eljárásokkal foglalkozó adatbiztonság és az összeolvadó távközlő és számítástechnikai hálózatokat egyaránt lefedő gráfelmélet és alkalmazásai, valamint a hálózatok funkcionális rétegmodellje a konvergencia jegyében született további öt alfejezet témája. Az utolsó előtti alfejezet a távközlés gazdaságtan, az utolsó pedig a hálózat- és szolgáltatásminőség alapvetése.

Reméljük, hogy a kiváló szakértő szerzők és lektorok munkája hosszú időn át fogja segíteni a tisztelt Olvasót az információs társadalom legkarakterisztikusabb infrastruktúrájában való eligazodásban.

Dr. Gordos Géza, fejezet szerkesztő

1.1. Jelelmélet

Szerző: dr. Levendovszky János

Lektor: dr. Dallos György

A kommunikációban a jeleknek jut a fő szerep. Beszédjeleket továbbítanak a telefonhálózatok rézvezetékein, míg a mobil telefónia digitális üzeneteit a rádióhullámok térerőssége vezérelt változása közvetíti. Ugyanakkor szélessávú bitfolyamok áramlanak a számítógépek közötti optikai hálózatokon.

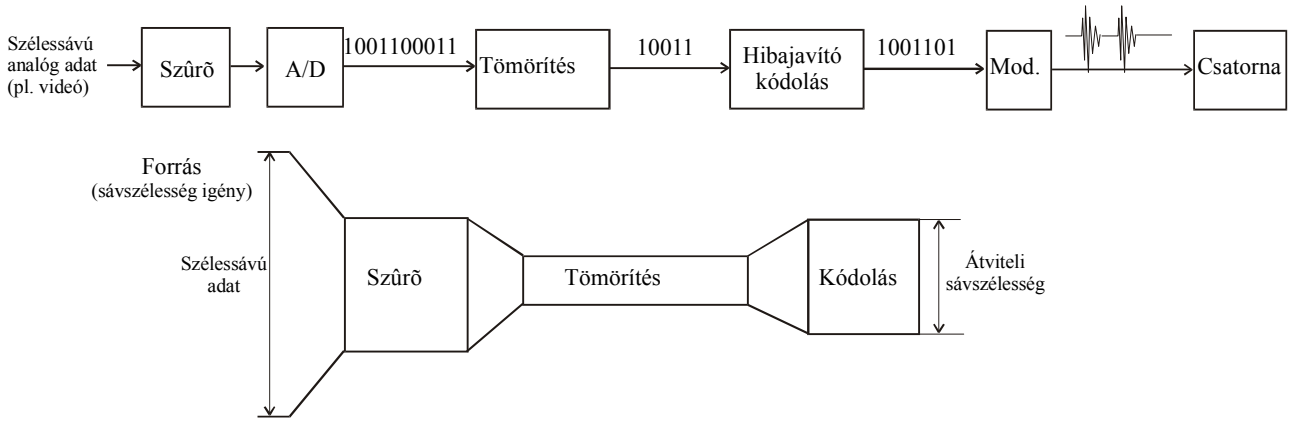
Az információ közlése, függetlenül az információ valódi természetétől (hang, kép, vagy adat) egy kommunikációs rendszer alapvető feladata, ennek része a jelek átvitele, átalakítása és feldolgozása a megkövetelt szolgáltatásminőség (Quality of Service = QoS) biztosítása végett. QoS jelöli azokat a követelményeket (pl. jel-zaj viszony, bit-hibaarány, késleltetés ...stb.), amelyeknek a megbízható információ átvitelnek meg kell felelnie.

A fenti gondolatok alapján, az információra a továbbiakban úgy tekintünk, mint egy analóg, vagy digitális adatstruktúrára (pl. analóg beszédjel, vagy digitalizált képjel). A kommunikációs rendszerekben ezen adatstruktúrákat egymásba átalakítják a következő elvek szerint:

1. meg kell találni a jelek azon formáját, amelyek inkább alkalmasak az előírt minőségű szolgáltatás megvalósítására;
2. ki kell dolgozni azokat a transzformációs algoritmusokat, amelyekkel a jeleket hatékonyan (nagy megbízhatósággal és gazdaságosan) lehet egyik reprezentációból a másikba átalakítani.

Általában a "megbízható" átvitel és kezelés megfelelést a kritériumoknak, míg "gazdaságosság" a lehető legkisebb erőforrásigényt (sávszélesség-, vagy processzorigény) jelenti.

A jelfeldolgozás, mint tudományos diszciplína a jelek leírásával, és azon absztrakt tulajdonságainak a feltárásával foglalkozik, amelyek megadják, hogy egy adott reprezentáció valamely szempontból optimális-e. Ezért szinte minden kommunikációval foglalkozó könyv, és tanulmány egy jelelmélettel foglalkozó összefoglalóval kezdődik. A jelen fejezet ezen gazdag és szerteágazó témakört



1.1.1. ábra Az információ tipikus áramlása egy digitális kommunikációs rendszerben (különös tekintettel arra, hogyan manipulálják a jelet a keskeny sávszélességű adatátvitel érdekében)

meglehetősen vázlatosan foglalja össze. A fő cél néhány klasszikus eredmény előhívása, szabatos definíciók megadása, valamint a megfelelő jelölés- és terminológiarendszer előkészítése a könyv egésze számára. Az anyag tárgyalása során feltételezzük, hogy az olvasó elemi ismeretekkel rendelkezik a matematikai analízis, valószínűségelmélet és lineáris algebra területén.

1.1.1. A jelek alapvető fogalma és a jelfeldolgozás célja

Ebben a szakaszban a jelelmélet céljait foglaljuk össze, a következő területeket érintve:

- jelek reprezentációja;
- jelfeldolgozási architektúrák;
- a jelek legfontosabb tulajdonságai egy kommunikációs rendszer számára (pl. sávszélesség, redundancia ...stb.)

Jelfeldolgozás mint a jelek reprezentációjának a kérdése

A jeleket a továbbiakban mint időfüggvényeket kezeljük, feltételezve, hogy az információs folyamatok valamilyen fizikai jellemző időfüggését tükrözik. Fizikailag a jel lehet egy időben változó feszültség egy mikrofon kimenetén, vagy diszkrét feszültségszintek időbeli egymásutánja valamilyen digitális áramkörben. (Még a térbeli állóképi információ is lefordítható időtől függő jellé a kép letapogatásával.) A jelenlegi tárgyalás során azonban ezeket a jelenségeket, a konkrét fizikai tartalomtól független időfüggvények írják le. Az időfüggvény értelmezési tartományának és

értékkészletének megfelelően a következő elnevezéseket és jelöléseket alkalmazzuk:

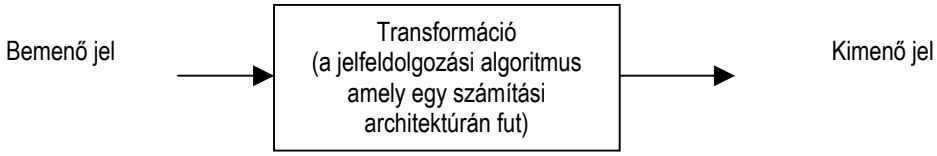
A jel típusa	Jelölés	Értelmezési tart. (idő)	Értékkészlet (amplitúdó)
Analóg jel	x(t), $0 \le t < \infty$	folytonos	folytonos
Mintavételezett analóg jelek	x_k, $k = 0,1,2...$	diszkrét	folytonos
Digitális jelek	$\hat{x}_k$, $k = 0,1,2...$	diszkrét	diszkrét
Analóg véletlen jelek (sztochasztikus folyamatok)	$\xi(t)$, $0 \le t < \infty$	folytonos	folytonos
Mintavételezett véletlen jelek (sztochasztikus sorozatok)	ξ_k, $k = 0,1,2,...$	diszkrét	folytonos
Digitális véletlen jelek	$\hat{\xi}_k$, $k = 0,1,2,...$	diszkrét	diszkrét

A kommunikáció tanulmányozása során a véletlen jelek leirása a fontos, hiszen az információ alapvetően az átvitt üzenetekben szereplő "véletlenség" mértékével azonosítható (a precíz definíció az 1.3 fejezetben található). Mivel a modern kommunikációs rendszerekben az információt bináris sorozatokban (gyakran csomagokba szabdalva) továbbítjuk, az ilyen rendszerek vizsgálatánál a következő kérdések merülnek fel:

- Hogyan lehet az eredendően analóg információt (pl. kép, zene, beszéd) digitálissá alakítani ?
- Mekkora információveszteség lép fel egy ilyen átalakítás során ?
- Milyen tartományban érdemes a jeleket leírni egy kommunikációs rendszerben, azaz milyen matematikai leírás "illeszkedik" az átvitel fizikai természetéhez (pl. az idő-, vagy frekvencia-tartománybeli leírás...stb.) ?
- Mi a jelek optimális (legrövidebb) reprezentációja (pl. hogyan lehet multimédiás adatfolyamokat keskenysávú csatornákhoz illeszteni, a legtömörebb reprezentáció kiválasztásával) ?

Jelfeldolgozás mint a számítási komplexitás kérdése

Egy jelfeldolgozási feladat a bemeneti jelnek - egy meghatározott algoritmus szerint végrehajtott - transzformációját jelenti a kimeneti jelbe (lásd 1.1.2 ábra). Ezt a



1.1.2. Ábra Jelek transzformációja egy adott számítási architektúrán

transzformációt egy adott számítási architektúrán valósítja meg (pl. egy PC CPU egysége, egy DSP, vagy egy analóg szűrő).

A következő táblázat néhány jelfeldolgozási számítási architektúrára ad példát a hozzájuk tartozó jelreprezentációval és számítási mechanizmussal együtt:

A jel típusa	Matematikai reprezentáció	Jelfeldolgozási algoritmus	Számítási architektúra	Számítási modell	Számítási idő .
analóg	idő tartomány, frekvencia tartomány,	konvolúciós integrál	analóg szűrők	Newton gép	microsec
digitális	diszkrét tartomány (pl. diszkért idő, DFT, z-transzformáció ...stb.)	formális nyelvekkel leírható szekvenciális algoritmusok analóg konnekciós algoritmusok parciális differenciálegyenletekkel leírva	CPU, DSP CNN chips	Turing gép Trajektória gép	microsec nanosec

A fenti táblázat alapján látható, hogy mind az analóg, mind a digitális jelfeldolgozás fontos szerepet tölt be a modern kommunikációs technológiákban. Annak ellenére, hogy CPU alapú eszközök (PC-k, vagy Work Station-ok), valamint DSP-k széleskörűen elterjedtek az információ digitális feldolgozására, ezen eszközök sebessége néha kritikus ponttá válhat a szélessávú kommunikáció során. Ezért a jelfeldolgozás újszerű megközelítései (konnekciós modellek, analóg számítási trajektóriák), amelyek analogikai chip-eken implementálhatók, vonzó alternatívái lehetnek a hagyományos digitális rendszereknek, ha a sebesség további növelése a cél. Ebben a fejezetben azonban csak a jelfeldolgozás klasszikus apparátusára összpontosítunk. Az analogikus számítások elméletének néhány új eredményét az olvasó Roska Tamás és Leon Chua műveiben [1.1.1,1.1.7] találja meg.

Jelfeldolgozás mint matematikai módszerek alkalmazása

A jelelmélet matematikailag szigorú, formális tárgyalása helyett, a fejezet célja a mérnöki tervezés számára fontos módszerek bemutatása. A következő táblázat azt mutatja, hogy a jelelmélet melyik diszciplináris eleme szükséges a kommunikációs hálózatok és szolgáltatások tervezéséhez.

Matematikai módszer	A mérnöki tervezéshez kapcsolódó fogalom
Fourier transzformáció	Determinisztikus jelek (analóg) sávszélessége
A sztochasztikus folyamatok négyzetes közép elmélete (gyengén stacioner folyamatok spektrális sűrűsége)	Véletlen jelek (analóg) sávszélessége
Mintavételezési tétel + kvantálás + kódolás	Digitális jelek sebessége (sávszélessége) és A/D átalakítás
Véletlen folyamatok ergodelmélete	Adaptáció, tanulás és predikció
Centrális határeloszlási tétel és Gauss folyamatok	A termikus zaj és multiplexált információs folyamok modellezése

1.1.2. A determinisztikus jelek rövid összefoglalása

A jelek alapvető megértéshez először a determinisztikus jelelmélet néhány alapvető fogalmát kell feleleveníteni. Ezután kerül sor a sztochasztikus (véletlen) jelek tárgyalására a négyzetes középelmélet keretein belül.

Az analóg jel egy időfüggvényként írható le, amelyet a továbbiakban *x(t)* jelöl. Az *x(t)* alapvető tulajdonsága, az átlaga

$$\bar{x} = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)dt$$

és az 1 ohmra eső átlagteljesitménye

$$\bar{x}^2 = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x^2(t)dt.$$

Ha $x(t)\begin{cases} \neq 0 \text{ ha } t \in [T_1, T_2] \\ = 0 \text{ egyébként} \end{cases}$ akkor a $[T_1, T_2]$ intervallumot az *x(t)* "tartójának" hívjuk.

A periodikus jelre teljesül, hogy $x(t) = x(t + kT)$, ahol $k = 0,1,2....$ és *T* a periódusidő. Ezek a jelek Fourier sorba fejthetők, az alábbi formulák alapján:

$$x(t) = \sum_{k=-\infty}^{\infty} c_k e^{jk2\pi f_0 t} \text{ ahol } c_k := \frac{1}{T} \int_{-T/2}^{T/2} x(t) e^{-jk2\pi f_0 t} dt \text{ és } f_0 = \frac{1}{T}$$

A kommunikációs csatornák átviteli képességét gyakran a sávszélességük jellemzi. Ezért, gyakorlati szempontból, az analóg jeleket néha hasznosabb leírni a Fourier transzformáltak tartományában, ahol a jel sávszélessége a jel spektruma (Fourier transzformáltja) alapján értelemszerűen megállapítható. Ha $\int_{-\infty}^{\infty} |x(t)| dt < \infty$ teljesül, akkor *x(t)*-nek létezik a következő előállítása:

$x(t) = \int_{-\infty}^{\infty} X(f) e^{j2\pi ft} df$, ahol $X(f) = \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft} dt$ a jel Fourier transzformáltja (spektruma).

Bevezetvén a $s = \sigma + j\omega$ komplex frekvenciatartományt, a "belépő" jelekre ($x(t) = 0$ ha $t \le 0$), a jel Laplace transzformáltja is definiálható a következőképpen:

$$X(s) = \int_{0}^{\infty} x(t) e^{-st} dt .$$

(A Laplace transzformáció ereje abban rejlik, hogy minden függvénynek, amelyre $\int_{-\infty}^{\infty} \left| x(t) e^{-\sigma t} \right| dt < \infty \bigcap \sigma \ge 0$ teljesül létezik Laplace transzformáltja, míg a Fourier transzformált létezése a szigorúbb $\int_{-\infty}^{\infty} |x(t)| dt < \infty$ feltétel kielégülését követeli meg.) A Laplace transzformáció inverzét a következő integrál definiálja: $f(t) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} F(s) e^{st} ds$.

A gyakorlatban a reziduum tétel alkalmazásával számos módszer létezik az inverz Laplace transzformált algebrai meghatározására a komplex vonalintegrál nehézkes kiértékelése helyett (az érdekelt olvasó bővebb részleteket talál Proakis munkájában [1.1.5]).

A szűrés az analóg jelek lineáris invariáns transzformációja. Az analóg jelfeldolgozás alapvető eszköze a szűrés, amely általában a jelek spektrális formálását jelenti. Ha a szűrő impulzusválasz függvényét *h(t)* jelöli, akkor a lineáris transzformáció a szűrő kimeneti és bemeneti jele között egy konvolúciós integrállal adható meg $y(t) = \int_{-\infty}^{\infty} h(\tau) x(t-\tau) d\tau$ vagy a frekvenciatartományban $Y(f) = H(f) X(f)$ (ahol *X(f), Y(f)* és *H(f)* a megfelelő időfüggvények Fourier transzformáltjai). A $H(f)$ függvény jellegétől függően alul-, felül- vagy sáváteresztő szűrést valósít meg.

A továbbiakban a Mintavételezett jelek leírását és reprezentációit ismertetjük. Ahogy a fentiekben már jeleztük, a modern kommunikációs rendszerekben az információt nem analóg formában továbbítják. Azért, hogy a digitális jelfeldolgozás teljes apparátusa bevethető legyen az analóg jelet digitalizálni kell. A digitalizálás első lépése, hogy az időben folytonos jelet mintavételezzük és csak a megfelelő, $t_k = kT + t_0$ időpillanatokban vett mintákkal foglalkozunk. Itt t_0 a mintavételezés kezdőpillanatát jelöli, míg *T* a mintavételi periódus. Az általánosság korlátozása nélkül feltételezhetjük, hogy $t_0 = 0$ és $T = 1$, ami jelölések szintjén annyit jelent, hogy az *x(t)* analóg jelet az x_k mintasorozattal helyettesítjük.

Természetesen a fenti mintavételezés pusztán matematikai absztrakció, amely "végtelenül gyors" kapcsolókat igényelne. A valódi mintákat mindig egy véges hosszúságú mintavételi időtartam (semmint egy mintavételi időpillanat) alatt kapjuk, amíg a mintavételező kapcsoló zárt állapotban van. Ez az $\sum_k x_k h(t - kT)$,jelet eredményezi, ahol $h(t)$ a mintavételező áramkör által determinált időfüggvény. Az $\sum_k x_k h(t - kT)$ jel azonban úgy is felfogható, mint egy $h(t)$ impulzusválaszú szűrő kimeneti jele az $\sum_k x_k \delta(t - kT)$ bemeneti jelre. (Itt $\delta(t)$ a jól ismert Dirac impulzus függvény, a $\int_{-\infty}^{\infty} \delta(t)dt = 1$ tulajdonsággal). Ezért egy valódi mintavételező rendszer esetén a mintavett jel másik matematikai absztrakciója az $\tilde{x}(t) = \sum_k x_k \delta(t - kT)$ jel.

Az x_k sorozat Diszkért Fourier Transzformációja (DFT) a következőképen van definiálva: $\tilde{X}(f) = T\sum_k x_k e^{j2\pi kfT}$ amelyből $x_k = \frac{1}{2\pi}\int_0^{1/T} X(f)e^{-j2\pi kfT}df$.

(A fenti definíció alapján könnyű észrevenni, hogy az x_k minták az $\widetilde{X}(f)$ függvénynek a Fourier együtthatói. Az $\widetilde{X}(f)$ függvény a frekvenciának periodikus függvénye 1/T periódushosszal). Ezek alapján levezethető (lásd [1.1.5, 1.1.6]), hogy

$$\widetilde{X}(f) = \sum_{k=-\infty}^{\infty} X\left(f + \frac{k}{T}\right),$$

amiből direkt módon következik a híres mintavételezési tétel:

Ha az x(t) jel X(f) spektruma sávkorlátolt a (-B,B) intervallumban és $2B \le \frac{1}{T}$, akkor x(t) teljesen előállítható az x_k mintákból. Az $f_0 = \frac{1}{T} = 2B$ frekvenciát gyakran Nyquist frekvenciának hívják.

Az Nyquist frekvencia kitüntetett szerepe annak köszönhető, ezen frekvencia feletti mintavételezési frekvenciát használva ($f_{sampling} > f_0$) igaz, hogy $\widetilde{X}(f) = X(f)$ az $f \in (-B, B)$ intervallumon, amiből valóban következik, hogy *X(f)* előállítható $\widetilde{X}(f)$-ből egy sima aluláteresztő szűrő segítségével, amelynek sávszélessége *B*.

Vegyük észre azonban, hogy ha $2B > \frac{1}{T}$ akkor $\widetilde{X}(f) \neq X(f)$ az $f \in (-B, B)$ intervallumon. Ennek alapján, ha a mintavételi frekvencia kisebb mint a Nyquist ráta, akkor $\widetilde{X}(f)$ az átlapolódott spektrumokat tartalmazza, azaz *X(f)* nem rekonstruálható $\widetilde{X}(f)$-ből.

(A gyakorlatban - azért, hogy implementálható aluláteresztő szűrő kerüljön megvalósításra - a szigorúbb $2B < \frac{1}{T}$ feltételt szokták alkalmazni, ahol az $f \in \left(B, \frac{1}{T} - B\right)$ szakasz a szűrő átviteli karakterisztikájának a véges meredekségéhez illesztendő.)

A fentiek alapján megállapítható, hogy a mintavételezés nem okoz információveszteséget, feltéve, ha sávhatárolt folyamatot mintavételezünk, a Nyquist frekvenciánál nagyobb mintavételezési frekvenciával.

1.1.3. Kvantálás

Az analóg jelek digitalizálásnak a mintavételezés után következő lépése a kvantálás. Ez végső soron a minták amplitúdójának a diszkretizálását jelenti. Ennek végeredményeként a jel mind időben, mind amplitúdóban diszkrétté válik ezért megfelelő hosszúságú bináris sorozatokkal reprezentálható. Így viszont a jelek további transzformációjára a bináris adatstruktúrákon definiálható teljes algoritmikus "fegyvertár" bevethető, ami a mai nagysebességű DSP technológiában igen tág horizontokat jelent.

Mindazonáltal, a kvantálás során az eredetileg "értékben" folytonos amplitúdók kerekítése helyrehozhatatlan hibát eredményez az analóg információn. Ezt a jelenséget kvantálási zajnak hívják és ez az ára a jel digitalizálásnak. Amennyiben a kvantálási zaj kis értékű marad az ebből származó hiba elhanyagolható. A kvantálási zaj vizsgálatához feltételezzük, hogy a jel lehetséges x_k mintái az *X* intervallumba esnek. Az *X* intervallumot lefedjük egy diszkrét rácshálózattal, amelyet az

$\hat{X} = \{C_{-M}, C_{-M+1}, ..., C_{-1}, 0, C_1, ..., C_{M-1}, C_M\}$ diszkrét halmaz jelöl. A kvantálás annyit jelent, hogy a mintákat a legközelebbi társ-szabály alapján kerekítjük a rács által meghatározott értékekre. Pontosabban az *x* mintát helyettesítjük a C_i értékkel, ahol $i = \arg\min_n |x - C_n|$.

(A fentiekben már elhagytuk a mintavételi időpont kurrens értékére vonatkozó *k* indexet az x_k-ből, hiszen a kvantálást minden mintán végrehajtjuk a *k*-tól függetlenül).

A kvantálási zaj (vagy más nevein kvantálási hiba, illetve torzítás) definíciója $\varepsilon := x - C_i$ ha az *x* éppen a C_i kvantálási értékhez esett a legközelebb. A kvantálás minőségét a kvantálási jel-zaj viszony határozza meg, amelynek definíciója: $SNR = \frac{E(x^2)}{E(\varepsilon^2)}$, ahol *E* a várhatóértéket jelöli, hiszen a mintavett jel amplitúdója, és a kvantálási zaj is véletlen mennyiségek.

A fenti formula a jel átlagos energiáját hasonlítja a kvantálási zaj átlagos energiájához. Az alapvető kérdés, hogyan válasszuk meg a kvantálási szinteket (az $\hat{X}$ halmazt), hogy a lehető legkisebb információ-veszteség lépjen föl. Ez ekvivalens

azzal, hogy a kvantálási jel-zaj viszonyt maximalizáljuk. Ahhoz, hogy a feladatot precízen is meg tudjuk fogalmazni, feltételezzük, hogy az $\hat{X}$ halmaz szintjei egy egyenletes szintű kvantálást jellemző $\tilde{\hat{X}}$ halmaz szintejeinek (ahol $\tilde{\hat{X}}$-ben a kvantálási szintek azonos távolságra vannak egymástól) egy *f(u)* függvény szerinti transzformációja alapján adódnak.

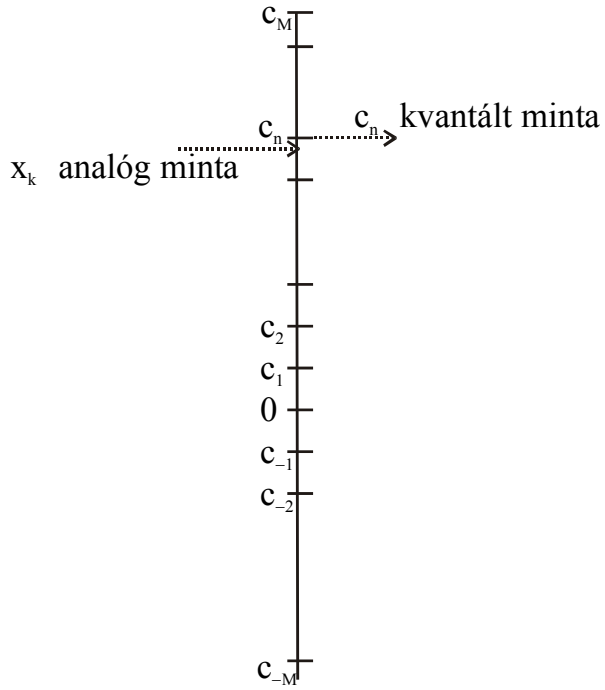
Precízebben $C_i = f(i\Delta)$ és $\tilde{\hat{X}} = \{-M\Delta, (-M+1)\Delta, ..., -\Delta, 0, \Delta, ..., (M-1)\Delta, M\Delta\}$, ahol Δ az egyenletes szintű kvantálás lépésköze, míg $\hat{X} = \{f(-M\Delta), f((-M+1)\Delta), ..., f(-\Delta), f(0), f(\Delta), ..., f((M-1)\Delta), f(M\Delta)\}$.

Jelölje *SNR(f)* azt a kvantálási jel-zaj viszonyt, amely az *f* függvényhez tartozik (különbözően választva az *f* függvényt más-más jel-zaj viszonyhoz juthatunk). Ennek megfelelően az optimális kvantálás a következő feladat megoldását jelenti:

$f_{opt} : \max_f SNR(f)$, vagy alternatívan fogalmazva

$$f_{opt} : \min_f \sum_{i=-M}^{M} E\left((x - C_i)^2 \middle| i = \arg\min_n |x - C_n| \right) p_i \text{, ahol } p_i := P(i\Delta \le x < (i+1)\Delta).$$

Mielőtt az általános feladatot oldanánk meg, határozzuk meg az egyenletes szintű kvantálás jel-zaj viszonyát!



1.1.4. Ábra A minták egyenletes kvantálása

Feltételezve, hogy a kvantálási zaj ε egyenletes eloszlást követ

$$E\left(\varepsilon^2\right)=\int_{-\Delta/2}^{\Delta/2} x^2 \frac{1}{\Delta}dx,$$

ugyanakkor a jel átlagos energiája $C_M^2/2$, C_M amplitúdójú szinuszos jelet feltételezve. Ez $SNR=6(C/\Delta)^2$ jel-zaj viszonyt eredményez, ami az $n:=\log_2(2M+1)$ jelöléssel $SNR=\frac{3}{2}2^{2n}$ jel-zaj viszony ad. Az utóbbi formula kifejezi azt a fontos kapcsolatot, amely a kvantálási szintek bitben kifejezett száma valamint a kvantálás minősége közt van.

A nem egyenszintű kvantálás általánosabb esetét a következőképpen vizsgáljuk:

Meg akarjuk találni azt az *y=f(x)* függvényt (amit gyakran kompressziós karakterisztikának hívnak), amely maximalizálja a jel-zaj viszonyt. Feltételezvén, hogy Δ elegendően kicsiny, $C_i-C_{i-1}\approx\frac{\Delta}{f^{(-1)}{}'(i\Delta)}=\frac{2}{Nf^{(-1)}{}'(i\Delta)}$ (ahol $f^{(-1)'}(x)$ az inverz függvény deriváltját jelöli). Az átlagos kvantálási zajenergia az *i*-ik intervallumban $E\left(\varepsilon^2\middle|x\in(C_{i-1},C_i)\right)=\frac{1}{3N^2\left(\frac{df(x)}{dx}\right)^2}$, ami következő jel-zaj viszonyt eredményezi

$SNR(f)=\frac{3N^2\int_{-1}^{1}p(x)x^2dx}{\int_{-1}^{1}p(x)\left(\frac{df^{-1}(x)}{dx}\right)^2dx}$. Ennek megfelelően az optimális kvantálás, egy jól-meghatározott optimalizálási (variációszámítási) feladat megoldást jelenti:

$$f_{opt}:\min_f\frac{3N^2\int_{-1}^{1}p(x)x^2dx}{\int_{-1}^{1}p(x)\left(\frac{df^{-1}(x)}{dx}\right)^2dx}.$$

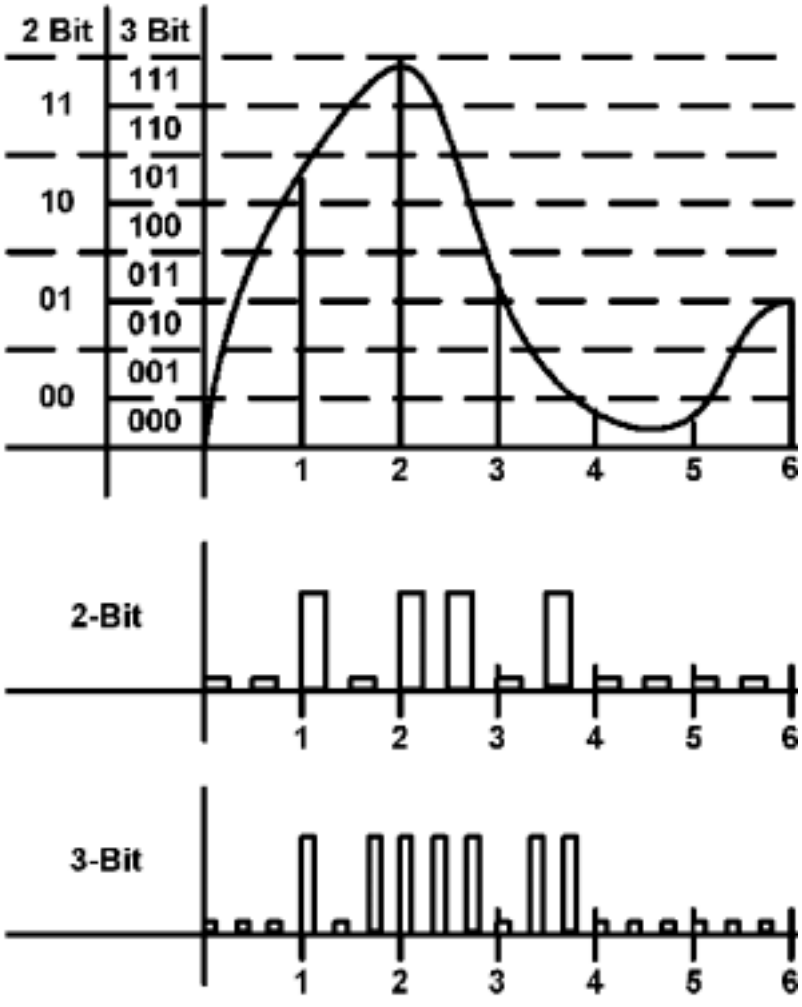
Sajnos f_{opt} analitikus meghatározása nehézkes. Ugyanakkor a cél gyakran azonos SNR biztosítása a bejövő jel szintjétől függetlenül (kb. ugyanolyan minőségű legyen a kvantált jel, ha az analóg jel max amplitúdója mindössze a fele a maximális kivezérelhetőségnek). Ekkor könnyű levezetni, hogy az optimális kompressziós

függvény (vagy másnéven karakterisztika) $f(x) \approx \log x$. Ez gyakran használt karakterisztika a beszéd hírközlésben (lásd [1.1.5]).

1.1.4. A mintavételezés és kvantálás egy szabványos megvalósítása - a PCM

A PCM (Pulse Code Modulation) először a hetvenes években terjedt el a beszédjelek digitalizálására, majd maga a rendszer az időosztásos multiplexálási hierarchiák alapkövévé vált. A rendszer jelenlegi specifikációját az ITU-T G.711 szabvány írja le.

A PCM rendszer megvalósításának első lépése az analóg beszédjel szűrése a 3.4KHz feletti komponensek fokozatos elnyomása érdekében (a beszéd jó érthetőségű és hanghűségű átviteléhez elég a kb. 3.4KHz-ig terjedő sáv átvitele).



1.1.5. Ábra. Analóg jel átlakítása digitális impulzussorozattá

Ezután a jel mintavételezése következik 8kHz frekvenciával, amit egy egyenletes lépésközű, 8 bites kvantáló követ (a kvantálási szintek száma $2^8 = 256$). Ezért a PCM jel sebessége 64 Kbit/sec. Egy analóg jel digitalizálását (2 és 3 bites kvantálás szerint) az 1.1.5. ábra mutatja.

Mivel az egyeneletes lépésközű kvantálás nem jól használja ki a beszédjel statisztikai tulajdonságait (a kvantálási jel-zaj viszony értéke nem a legjobb), ezért a nemlináris kvantálást hajtanak végre. A nemlineáris kvantálás megvalósításának az a módja, hogy a jelet előtorzítják („összenyomják"), majd a vételi oldalon, a digitál-analóg átalakítás során, helyreállítjak egy inverz nemlinearitással („kitágítják"). E két szó (kompresszió és dekompresszió) ötvözetéből jött létre a gyakran használt „kompandálás".

Az általános megfontolásoknak megfelelően a kompressziós karakterisztika logaritmikus. Az Európában elterjedt kompressziós karakterisztikát „A-szabálynak" hívják és alakja a következő:

$$y = \frac{1 + \log(Ax)}{1 + \log(A)} \text{ ha } \frac{1}{A} < x \le 1 \text{ vagy } y = \frac{Ax}{1 + \log(A)} \text{ ha } 0 \le x \le \frac{1}{A}$$

Az Egyesült Államokban és Japánban a „μ-szabályt" alkalmazzák, a következő függvény szerint:

$$y = \frac{\ln(1 \pm \mu x)}{\ln(1 + \mu)}.$$

+ ha a jel pozitív
- ha a jel negatív

Az általánosan használt értékek $A = 87.6$ és $\mu = 1000$.

A kvantálás során a jelamplitúdó-tartományát 16 szegmensre bontják (8 pozitív és 8 negatív szegmens). Mindegyik szegmens az előző szegmens kétszerese. Mindegyik szegmensben egyenletes a kvantálás. A kvantálási szintekhez tartozó 8-bites kódszavak első bitje reprezentálja az amplitúdó előjelét, a másidok, harmadik és negyedik bit identifikálja melyik szegmensben esik éppen a minta, végül az utolsó 4 bit jelöli ki az adott a szegmensbeli kvantálási szint értékét.

Az „A-szabály" nagyobb dinamikus tartományt biztosít mint a „μ-szabály", ugyanakkor a „μ-szabály" kis szintű jelekre jobb kvantálási jel-zaj viszonyt ad. A nemzetközi távbeszélő kapcsolatok felépíteskor természtesen meg kell oldani az $A - \mu$ konverziót (ez egyezmény szerint ez a „μ országok" felelőssége).

A PCM-en alapuló beszédkódolás továbbfejlesztése a DPCM, ADPCM, valamint Delta Moduláció részletes leírása az 1.4-es alfejezetben találhatók.

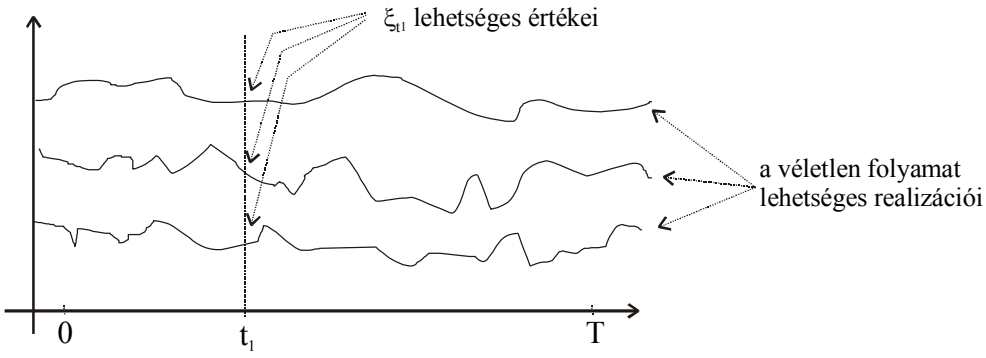
1.1.5. Véletlen folyamatok

Az előzőek szerint az információ a jelekben fellelhető "véletlenséggel" van kapcsolatban, hiszen egy előre ismert jel átvitele semmilyen plusz tudást nem eredményez a vételi oldalon. Ezért a véletlen, vagy más néven sztochasztikus folyamatok leírása központi szerepet játszik a kommunikáció elméletében. Az alapvető cél itt is az, hogy néhány mérnökileg jól definiált jellemzőt (sávszélesség, átlagos energia ...stb.) kapjunk, ami lehetővé teszi információátviteli rendszerek tervezését.

Ahhoz, hogy a véletlen folyamatok fogalmát egyelőre intuitíven megalapozzuk, vissza kell térni az elemi valószínűségszámításhoz. Emlékezvén arra, hogy a kockadobás eredménye egy valószínűségi változó amely a {1,2,3,4,5,6} halmazból veszi fel az értékeit, valaki azt mondhatná, hogy a sztochasztikus folyamat egy "általánosabb" valószínűségi változó, amely nem egy számot, hanem egy függvényt ad valamilyen belső, véletlen sorsolási mechanizmus eredményeként. Pl. a hőmérséklet alakulása az idő függvényében egy véletlen függvény. Sajnos a sztochasztikus folyamatoknak mint "véletlen függvényeknek" a leírása matematikailag nehezen kezelhető apparátushoz vezet. Ezért a sztochasztikus folyamatot, inkább mint valószínűségi változók egy seregét fogjuk értelmezni, a következőképpen (lásd még [1.1.3], [1.1.6]):

Definíció szerint, egy véletlen folyamat a valószínűségi változóknak egy halmaza $\{\xi_t, t \in [0,T]\}$, ahol egy adott $t_1 \in [0.T]$ időpillanatban az ehhez tartozó ξ_{t_1} valószínűségi változó kifejezi a folyamat helyettesítési értékének (amplitúdójának) a t_1-ben mutatott véletlenségét.

Ennek alapján a ξ_{t_1} valószínűségi változót gyakran a $\{\xi_t, t \in [0,T]\}$ folyamatnak a t_1 időpillanatban vett egydimenziós projekciójának hívják.



1.1.6. ábra. Egy véletlen folyamat különböző realizációi a t_1 időpillanatban vett egydimenziós projekcióval, mint valószínűségi változóval

Hasonlóképpen az egydimenzióshoz projekcióhoz, be lehet vezetni az $(\xi_{t_1}, \xi_{t_2}, ..., \xi_{t_n})$ többdimenziós projekciókat, amelyek a $(t_1, t_2, ..., t_n)$ időpontokhoz tartozó vektor értékű valószínűségi változókat jelentik. A végesdimenziós projekció eloszlásfüggvényének a definíciója a következő:

$F_{t_1,t_2,...,t_n}(x_1, x_2, ..., x_n) = P(\xi_{t_1} < x_1, \xi_{t_2} < x_2, ..., \xi_{t_n} < x_n)$, illetve az ebből nyerhető valószínűségsűrűségfüggvény $f_{t_1,t_2,...,t_n}(x_1, x_2, ..., x_n) = \dfrac{\partial^n F_{t_1,t_2,...,t_n}(x_1, x_2, ..., x_n)}{\partial x_1 \partial x_2 ... \partial x_n}$.

A sztochasztikus folyamatok gyakorlati leírására használhatók az első és másodrendű statisztikák (ezek természetesen nem olyan statisztikai mélységben jellemzik a folyamatot, mint a végesdimenziós projekciók eloszlásai):

$$m(t) := \int_{-\infty}^{\infty} x f_t(x) dx \text{ és } R(t_1, t_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 f_{t_1 t_2}(x_1, x_2) dx_1 dx_2$$

A fentiekből látható, hogy a véges dimenziós projekciók fontossága abban rejlik, hogy végesdimenziós valószínűségi vektorváltozókról lévén szó, a klasszikus valószínűségszámítás apparátusával (pl. eloszlásfüggvény) leírhatóak. Ha sztochasztikus folyamatokat véletlen függvényekként értelmeztük volna, akkor ezeket a fogalmakat a realizációk, mint függvények egy terén kellett volna bevezetnünk, ami komplikált leíráshoz vezet (további részletekért lásd [1.1.3], [1.1.2]).

A sztochasztikus folyamatok egyik legfontosabb alosztályát a *stacionárius* folyamatok képezik. A stacionaritás a statisztikai jellemzők időfüggetlenségével (időbeli eltolásra való invarianciájával) van kapcsolatban. Pl. ha valaki egy adott időpillanatban ismeri a folyamat egydimenziós projekciójának az eloszlását, akkor

vajon milyen hosszú ideig marad "érvényes" ez az ismeret ? Időben ugyanaz marad-e, vagy megváltozik a folyamat viselkedését meghatározó "véletlen sorsolási mechanizmus" ? Egy stacioner folyamatnál érvényes, hogy amit statisztikailag ismerünk a folyamatról egy adott időpillanatban az igaz lesz bármilyen jövőbeli időpillanatban. (Pl. egy kockadobás eredménye mindig egyenletes eloszlást követ, függetlenül attól, hogy vasárnap 5 órakor, vagy szerdán 10 órakor végezzük el a kísérletet).

A stacionaritás fogalma attól is függ, hogy a folyamatot milyen mélységig jellemző statisztikai jellemzőre teljesül az időbeli eltolással szembeni invariancia. Ebben két szintet különböztetünk meg:

- *gyenge stacionaritás* amikor csak a várhatóérték és a korrelációs függvény időeltolással szembeni invarianciája teljesül;
- *erős stacionaritás* amikor az összes véges dimenziós projekció valószínűség-eloszláfüggvénye invariáns az időbeli eltolásra.

A fentiek alapján, egy $\xi(t)$ folyamat gyengén stacioner ha

$m(t) = m(t+\tau) \quad \forall \tau > 0$ és $R(t_1, t_2) = R(t_1 + \tau, t_2 + \tau) \quad \forall \tau > 0$, amiből a következő tulajdonságok vezethetők le:

$m(t) = \text{constans}$ és $R(t_1, t_2) = R(t_1 - t_2) = R(\tau)$.

Egy folyamat erősen stacionárius ha $F_{t_1, t_2, \ldots, t_n}(x_1, x_2, \ldots, x_n) = F_{t_1+\tau, t_2+\tau, \ldots, t_n+\tau}(x_1, x_2, \ldots, x_n) \quad \forall \tau > 0$.

Vegyük észre, hogy az erős stacionaritás maga után vonja a gyenge stacionaritást is (a várhatóérték- és korrelációfüggvények időeltolással szembeni invarianciája könnyen levezethető az eloszlásfüggvények időeltolással szemben mutatott invarianciájából). Ugyanakkor az állítás megfordítása általában nem igaz.

Az erősen staiconáris folyamatokon belül fontos alosztályt képeznek az ergodikus folyamatok.

Egy folyamat statisztikájának a felderítéshez, valaki megfigyeléseket végezhet az idő folyamán (különböző időpillanatokban mintavételezvén a folyamatot), majd ezen minták alapján időbeli átlagokkal próbálja közelíteni a folyamat statisztikai jellemzőit. Pl. az egydimenziós projekció alapján a várhatóértékét úgy próbálja megállapítani, hogy a folyamat tíz különböző időpillanatban megfigyelt értékét

átlagolja. Ez az eljárás a következő fundamentális kérdést veti fel: mennyire megbízhatók az időátlagolással kapott becslései a folyamat valódi statisztikájának? A kérdés hasonló ahhoz a kísérlethez, mikor valaki egyszerre dob fel tízezer kétforintost, vagy időben egymásután tízezerszer dob fel egy darab kétforintost azért, hogy empirikusan határozza meg a fej vagy írás valószínűségi változó várhatóértékét (vagy bármilyen más statisztikai jellemzőjét, pl. eloszlását). Van-e bármilyen különbség a két kísérlet között ? (Azt leszámítva, hogy az első kísérlethez jóval nagyobb "tőkeberuházás" kell, mint a másodikhoz.) . A választ az ergodicitás fogalma adja meg. Egy sztochasztikus folyamat ergodikus ha szinte bármilyen *g* függvényre (pontosabban bármilyen *g* Borel mérhető függvényre, lásd [1.1.2], [1.1.3]) az időbeli és statisztikai átlagok azonosak, azaz:

$$\lim_{T\to\infty}\frac{1}{T}\int_{-\infty}^{\infty} g\big(x(t)\big)dt = E\big(g(\xi_t)\big)$$

Mérnöki szempontból a *g* függvénynek két konkrét választása érdekes. Mikor *g* az identitás függvény, az maga után vonja, hogy

$$\lim_{T\to\infty}\frac{1}{T}\int_{-\infty}^{\infty} x(t)dt = E\big(\xi_t\big).$$

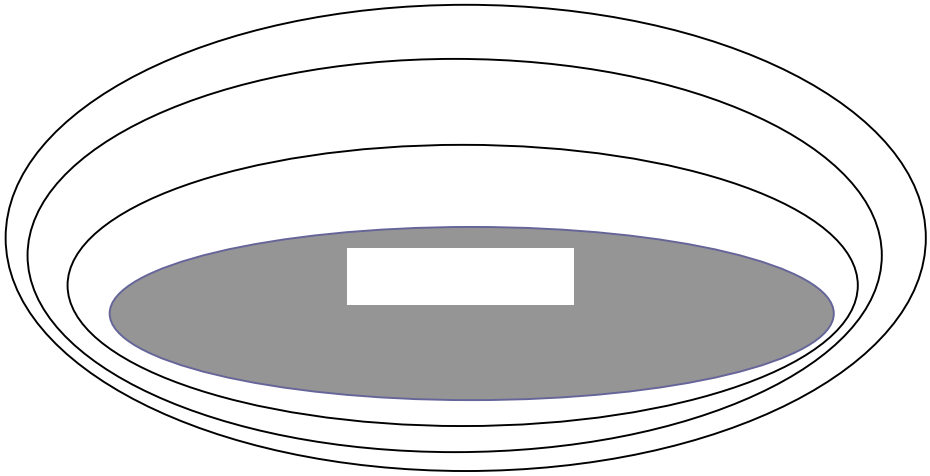
Mikor *g* a négyzetes függvény, az maga után vonja, hogy

$$\lim_{T\to\infty}\frac{1}{T}\int_{-\infty}^{\infty} x^2(t)dt = E\big(\xi_t^2\big).$$

Ez annyit jelent, hogy a folyamat első és másodrendű statisztikája időben egymásutáni megfigyelésekből és átlagolásokkal rekonstruálható.

Vegyük észre, hogy $\lim_{T\to\infty}\frac{1}{T}\int_{-\infty}^{\infty} x^2(t)dt = E\big(\xi_t^2\big) = R(0)$, ami annyit jelent, hogy ergodikus folyamatok esetén a korrelációs függvény nulla pontban felvett értéke megadja a folyamat átlagos teljesítményét.

Könnyű megmutatni, hogy az ergodikus folyamatok erősen stacionáriusak, ugyanakkor a gyengén stacionárus folyamatok osztálya tartalmazza az erősen stacionárius folyamatok osztályát. Ezt a kapcsolatot az 1.1.7 ábra demonstrálja.



1.1.6. Gyengén stacionárius folyamatok spektrális sűrűsége és sávszélessége

Ahhoz, hogy megvizsgáljuk, hogyan oszlik el egy folyamat energiája a frekvenciatartományban, bevezetjük a spektrális sűrűségfüggvényt (amire sok szerző a *teljesítménysűrűség spektrum* elnevezést is használja), mint a korrelációs függvény Fourier transzformáltját:

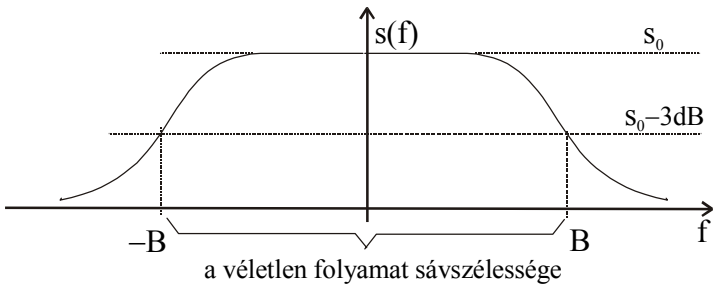
$$s(f) = \int_{-\infty}^{\infty} R(\tau) e^{-j2\pi f\tau} d\tau \,.$$

Ez valóban meghatározza az energia "spektrális eloszlását" hiszen

$\lim_{T\to\infty} \frac{1}{T} \int_{-\infty}^{\infty} x^2(t)dt = E\left(\xi_t^2\right) = R(0)$ ami inverz Fourier transzformálttal $\lim_{T\to\infty} \frac{1}{T} \int_{-\infty}^{\infty} x^2(t)dt = R(0) = \int_{-\infty}^{\infty} s(f)df$. Ez megmutatja a teljesítmény és a spektrális sűrűség kapcsolatát.

A spektrális sűrűség alapján lehetőség van véletlen folyamatok sávszélességét is definiálni, azon meggondolás alapján, hogy nincs szükség a

véletlen folyamat átvitelére azokon a frekvenciákon, ahol az energiája zérus. Ezért a sávszélesség az *s(f)* függvény szerint pl. az 1.1.8. ábra alapján definiálható.



1.1.8. ábra. Egy sávkorlátozott véletlen folyamat sávszélessége

1.1.7. Véletlen folyamatok lineáris invariáns transzformációja (szűrése)

Ha egy $\xi(t)$ véletlen folyamatot egy *h(t)* impulzusválaszú szűrőn bocsátunk keresztül, akkor a kimeneti jelet a következő összefüggés adja meg:

$$\eta(t) = \int_{-\infty}^{\infty} h(t-\tau)\xi(\tau)d\tau\,.$$

Könnyű bizonyítani, hogy a kimeneti és bementi folyamat spektrális sűrűsége közötti összefüggés a következő:

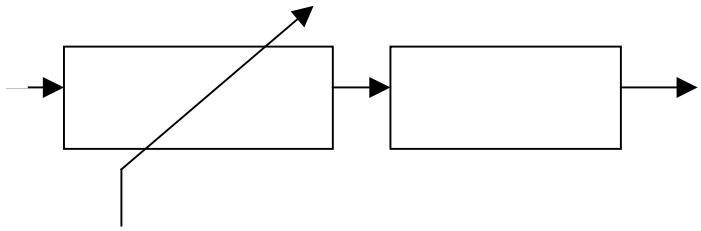
$s_\eta(f) = |H(f)|^2 s_\xi(f)$, ahol *H(f)* a szűrő frekvencia karakterisztikája ($H(f) = \int_{-\infty}^{\infty} h(t)e^{-j2\pi ft}dt$). Ezen összefüggés alapján egy ergodikus véletlen folyamat spektrális sűrűsége mérhetővé válik, ha bevezetjük a tűszűrő fogalmát:

$$H(f) = \begin{cases} 1 \text{ ha } f_0 - \dfrac{\Delta}{2} \le |f| \le f_0 + \dfrac{\Delta}{2}, \\ 0 \text{ máskülönben} \end{cases}$$

ahol Δ egy infinitezimálisan kicsi mennyiség. Ha a tűszűrő $\eta(t)$ kimeneti folyamatának (aminek az éppen mért realizációját *y(t)* jelöli), mérjük az átlagteljesítményét, ez arányos lesz a bementi folyamat spektrális sűrűségfüggvényének az f_0 helyen vett értékével:

$$\lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2}y^2(t)dt = E\left(\eta_0^2\right) = R_\eta(0) = \int_{-\infty}^{\infty}s_\eta(f)df = \int_{-\infty}^{\infty}\left|H^2(f)\right|s_\xi(f)df = 2\int_{f_0-\frac{\Delta}{2}}^{f_0+\frac{\Delta}{2}}s_\xi(f)df \approx 2\Delta s_\xi(f_0)$$

Ez a formula azt vonja maga után, hogy egy tűszűrővel és "átlagteljesítménymérővel" felszerelkezve, egy ergodikus folyamat spektrális sűrűségfüggvénye mérhető. A mérési elrendezést az 1.1.9. ábra mutatja.



$s_\xi(f)$-ből a korrelációs függvény inverz Fourier transzformációval meghatározható, tehát minden eredmény adott, hogy a folyamatot "négyzetes közép szintjén" jellemezzük.

1.1.8. Gaussi folyamatok és a fehérzaj

Hasonlóan az elemi valószínűségszámításhoz, a gaussi (vagy más néven normális) folyamatok főszerepet játszanak a sztochasztikus folyamatok elméletében. Ezek fontossága kettős:

- egy gaussi folyamatot teljesen leír a várhatóérték- és korrelációs függvénye.
- a centrális határeloszlási tétel általánosítása nyomán, sok, egymáshoz hasonló értékű, független véletlen folyamat összege a gaussi folyamathoz tart.

Egy $\xi(t)$ folyamat gaussi folyamat, ha minden $(t_1,t_2,...,t_n)$ n=1,2,.... végesdimenziós időbeli mintavektorhoz tartozó $\overline{\xi}=\left(\xi_{t_1},\xi_{t_2},...,\xi_{t_n}\right)$ projekció mint valószínűségi vektorváltozó normális eloszlást követ a következő valószínűség-sűrűségfüggvénnyel:

$$f\left(x_1,x_2,...,x_n\right)=\frac{1}{\sqrt{(2\pi)^n\,\det(\mathbf{K})}}e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^{\mathrm{T}}\mathbf{K}^{-1}(\mathbf{x}-\mathbf{m})},$$

ahol **K** jelöli a kovarianciamátrixot, amelynek elemei $K_{ij} = E(\xi_{t_i}\xi_{t_j}) - E(\xi_{t_i})E(\xi_{t_j})$, valamint **m** a várhatóértékvektort.

Ez egyúttal maga után vonja, hogy $K_{ij} = R(t_i, t_j) - m(t_i)m(t_j)$. Azaz egy gaussi véletlen folyamatot valóban teljesen meghatároz a várhatóérték- és korrelációs függvénye. A várhatóérték- és korrelációs függvényeknek a valószínűség-sűrűségfüggvényekkel való kapcsoltából következik, ahogy amelyik gaussi folyamat gyengén stacionárius az egyúttal erősen is stacionárius (általános folyamat esetén ez nem feltétlenül igaz).

A fehér gaussi folyamat (amelyet gyakran fehérzajnak hívnak) egy olyan speciális gaussi folyamat amelynek a spektrális sűrűsége állandó a frekvencia függvényében, azaz $s(f) = const, \quad f \in (-\infty, \infty)$. Ezért ezen folyamat korrelációs függvénye a Dirac deltával arányos, $R(\tau) = const\,\delta(\tau)$. Ezért a fehér gaussi folyamat mintái korrelálatlanok. A fehér gaussi folyamat inkább egy hasznos matematikai absztrakció, semmint fizikai valóság (hiszen a folyamat átlagos teljesítménye végtelen). Ugyanakkor egy szűrt (színes) gaussi folyamat tulajdonságai könnyen számíthatók a fehér gaussi folyamat alapján.

A gaussi folyamatok fontossága a centrális határeloszlási tételben rejlik, hiszen sok, független folyamat eredője gaussi folyamathoz tart, ezért ez a folyamat jó approximációja a termikus zajnak (amikor sok független részecske véletlen hőmozgásának az eredője okozza a zajt). Sőt, nagyszámú információs folyamat multiplexálásaként kapott eredő folyam (pl. multiplexált beszédfolyamok) szintén jól közelíthetőek gaussi folyamattal. Ez megmagyarázza, hogy a gaussi folyamatoknak miért jut főszerep a manapság divatos "traffic engineering"-ben (a távközlő hálózatok forgalmi leírásában és méretezésében).

1.1.9. Összefoglaló

Ebben az alfejezetben a jelek fő tulajdonságait és a jelelmélet egyes elemeit mutattuk be. A fő hangsúly az analóg információ digitálissá való átalakításán volt, valamint az információs folyamatok (véletlen jelek) leírásán. Ennek megfelelően a stacioner és ergodelmélet rövid taglalására is kitértünk.

Irodalomjegyzék

[1.1.1] Chua, L.O. and Roska, T.: "The CNN Paradigm", *IEEE Trans. on Circuits and Systems*, Vol. 40., March, 1993.

[1.1.2] Doob, J. *Stochastic Processes*, New York-London, 1953

[1.1.3] Gihman, I., Szkorohod, A: *Introduction to stochastic processes*, Technical Press, Budapest 1975 (in Hungarian)

[1.1.4] Haykin, S.: Neural networks - a comprehensive foundation, Prentice-Hall, 1999

[1.1.5] Proakis, J, Manolakis, D..: *Digital signal processing*, McGraw-Hill, 1996

[1.1.6] Papoulis, A.: Probability, random variables and stochastic processes, McGraw-Hill, 1984

[1.1.7] Roska, T. and Chua, L.O.: "The CNN Universal Machine: an Analogic Array Computer", *IEEE Trans. on Circuits and Systems*, Vol. 40., March, 1993.

1.2. Híranyagok jellemzői

Szerző: dr.Levendovszky János

Lektor: dr.Dallos György

A kommunikációs rendszerek tervezésének alapvető célja az információ megbízható közlése. A megbízható átvitel általában előírt késleltetési korlát, hibaarány, jel-zaj viszony betartását jelenti. Egy gazdaságosan működő rendszer tervezésénél a cél az, hogy előírt QoS paraméterekkel működő összeköttetéseket a lehető legkisebb erőforrásigénnyel valósítsuk meg. Ehhez a valós források információs folyamainak (hang, kép, adat) tanulmányozása szükséges. A vizsgálat célja, hogy felfedjük mekkora erőforrásigény jelentkezik az egyes híranyagok előírt minőségű átvitele esetén, ahol az erőforrásigény, pl. a szükséges sávszélességben testesül meg.

Az ilyen vizsgálatok jelentőségét az integrált szolgáltatások elterjedése is alátámasztja, hiszen ebben az esetben sokfajta, különböző szolgáltatás számára kell egyazon hálózaton belül hatékony sávszélesség-gazdálkodást biztosítani.

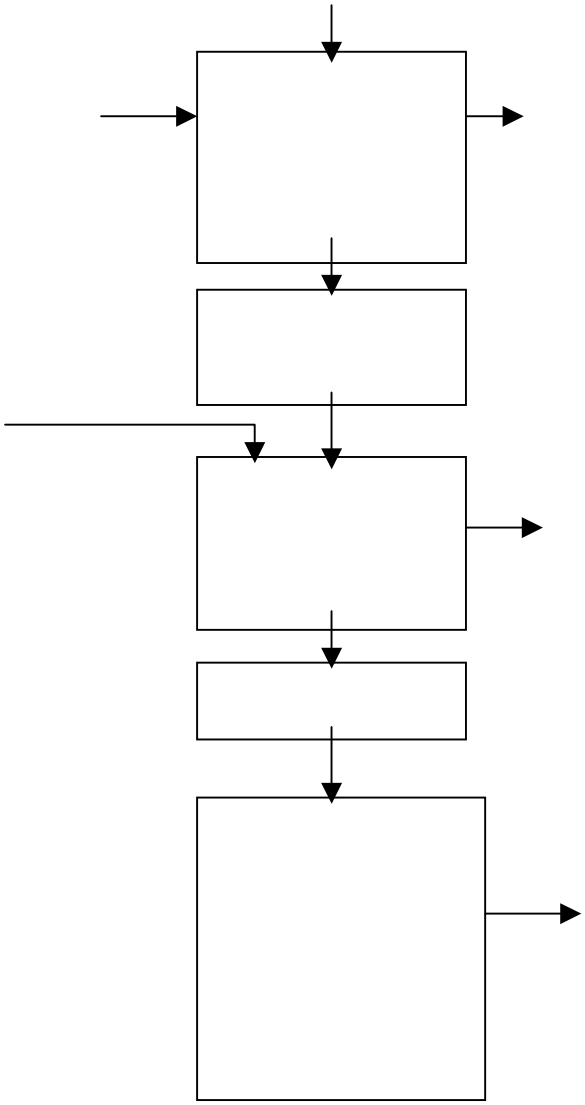
A források modellezése szempontjából két esetet különböztetünk meg: (i) analóg források (beszéd, zene, kép) modellezése, amikor az információközlés végállomása valamelyik emberi érzékszerv (fül, vagy szem); (ii) adatfolyamok leírása (pl. fájlok, elektornikus levelek ..stb.), amikor az információ már eredetileg is digitális formában keletkezett. Ugyanakkor meg kell említeni, hogy a "konvergencia" korszakában (amikor a hagyományosan különálló távközlési és műsorszórási világ az adatátviteli világhoz közeledik) gyakorlatilag minden információ "adatnak" minősül. Ezért az információs forrásokat a csomagkapcsolt hálózatok platformján, mint forgalmi folyamokat kell modellezni, ahol az átvitt információ minőségét rá vonatkozó QoS követelmények határozzák meg.

Analóg források esetén az átvitt információ "élvezhetősége" a legfontosabb paraméter. Az "élvezhetőség" azonban szubjektív kritérium, amelynek kvantifikálása, csak kísérletileg és statisztikai átlagok alapján történhet. Ugyanakkor a mérnöki tervezés számára fontos paraméterek azok, amelyek kifejezik egy adott információs folyam erőforrásigényét. Ezért a való életbeli folyamatok leírásánál az a feladat, hogy

a szubjektív élvezhetőséget, mérnökileg szabatosan definiált paraméterekbe transzformáljuk. Ennek érdekében a továbbiakban röviden ki kell térnünk az emberi érzékszervek leírására is.

Ahhoz, hogy egy csomagkapcsolt hálózatban adott minőségű szolgáltatást tudjunk nyújtani, a beszéd-, zene-, kép- és adatforrások, mint "véletlen csomagkibocsátó egységek" statisztikai tulajdonságait kell áttekinteni. Pontosabban a statisztikus multiplexálás szempontjából fontos paraméterek, a statisztikus sávszélesség, teljesítményeloszlás felfedése a cél.

A fenti gondolatoknak megfelelően, a híranyagok jellemzőinek tárgyalása, a következő, rétegelt modell szerint történik:

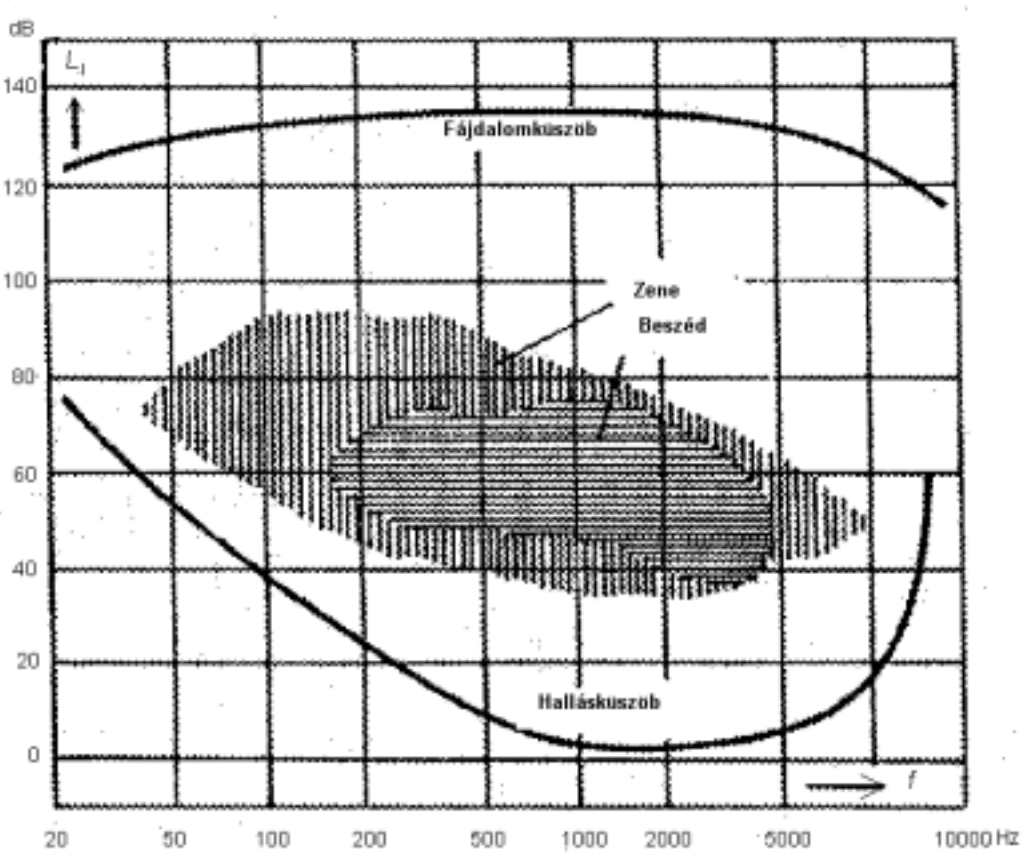


1.2.1. Audio jelek (beszéd és zene)

Ebben a pontban a beszéd és zene jeleket írjuk le, sávszélesség, jel-zaj viszony és egyéb kritériumok alapján.

Érzékszervi leírás

A beszéd- és zeneátvitel karakterizálásához először az emberi hallás rövid leírására van szükség. A fül „frekvenciakarakterisztikája”, azaz a hallás szinuszos hangokra értelmezett két határtulajdonsága, a beszéd és a zene tipikus spektrumával együtt az 1.2.1. ábrán látható:



A fenti ábra alapján a beszéd (20 Hz, 5000 Hz) míg a zene a (10 Hz, 20 KHz) frekvenciasávba esik.

A beszédátvitel minőségi követelményei

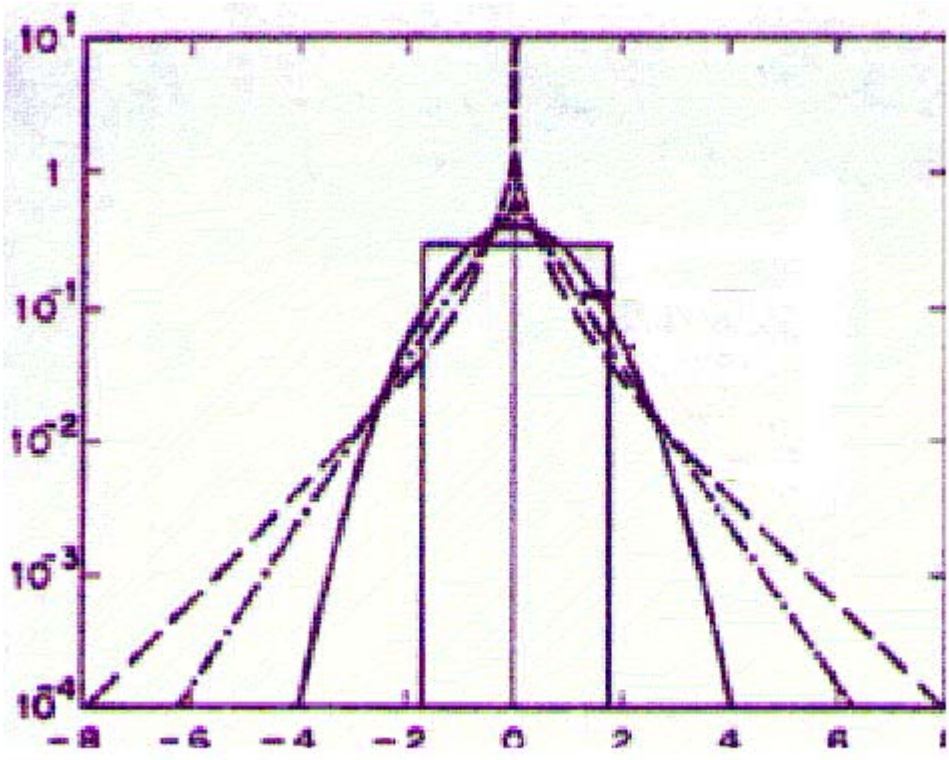
Az analóg beszédátvitelt a frekvenciasávon p_a kívül a jel-zaj viszony is minősíti, melynek határértéke a klasszikus telefóniában, a halk beszédnél 20 - 25 dB.

Egy másik alapvető jellemző a tolerálható nemlineáris torzítás mértéke. A nemlineáris torzítás egy tiszta szinuszos jellel gerjesztett nemlineáris elem (pl. szénmikrofon) kimenetén megjelenő felharmonikus tartalommal mérhető, a következő módon:

$k = \frac{\sqrt{\sum_{i=2}^{\infty} c_i^2}}{\sqrt{c_1^2}}$, ahol c_1 az f_0 frekvenciájú alapharmonikus együtthatója, míg c_i az if_0, *i=2,3....* frekvenciához tartozó felharmonikus amplitúdó. Az analóg műsorszórásban a *k*=1%, amíg az analóg telefóniában *k* =5-10 a megengedett érték. Az érték attól függ, hogy a torzítás termékek a hallható tartományba esik-e.

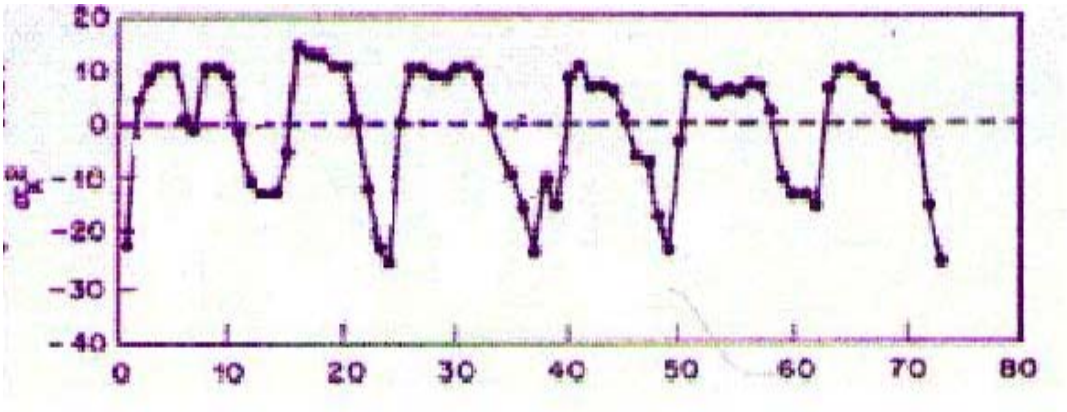
A beszédjel statisztikai tulajdonságai

A beszédjel statisztikájának a feltárására rengeteg kísérletet végeztek, amit az



optimális hangkódolási eljárások kifejlesztése motivált. A beszédjelek nemstacionárius véletlen folyamatok, ezért a rövidtávú statisztikák (amelyeket általában 32msec hosszúságú időintervallumon vesznek fel) alapvetően különböznek a hosszútávú statisztikától (amelyet 6400 msec hosszúságú időintevallummon számolnak). A legtöbb kutató a hossztávú statisztikát Laplace eloszlással közelítette, amelynek sűrűségfüggvénye: $p(x) = \frac{1}{\sqrt{2}\sigma} e^{-\sqrt{2}|x|/\sigma}$, ahol σ a szórás (gyakorlatilag az effektív érték, lásd [1.2.1], [1.2.9]).

A modellezéshez használt sűrűségfüggvények alakját a következő ábra mutatja: A beszédjel szórásának az időfüggését az 1.2.3. ábra szemlélteti (lásd [1.2.2]).



A digitális beszédjel jel-zaj viszonyát az $SNR = 10 \lg \frac{\sigma_x^2}{\sigma_z^2}$ összefüggés definiálja, ahol σ_x az analóg beszédminták szórását, míg σ_z a kvantált beszédminták szórását jelöli.

Digitális telefónia

A szabványos digitális telefóniában (PCM rendszerek) a beszédjel sávszélességét (300 Hz, 3400 Hz)-ben határozták meg. A PCM kódoló 8 KHz-es mintavételi frekvenciát használ, amit egy 8 bites logaritmikus kvantáló követ. Ezért a beszéd által igényelt digitális sávszélesség 64 Kb/s. Természetesen ez a sávszélesség a kódoló algoritmikus komplexitásától is függ. (Pl.a vokóder sokkal kisebb sávszélességben tud jó minőségű beszédátvitelt garantálni, azonban ennek

ára a sokkal bonyolultabb kódolási és dekódolási eljárás.) A PCM rendszerekben a jel-zaj viszony minden bit hozzáadásával közel 6 dB-vel javul, pontosabban $SNR = 6n - 10\lg a$, ahol *a* egy empirikus konstans 1<*a*<10, valamint *n* a kvantálásnál használt bináris kódszó hossza (lásd [1.2.2]).

A CCITT 30 csatornás "*A*-törvényű" multiplex rendszerben, mindegyik TDM keret 32 csatornát tartalmaz (az időrések 0-tól 31-ig számozottak, amiből 30 csatorna beszédet, kettő pedig jelzéseket visz át). Mindegyik csatorna időrés 8 bitet tartalmaz, amely 125microsec/32=3.9 microsec hosszú.

Az IP alapú ás egyéb digitális hangátviteli rendszerek jelenlegi kódolói és dekódolóit a G.711, G.723, G.726, G.729, G.728-el jelölt ITU ajánlások definiálják. Az alkalmazott tömörítéstől függően a beszéd sávszélessége a 6.4-64 Kb/s tartományban változik. Az alacsonyabb sávszélesség elérése csak nagyteljesítményű DSP-kel lehetséges.

A mobil telefóniában a GSM 06.10 transzkódolás használatos Európa szerte. Ez 13 kb/s-os sávszélességet ér el hosszú távú predikciót is használva.

A zene leírása

A zene hasonló műszaki paraméterekkel jellemezhető, mint a beszédjelé. Az idevágó számértékeket a következő táblázat tartalmazza:

A zene paraméterei	
Sávszélesség	15 KHz (FM rádió), 20 kHz (CD)
Jel-zaj viszony	40 dB (FM rádió) , 96 dB (CD)
Nemlineáris torzítás	1% (FM rádió), 0.005% (CD)
Mintavételezési frekvencia	44,1 KHz (CD, MPEG)
Kvantáló	16 bit
Digitális sávszélesség	705.6 Kb/s (mono), 1.411 Mb/s (sztereo)

1.2.2. A képi információ jellemzése

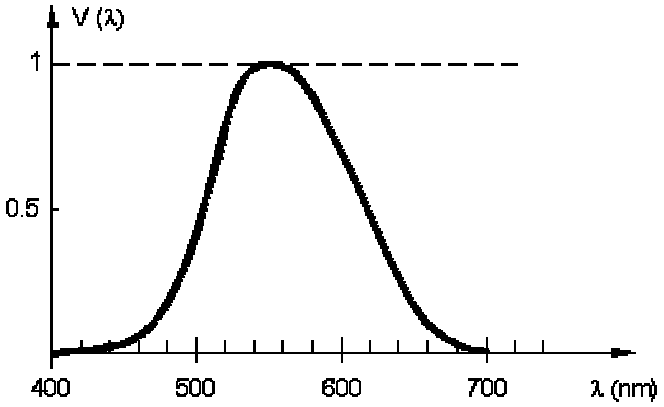
Ez a szakasz a hanghoz hasonlóan a képi inofrmáció bemutatására törekszik.

A vizuális érzékelés

A videojel jellemzéséhez az emberi látás néhány tulajdonságát kell összefoglalnunk. Az emberi szem a fényt a 400 nm-től a 700 nm-ig terjedő

tartományban érzékeli. A látás érzékenységét ebben a tartományban az 1.2.4. ábra szemlélteti.

Ugyanakkor a kísérletek azt bizonyították, hogy a színérzékelés során elegendő olyan színeket figyelembe venni, amelyek a három alapszín, pl. a vörös ($\lambda_R = 700\,nm$), a zöld ($\lambda_G = 564.1\,nm$) és a kék ($\lambda_B = 435.8\,nm$) kombinációjaként állnak elő. Ezt a fajta előállítást, amely az adott színt a benne szereplő vörös, zöld és kék intenzitásaival, egy $\mathbf{x} = (x_R, x_G, x_B)$ háromdimenziós vektorként jellemzi az angol nyelvű rövidítés alapján gyakran RGB reprezentációnak nevezzük.



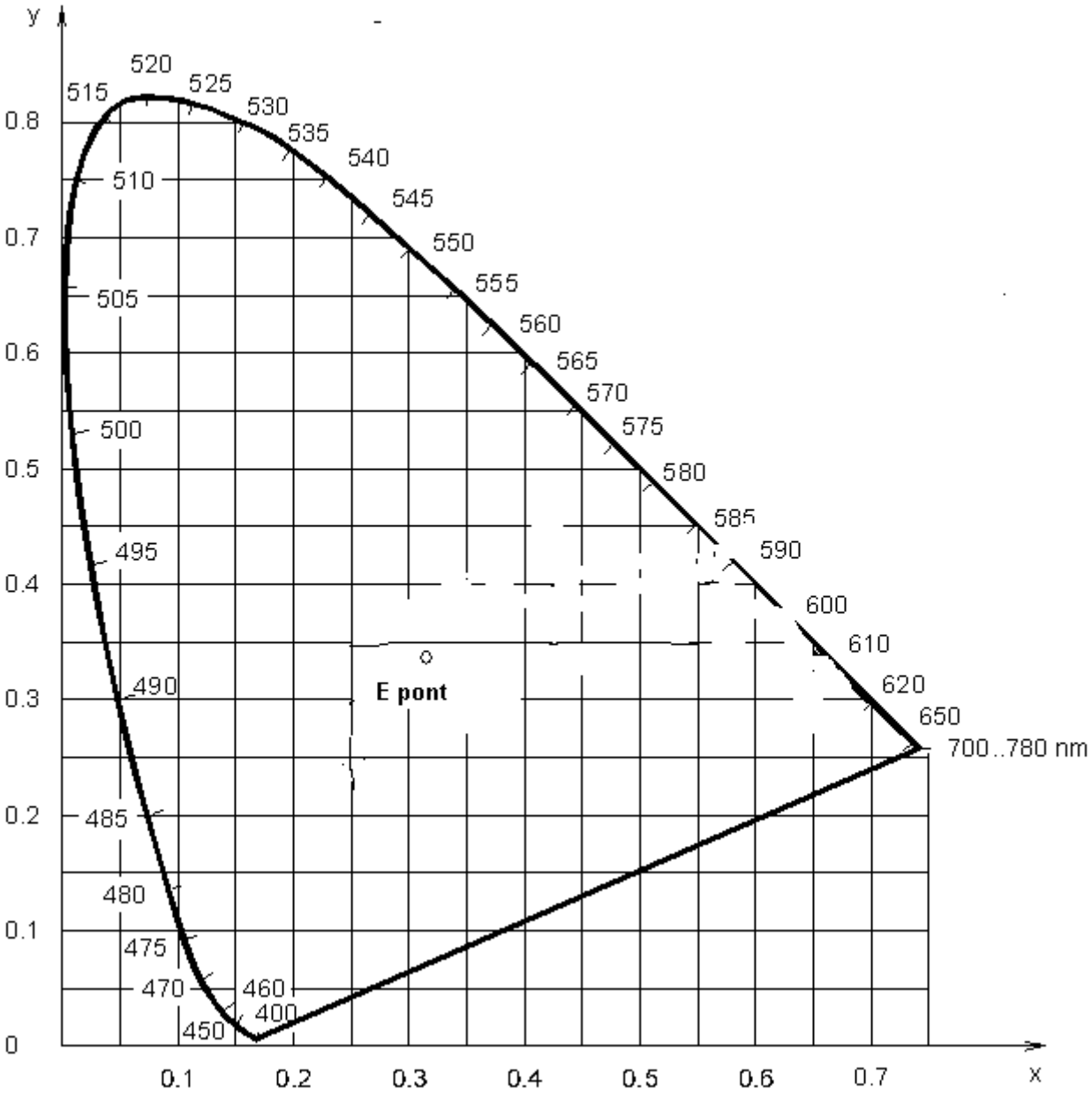
A háromdimenziós leírás egyszerűsítése végett a színteret egy kétdimenziós diagrammba képezzük le megfelelő koordináta-transzformációkkal, amelynek végeredménye az ún. Színpatkó (1.2.5. ábra). A színptakó görbe része reprezentálja a spektrális színeket, amíg a lezáró egyenes az úgynevezett "bíboregyenes" a vörös és kék kombinációjából kiadódó színeket tartalmazza. Az *E* fehér pont koordinátái a színpatkó belsejében (x=0.33 y=0.33) találhatóak. A színpatkó belső pontjai olyan színek, amelyek előállíthatóak a fehér szín és a megfelelő spektrálszín keverékéből. Az adott színhez tartozó spketrálszín úgy áll elő, hogy az adott színnek megfelelő pontot és a fehér pontot összekötő egyenest a színpatkóra vetítjük.

A televízió jel

A televíziójel leírásánál figyelembe vesszük, hogy az emberi szem egy 4:3 arányú téglalap alakú képre tud optimálisan fókuszálni, függőlegesen kb. 20 fokos szögben. Mivel a szem felbontási képessége 2', ezért a képnek 800x600=480000

képpontot kell tartalmaznia. A fekete-fehér televíziójel ezen képpontok világosságértékének jobbról-balra történő soronkénti leolvasását tartalmazza, másodpercenként 25 képsebességgel (az ilyen frekvenciával egymásután vetített állóképeket a szem mozgóképként érzékeli). Ennek ellenére a televízió 50 Hz-el dolgozik, először a páros, majd a páratlan sorokat megjelenítve a képernyőn. Az amplitúdók aránya alapján a fekete-fehér kép, kb. 70% videó-információt és 30% szinkron információt tartalmaz. A színes tévé esetén az Y=0.3R+0.59G+0.11B összefüggéssel definiált világosság jel és a *R-Y* és *B-Y*, színkülönbségi jel hordozza.

Történetileg három szabványa alakult ki a színes televíziónak: az NTSC, PAL



és a SECAM. Az NTSC a színjelet QAM segítségével viszi át, ezért ez nagyon érzékeny a fázistorzításokra. A SECAM figyelembe veszi az erősen korrelált színjelet és ezért csak egy színkülönbségi jelet továbbít soronként (a másik az előző sorhoz tartozik és a memóriából előhívandó). Ezért nincs szükség QAM-re, hiszen csak egy színkülönbségi jelet kell adott időben frekvenciamodulációval átvinni. A PAL rendszer váltogatja a vörös jel fázisát, amely úgy javítja meg a QAM jel spektrumát, hogy kevésbé lesz érzékeny a fázistorzításokra.

Napjainkban nagyfokú érdeklődés mutatkozik a nagyfelbontású digitális televíziók (HDTV= High Definition TeleVision) iránt. Ez a letapogatott sorok majdnem kétszeresére növelésével élesebb képet nyújt. A képméretek aránya 16:9, amely jobban illeszkedik a mozifilmekhez. A HDTV adaptív kódolási technikákat alkalmaz (nagyobb felbontásban kódolja a kép állórészeit, míg kisebb felbontásban a kép mozgórészeit). A sávszélesség jobb kihasználása érdekében egy mozgásvektort visz át a mozgás helyett (részletesebb leírás az [1.2.8]-ban található).

A digitális képjel sávszélessége

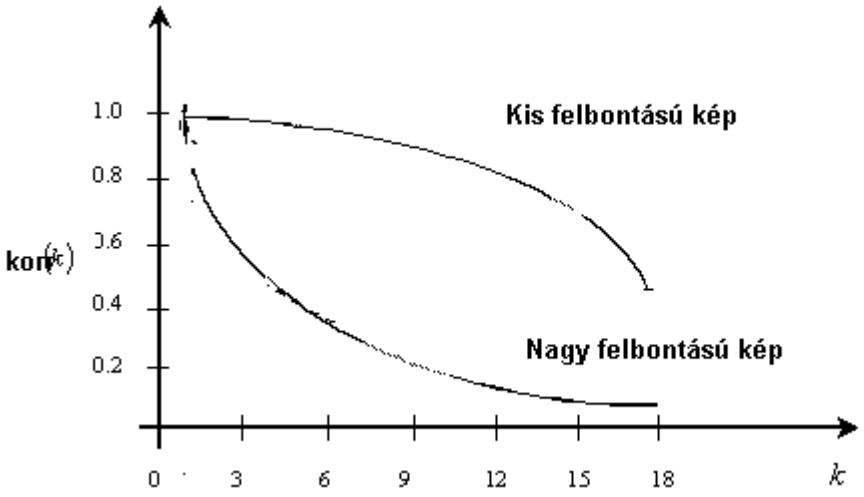
A kép 800x600=48,0000 képelemet tartalmaz, képpontokhoz tartozó világosságértéket 100 szintben kvantálva. Ezért egy fekete-fehér kép 48,0000 ld. 100 = 3,19 Mbit információnak felel meg. Mivel az emberi szem 25 Hz-es frekvenciájú állóképsorozatokat mozgóképpé átlagol, ezért a fekete-fehér mozgókép átviteléhez szükséges digitális sávszélesség: $3{,}19 * 10^6 * 25 \approx 80 Mbps$.

Színes képek átvitele esetén csak a színkülönbségi jeleket kell még átvinni, hiszen a világosságjelet már tartalmazza a fekete-fehér képjel. Ráadásul az emberi szem színfelbontása rosszabb mint a világosság-felbontása, ezért elegendő a fekete-fehér képhez képest egy- ötödös felbontást alkalmazni, ami 166 Kbit többlet információt jelent színes állókép továbbítása esetén. A színes mozgóképek estén a szükséges sávszélesség másodpercenként: 25*(3,19 Mbit+0,166 Mbit)=84 Mbps. Ebből látszik, hogy a színinformáció átvitele csak alig növeli a képjel sávszélességét.

A televízión kívül a számítógép-képernyőn megjelenő kép továbbítása isszükséges. A számítógépes képekre a következő szabványok érvényesek: 640x480 VGA, 800x600 SVGA, 1024x768 XGA. Egy XGA képernyő 24 bit per képelem felbontással és 25 keret per másodperc átvitelével 472 Mb/s-os sávszélességet igényel.

A képjelek statisztikai tulajdonságai

A képjelek statisztikai leírásánál felmerül a "hosszúidejű függőség" jelensége, amely a korrelációs függvény viszonylagosan nagy értékeire utal, még időben távoli mintapárok esetén is. Ezen jelenség a keretekben ismétlődő ***képminták*** jelenlétével magyarázható. Ezért a videojelek kváziperiódikusak a keretfrekvencia szerint. Néhány tipikus korreláció látható az 1.2.6. ábrán:



A videojelek statisztikájáról további információk az [1.2.4], [1.2.5] művekben találhatók.

1.2.3. Az adatforrások modellezése

A modern telekommunikációs rendszerekben minden információt digitális formában visznek át. Ezért fontos, hogy a forrásokat vagy mint állandó sávszélességű adatfolyamokat modellezzük a vonalkapcsolt hálózatok számára, vagy mint véletlen csomagsorozatok a csomagkapcsolt hálózatok számára.

Forrásmodellezés vonalkapcsolt hálózatokhoz

A vonalkapcsolt hálózatok esetében bizonyos erőforrások a teljes hívás alatt a rendelkezésre állnak (pl. egy telefonhívás esetén állandó kapcsolat épül ki a hívó és hívott között), ezért a legtöbb szolgáltatás állandó sávszélességet biztosít. Ennek

megfelelően az egyes források által kibocsátott információs folyamatokat a sávszélességükkel és időtartamukkal lehet jellemezni.

Az 1.2.7. ábra néhány tipikus forrásmodellt mutat sávszélesség és időtartam szerint (néha forrásmodell helyett a "szolgáltatás" terminológiát használják, hiszen a szolgáltatások adott típusú forrásokhoz kapcsolódnak). [1.2.6.]

Forrásmodellezés csomagkapcsolt hálózatokban

Ha az információt bitekből összeálló csomagok formájában továbbítjuk, akkor a forrásmodellezésnek tovább kell lépnie és véletlenszerű csomagfolyamot generáló forrást kell figyelembe vennie. Az ilyen modelleket forgalmi modelleknek nevezzük. Ezek fontosságát az is tanúsítja, hogy a forgalmi modellezés lassan önálló diszciplínává növi ki magát, egyesítve a véletlen folyamatok, sorálláselmélet és a káoszelmélet eredményeit. Ebben a szekcióban néhány elemi modellt vezetünk be a videó, hang és adatfolyamok csomagszintű leírására. Az itteni fogalmak mélyebb és kiterjedtebb tárgyalása az 1.7. alfejezetben található.

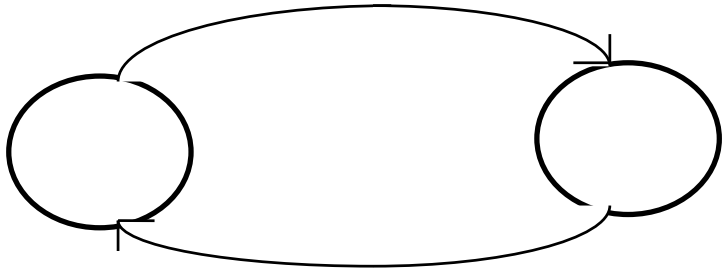
Az információs folyamatok csomagszintű modellezésére különböző mélységű modellek születtek, amelyek Markov Modulált Poisson Folyamatokon, fraktálanalízisen, vagy egyéb módszeren alapulnak. Jelen esetben két alapvető modellel foglalkozunk:

1. On/Off modell a zérus puffer közelítés esetén (ebben az esetben a csomagkapcsolt hálózat úgy van dimenzionálva, hogy a pufferek hossza relatíve elhanyagolható, a valós idejű szolgáltatások támogatása végett). A modellezés alapja, hogy a forrás egy *X* valószínűségi változó, amely $p_i = P(X = i)$, $i = 0,...,h$ valószínűségeloszlása szerint generál időegység alatt *i* db. csomagot. Jelölje $m = \sum_{i=1}^{h} ip_i$ az *X* forrás működését jellemző várhatóértéket. Az "On/Off" modellezés annyit jelent, hogy az eredeti forrást egy kétértékű $\widetilde{X}$ valószínűségi változóval helyettesítjük, amelynek eloszlása $P(\widetilde{X} = 0) = 1 - \frac{m}{h}$ és $P(\widetilde{X} = h) = \frac{m}{h}$. Ily módon $\widetilde{X}$ az úgynevezett "börsztös" (vagy On/Off) megfelelője *X*-nek, amely vagy nulla, vagy csúcssebességgel ad.

2. Markov vezérelt folyamatok, amelyek az állapotaikkal, az adott állapotokban való tartózkodási idejükkel és az állapot-átmeneti valószínűségeikkel vannak leírva.

Az On/Off modellre a hang esetén a tipikus paraméterek m= 10 Kb/s és h=64 Kb/s. A hangforrások statisztikai multiplexálása ezért binomiális eloszlást követ. Érdemes még megjegyezni, hogy az On/Off források statisztikai sávszélessége (amelyet Kelly vezetett be először annak mérésére, hogy mekkora véletlen terhelést jelent egy On/Off forrás a hálózat számára), a $\mu(s) = \frac{1}{s}\log\left(1 - \frac{m}{h} + \frac{m}{h}e^{sh}\right)$ összefüggéssel van megadva, ahol *s* a Chernoff határ optimalizálásból származó paraméter (lásd [1.2.3]). Ez a koncepció azon alapul, hogy a multiplexált forgalom "farokeloszlását" a log moment generáló függvényekkel becsüljük felülről.

A beszédet modellező Markov vezérelt folyamat a következő:

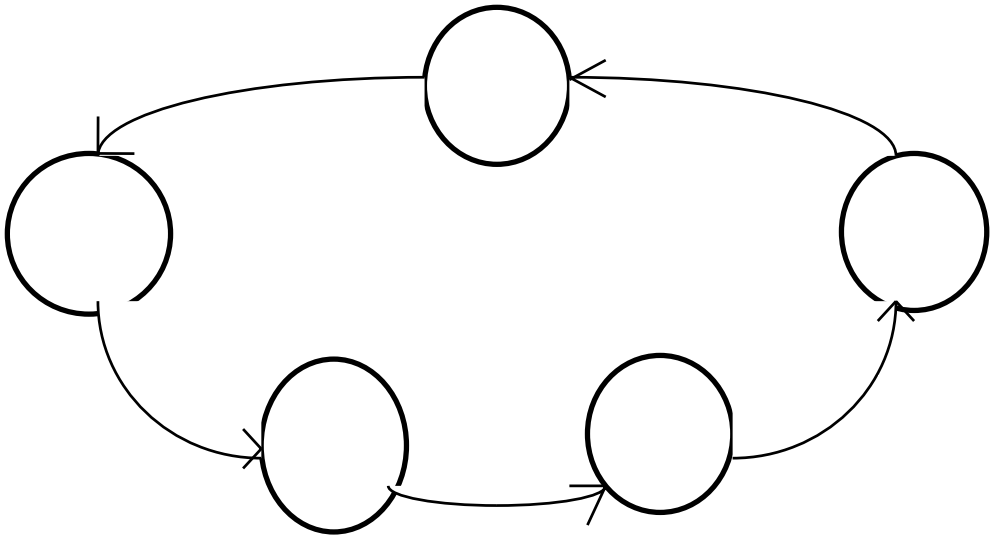


A videoforrás szintén modellezhető On/Off forrásként, habár ez nem tükrözi a képjel nagyfokú korreláltságát. A megfelelő On/Off modell paraméterei a következőek: m = 2,2121 Mbit/s és h = 11,02 Mbit/s.

Egy másik megközelítése a videó forrásoknak Markov Modulált Poisson Folyamatokon alapul, ami szintén alkalmatlan a hosszúidejű függőség reprodukálásra. Ezért a csomagkapcsolt hálózatok forgalmi modellezésének egyik jelenlegi kihívása a videó, illetve multiplexált videó folyamok hiteles, analítikus leírása.

Az ún. gyors-adat modell, amely *ftp*-s folyamatok modellezésére került kifejlesztésre a következő paraméterekkel bír: m = 0,4356 Mbit/s, h =11,01 Mbit/s

A Markov vezérelt modell (lásd [1.2.7]) a következő:



1.2.4. Összefoglaló

Ebben a fejezetben az információs források tulajdonságait vizsgáltuk. A bevezetőben vázolt rétegelt megközelítés alapján először az analóg beszéd, zene és képjel jellemzőit taglaltuk, amit a digitális beszéd-, kép- és adatfolyamok leírása követett. Végül a fenti folyamatokat csomagok szintjén modelleztük, amelyek a való életbeli források forgalmi statisztikáját írják le.

Irodalomjegyzék

[1.2.1] Flanagan, J.L.: Speech analysis, synthesis and perception, Springer-Verlag, 1972

[1.2.2] Jayant, N.S. , Noll, P.: Digital coding of waveforms, Prentice Hall, 1984

[1.2.3] Kelly, F.: "Notes on effective bandwidths", *www.statslab.cam.ac.uk./~frank/*

[1.2.4] Kretzmer, E. R.: "Statistics of Television signals", Bell System Tech. J., pp. 751-763.

[1.2.5] Kummerow: "Statistics for efficient linear and nonlinear picture coding", Int. Telemetering Conf, pp.149-161, October 1972.

[1.2.6] Lee, B., Kang, M., Lee, J.: *Broadband telecommunications technology*, Artech House, 1996.

[1.2.7] Levendovszky, J., Elek, Zs., Végsõ, Cs.: "Validation of novel CAC algorithms", *ICAM- IEEE 1999*, pp. 195-211

[1.2.8] Nimoneya, et al.: "An HDTV Broadcasting System Utilizing a Bandwidth Compression Technique MUSE", *IEEE Trans. on Broadcasting*, pp 130-160, 1987

[1.2.9] Zelinski, R., Noll, P.: "Optimal quantization for PCM", Technical report , Heinrich Hertz Institute, Berlin, 1974 (in German)

[1.2.10] D.L:Richards: Telecommunication by speech, transmission performance of telephone networks. Butterworth 1973. London.

[1.2.11] Kleinrock, Leonard: Sorbanállás-kiszolgálás, bevezetés a tömegkiszolgálási rendszerek elméletébe. Műszaki Könyvkiadó, Budapest, 1979

1.3. Forráskódolás és csatornakódolás

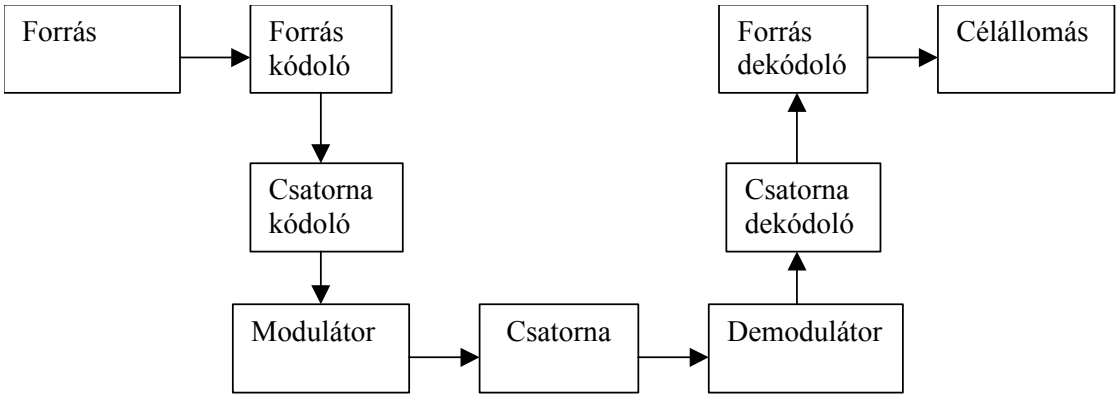
Szerző: dr. Vajda István

Lektor: Osváth László

A kommunikációs rendszerek az adatokat az adatforrás és a célállomás között továbbítják (lásd 1.3.1.ábra). Az adatokat először a *forráskódoló* kezeli, amelynek eredményeképp az adatok tömörebb formában kerülnek ábrázolásra. A kódoló kimenetén forráskódbeli kódszavak sorozata jelenik meg, amely kódszavak forrásábécé feletti sorozatok. A forráskódoló kimenete kerül a *csatornakódoló* bemenetére, amely a kimenetén csatonakódszavak sorozatát állítja elő. A csatornakódoló, a forráskódolóval ellentétes műveletet végez, meghosszabbítja az input sorozatokat, redundanciát ad hozzájuk. A modulátor a csatornakódszó szimbólumait (karaktereit) jelekké (analóg szimbólumokká) transzformálja, amelyek átvihetők a csatornán. A csatornában ezen jelek különböző változásokat szenvednek el a zaj, torzítás és interferencia kapcsán. A demodulátor az az egység, amely a csatorna kimeneti jeleit csatornakód szimbólumokká alakítja vissza. Egy csatonakódszónak megfelelő demodulált sorozat, a *vett szó*. A csatornadekódoló feladata az, hogy a vett szó alapján a legjobb döntést hozza az átküldött kódszóra. A forrásdekódoló kitömörítést végez, azaz dekódolja a forráskód szavakat.

Az adattömörítés nem függ a csatornától, a csatornakódolás pedig nem függ a forrástól. Bebizonyítható, hogy a fentiekben vázolt két lépésben végzett kódolás (forráskódolás, csatornakódolás) ugyanolyan jó lehet, mint bármely más módszer, adatoknak zajos csatornán történő továbbítására ([1.3.1.],[1.3.4.]). Következésképpen ezen kódolási algoritmusokat szeparáltan lehet tervezni. A cél a leghatékonyabb adatátvitel. Ez, például, azt jelenti, hogy adott szolgáltatási minőség mellett a lehető legnagyobb átviteli sebességet szeretnénk elérni.

Az információelmélet adta meg a forráskódolás és a csatornakódolás elvi lehetőségeinek határait. A klasszikus információelméleti eredmények között találjuk az első adattömörítő algoritmusokat. A csatornakódolás terén mai napi gyakorlatban alkalmazott módszerek javarésze az algebrai kódelmélet eredménye.



1.3.1.ábra. Digitális kommunikációs csatorna blokkvázlata

Természetesen, egy rövid fejezet csak arra ad módot, hogy dióhéjban összegzésre kerüljenek a főbb gondolatok, alapvető definíciók ezen két nagy tudományterületről.

Az alábbiakban először a forráskódolás témájába adunk egy bevezetőt. Az 1.4 fejezet részletes áttekintést ad a mai, praktikus adattömörítő módszerekről. A jelen fejezet nagyobb részét a csatornakódolásnak szenteljük.

1.3.1. Forráskódolás: entrópia és tömörítés

Az információ fogalma túl széles ahhoz, hogy jól megfogható legyen egyetlen definícióval. Az entrópia, amelyet tetszőleges valószínűségeloszláshoz ki tudunk számolni, sok olyan tulajdonsággal rendelkezik, amely jól megragadja az információ menyiségével kapcsolatos intuíciónkat.

Legyen X egy diszkrét valószínűségi változó, amely egy X halmazból veszi értékeit $p(x)=\Pr(X=x)$, $x\in X$ valószínűségeloszlás szerint. Az X diszkrét valószínűségi változó $H(X)$ entrópiájának definíciója:

$$H(X) = -\sum_{x} p(x)\log_2 p(x) \quad [bit]$$

Bináris valószínűségi változó esetén, ahol p(1)=q és p(0)=1-q, az *entrópia* formula a következő egyszerű alakot ölti:

$$h(q) = q\log_2 q + (1-q)\log_2(1-q)\,.$$

Például, pénzfeldobás kísérlet esetén a véletlen kimenetel entrópiája 1 bit. Az entrópia nem negatív értékű, ami könnyen látható, mivel $0\le p(x)\le 1$ implikálja, hogy $\log_2 p(x)\ge 0$.

Egy további példaként tekintsünk egy 4 kimenetelű *X* valószínűségi változót, ahol a kimenetelek a,b,c,d, továbbá a hozzájuk rendelt valószínűségek rendre 1/2, 1/4, 1/8, 1/8. Tegyük fel, hogy egy A személy ezen eloszlás alapján megfigyel egy kimenetelt. Egy tőle különböző B személy feladata, hogy kitalálja ezt a kimenetelt minimális számú kérdést feltéve A-nak. Első kérdése: " A kimenetel 'a'? ". Ha a válasz igen, akkor nem kérdez tovább, egyébként a következőt kérdezi: " A kimenetel 'b'? ". Ha a válasz igen, akkor nem kérdez tovább, egyébként a harmadik kérdése: " A kimenetel 'c'? ".

Ezen a módon B feloldja a bizonytalanságot, legfeljebb 3 kérdéssel. A megadott eloszlás alapján a kérdések számának várható értéke: 0.5*1+0.25*2+0.25*3=1.75. Ami itt igen érdekes, az az, hogy az *X* entrópiája is ugyanezen érték, azaz *H(X)*=1.75! Általában igaz, hogy a minimális kérdésszám várható értéke *H(X)* és *H(X*)+1 közötti.

Ezen észrevétel elvezet az adattömörítés elvi korlátjának megállapításához. Az adattömörítés során a legnagyobb valószínűségű kimenetelekhez a legrövidebb leírást (bináris sorozatot, kódszót) rendeljük, s szükségszerűen hosszabbakat a kisebb valószínűségű kimenetelekhez. A híres Huffman-kódolás ([1.3.1]) ezen az alapelven a minimális várható kódszóhoszat biztosítja, ahogy azt az alábbiakban bemutatjuk, némi formalizálás után.

Tekintsünk egy *C* forráskódot *s* elemszámú, $Y^*=\{y_1,y_2,...,y_s\}$ forráskód ábécé felett (s=2 esetben bináris kódot készítünk). Jelölje *C(x)* az *X*=*x* kimenetelhez rendelt kódszót, amelynek hossza l*(x).* A *C* kód *L(C)* várható hossza:

$$L(C)=\sum_x p(x)l(x)\,.$$

A fenti számpéldánk esetén *C*={0,10,110,111} választással (*C*(a)=0, *C*(b)=10, *C*(c)=110, C(d)=111) *L(C)*=1.75 adódik.

A *C* kód *prefix*, ha egy kódszó sem prefix egy másik kódszóban. Példánk kódja prefix. Ha prefix kód kódszavait egymás után írja egy A személy, akkor egy tőle különböző B személy dekódolni képest azt, azaz egyértelműen meg tudja állapítani a kódszóhatárokat abban pillanatban, amikor - előről kezdve sorosan olvasva a kódszavak alkotta sorozatot - egy kódszó végére ér. Ezt a tulajdonságot önszinkronizációs képességnek hívjuk. Általában pedig, ha a kódszavak sorozatát

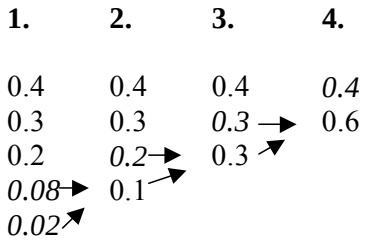
egyértelműen szét tudjuk bontani komponens kódszavakra, akkor a kódot *egyértelműen dekódolható*nak hívjuk.

Megmutatható, hogy tetszőleges egyértelműen dekódolható *C* kód esetén a várható kódszóhossz nem lehet kisebb, mint $H(X)/\log_2 s$. A már említett *Huffman-kód* optimális egyértelműen dekódolható kódot eredményez, ahol az optimalitás a minimális várható kódhosszat jelenti:

$$H(X)/\log_2 s \le L(C) < H(X+1)/\log_2 s + 1,$$

azaz - ahogy már fentebb is említettük - bináris ábécé (s=2) esetén a várható hossz az entrópiát 1 biten belül megközelíti.

Alább egy példát mutatunk, ahol az *X* valószínűségi változó eloszlása: {0.4, 0.3, 0.2, 0.08, 0.02}.



A módszer során lépésenkét a két legkisebb valószínűséget kombináljuk. Ezután visszafele lépünk, az ábrán a 4. lépéstől kezdve. Először $C(x_1)$=0 kódszót rendeljük a 0.4 valószínűségű kimenetelhez, s a többi valószínűségekhez - amelyek a 0.6 valószínűségbe akkumulálódtak - tartozó kódszavak első karakntereként (prefixeként) az 1 bitet rendeljük. A nyilakon visszafele lépve, a 0.3 valószínűségű kimenetelhez $C(x_2)$=10 kódszót rendeljük, míg a fennmaradó kimenetelek kódszavainak prefixét 11-re állítjuk. A 2. lépéshez visszalépve a 0.2 valószínűségű kimenetelhez $C(x_3)$=110 kódszót rendeljük, míg a fennmaradó kimenetelek kódszavainak prefixét 111-re állítjuk. Az 1. lépéshez visszalépve a $C(x_4)$=1110, $C(x_5)$=1111 kódszavakat rendeljük a 0.08 illetve 0.02 valószínűségű kimenetelekhez.

Az önszinkronozó képességben rejlő előny egyben hátrány is, ha ezen kódszavakat elemei meghibásodnak a csatornán való továbbítás során. Átviteli hibák okozhatják azt, hogy a szinkronizációs képesség megszűnik a kódszavak alkotta sorozat egy pontján, s attól a ponttól kezdve a további kódszavak is helytelenül dekódolódnak. Ekkor, sajnos, fel kell adni az önszinkronizációs előnyt, s előre (a

sorozat elején) rendelkezésre álló szinkroninformáció alapján szétválogatható konstans szóhosszra kell áttérni. Elvesztjük továbbá a fentemlített tömörítés optimalitást is, amennyiben az egyértelmű dekódolhatóság kritériumát meg kívánjuk tartani, vagyis ha továbbra is különböző - de konstans hosszú - kódszavakat kívánunk használni a kódunkban. Konstans szóhosszak esetén hatékonyabb kompresszióra csak úgy van lehetőség, ha feladjuk az egyértelmű dekódolhatóságot, azaz torzítással kódolunk.

Ilyen torzítás természetesen nem engedhető meg, ha például programokat tömörítünk, de megengedhető például beszéd- illetve képtömörítés esetén, az előírt szolgáltatási minőség mértékében. (lásd még az 1.4. fejezetet)

1.3.2. Csatornakódolás

Alapvető koncepciók

Tegyük fel, hogy adataink bináris formában állnak rendelkezésre. Az adatok *k* bites szegmenseit üzenetnek nevezzük. Egy bináris csatornakód *M* kódszót tartalmaz, ahol $M = 2^k$, továbbá a kódszó hossza *n* bit. A kód két dimenzionális paramétere *(n,k).* Például a

C={ 00000, 10101, 01111, 11010} (1)

kód egy (5,2) paraméterű bináris kód, azaz bitpárokat kódulunk 5 bites szavakba. Általában a kódokat {0,1,2,...,*q*-1} véges ábécé felett értelmezzük. Bináris esetben *q*=2.

Az itt bevezetett kód a *blokk kód*, azaz amelynél a forrás kimeneti sorozatának szeleteiből képzett üzenetblokkokat kódszóblokkokba kódolunk. Az output blokk csak a pillanatnyi input blokk függvénye. Az *R* kódolási sebesség definíciója $R = k/n,\ 0 \le R \le 1.$ Kisebb sebesség nagyobb redundanciával jár együtt, s megfordítva.

A másik nagy kódcsalád, a *konvolúciós kódok*. A forrás kimenetéről ekkor rövid, k_0 hosszúságú szeleteket, input kereteket vágunk. A kódoló egy lépésekor egy ilyen input keretet léptetünk a kódolóba, amelynek kimenetén egy n_0 hosszú output keret jelenik meg. A kódolónak állapota van, s az outputot a kódoló inputja és az

állapota határozzák meg. A kódolási sebesség definíciója értelemszerűen $R = k_0 / n_0,\ 0 \le R \le 1.$ A konvolúciós kódokra a továbbiakban még visszatérünk.

A blokk kódok harmadik fontos paramétere a *kódtávolság*, d_{min}. Ennek megfelelően a $C(n,k,d_{min})$ paraméterhármassal hivatkozunk egy kódra, megadva továbbá az kód-ábécé *q* méretét is. Két kódszó *d(x,y)* Hamming-távolsága azon pozíciók darabszáma, amelyen a két kódszó különbözik. A minimális kódtávolság a legkisebb Hamming-távolság a *C* kód különböző kódszavai között. Például az (1) kód esetén d_{min}=3.

Tegyük fel, hogy egy c kódszó került továbbításra és r a vett szó:

$$(r_0, r_1, ..., r_{n-1}) = (c_0, c_1, ..., c_{n-1}) + (e_0, e_1, ..., e_{n-1})$$

ahol *c* és *r* vektorok differenciája a hibavektor. Például ha a *c*=(01111) kódszó került továbbításra és *r*=(01011) a vett szó, akkor ez egy hibát jelent a második pozícióban, s értéke 1, azaz *e*=(00100).

A csatornakódok két alapvető használata a *hibajavítás* és a *hibadetekció*. Hibadetekció esetén a vételi oldalon azt szeretnénk megállapítani, hogy hibás a vett szó, azaz különbözik az átküldött kódszótól. Detektált hiba esetén a vevő a kódszó újraküldését kéri, mindaddig míg az helyesen meg nem érkezik (vagy egy időzítés le nem jár). Hibajavítás esetén maga a vevő próbálja meg korrigálni a hibát.

Nem nehéz látni, hogy a *C* kód maximum $t_{\det} = d_{\min} - 1$ hibát képes detektálni: ha ugyanis az átküldött kódszó kevesebb pozícióban változik meg, mint a legközelebbi kódszó távolsága, akkor a vett szó mindenképpen különbözik bármely kódszótól, s ennek alapján a meghibásodás detektálható. Például az (1) kód 2 hibát tud detektálni.

Hasonlóan, a *C* kód maximum $t_{corr} = \text{int}[(d_{\min} - 1)/2]$ hibát tud javítani, ahol int(x) az x egész része: amíg a vett szó közelebb van az átküldött kódszóhoz, mint bármely más kódszóhoz, addig helyesen dekódolni tudjuk azt.

Tegyük fel, hogy egy - nembináris - kód esetén tudjuk a hiba pozícióját, de nem ismerjük annak értékét. Ezt törléses hibának nevezzük. Egy *C* kód maximum $t_{eras} = d_{\min} - 1$ ***törlés***t képes javítani: ha az átküldött szó kevesebb számú pozícióban törlődik, mint a legközelebbi kódszó távolsága, akkor a nem-törlődött pozícióban álló

értékek az átküldött kódszóra illeszkednek a legjobban. Például az (1) kód 2 törléses hibát tud javítani.

A legegyszerűbb kód az egyszerű paritáskód, *C(n,n*-1,2), *q*=2, amelynél egyetlen paritásbitet illesztünk az üzenethez. Ez a kód - a fentiek alapján - 1 hibát tud jelezni. Például a *C*={0000, 0011, 0101, 0111, 1001, 1010, 1100, 1110} kód paraméterei C(4,3,2), q=2.

A legegyszerűbb hibajavító kód az egyszerű ismétléses kód, *C(n,*1*,n), q*=2. Ez a kód két kódszót tartalmaz a csupa 0 és a csupa 1 kódszót, s az üzenet egyetlen bit. Például *C*={00000, 11111} paraméterei C(5,1,5) , q=2, amely kód képes 2 hibát javítani és 4 hibát jelezni. Sebessége alacsony: *R*=1/5.

A 2-dimenziós paritáskód szintén elemi kód. Az üzenetet mátrixba rendezzük. A sorait illetve oszlopait egy-egy paritásbittel egészítjük ki. Tekintsük az alábbi kódszót, ahol 9=3×3 bites az üzenet:

1	1	0	*0*
0	1	1	*0*
1	1	1	*1*
0	*1*	*0*	*1*

A jobb alsó sarokba írt bit a paritások paritása, amelyet akár a sorok, akár az oszlopok paritása alapján is számolhatunk, azonos eredményre jutunk. A kód paraméterei *C*(16,9,4),*q*=2. A kód tehát egy hibát képes javítani. A javítás algoritmusa igen egyszerű: a sorparitás-hiba indexe valamint az oszlopparitás-hiba indexe egyértelműen kijelölik a hiba pozícióját a mátrixban.

Nembináris aritmetika

Az algebrai kódok intenzíven használnak algebrai eszközöket. A véges test, az úgynevezett Galois-test, GF(*q*) az egyik leggyakrabban használt algebrai struktúra nembináris blokk kódok esetén.

Egy test testelemek egy F halmaza szorzás és összeadás művelettel, amelyek rendelkeznek a szokásos szabályokkal (asszociativitás és kommutativitás, a disztributivitási szabály, az additiv egység (0) léte, a multiplikatív egység (1) léte, az

additív inverzek léte, valamint a 0 elem kivételével az elemek multiplikatíve inverzének a léte).

A GF(q) Galois-test véges test, elemszáma q. A q elemszám prímszám vagy prímhatvány, lehet csak $q = p^m$, m=1,2,... .

Egyszerűbb az aritmetika q=p esetben, amikor modulo p összeadás illetve modulo p szorzás a két testművelet. Tekintsük például GF(5) esetét. Ekkor 5 testelem van: 0,1,2,3,4. A műveletek modulo 5 szerintiek. Például 3+4=7=5+2=2 mod 5, vagy 4 inverze 4, azaz 4^{-1}=4 mod 5, mivel 4*4=3*5+1 =1 mod 5. A legegyszerűbb eset a GF(2) test a szokásos bináris aritmetikával. A GF(p) test kiterjesztésével jutunk az általános esethez, azaz amikor m>1:

Egy GF(p) feletti polinom a következő matematikai kifejezés

$$f(x) = f_{n-1}x^{n-1} + f_{n-2}x^{n-2} + ... + f_1 x + f_0 ,$$

ahol az $f_{n-1},..., f_0$ együtthatók GF(p)-beli elemek. Egy $p(x)$ polinom irreducibilis, ha csak $\alpha p(x)$ vagy α, $\alpha \in$GF(p) az osztója, azaz valódi osztója nincsen. A $p(x)$ polinom főpolinom, ha f_{n-1}=1. Tekintsük GF(p) feletti polinomok gyűrűjét modulo $p(x)$, azaz ahol a két művelet a polinom szorzás és összeadás modulo $p(x)$. Ha $p(x)$ tetszőleges GF(p) feletti polinom, akkor gyűrű kielégíti a testek fentebb említett szabályainak mindegyikét, kivéve a multiplikatív inverzek létét. Ha azonban további megszorítással élünk a $p(x)$ polinomra, akkor megmutatható (lásd például [1.3.2]):

A GF(p) feletti polinomok gyűrűje modulo $p(x)$, $p(x)$ főpolinom esetén test, akkor és csak akkor, ha $p(x)$ irreducibilis polinom.

Tekintsük például GF(2^3) (=GF(8)) testet. A testelemek 0,1,2,3,4,5,6,7, amelyeket bináris polinomokként tekintünk, legfeljebb 2 fokszámmal:

GF(8)={0, 1, x, x+1, x^2, x^2+1, x^2+x, x^2+x+1} ,

ahol a természetes hozzárendelés: {0 <--> 0, 1 <--> 1, 2 <--> x, 3 <--> x+1, 4 <--> x^2, 5 <--> x^2+1, 6 <--> x^2+x, 7 <--> x^2+x+1}, tehát a polinomok bináris együtthatóinak decimális megfeleltetésével. A műveleteket ekkor modulo x^3+x+1 végezzük, ahol az x^3+x+1 polinom egy bináris, irreducibilis harmadfokú polinomként került kiválasztásra. Az összeadás nem nehéz, például: 3+5=(x+1)+(x^2+1)=x^2+x=6. A modulo polinom aritmetika igazából a szorzásnál kerül felhasználásra. Például 3*6=(x+1)*(x^2+x)= x^3+2x^2+x=x^3+x=(x+1)+x=1 mod x^3+x+1, így 1/3=6.

Az algebrai blokk kódok esetén *q*-áris kódábécéként GF(*q*) kerül választásra.

Lineáris blokk kódok

Legtöbb praktikus kód lineáris: ha a kódszavak tetszőleges lineárkombinációja is kódszó, akkor a kód lineáris. Egy lineáris kódok algebrai struktúra szempontjából egy lineáris altér az *n* hosszúságú *q*-áris szavak terében. A $C(n,k,d_{\min})$, *q* lineáris kód szavainak száma q^k , lineáris altérként dimenziója k.

Mátrixok segítségével definiálhatjuk a lineáris kódot; ezek a generátormátrix, valamint a paritásellenőrző mátrix. A *G generátormátrix* dimenziója *kxn*, ahol a *k* sort *k* számú lineárisan független kódszó alkotja, így a *G* sorai mint bázisvektorok kifeszítik a kódszavak terét:

$$(c_0, c_1, ..., c_{n-1}) = (m_0, m_1, ..., m_{k-1}) \cdot G$$

ahol $m = (m_0, m_1, ..., m_{k-1})$ üzenet, mint vektor balról szorozza a generátormátrixot, lineárisan kombinálva azok sorait. A csupa nulla kódszó mindig eleme egy lineáris kódnak: a csupa nulla üzenethez tartozik.

A *H paritásellenőrző mátrix* dimenziója *(n-k)xn*. Tulajdonsága a következő: a kódszavak és csak azok, ortogonálisak a *H* soraira; formálisan tetszőleges $c \in C$ kódszóra: $Hc^T=0$, ahol c^T jelöli a *c*, mint sorvektor transzponáltját.

Szisztematikusnak nevezzük a *C* kódot, ha kódszavai az alábbi struktúrájúak: $c = (m_0, m_1, ..., m_{k-1}, p_0, p_1, ..., p_{n-k-1})$, azaz az üzenet tisztán megjelenik a kódszó szeleteként. A szisztematikus kódelőállítás praktikus: a csatorna dekódoló első lépésként döntést hoz arra, mely kódszó került továbbításra, majd a második lépés ezen dekódolt kódszó alapján az üzenet rekonstruálása, amely művelet szisztematikus alak esetén triviális.

Az (1) kódot tekintve

$$G = \begin{pmatrix} 10101 \\ 01111 \end{pmatrix} \text{ and } H = \begin{pmatrix} 11100 \\ 01010 \\ 11001 \end{pmatrix} \qquad (2)$$

ahol továbbá azt is észrevehetjük, hogy a kód szisztematikus: a G mátrix egységmátrix-szal kezdődik.

Lineáris kódok esetén a minimális kódtávolság a minimális nemzéró Hamming-súllyal egyezik meg. Ez a megállapítás három egyszerű tény következménye: a linearitás miatt két kódszó különbsége is kódszó; két kódszó Hamming-távolsága a kódszavak különbségének Hamming-súlya; bármely kódszó triviálisan előállítható, két kódszó különbségeként, ahol az egyik kódszó önmaga, s a másik a zérus kódszó.

Ez az észrevétel jelentősen egyszerűsíti a minimális távolság kiszámítását.

Alsó korlátot kaphatunk a minimális kódtávolságra a H mátrix egyszerű algebrai tulajdonsága alapján: a H oszlopavektorai halmazát tekintve, a lineárisan függő oszlopok minimális számánál nem lehet kisebb a minimális távolság. Például a (2) alatti H mátrixot tekintve, annak oszlopai különbözők, következésképp - bináris lévén - bármely két oszlopa lineárisan független, s három oszlop alakothat csak lineárisan függő részhalmazt. Következésképp legalább 3 a minimális távolság (adott esetben ez pontosan 3).

Ezen észrevételünk egyben kódkonstrukció alapötlete is. A híres C(7,4,3), GF(2) Hamming-kód egy paritásellenőrző mátrixa:

$$H = \begin{pmatrix} 0111100 \\ 1011010 \\ 1101001 \end{pmatrix} \qquad (3)$$

Ciklikus lineáris blokk kódok

A legelterjedtebb algebrai blokk kódok lineárisak és ciklikusak. Egy kód *ciklikus*, ha tetszőleges kódszavának ciklikus forgatásával újra kódszót kapunk az adott kódban. A ciklikus kódok szavainak alkalmas reprezentálását adják a polinomok:

$$c = (c_0, c_1, ..., c_{n-1}) \; <---> \; c(x) = c_0 + c_1 x + ... + c_{n-1} x^{n-1}$$

Két alapvető polinom tartozik egy kódszóhoz: a g(x) generátor polinom és a h(x) paritásellenőrző polinom, ahol g(x) fokszáma n-k, h(x) fokszáma k. A generátorpolinom segítségével a következőképp állíthatók elő a kódszavak:

$$c(x)=m(x)g(x) \qquad (4)$$

ahol $m(x) = m_0 + m_1 x + ... + m_{n-1} x^{n-1}$ az üzenet polinom alakban. A két alap-polinom között egyszerű kapcsolat áll fenn : g(x)h(x)=x^n-1.

A C(7,4,3), GF(2) paraméterű ciklikus Hamming-kód generátorpolinomja g(x)=x^3+x+1. Például m=(1011) <---> m(x)=1+x^3+x^4 üzenetet c(x)=1+x+x^5+x^6+x^7 <---> c=(1,1,0,0,0,1,1,1) kódszóba kódoljuk.

A (4) előállítás nem-szisztematikus. Szisztematikus generálás elve a következő: számítsuk ki az alábbi osztási maradékot

r(x) = - $x^{n-k} m(x)$ mod g(x), (5)

amelynek alapján a szisztematikus alakú kódszó

c(x)= $x^{n-k} m(x)$ + r(x). (6)

A (6) előállítás a CRC (*Cyclic Redundancy Check*) generálás, ahol a g(x) polinomot CRC generátor polinomnak nevezzük.

Reed Solomon kódok

A Reed-Solomon kódok hatékony nonbináris ciklikus lineáris blokk kódok. Velük találkozhatunk leggyakrabban: ez a CD felvételeknél alkalmazott kódolás. Egy C(n,k,d_{min}), GF(q), n=q-1 Reed-Solomon kód generátorpolinomja:

$$g(x) = (x-\alpha)(x-\alpha^2)...(x-\alpha^{n-k-1}) \quad (7)$$

ahol $\alpha \in GF(q)$ egy n-edrendű testelem, ami azt jelenti, hogy n az a legkisebb pozitív egész kitevő, amelyre $\alpha^n = 1.$ A Reed-Solomon kódok optimálisak, MDS (Maximum Distance Separable) tulajdonságú kódok: a legoptimálisabban használják fel a redundanciát. Tetszőleges kód esetén - a Singleton-korlát alapján - $d_{min} \le n-k+1$, míg a Reed-Solomon kódok esetén d_{min}=n-k+1, azaz a maximális távolság eléretik az n-k paritáshossz adott értéke mellett.

Például a C(12,8,5), GF(13) Reed-Solomon kódot: $g(x) = (x-2)(x-4)(x-8)(x-3)$ generátorpolinom állítja elő, s ezen kód két hiba javítására alkalmas. Gyakorlati paraméterek például a következők: C(255,255-2t,2t+1), GF(256), ahol t a megkívánt javítóképesség, továbbá a kód byte ábécé feletti.

Dekódolási koncepciók

A dekódoló célja a téves dekódolás hibavalósínűségének minimalizálása. Az optimális eljárást a maximum likelihood dekódolás (ML) adja.

Tegyük fel, hogy ismert a csatorna egy valószínűségi modellje a P(r|c) feltételes valószínűségek formájában. r vett szó esetén az ML dekódoló azt a c' kódszót dekódolja, amelyre a maximális, P(r|c') feltételes valószínűség adódik. Sajnos gyakorlati csatornák esetén ilyen mélységű leírás nem áll rendelkezésre.

Ezért a dekódoló egyszerűbb algoritmust alkalmaznak: a vett szóhoz Hamming-távolságban legközelebbi kódszóra döntenek. Ez az algoritmus ML dekódolással ekvivalens, ha a csatorna bináris, szimmetrikus hibázású és a meghibásodások statisztikailag függetlenek az kódszó különböző pozícióiban (DM-BSC, Discrete Memoryless - Binary Symmetric Channel). Sajnos általában még ez az algoritmus sem kivitelezhető közvetlenül: praktikus kódméretek esetén a kódszavak száma csillagászati, így nem tudjuk megkeresni a legközelebbi kódszót. Ezen problémát enyhíti a szindróma dekódolás.

Az r vett szó s szindrómájának definíciója: $s=Hr^T$. Következésképp $s=Hr^T=H(c+e)^T= Hc^T+He^T= He^T$. Ha nincsen hiba, azaz e=0, akkor s=0. A szindrómákat és a nekik megfelelő javítható hibát - elvileg - egy táblázatba, a szindróma dekódolási táblázatba rendezzük. Javítható hibamintának azt a legkisebb súlyú szót választjuk, amelynek szindrómája az adott szindróma. Az (1) példa-kódunkra, amelynek H mátrixa a (2) alatti, a következő szindróma dekódolási táblázat adódik:

s	e
000	00000
001	00001
010	00010
011	01100
100	00100
101	10000
110	01001
111	01000

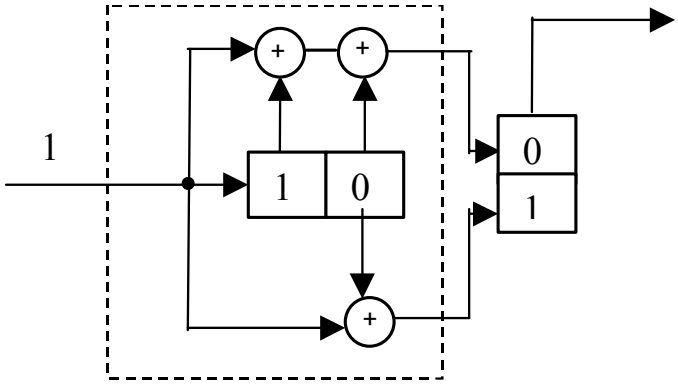
Tegyük fel, hogy a vett szó szindrómája 111. A dekódoló a táblázatba tekint s kiválasztja a megfelelő hibavektort, ami ezesetben 01000, következésképpen a dekódolt kódszó úgy adódik, hogy a vett szó bitjét invertáljuk. Praktikus kódméretek

mellett ez a táblázatos megoldás sem kivitelezhető a nagy tárigény miatt (a táblázat sorainak száma bináris kódok esetén 2^{n-k}). Ezért táblázat helyett algoritmust használunk, minden vett szóra újra és újra kiszámolva a javító hibamintát. A klasszikus szindróma dekódolási algoritmus a Peterson-Gorenstein-Zierler-algoritmus (PGZ), s egy gyorsabb megoldás a Berlekamp-Massey algoritmus.

Konvolúciós kódok

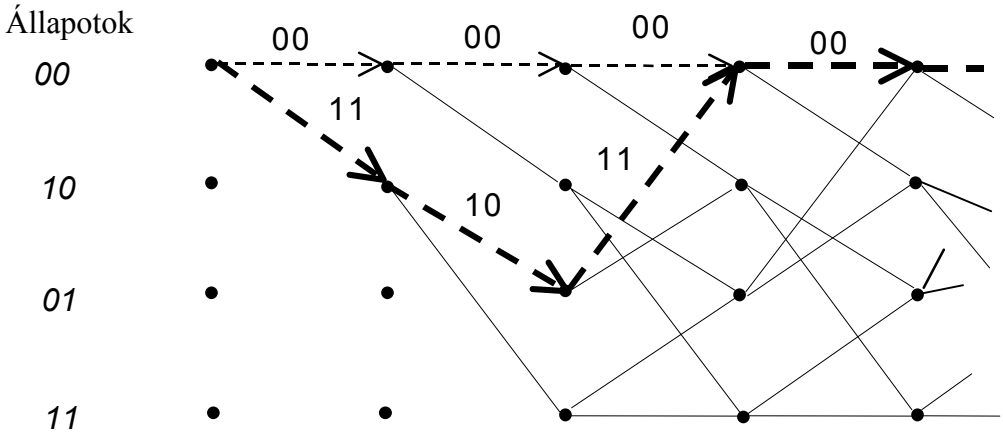
A konvolúciós kódok kódolója egyszerű elemekből épül fel: shiftregiszterekből és modulo 2 összeadókból:

A 1.3.2.ábra kódolója kimenetén egy lépésben $n_0=2$ csatornabit lép ki, annak hatására hogy a bemenetére $k_0=1$ adatbit kerül, azaz a kódsebesség R=1/2. Az input bit belép a 2 bitcellás shiftregiszterbe. A blokk kódokkal ellentétben az üzenethossz nem rögzített, elvileg félig végtelen. A pillanatnyilag tárolt bitek és az input bit modulo 2 összegei határozzák meg az output biteket. A kódolót kettő generátor polinom írja le, amely polinomok együtthatói a shiftregister megcsapolási pozícióinak felelnek meg természetes hozzárendeléssel. A konvolúciós kódoló lineáris. Láttuk, hogy lineáris kód esetén a minimális kódtávolság megegyezik a minimális nemzéró kódszósúllyal. Ha csupa nulla input lép a nulla kiinduló állapotú kódolóba, akkor az output is csupa nulla lesz. Példánk esetén a minimális kódszósúlyra vezető input az 1000... a 00 kezdeti állapot mellett. Ekkor az output 11 10 11 00 00.... lesz, azaz a minimális kódtávolság (itt tradicionális neve szabad távolság) értéke 5.



1.3.2.ábra. Konvolúciós kódoló [x^2+x+1 , x^2+1]

Konvolúciós kódok esetén ismert a gazdaságos maximum likelihood dekódoló, a *Viterbi-dekódoló* ([1.3.2],[1.3.7],[1.3.9]). Ezen dekódoló a kód speciális gráf reprezentációja, a trellis alapján dolgozik:

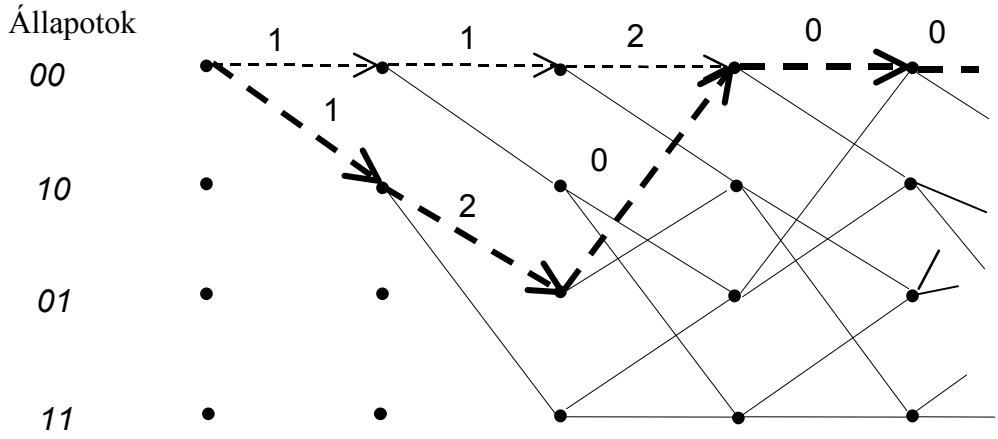


A trellis csomópontjai a shiftregiszter állapotainak, az irányított élek az állapotátmeneteknek, a kódoló egy lépésének felelnek meg. Az élek sorozata alkotta utak értelemszerűen kódszavaknak feleltethetők meg. Az élekre írt cimkék (itt bitpárok) az adott lépés során kapott kimenetet mutatják.

Ha a csatorna jellemezhető DM-BSC modellel, akkor - ahogy azt már említettük -az ML dekódolás a minimális Hamming távolságban dekódolással ekvivalens. Tegyük fel, hogy a csupa zérus kódszó került átvitelre és a vett szó 10 01 11 00 00 ... .

Az éleket súlyokkal címkézzük fel, ahol a súly az adott él Hamming távolsága a vett szó megfelelő szeletétől (1.3.4.ábra). Egy út súlya az azt alkotó élek súlyainak összege. A Viterbi-algoritmus súlyozott éleket tartalmazó trellisen a szélsőérték súlyú út megkeresését végzi; esetünkben minimális Hamming súlyú utat keresi. Ezt a feladatot az üzenethosszban lineáris bonyolultság mellett, s nem pedig - a blokk kódoknál említett - exponenciális bonyolultság mellett végzi.

A súlyozás általánosítható általános - azaz nemcsak bináris szimmetrikus - emlékezetnélküli csatornák esetére: egy él súlyának log[p(y|x)] , feltételes valószínűség logaritmusa kerül választásra, ahol x és y a trellis éle és a vett szó megfelelő szelete. Additív fehér gaussi zajú csatornák esetén ezen számítások jól követhetők.



1.3.4.ábra: Trellis súlyozott élekkel (részlet)

Gyakorlati kódgenerálási technikák

A csatornakódok módosításával, illetve több kód kombinálásával új, kedvező tulajdonságú kódok generáhatók. A legfőbb módszerek a következők ([1.3.1],[1.3.6],[1.3.8]):

- kódrövidítés
- paritásbittel bővítés
- interleaving (kódátfűzés)
- szorzat kód generálás
- kaszkádosítás

Tekintsünk egy $C(n,k,d_{min})$ szisztematikus lineáris kódot. *L-bites rövidítés* során a kódszavak halmazának azon részhalmaza kerül kiválasztásra, amelyben a kódszavak L darab zéró bittel kezdődnek. Az L zérus nem kerül továbbításra. Az eredményül kapott kód paraméterei: $C'(n-L,k-L,\geq d_{min})$.

Tekintsünk egy $C(n,k,d_{min})$ lineáris kódot, ahol d_{min} páratlan szám. Egy paritásbittel bővítjük a kódszavakat. Az eredményül kapott kód paraméterei: $C'(n+1,k, d_{min}+1)$. Például a C(7,4,3) Hamming-kódból kiindulva C(8,4,4) *teletext kód*ot kapjuk.

Csomós hibák ellen a kódátfűzés technika hatékony. Gyakran használt módszer rádiócsatornán történő digitális átvitelnél. *h-szoros kódátfűzés* esetén h kódszót továbbítás előtt egy mátrixba rendezünk sorokként, majd a mátrixik oszloponként továbbítjuk a csatorába. A vevő újra felépíti ezt a mátrixot, majd

soronként dekódolja azt. A javítható hibacsomó-hossz a komponens kód kódtávolságának h-szorosa.

A *szorzat-kód* technika két kódot használ: C1(n_1,k_1,$d_{1,min}$) és C2(n_2,k_2,$d_{2,min}$) kódokat. Az üzenetbiteket k_1xk_2 méretű mátrixba rendezzük, majd soronként a C1, oszloponként a C2 kóddal kódoljuk szisztematikusan. A paritások paritása jobb alsó szegmenst bármelyik kóddal kódolva számíthatjuk. Az eredményül kapott kód paraméterei: C'(n_1n_2, k_1k_2, $d_{1,min}d_{2,min}$), azaz a komponens kódok paramétereit összeszorozzuk.

A *kaszkád-kód* technika hatékony kevert hibázás esetén, azaz amikor mind csomós mind pedig egyedi, véletlen hibákra is fel kell készülni. Két komponens kódot használunk, az egyik neve belső kód, a másik neve külső kód. A külső kódoló inputját adja a forrás. A külső kód, egy nembináris kód, amelynek szavai érkeznek a belső kódolóhoz. A belső kódoló a külső kódoló egy-egy nembináris karakterét binárisba fejtve, mint egy-egy önálló bináris üzenetet kezel, s ezt kódolja tovább. Így a vételi oldalon a belső dekódoló által rekonstruált üzenetek a külső kódoló egy kódszavának egy karakterét jelentik. A belső dekódolót úgy tervezzük, hogy az egyedi hibákat tudja dekódolni, míg a csomós hibák a külső dekódoló számára javítható karakterhibákként jelenkezzenek. Így a két kód együttélése optimális.

Spektrális bitsebesség és bitenergia per zaj viszony

Az eddigiekben a demodulátort és a dekódert külön egységként kezeltük. Tekintsük most őket egy egységként, amely analóg jeleket fogad a csatornából, s üzeneteket dekódol belőle közvetlenül. Legyen S a jelteljesítmény, továbbá R_b [bit/sec] jelölje a - fizikai - bitsebességet. Nyilván E_b=S/R_b alapján adódik a bitenergia. Shannon adta meg a C [bit/sec] kapacitást additív gaussi zajú csatorna esetén:

$$C = W \log_2\left(1 + \frac{S}{N_0 W}\right) \qquad (8)$$

ahol W a sávszélesség, ami azt jelenti, hogy s(t) a jelalak teljesítménye elhanyagolható a [-W/2,W/2] frekvenciasávon kívül, továbbá N_0 [watts/Hz] jelöli az egyoldalas zaj teljesítménysűrűséget. A (8) formula megadja azt a legnagyobb fizikai bitsebességet, amely mellett - elvileg - tetszőlegesen kicsi dekódolási hibavalószínűség valósítható meg a vételi oldalon.

(8) formula alapján hasznos - és áttekinthető - korlátok vezethetők le. Tekintsük azt az esetet, amikor a sávszélesség tetszőlegesen nagy lehet (elméletileg végtelen). Kiderül, hogy a legkisebb bitenergia per zaj viszony, E_b/N_0 elvileg -1.6dB-ig csökkenthető!

A tapasztalat azt mutatja, hogy könnyen tervezhetők olyan átviteli rendszerek, amelyek gyakorlatilag kicsi bithibaarányt (BER, Bit Error Rate) érnek el 12dB E_b/N_0 mellett (szokásos digitális modulációs eljárások külön csatornakódolás nélkül). A csatornakódolók az úgynevezett *kódolási nyereség*ükkel (coding gain) [dB] csökkentik ezt az E_b/N_0 értéket ([1.3.6-9]).

Tehát különböző modulációs-kódolási rendszerek az alapján vethető egybe, hogy adott BER eléréséhez (például BER=10^{-5}), mekkora E_b/N_0 érték szükséges. Az utóbbi évek igéretes kódolási eljárása a *turbó kód* ([1.3.11]). Ezen kódokkal már demonstráltak 1dB körüli E_b/N_0 értékeket praktikus BER mellett. A turbó kódok a konvolúciós kódok továbbfejlesztései és igen komplex, iteratíve, valószínűségi modellekre támaszkodó dekódert használnak.

Ha két, különböző sávszélességet használó redszert vetünk egybe, akkor az egyszerű bitsebesség helyett az r= R_b/W [bit/sec / Hz] *spektrális bitsebesség* használható.

(8) formula alapján, ha tetszőlegesen kicsi BER elérésének lehetőségét vizsgáljuk

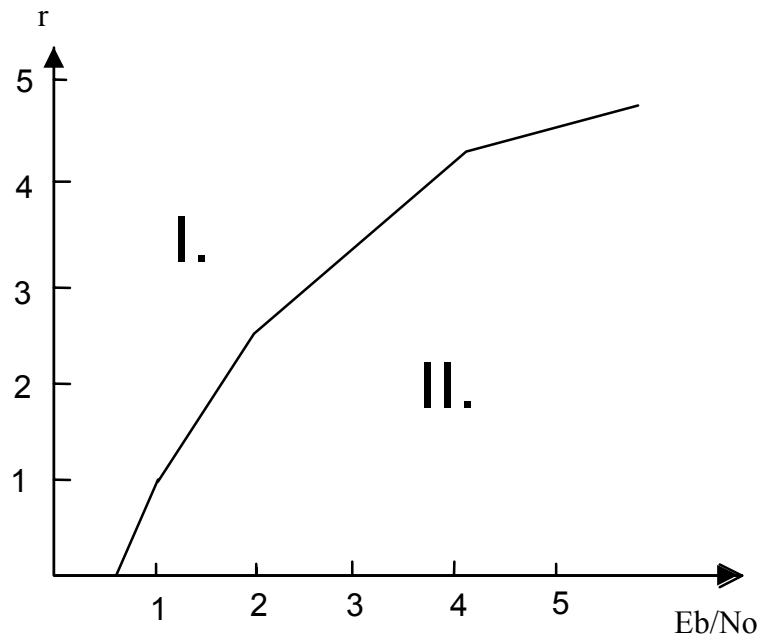
$$\frac{E_b}{N_0} \geq \frac{2^r - 1}{r} \qquad (9)$$

összefüggés adódik. Ezt illusztrálja az 1.3.5.ábra (lineáris skálák).

A II. tartomány felel meg az [r, E_b/N_0] elérhető pároknak, azaz tetszőleges átviteli rendszer belül kell maradjon ezen tartományon Például ha r tart a nullához (W tart a végtelenhez) a legkisebb elérhető bitenergia per zaj viszony a már említett 0.69 (-1.6dB).

Számos fontos témakörre ilyen rövid anyagban nem tudtunk kitérni, ahol elsősorban a kódolási alapfogalmak bevezetésére törekedtünk, segítve az olvasót a további speciális témakörök, anyagok önálló feldolgozásában. Ilyen további témakörök között említjük meg a sávszélességkorlátozás melletti kódmodulációs

technikákat ([1.3.13]), fadinges csatornák speciális kódolási technikáit a mobil kommunikáció kapcsán([1.3.12]), a többszörös hozzáférésű kódosztásos csatornák kódolási és dekódolási problémáit ([1.3.7],[1.3.9]). A korszerű kommunikációs rendszerek tervezésében a modulációs, az csatornakiegyenlítési, az adaptív



1.3.5.ábra. Az elérhető és a nem leérhető tartomány

teljesítményszabályozási valamint a kibakontroll kódolási technikák együttes szempontok szerinti tervezése kerül előtérbe.

Irodalomjegyzék

[1.3.1] Györfi L, Györi S., Vajda I.: *Információ és kódelmélet*, Typotex Kiadó, 2000.

[1.3.2] Blahut, R.E.: *Theory and Practice of Error Control Codes*, Addison wesley, 1986.

[1.3.3] van Lint,J.H.: Introduction to Coding Theory, Spinger Verlag, 1982

[1.3.4] McElice, R.J. *The Theory of Information and Coding*, Addison wesley, 1977.

[1.3.5] MacWilliams,F.J. and Sloane, N.J.A: *The theory of Error-Correcting Codes,* Part I-II., Noth-Holland Publishing Co., 1977

[1.3.6] Steele,R and Hanzo,L.: Mobile radio Communications, 2001

[1.3.7] Viterbi,A.J.: CDMA.Principles of Spread Spectrum Communication, Addison Wesley, 1995

[1.3.8] Proakis,J.: Digital Communications, McGraw Hill, 1992

[1.3.9] Simon,M. et.al.: Spread Spectrum Communications, Computer Science Press, Vol I-III, 1985

[1.3.10] Shannon,C.E.: *A mathematical theory of communication* ,Bell System Tech. J., Vol. 27, 1948, pp. 379-423 (Part I), pp. 623-656 (part II).

[1.3.11] Berrou,C., Glavieux,A. and Thitimasjshima,P: *Near Shannon limit error-correcting coding and decoding:Turbo codes*', Proceedings IEEE Int.Conf. on Communications, Geneva, May 1993, pp.1064-1070.

[1.3.12] Biglieri,E., Proakis,J and Shamai,Sh.: *Fading channels: information-theoretic and communication aspects*, IEEE Trans. Information Theory, Vol. 44, No. 6, October 1998, pp. 2619-2692.

[1.3.13] Forney D. and Ungerboeck,G: *Modulation and coding for linear Gaussian channels*, IEEE Trans. Information Theory, Vol. 44, No. 6,October 1998, pp. 2384-2415.

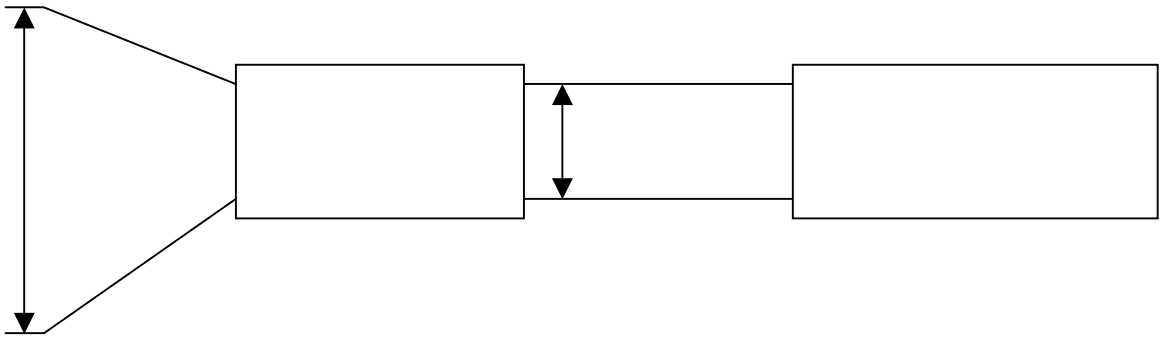
1.4. Adattömörítés

Szerző: dr. Levendovszky János

Lektor: dr. Dallos György

Ez a fejezet a modern adattömörítési algoritmusokkal foglalkozik, amelyek célja az átviendő (vagy feldolgozandó) információs folyam sávszélességének a csökkentése. A modern informatikában az adattömörítésnek központi szerepe jut, amelyet a tároló kapacitások jobb kihasználása, vagy a jelfeldolgozási igény csökkentése motivál. Kommunikációs hálózatoknál azonban az elsődleges cél a sávszélesség-gazdálkodás optimalizálása annak érdekében, hogy a szélessávú adatátvitel, keskenysávú hozzáférési hálózatokon is megvalósítható legyen (pl. multimédiás adatfolyamot átvitele a Számítógépes Világhálózatra kapcsolódó MODEM-es telefonvonalakon keresztül).

Az alapelvet az 1.4.1. ábra szemlélteti.



A feladat fontosságát az is indokolja, hogy az adattömörítés (amely tradicionálisan az információelmélettel foglalkozó tankönyvek egy-egy fejezete volt) ma már önálló "tudományággá" nőtte ki magát, számos kiemelkedő tudóssal, cikkel és szakmai monográfiával [1.4.6, 1.4.3, 1.4.7]. Sőt, a kidolgozott algoritmusok egy része olyan jól ismert szabványokká érlelődött, mint JPEG, GnuZIP, MPEG ....stb.

Az adattömörítéssel foglalkozó algoritmusok két részre oszthatók:

1. Ha az adat mintavételezett analóg minták formájában adott, akkor transzformációs, vagy adaptív prediktív technikákkal lehet tömöríteni. Mivel a

módszer alkalmazása során információvesztés lép fel, ezért ezeket a módszereket "veszteséges" módszereknek hívják. Ha az információvesztés "kontrollált" formában történik (ennek mértéke egy előírt szint alatt marad), akkor ez az információ minőségét nem rontja el. Ezeket az eljárásokat általában beszéd és képjelekre alkalmazzák. Segítségükkel aránylag nagy sávszélességcsökkenés érhető el ([1.4.6, 1.4.7]).

2. Ha az adat digitális (pl. fájl) formában adott, akkor információelméleti módszerek (pl. Huffman kódolás) vagy szótáralapú metódusok (pl. Lempel-Ziv algoritmus) alkalmazhatóak az információ tömörítésére. Mivel ebben az esetben a tömörített fájlból a teljes információs folyam, hiba nélkül visszanyerhető, ezért ezeket a módszereket "veszteségmentes" eljárásoknak szokták nevezni. Ezen módszerek részletes áttekintését az olvasó a [1.4.1, 1.4.3, 1.4.5] irodalmakban találhatja meg.

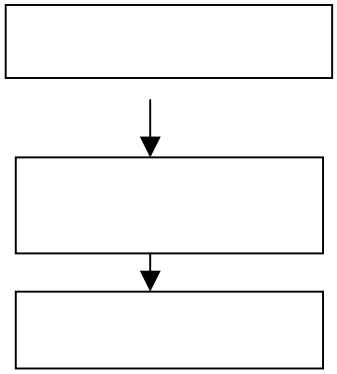
Mivel digitális kommunikáció esetén minden információs folyamat végül digitális formában jelenik meg (legyen ez hang vagy egy szövegszerkesztő által generált szövegfájl), ezért a második módszer univerzálisan alkalmazható. Ugyanakkor a "veszteségmentes" módszerek nem biztos, hogy jól illeszkednek az analóg információhoz (kép és hang), mert nem mindig tudnak nagy kompressziós arányt elérni, valamint túl nagy az algoritmikus komplexitásuk. Ezért a tradicionális szereposztás szerint a "veszteségmentes" tömörítést tipikusan digitálisan generált fájlokra, míg a "veszteséges" módszert hangra és képre alkalmazzák. Természetesen hibrid módszerek is elképzelhetők.

Az adattömörítő algoritmusok másik fontos jellemzője az algoritmikus komplexitás, amit gyakran „rejtett költségnek” neveznek. Ezt a költséget két tényező határozza meg:

- mekkora mennyiségű információra van szükség a tömörítő algoritmus lefuttatásához (pl. egy teljes szótár tárolására és átvitelére is szükség van szótáralapú eljárásoknál);
- mi a tömörítő és a visszaállító algoritmus numerikus komplexitása (pl. összeadások és szorzások száma)

Az utóbbi faktor kritikussá válhat, mikor a tömörítendő információ több Mbps sebséggel kerül a tömörítő bementére és valós-idejű tömörítés a cél. Ez megköveteli az algoritmikusan egyszerű módszerek használatát, amelyek kereskedelmileg elérhető hardver elemeken (pl. DSP-n) is megvalósíthatóak.

A különböző adattömörítő algoritmusok tárgyalása a következő szerkezet szerint történik:



1.4.1. Alapelvek

Az adattömörítés alapgondolata a forrásinformáció statisztikus függőségi viszonyainak a kiaknázásában rejlik. Pl. egy szó kiejtésénél a szó eleje gyakran meghatározza a végét. Ezért a beszéd során, vagy egy írásos dokumentumot olvasva, könnyű megjósolni, hogy egy adott szakasz után mi következik. A képi információban sok visszatérő motívum és periodicitás van, amitől a tömörítés során "megszabadulhatunk". A tömörítés hatékonysága azon múlik, milyen "ügyesen" tudjuk a primer információ korrelációs tulajdonságait kihasználni. A statisztikus függőség kihasználására különböző módszerek léteznek:

- Tervezhetünk egy prediktort, amely a múlt alapján "megjósolja" a jövőt. Ebben az esetben csak a prediktor együtthatói, a kezdeti állapot, valamint a kvantált predikciós hiba átvitelére van szükség, amely jelentős sávszélesség csökkenést eredményezhet. Mivel a kvantálás információvesztéssel jár, ezért ez egy "veszteséges" módszer [1.4.6, 1.4.8].
- Az információs folyamat korrelációs tulajdonságait analizálva, eldönthető, hogy mi a fontos része az információnak és mi kevésbé fontos. A fontos részek (szakmai terminológiával élve: a főkomponensek) meghagyásával és a mellék komponensek eldobásával szintén tömöríthető az eredeti információ. Ezek a módszerek szintén veszteségesek és matematikailag a főkomponens analízisből, illetve a transzformációs alapú eljárásokból erednek [1.4.4].
- Megfigyelvén az egyes bitsorozatok relatív gyakoriságát egy forrásfájlon belül, az információelméletből ismert adaptív forráskódolási (más néven entrópiakódolási) algoritmusok alkalmazhatóak [1.4.3].
- Azért, hogy egy forrásfájlban az ismétlődő sztringektől megszabaduljunk, valaki felállíthat egy "szótárt", amely a szövegben leggyakrabban előforduló sztringeket tartalmazza, majd a különböző sztringeket helyettesítheti a szótár megadott helyére utaló referenciákkal (pl. hely és hosszúság) [1.4.9, 1.4.6].

1.4.2. Algoritmusok

Ebben a fejezetben különböző tömörítő eljárásokat fogunk bevezetni. A hangsúly inkább az alapgondolat világossá tételén van, semmint a formálisan precíz tárgyaláson. Az algoritmusok részletes leírása megtalálható [1.4.6, 1.4.7] munkáiban.

Az adattömörítés veszteséges módszerei

Ebben a részben az analóg adat sávszélességének a csökkentésére szolgáló eljárásokat vesszük sorra, ahol a tömörítés ára az információveszteség.

Differenciális PCM

Mivel az információs folyamok (pl. beszédminták) mintái között erős korreláltságot találtak, ez arra utal, hogy a múltbeli minták „többé kevésbé" meghatározzák a beszédminta jelenbeli értékét. Ebben az esetben viszont elég csak a beszédjel múltbeli értkeihez képesti változását kódolni, hiszen így valószínűleg sok bitet megspórolunk. (Mivel az erős korreláltság miatt a múltbeli értékek statisztikailag meghatározzák a jelenbeli mintát, a változás meglehetősen kis információ tartalmú lehet, ezért a változás kódolásához valóban kevesebb bitre van szükség, mint ha magát a mintát kódolnánk. Extrém esetben akár elég azt jelezni, hogy nő-e, vagy csökken-e a jel, ami 1 bit információt igényel, lásd „Delta moduláció").

Azt a kódolási eljárást, ami az információs folyamat időbeli korreláltságának a kihasználásán alapul, differenciális PCM-nek (D-PCM) hívják. A kódoló az AutóRegresszív (AR) modell alapján jósolja meg a folyamat értékét az egyes időpillantokban. Nevezetesen, minden minta az ezt megelőző, a *M* db. múltbeli minták alapján kerül becslésre, a következő formában $\hat{x}_n = \sum_{j=1}^{M} w_j x_{n-j}$. A $w_j, j = 1,...,M$ együtthatók úgy választandóak meg, hogy minimálissá tegyék az előrejelzés $E(x_n - \hat{x}_n)^2$ négyzetes középhibáját.

$$\mathbf{w}_{opt} : \min_{\mathbf{w}} E\left(x_n - \sum_{j=1}^{M} w_j x_{n-j} \right)^2$$

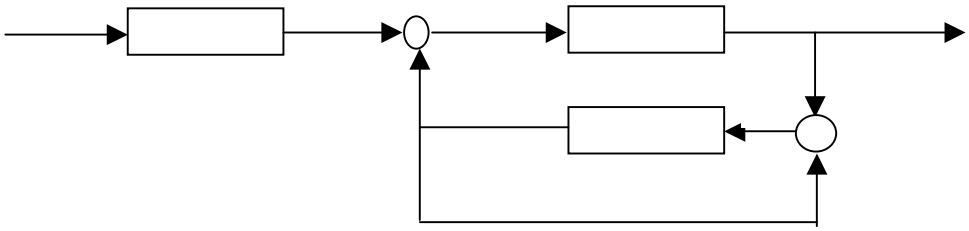
Ez a feladat egy lineáris egyenletrendszer megoldásához vezet (amit gyakran Yule-Walker egyenleteknek neveznek)

$$\mathbf{R}\mathbf{w}_{opt} = \mathbf{r}\,, \qquad (1.4.1)$$

ahol $r_i = E(x_n x_{n-i})$ és $R_{ij} = E\left(x_{n-i} x_{n-j}\right)$, i,j=1,2...M

A fenti egyenletrendszer megoldására Levinson és Durbin fejlesztett ki hatékony, rekurzív megoldást [1.4.2].

A DPCM kódolót az 1.4.2. ábra mutatja.



A DPCM kódolót/dekódolót leíró egyenletek a következőek:

a kvantáló bemeneti jele $e_n = x_n - \hat{\tilde{x}}_n = x_n - \sum_{j=1}^{M} w_j \tilde{x}_{n-j}$, míg $\tilde{e}_n$ jelöli a kvantált hibát és $\hat{\tilde{x}}_n = \sum_{j=1}^{M} w_j \tilde{x}_{n-j}$.

(A következő tárgyalás során a változó fölé írt "hullám" mindig a kvantált mintákat jelöli. Figyeljük meg, hogy a prediktor kvantált minták alapján dolgozik, ezért van szükség a "kalap" és "hullám" együttesére.)

A DPCM-el a beszéd eredetileg 64 Kbit/sec-es sávszélessége (PCM), 48 Kbit/sec-re, vagy ez alá csökkenthető.

Adaptív DPCM (ADPCM)

Eddig gyengén stacionárius folyamatokat feltételeztünk, amelyeknek a korrelációs tulajdonságai ismertek. Az (1.4.1)-es egyenlőség ezen feltétel mellett igaz. A valóságban azonban a beszéd és videó folyamatok nem stacionerek, ezért időben változó korrelációs tulajdonságaik vannak. A stacionaritás hiányának egy következménye, hogy az egyenletes lépésközű kvantáló időben változó jel-zaj viszonyt produkál, attól függően, hogyan változik a bemenetére kerülő véletlen jel

statisztikája. Ennek kiküszöbölésére, adaptív kvantálót használnak, amelynél a kvantálási lépcső nagysága (amit Δ jelöl) nem állandó, hanem az idő függvényében változik Ezt a szabályt a $\Delta_{n+1} = \Delta_n M(n)$ formula írja le, ahol *M(n)* egy szorzófaktor, ami az éppen kvantált minta amplitúdójától függ. A beszédejelre vontakozó *M(n)* értékeket Jayant adta meg [1.4.6]. Ezeket 2, 3 és 4 bites adaptív DPCM-re a következő táblázat foglalja össze:

M(n) értéke	2 bites kvantáló	3 bites kvantáló	4 bites kvantáló
M(1)	0,80	0,90	0,90
M(2)	1,60	0,90	0,90
M(3)		1,25	0,90
M(4)		1,70	0,90
M(5)			1,20
M(6)			1,60
M(7)			2,00
M(8)			2,40

A (1.4.1)-es egyenlőség szintén csak a gyenge stacionaritás feltétele mellett igaz. Ahogy az előbb említettük a beszéd és videó folyamatok nem stacionérek, ezért időben változó korrelációs tulajdonságaik vannak. Ezért nemcsak a kvantálási lépésköznek, hanem a pediktornak is adaptívnak kell lennie, azaz konkrét korreláció ismeretének hiányában ezt a bejövő minták alapján kell megbecsülnie. Ez a következő algoritmus szerint történik:

$$w_l(k+1) = w_l(k) - \delta\left(x_n - \sum_{j=1}^{M} x_{n-j} \right) x_{n-l} \quad l = 1,2..M \ , \qquad (1.4.2)$$

ahol δ az ún. "tanulási paraméter". Itt az *n* index továbbra is a megfelelő idősor indexét reprezentálja, míg *k* tanulási algoritmus *k*-ik lépépésére utal. Ezt az algoritmust Robbins-Monroe típusú sztochasztikus approximációnak hívják. Az eljárásra vontakozóan az érdeklődő részletes konvergencia-analízist találhat [1.4.4] munkájában.

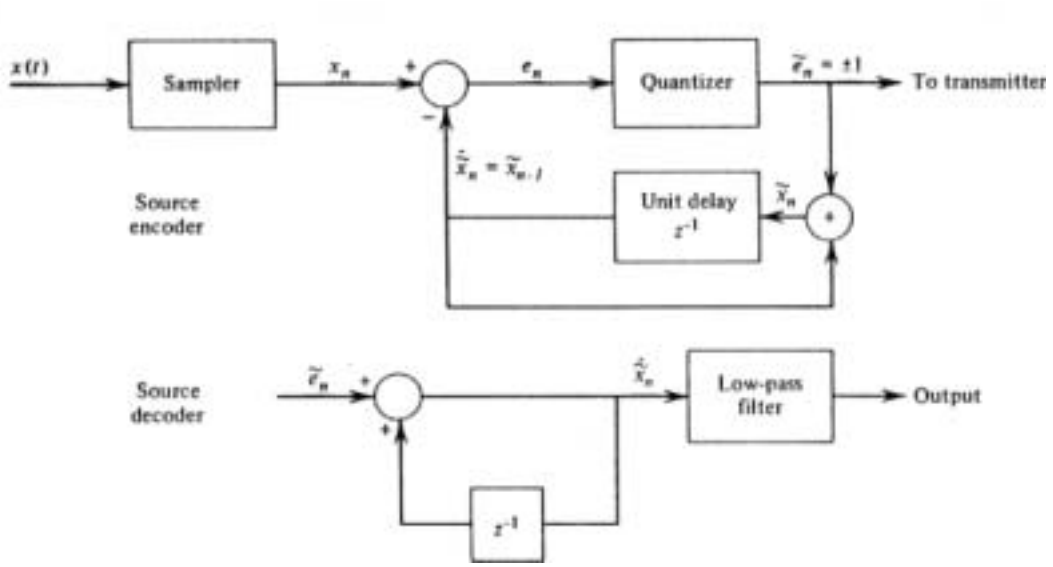
A (1.4.2) összefüggés alkalmazása során a prediktor együtthatókat is időről időre át kell vinni a vételi oldalra, ami sajnos plusz sávszélességet igényel. Ezért sokszor a vételi oldalnak van egy a „saját" prediktora, ahol a koefficensek adaptálása nem x_n, hanem a detektált $\tilde{x}_n$ minta lapján történik. Mivel $\tilde{x}_n$ a kvantálási zajjal

különbözik x_n-től, az ilyen típusú algoritmusok konvergenciájának a bizonyítása még ma is kutatások tárgyát képezi.

Az ADPCM modulációt nemcsak beszéd, hanem képtömörítésre is lehet használni. Itt a képjelek kettős korrelációját lehet kihasználni. Ugyanis nemcsak a szomszédos képpontok, hanem az időben egymás után jövő keretek is korreláltak. Ezért a DPCM mind intra-, mind interframe kódolásra használható. Az algoritmus leírása hasonló a beszédejelnél megismertekhez.

Delta moduláció

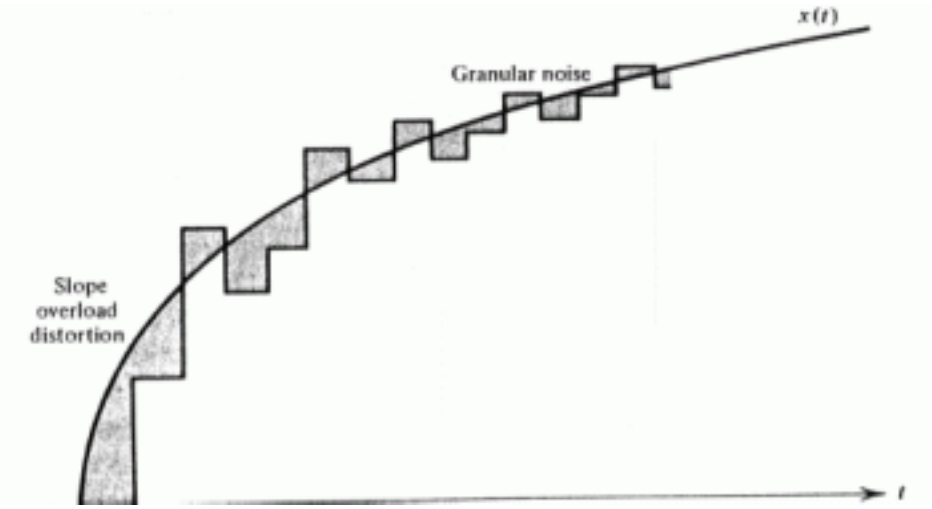
A legegyszerűbb módszer az időbeli korreláltság kihasználására a delta moduláció (DM). Ez az eljárás egy {0,1} kétállapotú kvantálón alapszik, illetve egy elsőrendű prediktoron (a megjósolt jel csak az eggyel előző minta alapján adódik). Mivel a kódoló egy diszkrét differenciáló gyanánt működik, a dekódoló egy diszkrét integrátor. A megfelelő architektúrákat az 1.4.3. ábra mutatja.



1.4.3. ábra A delta modulációs rendszer kódolója és dekódolója

A delta modulátor ekvivalens az eredeti jelformának egy lépcsős approximációjával. Ahhoz, hogy jó minőségű közelítést kapjunk az eredeti jelformának lassan kell változnia a mintavételi frekvenciához képest. Ez maga után vonja, hogy a deltamodulátorban a Nyquist frekvencia négyszeresére vagy ötszörösére állítják be a mintavételi frekvenciát.

A deltamodulátor esetén kétfajta torzítás léphet fel. Ez elsőt "granuláris zajnak" hívják, amely a kvantálási lépcső nagyságával kapcsolatos. A másik a véges "követési" képességből adódik, ezt "meredekség túlterheltségi" torzításnak szokták nevezni. Az 1.4.4. ábra ezen torzítások hatását demonstrálja.



1.4.4. Ábra. A DM kétféle torzításnak a szemléltetése

A fenti torzítások csökkentése végett adaptív delta modulátorokat alkalmaznak, amelyben a kvantálási lépcső mérete adaptívan változik a következő szabály szerint [1.4.8]:

$$\Delta_n = \Delta_{n-1} K^{\tilde{e}_n \tilde{e}_{n-1}}$$

ahol a $K \geq 1$ állandó a torzítások minimalizálása szerint van beállítva.

Az irodalomban számos más adaptív delta modulátor ismert [1.4.8], pl. az ún. „folytonosan vátozó meredekségű” deltamodulátor, ahol a kvantálási lépcső állítási szabálya a következő: $\Delta_n = \alpha\Delta_{n-1} + \beta_1$ ha $\tilde{e}_n$, $\tilde{e}_{n-1}$ és $\tilde{e}_{n-2}$ előjele megegyezik, vagy $\Delta_n = \alpha\Delta_{n-1} + \beta_2$ máskülönben.

Transzformációs alapú algoritmusok képek tömörítésére

A transzformációs alapú algoritmusok alapvetően "veszteséges" módszerek, amelyek a forrásinformáció egy részét "eldobják" a sávszélesség csökkentése érdekében. Természetesen az információ "eldobása" kontrollált módon kell, hogy lezajlódjon (pl. nem sok felhasználó tolerálná, ha a híradó képeinek az alsó egynegyedét kivágnák). Ahhoz, hogy megtudjuk a képi információ melyik része

felesleges (és ezért eldobható), valamilyen fontossági sorrendet kell felállítani az eredeti információs folyamban. Azaz az eredeti információt egy olyan reprezentációba kell áttranszformálnunk, amely jól tükrözi, hogy "mi a fontos" és "mi a nem fontos" része az információnak. Ezek után a "nem fontos" részek eldobhatók, egészen addig, amíg el nem érjük a minőségromlásnak egy előre adott korlátját. A fenti elv szerint a dekompresszió (az eredeti információ visszaállítása a tömörített verzióból) a tömörítés szerint "csonkolt" reprezentáción az inverz transzformáció végrehajtását jelenti. Mivel ezt a módszert alapvetően a képi információ tömörítésére használják, ezért a következőkben mindig képtömörítésről fogunk beszélni.

A transzformációs módszer általános blokkdiagramját az 1.4.5. ábra szemlélteti:

A kép reprezentálására (ami alapján eldönthető, hogy mi képezi a fontos és kevésbé fontos részét a képi információnak) többfajta matematikai transzformáció szolgálhat. A jelen alfejezet ebből a két legismertebbet, a Karhunen-Loeve sorfejtést (Karhunen-Loeve Expansion = K LE), illetve a diszkrét koszinusz transzformációt (Discrete Cosine Transformation = DCT) fogja tárgyalni.

Adattömörítés főkomponens analízissel

Ez a módszer az eredeti információ Karhunen-Loeve sorfejtésén alapszik. A képet a továbbiakban egy **x** valószínűségi vektorváltozó jelöli (ahol **x** oszlopvektort, míg $\mathbf{x}^{\mathrm{T}}$ sorvektort jelöl), amelynek a dimenziója *N*. A korrelációs mátrix $\mathbf{R} = E\left(\mathbf{x}\mathbf{x}^{\mathrm{T}}\right)$ ($\mathbf{x}\mathbf{x}^{\mathrm{T}}$ a külső, vagy más szóval diadikus szorzatot jelöli). A korrelációs mátrix hermitikus, ezért a sajátértékei valósak, a sajátvektorai pedig ortogonálisak egymásra. A sajátértékek és sajátvektorok jelölése a következő:

$\mathbf{R}\mathbf{s}^{(i)} = \lambda_i \mathbf{s}^{(i)}, i = 1,.., N$. Mivel az $\mathbf{s}^{(i)}, i = 1,..., N$ sajátvektorok ortonormált bázist alkotnak, ezért **x** rekonstruálható, mint $\mathbf{x} = \sum_{i=1}^{N} y_i \mathbf{s}^{(i)}$, ahol $y_i = \mathbf{s}^{(i)T}\mathbf{x}$.

A kompresszió menete ezután a következő:

Határozzuk meg és rendezzük a korrelációs mátrix sajátértékeit monoton csökkenő sorrendbe $\lambda_1 > \lambda_2 > ... > \lambda_M > ... > \lambda_N$. Ezek után a megfelelő sajátvektorokkal határozzuk meg az $(y_1,..., y_M)$ komponenseket. Az információveszteséget az $(y_{M+1},..., y_N)$ komponensek eldobása okozza. Ezek után a de-kompresszált kép megkapható az $\tilde{\mathbf{x}} = \sum_{i=1}^{M} y_i \mathbf{s}^{(i)\mathbf{T}}$ összefüggés alkalmazásával.

Be lehet bizonyítani (lásd [1.4.4]), hogy $E\left(\left\|\mathbf{x} - \tilde{\mathbf{x}}\right\|^2\right) = \sum_{i=M+1}^{N} \lambda_i$. Mivel a sajátértékek monoton csökkenő sorrendben voltak elrendezve, ezért a négyzetes hiba, ami a komponensek eldobása miatt keletkezik, minimális ($\sum_{i=M+1}^{N} \lambda_i$ a legkisebb összeg, mert a legkisebb sajátértékeket tartalmazza).

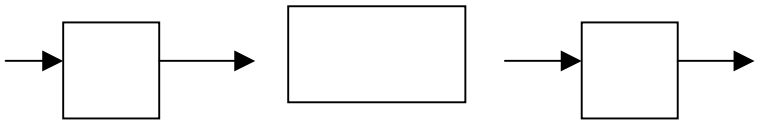
Azaz, a főkomponens analízis megvalósítja a bevezetőben vázolt általános célkitűzést, hiszen a KLT segítségével a transzformációs tartományban eldönthető, "mi a fontos" (a nagy sajátértékekhez tartozó sajátvektorok irányába mutató komponensek), illetve mi "nem fontos" (a kis sajátértékekhez tartozó sajátvektorok irányába mutató komponensek). Ezt a fontossági sorrendet a négyzetes hiba minimalizálása motiválta, amely alapján a megfelelő komponensek eldobhatók. A fentiek alapján, ha valaki egy meghatározott ε minőségű tömörítést akar végrehajtani (ahol a minőséget az $E\left(\left\|\mathbf{x} - \tilde{\mathbf{x}}\right\|^2\right) \le \varepsilon$ egyenlőtlenség teljesítése jelenti), akkor a következő eljárást kell megvalósítania:

1. Határozza meg az **R** korrelációs mátrix sajátvektorait $S = \left\{\mathbf{s}^{(i)}, i = 1,..., N\right\}$ és sajátértékeit $\Lambda = \left\{\lambda_i, i = 1,..., N\right\}$.
2. Rendezze a sajátértékeket csökkenő sorrendbe $\lambda_1 > \lambda_2 > ... > \lambda_M > ... > \lambda_N$.
3. Határozza meg azt az *M* számot, amelyre $\sum_{i=M+1}^{N} \lambda_i \le \varepsilon$.

4. Tömörítse a forrásinformációt az $y_i := \mathbf{x}^{\mathbf{T}}\mathbf{s}^{(i)}$ i=1,...,M komponensek kiszámolásával. (az $\mathbf{y} = (y_1,...,y_M)$ vektor most már a kép "keskenysávú" megfelelője, valamint az elért tömörítési arány *M/N*)

5. A rekonstruált kép az $\tilde{\mathbf{x}} = \sum_{i=1}^{M} y_i \mathbf{s}^{(i)\mathbf{T}}$ inverz transzformáció alkalmazásával állítható elő.

A főkomponens alapú tömörítés blokkvázlatát az 1.4.6. ábra mutatja.



A fenti sémában a **K** és **L** mátrixok, amelyeknek a típusai *MxN* illetve *NxM*, míg a mátrixok egyes elemei a következőképpen vannak definiálva: $K_{ij} = s_j^{(i)}; i = 1,...,M\ ;\ j = 1,...,N$ és $L_{ji} = s_j^{(i)}\ \ j = 1,...,N\ \ i = 1,...,M$. Mivel a transzformációk lineárisak, ezért a nagysebességű képtömörítés a modern DSP technikák és architektúrák alkalmazásával lehetővé válik.

Mindazonáltal az előző módszerrel kapcsolatban felmerülhet néhány nehézség:

- a forrásinformáció korrelációs mátrixa nem ismert a priori, hanem ezt az egymásután következő keretekből kell megbecsülni;
- a főkomponens analízis a sajátértékek és sajátvektorok meghatározását igényli, ami numerikusan igen fárasztó feladat.

Ezen nehézségek elkerülésére számos módszert dolgoztak ki, pl. adaptív, "kernel-alapú" főkomponensanalízis, vagy az Oja által kifejlesztett hebbi alapú neurális főkomponens analízis. Az olvasó egy részletes áttekintést talál ezekről a módszerekről [1.4.4] munkájában.

Képtömörítés a diszkrét koszinusz transzformáció segítségével

Ebben az esetben a transzformáció, amivel a forrásinformációt új tartományba képezzük le a 2-dimenziós DCT. Az alapötlet az, hogy a transzformált tartományban megszabaduljunk a felesleges, nagy térbeli frekvenciával bíró komponensektől.

A DCT definíciója a következő:

$$\hat{f}(u,v) = \sum_{j,k=0}^{N-1} f(j,k)C(u)\cos\frac{(2j+1)u\pi}{2N}C(v)\cos\frac{(2k+1)v\pi}{2N}$$

ahol $f(j,k)$ reprezentálja a képi információ síkbeli amplitudóját (a *(j,k)* pixelhez tartozó képjel amplitúdója), valamint $C(0) = 1/\sqrt{N}$, $C(u) = 2/\sqrt{N}$ és $C(v) = 2/\sqrt{N}$ ha $u \neq 0$. Az inverz transzformáció a következőképpen számolható:

$$f(i,j) = \sum_{j,k=0}^{N-1} \hat{f}(u,v)C(u)\cos\frac{(2j+1)u\pi}{2N}C(v)\cos\frac{(2k+1)v\pi}{2N}.$$

Jelen esetben adattömörítés úgy érhető el, hogy a DCT "farkát" (nagyfrekvenciás komponenseit) agresszívebben dobjuk el, azon feltételezésen alapulva, hogy a nagyfrekvenciás komponensek nem számottevő részletet adnak csak hozzá a képhez. Pl. a JPEG kódolás (a szabvány részletesebb leírását lásd később) ezen az alapelven alapul és a következő lépéseket követi:

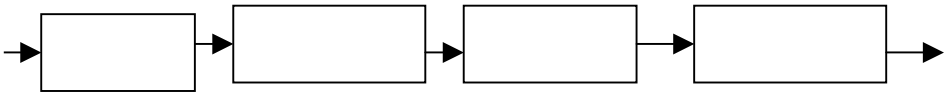
1. Az információ lokális feldolgozása: Ossza fel az eredeti képet 8x8-as blokkokra.

2. Transzformáció: Minden blokkon hajtson végre egy DCT-t a síkbeli frekvenciák eloszlásának a meghatározására.

3. Kvantálás: Kerekítse a DCT együtthatóit, úgy hogy "agresszívebb" kerekítést szenvedjenek a nagyfrekvenciás komponensek együtthatói.

4. Kódolás: Hajtson végre Huffman kódolást a digitalizált együtthatókon további tömörítés végett (ebben a lépésben nincs további információvesztés).

A dekompresszálás a következő lépéseket tartalmazza:



A JPEG részletes tárgyalásához lásd [1.4.5].

Az adattömörítés veszteségmentes algoritmusai

Ebben a szakaszban az adattömörítés veszteségmentes eljárásait foglaljuk össze. Ezen eljárások legfőbb tulajdonsága, hogy a forrásadat teljes mértékben visszaállítható a tömörített verzióból, azaz nincs információveszteség. Ezen módszerek teljes tárházat részletesen tárgyalják a [1.4.1, 1.4.3 ,1.4.7] irodalmak. Két alapvető megközelítés létezik "veszteségmentes" tömörítés esetén:

- helyettesítési módszerek, amelyeket az információelmélet statisztikus eredményei motiválnak;
- szótáralapú módszerek, amelyek az eredeti információra egy szótár, valamint alapján, pointerek és hosszúságok szerint referálnak.

Helyettesítési eljárások

Tételezzük fel, hogy az eredeti forrásinformáció egy bináris fájl, amelyet *W*-vel jelölünk. A tömörített fájlt jelölje *U.* A veszteségmentes tömörítés annyit jelent, hogy *W* teljesen (hiba nélkül) visszanyerhető *U*-ból. Néha *W* akár napokig rendelkezésre áll, hogy hosszú meditáció után döntsük el mi a megfelelő tömörítési eljárás (pl. archivált anyagoknál). Máskor viszont *W* egy nagysebességű adatfolyam, ami több Mbps sebességgel "kerül" a tömörítő bemenetére. Ezért egy kompressziós algoritmust nemcsak az elért tömörítési arány minősít, hanem a tömörítés, mint jelfeldolgozási műveletnek a működési sebessége, illetve az algoritmus futtatásához szükséges adatok mennyisége is. Nem jó az a tömörítési módszer, amelynél a dekompresszióhoz legalább annyi adatot kell tárolni, mint maga *W*.

A helyettesítésen alapuló tömörítés algoritmusa a következő:

Tételezzük fel, hogy az ún. "elválasztószavak{mert sokfajat módon feloszthtató egy szíveg elválasztószavakra }" $S = \{s_1,...,s_m\}$ halmaza adott. Azt mondjuk, hogy *W* elválasztó szavakra van felosztva (elválasztva), ha megadható mint a következő sorozat $W = (s_{i_1}, s_{i_2},..., s_{i_v}),\ s_{i_j} \in S$. Feltételezzük továbbá, hogy adott egy kódolási séma $C : S \to R$, amely az elválasztószavakhoz kódszavakat rendel hozzá (ahol a kódszavak halmazát *R* jelöli, $R = \{r_1,...,r_m\}$). Helyettesítvén az eredeti fájl minden s_{i_j} elválasztószavát a megfelelő r_{i_j} kódszóval, a tömörített $U = (r_{i_1}, r_{i_2},..., r_{i_v}),\ r_{i_j} \in R$ fájlt kapjuk. Ha az r_i kódszavak átlagos hossza kisebb mint az s_i elválasztószavak átlagos hossza, akkor *U* jelentősen rövidebb lehet, mint *W*.

Például tételezzük fel a következő elválasztószavakat: $s_1 = 0\ \ s_2 = 10\ \ s_3 = 110$ $s_4 = 1110\ \ s_5 = 11111$. Ugyanakkor legyen *W* egy 30 bit hosszúságú fájl

$W = 11111011111110111010111101110110$. Ekkor *W* "elválasztható" a következő sorozatba $W = s_5 s_2 s_5 s_4 s_4 s_5 s_1 s_4 s_3$. Ha következő kódolást alkalmazzuk, $r_1 = 1111\ r_2 = 1110\ \ r_3 = 110\ \ r_4 = 10\ \ r_5 = 0$, akkor $U = 01110010100111110110$. Látható,

hogy *U* 20 bit hosszúságú, azaz ezen az elemi példán is demonstrálható a tömörítés ténye.

A helyettesítési módszer hatékony alkalmazásához információelméleti megfontolásokat alkalmazhatunk. Pontosabban a veszteségmentes és hatékony tömörítés eléréshez a következő feltételeknek kell teljesülnie:

1. A kódszavak *R* halmaza olyan kódszavakat kell, hogy tartalmazzon, amelyek kielégítik a prefix mentesség feltételét (egyik kódszó sem előtagja a másiknak, a veszteségmentesség elérése érdekében)
2. Olyan kódot kell konstruálni, amely a gyakran előforduló elválasztószavakhoz rövid, míg a ritkán előforduló elválasztószavakhoz hosszú kódszavakat rendel.

Ennek megfelelően, az információelmélet bármelyik tradicionális forráskódolási algoritmusa (Shannon-Fano, Huffman, aritmetikus kódok) alkalmazható, amennyiben a szükséges statisztikák a forrásfájlból feldolgozás közben becsülhetőek.

Ezért ezen módszerek alapvető kérdése, hogyan becsülhető meg az elválasztószavak valószínűség-eloszlása. A legegyszerűbb módszernek a *W*-ben való előfordulások relatív gyakoriságának a meghatározása tűnik. Ennek alapján a következő statisztika nyerhető:

$$f_i = \frac{\text{az } s_i \text{ elválasztószó előlőfordusainak száma a forrásfájlban}}{\text{a forrásfájlban lévő elválasztószavak teljes száma}}$$

Az előző példában $f_1 = \frac{1}{9}$ $f_2 = \frac{1}{9}$ $f_3 = \frac{1}{9}$ $f_4 = \frac{3}{9}$ $f_5 = \frac{3}{9}$. Sajnos ennek a módszernek a hátránya, hogy a becslés nem biztosan ad pontos értéket, főleg rövid fájlok esetén. Nagysebességű, valós idejű tömörítés estén a relatív gyakoriságokat célszerű rekurzív módon becsülni. Adaptív Huffman kódolás esetén a kódfa adaptív "karbantartására" is szükség van, amely Gallager és Knuth algoritmusai alapján történhet. Az olvasó részletes leírást [1.4.3] munkájában talál.

Szótáralapú módszerek

A szótáralapú módszerek a kifejezések egy listáját (a szótár) használják és a forrásszöveg részeit a listára vonatkozó mutatókkal helyettesítik. Tömörítés akkor történik, ha a mutatók kisebb tárolási helyet igényelnek, mint maga a szöveg. Természetesen a mutatók mellett magának a szótárnak a tárolásáról, vagy átviteléről

is gondoskodni kell. Mivel egy kommunikációs rendszerben gyakran a valós idejű tömörítés a cél, ezért a szótárkészítésnek és a mutatók elhelyezésének adaptív módon kell történnie.

Az első adaptív szótáralapú módszerek a hetvenes évekre nyúlnak vissza, amikor Ziv és Lempel "mérföldkőnek" számító eredményeiket publikálták. A megfelelő algoritmusokat LZ77-nek és LZ78-nak nevezték el. Számos alkalmazás alapul az LZ77-en (pl. LHarc, PKZIP, GNU zip, vagy Info-Zip), illetve az LZ78-on (MODEM kommunikáció tömörítő algoritmusa, valamint a GIF grafikai formátum). Az alapötlet bemutatásához az LZ77 típusú mozgóablakos sémát használjuk. Ebben az esetben egy két részből álló ablak "halad" végig a szövegen, amelynek az első részét "históriának", míg a másodikat "előrenézőnek" nevezik. Pl.:

```
                    history          lookahead
...She sells sea shells by the seashore...
       09876543210987654321
```

A legegyszerűbb esetekben a história és az előrenéző fix hosszúságú és a szótár az összes olyan kifejezést tartalmazza, amelynek hossza nem haladja meg az előrenéző hosszát. A tömörítő úgy működik, hogy az előrenéző kezdeti szakaszát egy a szótár adott részére referáló mutatóval helyettesítjük, majd tovább mozgatjuk az ablakot. Az alábbi példában (lásd a következő ábrát) a kezdeti szegmens "he" illeszkedik a szótárban lévő "shell" szó első két karakteréhez (a 10-ik karakter a szótárban). Az algoritmus egy (*koordináta, hosszúság, karakter*) hármast ad eredményül, ahol a harmadik komponens a forrás következő karakterét adja. A jelen példában ez a számhármas a (10,2,_), ahol a "_" a szóköz karakterre utal. A karakter komponens engedi meg, hogy akkor is tovább tudjon haladni az algoritmus, ha nincs a históriában illeszkedő szó, amire referáló mutatót el lehetne helyezni

A következő ábra az LZ77 algoritmus működésének néhány lépését mutatja:

```
                    history          lookahead          output
...She sells sea shells by the seashore    (10,2)
.She sells sea shells by the seashore..    (18,4)
 sells sea shells by the seashore.....     (17,2)
ells sea shells by the seashore.....       'o'
       09876543210987654321
```

Az implementációhoz kapcsolódó néhány tipikus értéket talál az olvasó a következő táblázatban:

a história hossza	az előrenéző hossza	koordináta ábr.	hossz ábr.
4096 byte	4096 byte	12 bit	4 bit

A *koordináta* és a *hossz* megkülönböztetésére a karaktertől egy darab jelző bitet használnak. Ezért a *koordináta* és a *hossz* átvitelének a költsége 17 bit. Ezért csak olyan forrásszöveg részeket érdemes mutatóval helyettesíteni, amelyek hossza nagyobb mint 2 bájt.

A következő táblázat az egyes algoritmusok hatékonyságát mutatja (kompressziós arány szerint) az ún. "Calgary corpus"-on. A Calgary corpus a tipikus fájlok egy gyűjteménye, amelyen teljesítőképesség szempontjából össze szokták hasonlítani a különböző tömörítő algoritmusokat.

Fájl	KByte	Az LZ 77által elért kompressziós arány
bib	109	43 %
book1	751	51.8 %
Geo	100	80.3 %
Obj1	21	57.2 %
pic	501	20.9 %
progc	39	41.7 %
Átlag		49.2 %

Az LZ78 esetében a forrásfájl kifejezésekre van felosztva. Egy új kifejezésnek a szótárba foglalása esetén egy darab karaktert adunk hozzá egy, már létező kifejezéshez. Ezt a metódust a UNIX Compress, V.42*bis* MODEM kommunikációban, valamint a GIF grafikai formátumokban alkalmazzák.

Más technikák, pl. vektorkvantálás (lásd [1.4.4]), szintén hatásos kompressziót eredményeznek.

1.4.3. Szabványok és szabadalmak

A beszédjelet hagyományosan adaptív prediktív technikákkal tömörítik és az idevágó kódoló neve ADPCM. A gyakorlatban a 4 bites ADPCM-et használják a leggyakrabban, amely 32 kbps és 24 kbps alkalmazásokhoz illeszkedik, szemben az eredeti 64 kbps-el. A 32 kbps tömörítést az ITU G.726 szabvány írja le, ez 4 bites kvantálót tartalmaz.

Mobil telefóniában a GSM 06.10 transzkódolás a használatos Európa szerte, ez 13 kbps sávszélességet eredményez, a hosszú távú predikciót a "regular pulse excitation"-al kombinálva.

IP alapú hangátvitel esetén a G.7x családot alkalmazzák, mely néhány esetben akár 6,4 Kbps sebességet is elérhet (tizede a szabványos PCM hang

sávszélességnek), azonban ezek a tömörítő eljárások nagyteljesítményű jelfeldolgozó apparátust igényelnek.

A zene tömörítéséről beszélve a szabványos Music Instrument Digital Interface (MIDI)-t kell megemlíteni, amelyet a zeneipar nagyrésze adaptált. A MIDI üzenet zeneileg "szignifikáns" eseményeket visz át, pl. egy billentyű lenyomódott ...stb. Minden hangszer rendelkezik egy MIDI kóddal (pl. a zongora 0-ás, a hegedű a 40-es). Összesen 127 hangszernek van kódja. A MIDI rendszerben a számítógép a kapott üzenetek alapján egy szintetizátorral generálja a zenét. A szintetizátor mind a 127 hangszert ismeri és a megfelelő spektrumot generálja a kapott utasítások alapján. A MIDI előnye a hatalmas sávszélességbeli csökkenés. A hátránya viszont, hogy a generált zene minősége nagyban függ a szintetizátortól.

Az állóképek tömörítése a jól ismert JPEG (Joint Photographic Expert Group) szabvány alapján történik. A JPEG 640x480 pixel méretű 24-bit RGB videó jeleket tömörít. Először a világossági- és a színjelet határozza meg az RGB képből. A színjelet kétszeres faktorral tömöríti, hiszen az emberi szem kevésbé érzékeny a szín információra. Ezután a világossági és színmátrix 8x8-as blokkokra van felosztva, majd minden blokkon egy DCT-t hajtanak végre. A kevésbé fontos spektrális komponensektől kvantálással szabadulnak meg. Ezután egy "run length encoding", majd Huffman kódolás következik. A JPEG szabvány a komplikáltsága ellenére előnyös, hiszen képes a 20:1 tömörítési arány elérésére.

A mozgóképek tömörítésére az MPEG (Motion Picture Experts Group) szabvány terjedt el. Az első szabvány (MPEG-1) az NTSC képjelet 1.2 Mbps sávszélességre tömörítette. Az MPEG-1-et még filmek tárolására használják CD ROM -on a CD-I és CD-Video formátumban. Ezután az MPEG-2 került kifejlesztésre, amely a TV műsorszórásban az NTSC és PAL jeleket tömörítette 4-6 Mbps sávszélességre. Ezt a nagyfelbontású HDTV alkalmazásokra is kiterjesztették. Az MPEG-4-et közepes felbontású videokonferenciára használják. Az MPEG-ben a kép és hang együttesen kerül tömörítésre egy 90 kHz rendszer-órajel segítségével. A hangot 32 kHz, 44,1 KHz vagy 48 KHz frekvenciával mintavételezik, majd FFT, nemlineáris kvantálás, illetve Huffman kódolás következik. A képjel tömörítésénél az MPEG a JPEG-nél megismert módon szabadul meg a felesleges spektrális komponensektől. Mivel az időben egymásután következő keretek nem sokban különböznek egymástól, ezért az MPEG az időbeli korrelációt is kihasználja a

tömörítés során. Ez általában az egyes blokkok időben egymásból való kivonásával történik.

A szótáralapú módszerek esetén a GIF grafikai formátum az LZ78 algoritmuson alapul (Lempel, Ziv, Welch). Ez az algoritmus azért is "hírhedtté" vált, mert ebben az esetben az intellektuális tulajdonhoz kapcsolódó jogi viták is felmerültek. Az IBM és Unysis alkalmazta a módszert. A vita a körül forgott, vajon egy algoritmus szabadalmaztatható vagy sem. Ezideáig a US Patent Office szerint egy eljárás még nem, de annak egy implementációja szabadalmaztatható. Ez valóban egy ellentmondás, ami az algoritmikus fejlesztők körében jogi megoldást igényel.

1.4.4. Összefoglaló

Ebben a fejezetben a különböző adattömörítési eljárásokat foglaltuk össze. A "veszteséges" módszerek főleg a hang és a kép tömörítésére vonatkoznak, míg a "veszteségmentes" algoritmusokat eredetileg adattömörítésre fejlesztették ki. A kapcsolódó gondolatokat alapelvek, algoritmusok, szabványok és szabadalmak szintjén tárgyaltuk.

Meg kell jegyezni, hogy manapság nagy erőkkel folyik a kutatás a wavelet és fraktálanalízis alapján történő képtömörítés területén. Sajnos a fejezet limitált terjedelme miatt ezek a módszerek a fenti leíráson kívül estek.

1.4.5. Az adattömörítés szempontjából fontos web oldalak listája

Fraktális képtömörítéshez: http://inls.ucsd.edu/Research/Fisher/Fractals

GNU projektekhez: http://www.gnu.ai.mit.edu

Az Info-Zip-ről: http://www.cdrom.com/pub/infozip/

A JPEG-ről: ftp://ftp.uu.net/graphics/jpeg

A waveletalapú tömörítésről: http://www.cs.dramouth.edu/~gdavis.

A PNG és LZ77-ről: http://www.wco.com/~png/

Irodalom:

[1.4.1] Bell, T., Witten, J.: *Text compression*, Prentice-Hall, 1990

[1.4.2] Durbin, J.: "Efficient estimation of parameters in moving average models", *Biometrika*, Vol. 46, 1959

[1.4.3] Hankerson, D., Harris, A., Johnson, P.*: Introduction to information theory and data compression*, CRC Press, 1998

[1.4.4] Haykin, S.: *Neural networks - a comprehensive foundation*, Prentice-Hall, 1999.

[1.4.5] Independent JPEG Group, JPEG SW Release 6. ftp://ftp.uu.net/graphics/jpeg

[1.4.6] Jayant, N., Noll, P.*: Digital coding of waveforms, principles and applications to speech and video*, Prentice-Hall, 1984

[1.4.7] Nelson, M., Gailly, J.: *The data compression book*, M&T Books, 1996

[1.4.8] Proakis, J.: *Digital communications*, McGraw-Hill, 1983

[1.4.9] Ziv, J., Lempel, A.: "A universal algorithm for sequential data compression", *IEEE Trans. on Information Theory*, May, 1977, pp. 337-343

1.5. Modulációs eljárások

Szerző: dr. Frigyes István

Lektor: dr. Dallos György

1.5.1. Bevezetés

Az információ átvitele, tárolása és feldolgozása során, *elektromos vagy elektromágneses jel* alakjában jelenik meg. Fizikailag e jel lehet időben változó elektromos feszültség, áramerősség, elektromágneses térerősség, optikai teljesítmény vagy más. Az esetek egy részében az eredeti jelforrás elektromos jelet állít elő – e kategóriába tartoznak az adatforrások, az EKG (elektrokardiográf), az EEG (elektroenkefalográf) és mások. Más esetekben az átviendő, feldolgozandó, tárolandó jel az emberi érzékszervek által észlelhető alakban keletkezik – hang, látvány alakjában. Ekkor valamilyen átalakító hozza létre az elektromos, elektromágneses jelet.

Bármilyen fizikai alakban jelenik is meg, matematikailag *valós időfüggvényként* modellezhető. Az ilyen időfüggvények tulajdonságaival, matematikai leírásukkal az 1.1. szakasz foglakozik. A következőkben legtöbbször nem fogjuk megkülönböztetni tényleges fizikai folyamatként megjelenő jeleket az azokat matematikailag modellező absztrakt jelektől.

Ismeretes, hogy az elektromos források által előállított jeleket, illetve az elektromossá átalakított jeleket, ezek digitalizált változatát *alapsávi jeleknek* nevezik. Ezeket az jellemzi, hogy Fourier transzformáltjuk gyakran tartalmazza a 0 frekvenciát, vagy ha nem is, legkisebb frekvenciájú összetevőjük elég kicsi – noha ehelyütt nem foglalkozunk azzal, hogy mit tekintünk nagynak, mit kicsinek.

A jelek feldolgozását, tárolását legtöbbször vagy legalább is igen sokszor azok alapsávi alakján végzik. Néha a jeleket alapsávi alakban viszik is át. Igen gyakran azonban – vezetéken és vezeték nélkül egyaránt – kedvezőbbnek vagy egyenesen egyedül lehetségesnek az bizonyul, ha az alapsávi jeleket átalakítjuk olyan módon, hogy Fourier-transzformáltjuk egy valamivel vagy sokkal nagyobb frekvenciájú

sávban helyezkedjen el. Ezt a frekvenciasávot módosító műveletet nevezzük modulációnak illetve inverzét demodulációnak – mely műveletek, az azokhoz kapcsolódó fogalmak képezik tárgyát a jelen alfejezetnek.

1.5.2. Alapvető fogalmak

Legyen az információt tartalmazó valós időfüggvény *s*(*t*); ez lehet egy ismert – tehát determinisztikus – időfüggvény, de ugyancsak lehet sztochasztikus folyamat egy realizációja. Tekintsünk továbbá egy, az időben szinuszosan változó "jelet" amit *v*-vel jelöljük, a vivőhullám rövidítésére:

$$v(t) = \sqrt{2}A\cos(\omega_c t + \Phi) \qquad (1.5.1)$$

ahol *A* a jel effektív amplitúdója ($W^{1/2}$-ben),

$\omega_c = 2\pi f_c$ *a* körfrekvencia (rad/sec-ban; a *c* index a vivőfrekvenciára – angolul carrier – utal),

Φ a kezdőfázis.

A Bevezetésben körülírt modulációt legtöbbször egy szinuszos jel amplitúdójának vagy argumentumának módosításával érik el, olynanmódon, hogy az (1.5.1)-ben állandó amplitúdóba vagy az időtől lineárisan függő argumentumba az *s*(*t*) időfüggvényt "beírják" – azt modulálják. Ekkor a modulált paraméter(ek) az időnek *s*(*t*)-vel egyértelmű kapcsolatban levő függvénye(i) lesz(nek); ilyen módon lehetőség van a modulált jel demodulálására, e függvény inverzének képzése útján.

A modulációt a szinuszos vivőfrekvencia oldaláról nézve egy modulált jel általános alakja

$$x(t) = \sqrt{2}Am(t)\cos\left[\omega_c t + \Phi + \vartheta(t)\right] \qquad (1.5.2)$$
$$m(t) = f\left[s(t)\right];\ \vartheta(t) = g\left[s(t)\right]$$

ahol az *m*(*t*) időfüggvény az *amplitúdómodulációt*, $\vartheta(t)$ a *szögmodulációt* reprezentálja.

Ezzel a *moduláció*snak nevezett előállítással egyenértékű az úgynevezett *kvadratúra-előállítás*:

$$x(t) = Aa(t)\cos(\omega_c t + \Phi) - Aq(t)\sin(\omega_c t + \Phi) \qquad (1.5.3)$$

ahol $a(t)$ a jel fázisban-lévő,

$q(t)$ annak kvadratúrában vagy fázis-kvadratúrában levő összetevője.

A két függvény-pár közötti kapcsolat:

$$a(t) = m(t)\cos\vartheta(t); q(t) = m(t)\sin\vartheta(t)$$
$$m(t) = \sqrt{[a(t)]^2 + [q(t)]^2}; \operatorname{tg}\vartheta(t) = \frac{q(t)}{a(t)} \qquad (1.5.4)$$

A két leírás közül hol az egyik hol a másik bizonyul előnyösebbnek.

Megjegyezzük, hogy egyes esetekben a szinuszos vivő helyett más, az időnek ugyancsak periódikus függvényét használják vivőként. Ilyen vivő általános alakja

$$v(t) = \sum_{k=-\infty}^{\infty} Ap(t - kT + \tau_0); A = \frac{1}{T}\int_0^T [p(t)]^2 dt \qquad (1.5.1a)$$

A $p(t)$ függvény tartója $(0,T)$; viszonylag enyhe megkötéseknek kell eleget tennie, lényegében ugyanazoknak, melyek a Fourier-sor létezésének feltételei. Ezzel a változattal nem fogunk részletesen foglalkozni; a szinuszos vivőre elmondottak illetve elmondandók – a változók megváltoztatásával – erre az esetre is érvényesek.

E pont befejezéseként megemlítjük, hogy a következőkben többször szerepel jelek Fourier-transzformáltja; e fogalmazás determnisztikus jelet tételez fel. Az állítások legtöbbször sztochasztikus jelek spektrális sűrűségfüggvényére is igazak. Fordítva ez nem mindig áll fenn: a spektrális sűrűségfüggvényről mondandók nem feltétlenül érvényesek determinisztikus jel Fourier-transzformáltjára.

1.5.3. Az analitikus jel és a komplex burkoló

A komplex burkoló fogalma, mely modulált jelek leírására alkalmas, alapvető fontosságú nem csak e jelek vizsgálatában – az alábbiakban mi is használjuk – hanem hullámterjedési jelenségek vizsgálatában, a zaj leírásában is, szinte “sine qua non”-ja a távközlő rendszerek számítógépes szimulációjának, a digitális jelfeldolgozásnak és másoknak. Ezért kicsit részletesebben foglalkozunk vele.

A kvadratúra alakban megadott modulált jel így is felírható:

$$x(t) = A \cdot \operatorname{Re}\{[a(t) + jq(t)]e^{j\omega_c t}\} \qquad (1.5.5)$$

ahol Re a valós-rész képzés jele, a Φ kezdőfázist pedig, mint jelentéktelent elhagytuk.

Az *a(t)+jq(t)* komplex mennyiséget a jel *komplex burkolójának* nevezik.Most a komplex burkoló általános fogalmával, valamint a *keskenysávú jelekkel* ismerkedünk meg kissé precízebben illetve részletesebben.

Ismeretes, hogy egy valós *s*(*t*) jel Fourier-transzformáltja, *S*(ω), konjugált szimmetrikus. (azaz: *S*(-ω) = $S^*(\omega)$). Így például a pozitív ω-khoz tartozó *S(ω)* a jelet egyértelműen megadja – ismeretében tudjuk, hogy a Fourier-transzformáltat megkapjuk, ha a megfelelő negatív-frekvenciás részt is hozzáadjuk. Így a jelet jellemezhetjük *S*(ω) helyett $\hat{S}(\omega)$-val is ahol definíció szerint:

$$\begin{aligned}\hat{S}(\omega) &= S(\omega) + \operatorname{sign}(\omega)\cdot S(\omega) = \\ &= S(\omega) + j\left[-j\cdot\operatorname{sign}(\omega)\cdot S(\omega)\right]\end{aligned} \qquad (1.5.6)$$

$\hat{S}(\omega)$ pozitív ω-kra éppen 2*S*(ω)-val, negatív ω-kra 0-val egyenlő.

Persze képezhetjük $\hat{S}(\omega)$ inverz Fourier-transzformáltját; ehhez vegyük figyelembe, hogy a második tag két tényező szorzata - inverz transzformáltját konvolúcióval kapjuk meg:

$$\mathbf{F}^{-1}\left[-j\cdot\operatorname{sign}(\omega)\cdot S(\omega)\right] = s(t) * \mathbf{F}^{-1}\left[-j\cdot\operatorname{sign}(\omega)\right] \qquad (1.5.7)$$

(*: konvolúció).

A második tényező történetesen éppen 1/t. Így

$$\mathrm{F}^{-1}\left[-j.\operatorname{sign}(\omega).S(\omega)\right] = \int_{-\infty}^{\infty}\frac{s(\tau)}{t-\tau}d\tau = \hat{s}(t), \qquad (1.5.8)$$

az *s*(*t*) Hilbert-transzformáltja

Így írhatjuk:

$$\hat{s}(t) = \mathrm{F}^{-1}\left[\hat{S}(\omega)\right] = s(t) + j\cdot\hat{s}(t) \qquad (1.5.9)$$

Az $\hat{s}(t)$ komplex időfüggvényt az *s*(*t*)-hez tartozó *analitikus jelnek* hívják. (A név onnan származik, hogy ha a *t* változó helyett bevezetjük a *z = t + j.u* változót, abban $\hat{s}(t)$ kielégíti a Cauchy-Rieman parciális differenciál-egyenleteket, így

analitikus.) Az analitikus jel valós része az *s*(*t*) időfüggvény. Mint láttuk, azzal a tulajdonsággal rendelkezik, hogy Fourier-transzformáltja negatív frekvenciákon 0.

Az analitikus jelnek modulált jelek tárgyalásánál van jelentősége. Ugyanis $e^{j\omega_c t}$ a cos $\omega_c t$-hez tartozó analitikus jel. Ha e vivőfrekvencia amplitúdóját valamilyen *a*(*t*) *sávkorlátozott* jellel megmoduláljuk, *a*(*t*)$e^{j\omega_c t}$-t kapunk; ez analitikus lesz, ha a(t) sávszélessége kisebb $f_c=\omega_c/2\pi$ -nél; hasonlóan $e^{j\omega_c t}$ analitikus jel tartozik a *-j*.sin($\omega_c t$) vivőfrekvenciához is; ez analitikus marad, ha hasonló sávkorlátozott q(t)-vel moduláljuk. Így az adott feltételek között a modulált jel *valóban* előállítható

$$\mathrm{Re}\left\{\left[a(t)+jq(t)\right]e^{j\omega_c t}\right\} \qquad (1.5.10)$$

alakban, ha a moduláló jelek sávszélessége kisebb a vivőfrekvenciánál. Az ilyen jeleket *keskenysávú* moduláló jelnk nevezik. (Fontos hangsúlyozni, hogy a feltétel *a kisebb* és nem *a sokkal kisebb* állítás.) Az $\tilde{s}(t)=a(t)+jq(t)$ a modulált jel *komplex burkolója.*

(Zárójelben az ellenpélda: ha a sávszélesség $B>f_c$, $a(t)+jq(t)e^{j\omega_c t}$ *nem* analitikus, így valós része nem adja meg a modulált jelet.)

Ha a moduláló jel sávkorlátozott, szinte mindig keskenysávú; így számításokban elég az *a(t) + jq(t)* komplex burkolóval foglalkoznunk; tudjuk azután, hogy annak valós része a modulált jel koszinuszos, képzetes része annak szinuszos része. A komplex burkoló (komplex számértékű) *alapsávi* jel; Fourier-transzformáltját egyszerűen az analitikus jel transzformáltjának $-\omega_c$ -vel való eltolása útján kapjuk. Vagyis: a jelalak és a komplex burkoló közötti kapcsolat:

$$\tilde{x}(t)=\left[x(t)+j\hat{x}(t)\right]e^{-j\omega_c t};\; x(t)=\mathrm{Re}\left[\tilde{x}(t)e^{j\omega_c t}\right] \qquad (1.5.11)$$

Ismeretében a komplex burkolót megkapjuk, ha képezzük annak analitikus jelét és ezt exp[$-j\omega_c t$]-vel megszorozzuk; illetve fordítva: a komplex burkolót megszorozva exp[$j\omega_c t$]-vel, majd képezve a valós részt, a valós időfüggvényt kapjuk meg.

A komplex burkolóhoz hasonlóan definiálhatjuk egy sáváteresztő szűrő $\tilde{H}(\omega)$ *ekvivalens aluláteresztő szűrőjét*; ennek átviteli függvényét is -ω_c -vel való eltolás útján kapjuk; és a sávszűrőn átvitt modulált jelet megkapjuk, ha meghatározzuk az

$$\tilde{Y}(\omega) = \tilde{X}(\omega).\tilde{H}(\omega), \text{ illetve az} \qquad (1.5.12)$$
$$\tilde{y}(t) = \tilde{x}(t) * \tilde{h}(t)$$

kifejezéseket.

Egy fontos tulajdonság: ha mondjuk a vivő koszinuszos összetevőjét moduláljuk, $\tilde{X}(\omega)$ persze konjugált szimmetrikus; de $\tilde{H}(\omega)$ vagy az vagy nem (egy *sávszűrő* nem feltétlenül konjugált szimmetrikus a *vivő*frekvencia körül); utóbbi esetben $\tilde{y}(t)$ már nem lesz valós – vagyis áthallás keletkezik a szinuszos és koszinuszos vivők között; viszont nem keletkezik áthallás, ha $\tilde{H}(\omega)$ a vivőfrekvencia körül konjugált szimmetrikus.

1.5.4. Analóg modulációs rendszerek – amplitúdómoduláció

Amint az 1.1 szakasz részletesebben tárgyalja: *analóg*nak nevezett jelet olyan *s*(*t*) időfüggvény reprezentál, amely bizonyos specifikáción belül *tetszőleges* lehet. Ilyen specifikációs adat lehet a jel tartója vagy Fourier-transzformáltjának tartója, teljesítménye, dinamika-tartománya vagy hasonlók. (Fiatalabb olvasók kedvéért érdemes elmondani, hogy az elnevezés a számítástechnika korai éveiből származik, amikor a digitális számítógépek mellett még *analóg* számítógépek is léteztek. Ezek egy matematikai problémát úgy oldottak meg, hogy felépítették a problémával *analóg* – azonos differenciálegyenletet, integrálegyenletet stb. kielégítő – áramkört és vizsgálták a létrejövő jelalakokat. Ez az alapja a jelek analóg-digitális kategorizálásának: a folytonos értékkészletű, nem pontosan definiált jelalakok halmazát analóg jelnek nevezik) Az analógnak nevezett modulációs rendszerek stúdiuma lényegében azzal foglalkozik, hogy milyen módon modulálható egy (szinuszos) vivő analóg jellel.

Definíciók és spektrális tulajdonságok

i. Kétoldalsávos el-nem-nyomott vivőjű AM. Amplitúdó modulációban az (1.5.6) formulában szereplő $\vartheta(t)$ = 0 és $m(t)$ az $s(t)$ időfüggvénnyel legtöbbször lineáris kapcsolatban van. Így amplitúdómodulált szinuszos jel általános kifejezése

$$x(t) = \sqrt{2}A[1 + h \cdot s(t)]\cos\omega_c t \qquad (1.5.13)$$

ahol h a modulációs index.

Vizsgáljuk az amplitúdómodulált jel Fourier-transzformáltját (1.5.1. ábra). Az 1.5.3 szakasz megfontolásaiból látszik, hogy a fenti $x(t)$ komplex burkolója éppen $[1 + h \cdot s(t)]$; ennek Fourier-traszformáltja

$$\mathbf{F}[1 + hs(t)] = \delta(\omega) + hS(\omega) \qquad (1.5.14)$$

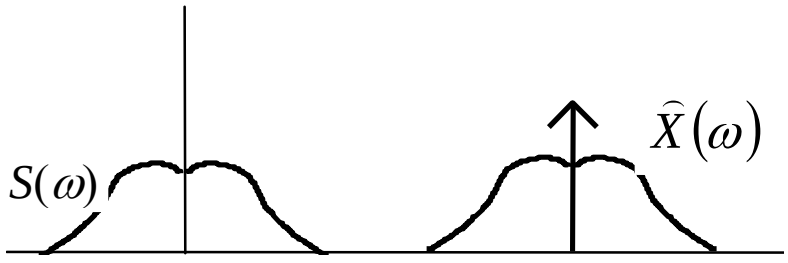
Így az analitikus jel Fourier-transzformáltja

$$\hat{X}(\omega) = \sqrt{2}A[\delta(\omega - \omega_c) + hS(\omega - \omega_c)] \qquad (1.5.15)$$

illetve a valós modulált időfüggvényé

(1.5.16)

Vizsgálva e formulákat egyrészt azt látjuk, hogy az (1.5.13) szerinti amplitúdómodulált jel *modulálatlan vivőfrekvenciát* tartalmaz; ez alatt azt értjük, hogy Fourier-transzformáltjában van egy vivőfrekvenciás spektrumvonal, melynek intenzitása független a moduláló jeltől. Továbbá látjuk, hogy $X(\omega)$ az $S(\omega)$-t változatlan, úgy is mondhatjuk: torzítatlan formában tartalmazza; az AM-et ezért szokták *lineáris modulációnak* nevezni. Folytatva vizsgálatainkat ne feledjük, hogy $S(\omega)$ – $s(t)$ valós időfüggvény lévén – konjugált szimmetrikus függvény. Így $X(\omega)$ a mudulálatlan vivőn kívül *két oldalsávot tartalmaz*, melyek a vivő körül konjugáltan szimmetrikusak. Látjuk, hogy amennyiben $S(\omega)$ tartójának szélessége W a modulált



jel 2*W* szélességű frekvenciasávot foglal el.

Az utóbbi megfontolásoknak megfelelően az amplitúdómoduláció (1.5.13) szerinti megvalósítását *két-oldalsávú, el-nem-nyomott vivőjű* AM-nek nevezik, angol rövidítéssel *AM-DSB-NSC*-nek (AM, Double-Side-Band, Non-Suppressed-Carrier). Az ilyen moduláció két szempontból is pazarol. Pazarló az elfoglalt frekvenciasávval, hiszen a teljes információ "benne van" a *W* frekvenciasávban, mégis elfoglaljuk ennek kétszeresét. Ugyancsak pazarló a felhasznált teljesítménnyel: a modulálatlan vivő-teljesítmény nem járul hozzá az információ-átvitel minőségéhez.

ii. Kétoldalsávos elnyomott vivőjű AM (AM-DSB-SC) Kívánatos lehet e pazarlás csökkentése. Az (1.5.13) szerinti $x(t)$ módosítható úgy, hogy ne legyen "elpazarolt teljesítmény"; akkor

$$x(t) = \sqrt{2}A.s(t)\cos\omega_c t \qquad (1.5.17)$$

Mint látjuk, ekkor a modulálatlan vivő egyszerűen meg van szorozva a moduláló jellel. Az imént végigtekintett *spektrumok* közül talán az analitikus jelé a legszemléletesebb; ez most

$$\hat{X}(\omega) = \sqrt{2}AS(\omega - \omega_c) \qquad (1.5.18)$$

iii. Egyoldalsávos AM (AM-SSB) Minthogy az átviendő információt egy oldalsáv is teljes egészében tartalmazza, indokolt, a frekvenciával való takarékosság érdekében, nem elfoglalni a másodikat. Az eddigiek ismeretében ilyen egy-oldalsávos AM-hez jutunk, ha a vivőt nem a moduláló jellel, hanem annak analitikus jelével moduláljuk (Ugyanis ekkor, mint tudjuk, $s(t)$-nek nincs negativ frekvenciájú összetevője, így $x(t)$-nek csak az egyik oldalsávja lesz meg.) Vagyis ekkor

$$x(t) = \sqrt{2}A.\left[s(t) \pm j\hat{s}(t)\right]\cos\omega_c t \qquad (1.5.19)$$

ahol a felső oldalsávot kapjuk, ha a pozitív, az alsót, ha a negatív előjelet választjuk.

Az (1.5.6) -ból látható, hogy $\hat{s}(t)$ képzéséhez az $s(t)$ jelet egy állandó abszolút értékű, $\pi/2$ fázistolású szűrőn kell átvezetni; ennek megvalósítása nem nagyon könnyű. (Persze előállíthatjuk az SSB jelet úgy is, hogy az egyik oldalsávot kiszűrjük; erre akkor van lehetőség, ha $S(\omega)$ nem tartalmaz kisfrekvenciás összetevőket – pl. a beszédjel).

iv. Csonka oldalsávos AM Hely hiányában nem részletezzük, csak megemlítjük, hogy az SSB létrehozásának, demodulálásának gyakorlati nehézségei jelentősen csökkenthetők, ha az egyik oldalsáv mellett a másiknak egy (akár elég kis) részét is átviszik.

AM jelek demodulálása

Az AM-DSB-NSC jel egyszerű burkolódetektorral demodulálható; ennek az egyszerűségnek fizetjük meg az árát amikor pazarlóan bánunk teljesítménnyel is, sávszélességgel is. Vegyük figyelembe, hogy ez feltételezi a modulált jel pillanatnyi amplitúdójáról, hogy az >0; vagyis szigorúan $h.s(t)<1$, csúcsértékben is.

Vivő nélküli AM demodulálása (akár egy, akár két oldalsávot tartalmaz) igényli a vivő *fázisának* ismeretét is. Vagyis a demodulátorban szükségünk van egy *referenciajelre*, mely a vett jellel – elvileg – koherens; ennek birtokában az AM jel szorzó-demodulátorral demodulálható. Pl AM-DSB-SC:

$$s_d(t) = \sqrt{2}Ahs(t)\cos\omega_c t \times \sqrt{2}R\cos\omega_c t = ARhs(t)(1+\cos 2\omega_c t) \qquad (1.5.20)$$

ahol R a referenciajel effektív amplitúdója. A kétszeres frekvenciájú összetevő persze könnyen kiszűrhető.

Kimutatható, hogy amennyiben a helyileg előállított referenciajel és a vett jel között δ fáziskülönbség van (vagyis a koherencia nem tökéletes)

- DSB-SC-nél a demodulált jel $\cos\delta$-val arányos; így a frekvenciáknak *pontosan* meg kell egyezni, hisz különben $\cos\delta$ folyamatosan változik ± 1 között
- SSB-nél a jelalak ugyan lényegesen eltorzul, azonban a teljesítmény (vagyis a komplex burkoló abszolút-érték-négyzete) állandó marad. Így ez a torzítás pl. beszédátvitelnél szubjektíven nem észlelhető.

1.5.5. Analóg modulációs rendszerek – szögmoduláció

Definíciók

Szögmodulációnál az (1.5.6) formulában $m(t) = 1$ és $g[]$ elvileg tetszőleges, a gyakorlatban lineáris operátor. Gyakorlati alkalmazásban háromféle g operátor illetve ennek megfelelő $\vartheta(t)$ függvény fordul elő:

i. Fázismoduláció (PM)

$$\vartheta(t) = 2\pi h s(t); x(t) = \sqrt{2}A\cos\left[\omega_c t + 2\pi h s(t)\right] \qquad (1.5.21)$$

Ekkor tehát a vivő fázisa arányos a moduláló jellel; *h* a modulációs index.

ii. Frekvenciamoduláció (FM)

$$\vartheta(t) = 2\pi\Delta F\int^{t} s(t)dt \qquad (1.5.22)$$

Ekkor a pillanatnyi frekvencia arányos a moduláló jellel; ΔF a frekvencialöket.

iii. FM preemfázissal

A moduláló jel kisfrekvenciás összetevőit elnyomják, a nagyfrekvenciásakat kiemelik majd ezzel az eltorzított jellel modulálják a vivő frekvenciáját. (Erre az FM zaj spektrumának alakja miatt lehet szükség; ez ugyanis ω^2-tel arányos.) Ekkor

$$\vartheta(t) = 2\pi.\Delta F.\int^{t} p(t)*s(t)dt; x(t) = \sqrt{2}A\cos\left[\omega_c t + 2\pi.\Delta F.\int^{t} p(t)*s(t)dt\right]$$

(1.5.23)

ahol *p*(*t*) a mondott spektrum-alakítást végző szűrő súlyfüggvénye.

Spektrális tulajdonságok – néhány példa

Az AM-mel ellentétben az FM nem lineárisan transzformálja *s*(*t*)-t *x*(*t*)-vé, így *x*(*t*) Fourier-transzformáltját közvetlenül nem tudjuk felírni. A következőkben néhány esetet áttekintünk.

i. Frekvenciamoduláció egyetlen szinusszal

Ekkor $s(t)=\cos \omega_m t$ és $\vartheta(t) = \Delta F/f_m.\sin \omega_m t$. Most a modulált jel periódikus és így Fourier-sorba fejthető; alkalmazva cos(sin *x*) ismert Fourier-sorát, *x*(*t*) Fourier-transzformáltja

$$X[\omega] = \frac{A}{\pi\sqrt{8}}\sum_{n=-\infty}^{\infty} \mathrm{J}_n\left(\frac{\Delta F}{f_m}\right)\left[\delta(\omega-\omega_c-n\omega_m)+\delta(\omega+\omega_c+n\omega_m)\right] \qquad (1.5.24)$$

ahol J_n az n-edrendű Bessel-függvény.

Mint látjuk, elvileg végtelen sok spektrumvonalból áll, melyek nagysága bonyolultan függ a modulációs indextől. Igen kis modulációs indexnél a vivőfrekvencián kívül mindössze két olyan oldalsáv van, melyek amplitúdója nem

elhanyagolható; ha $\Delta f/f_m$ éppen J_0 0-helyével egyezik meg, a vivőfrekvencia eltűnik; stb.

ii. Keskenysávú FM tetszőleges jellel

Ha a frekvencialöket a moduláló spektrum felső határfrekvenciájához képet kicsi, a moduláció lineárisnak tűnik: a két oldalsáv alakja megegyezik $\vartheta(t)$-ével:

iii. Szélessávú FM véletlenszerű jellel

Elég érdekes eredmény: a spektrális sűrűségnek két oldalsávja van, melyek alakja megegyezik a moduláló jel valószínűségi sűrűségével.

iv. Az elfogalt sávszélesség

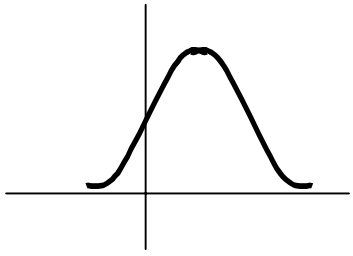
Amint egyebek közt (1.5.23)-ból látjuk, elméletileg a sávszélesség végtelen nagy. A gyakorlatilag elfoglalt sávszélességre elég jó közelítést ad az úgynevezett Carson-formula; eszerint

$$B = 2\left(\Delta F + f_{\max}\right) \qquad (1.5.24a)$$

Szögmodulált jelek demodulálása

Frekvenciamodulált átvitelnél a modulált jel pillanatnyi frekvenciája arányos a moduláló jellel. Így a demoduláláshoz frekvencia-detektor áramkört kell létrehoznunk, vagyis olyat, melynek kimenő jele a bemenő jel *frekvenciájával* arányos. E követelmény persze azt is magábafoglalja, hogy e kimenő jelnek függetlennek kell lenni a bemenő jel amplitúdójától

Frekvenciaváltozást amplitúdóváltozássá konvertáló áramkör megvalósítható félrehangolt rezgőkörrel – lásd az 1.5.2 ábra rezonanciagörbéjét. (A bejelölt abszcissza a vivőfrekvencia. Célszerűen a rezonanciagörbe inflexiós pontja erre illeszkedik.) Az így kapott jelet aztán burkolódetektorral detektálhatjuk.



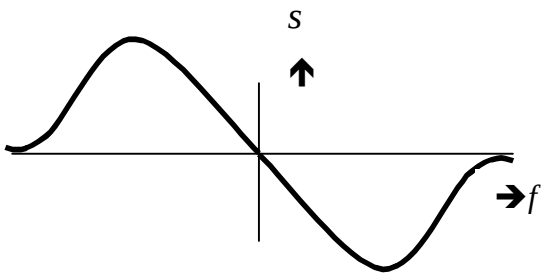
1.5.2. ábra Félrehangolt rezgőkör mint frekvencia-amplitúdó konverter

E frekvencia-amplitúdó konverzió persze nemlineáris torzítást okoz, mely legtöbbször nagyobb mint ami elfogadható. A konverzió linearitása javítható, ha két félrehangolt rezgőkört alkalmazunk és ezek detektált jelét kivonjuk egymásból. Ilyen – úgynevezett frekvencia-diszkriminátor – áramkör karaterisztikája az 1.5.3. ábrán látható.

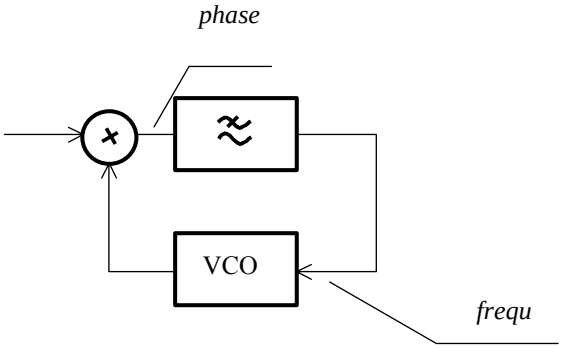
Annak érdekében, hogy a diszkriminátor kimenő jele valóban a frekvenciával legyen arányos, gondoskodni kell arról, hogy a bemenő jel állandó legyen; ez limiterrel érhető el. Így egy teljes FM demodulátor diszkriminátort és azt megelőző limitert tartalmaz.

FM jel ugyancsak detektálható PLL (fázisszabályozott hurok) áramkörrel, 1.5.4 ábra. Ha a PLL egyensúlyi állapotban van, a VCO frekvenciája *pontosan* megegyezik a bejövő jel frekvenciájával. Így ha az utóbbi változik, a VCO vezérlő jelének követnie kell a frekvenciaváltozást – vagyis e jel a detektált frekvenciát szolgáltatja. Továbbá a VCO modulációs karakterisztikája független a bejövő jel amplitúdójától, így a PLL a limiter szerepét is ellátja.

Részletezés nélkül megemlítjük, hogy a PLL analóg fázisdemodulátorként is



1.5.3. ábra Frekvenciadiszkriminátor karakterisztikája



1.5.4 ábra PLL mint analóg frekvencia- és fázisdemodulátor

használható, amit ugyancsak feltüntettünk az ábrán.

FM additív zaj jelenlétében

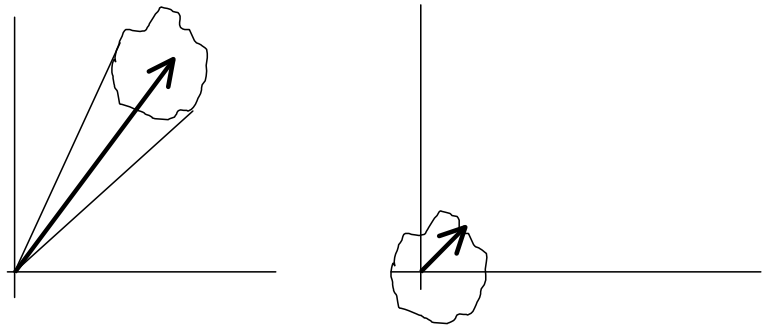
Az eddig figyelembe nem vett zaj természetesen módosítja a demodulált FM jelet – az is zajos lesz. Alább levezetés illetve különösebb indoklás nélkül közöljük a jel/zaj formuláját majd egy fontos jelenségről, a küszöbről ejtünk néhány szót.

Additív fehér zajjal terhelt FM jelet demodulálva a jel/zaj viszony az

$$S/N = \frac{C}{N}\left(\frac{\Delta F}{f_{\max}}\right)^2 \qquad \text{(1.5.24.b)}$$

formulából számítható, ahol *C/N* a (rádiófrekvenciás) vevő sávjában mért jel/zaj viszony. (Utóbbihoz vegyük figyelembe, hogy a vevő sávszélességét közelítőleg az (1.5.24.a.) formula határozza meg.) (1.5.24.b)-ből látható, hogy a frekvencialöketet növelve a jel/zaj viszony – első ránézésre tetszőleges mértékben – növelhető, a vett viviőfrekvenciás teljesítmény növelése nélkül. Frekvenciamoduláció alkalmazásának ez a fő előnye.

Az (1.5.24.b.) formula érvényességének van azonban két feltétele, melyeket először limiter-diszkriminátoros demodulátorra nézünk meg. Az első az, hogy a burkolódetektor sávszélessége ne legyen nagyobb f_{max}-nál – viszonylag könnyen teljesíthető feltétel. A másik az, hogy *C/N* legyen elég nagy, a gyakorlatban 7-10 dB-nél nagyobb. Ha ennél kisebb, a már említett küszöb jelenség lép fel; ekkor a diszkriminátor már nem működik ideális frekvenciadetektorként és a jel/zaj viszony kisebb, mint ami az (1.5.24.b)-ből adódna. E jelenség megértéséhez vegyük figyelembe az 1.5.5 ábrát. Nagy jel/zaj viszonynál a jel+zaj nagy valószínűséggel a jelvektor közelében van (mondjuk azt körbeveszi, mint az ábrán), ennek megfelelően változik annak fázisa (a változás határait az ábra mutatja); a változó fázis deriváltja



1.5.5. ábra. A zajos jel fázisának változása nagy és kis C/N esetén

pedig a (viszonylag kis) frekvenciazaj. Az (1.5.24.b) formulát ennek feltételezésével vezették le. Ha azonban a jel kicsi, a zaj-folyamat egy ugyanolyan realizációja a jel fázisának egy ciklusnyi csúszását eredményezi; ennek deriváltja egy igen nagy impulzus, vagyis a detektált jel a másik esetben nem túl nagy gaussi zaj helyett impulzuszajjal is terhelve lesz, jelentősen lerontva a jel/zaj viszonyt.

Elvileg hasonló jelenség fellép a PLL-es demodulátorban is, azonban nagy különbséggel. A PLL-nek csak az alapsávi jelet kell átvinnie, így sávszélességének nagyságrendje nem $2\Delta F+f_{max}$ hanem csak f_{max}. Cikluscsúszás csak sokkal kisebb jelnél lép fel. Így a PLL mint FM demodulátor az egyik legjobb megoldás a küszöbszint csökkentésére.

1.5.6. Digitális modulációs eljárások

Míg az analóg jeleket egy-egy meglehetősen tág függvényosztály egy-egy elemeként definiáltuk, *digiális* jelek kezelése – átvitele, tárolása, feldolgozása – során mindig egy *véges-elemszámú* jelkészlet egy-egy eleméről van szó. (Így a modulációs rendszer kiválasztása lényegében e jelalakok definiálását jelenti). A jelkészlet elemeinek száma *M* és mindegyik elemet az jellemzi, hogy *időtartama véges* (*T* vagy T_S) ideig tart és *energiája is véges*; egy jelalak energiáját annak négyzetes integráljaként definiálunk.

Így egy digitális jelfolyam alakja a kövekező:

$$x(t) = \sum_{k=-\infty}^{\infty} s_i(t - kT_s); i = 1,2,\ldots M \qquad (1.5.25)$$

ahol az $s_i(t)$ jelalakok lehetnek alapsávi, szinuszos vagy más vivőjű jelek.

A digitális jelek vektoriális ábrázolása

A jelkészlet elemeinek ismeretében definiálhatunk egy *D* dimenziós vektorteret, melyben minden jelalak egy-egy vektornak felel meg; kimutatható, hogy D≤M. A jel-időfüggvények és a jel vektorok között kölcsönös-egyértelmű kapcsolat van:

$$s_i(t) \Leftrightarrow \mathbf{s}_i; i = 1,2...,M$$
$$\mathbf{s}_i = (s_{i,1}, s_{i,2} ... s_{i,D}); s_i(t) = \sum_{j=1}^{D} s_{i,j}\varphi_j(t) \Leftrightarrow s_{i,j} = \int_0^T s_i(t)\varphi_j(t)dt \qquad (1.5.26)$$

Itt $\varphi_j(t)$; j=1,2…D az ortonormált bázisfüggvények.

Mint kimutatható, a jelvektorok abszolút érték-négyzete megegyezik a jel-időfüggvények energiájával. Ugyancsak kimutatható, hogy egy digitális jelsorozat által elfoglalt frekvenciasáv arányos a jelkészlet dimenzióinak számával.

A jelek vektoriális kezelése igen egyszerűvé teszi a modulációs rendszerek ismertetését, tárgyalását: egy-egy modulációs rendszer egyszerűen geometriai elrendezésként adható meg, vizsgálható. A vektoriális leírás általánosan lehetséges: tetszőleges – alapsávi vagy vivőfrekvenciás, szinuszos vagy másféle vivőjű – M darab véges energiájú időfüggvényhez egyértelműen definiálható egy vektortér, egy jelvektor-készlet. Mint látni fogjuk, igen sok, a gyakorlatban felhasznált modulációs eljárásban az összes jel frekvenciája azonos; ilyenkor ezen általános vektor-felfogás megegyezik a szinuszos jelek szokásos fazor-ábrázolásával.

Anélkül, hogy ebben a szakaszban foglalkoznánk e kérdéssel megemlítjük, hogy additív Gauss-zajos csatornán a hibaarány csak a vektorelrendezéstől függ és nem függ a konkrét jelalakoktól; ezen belül is a jel-vektor-végpontok távolságától.

Bináris (kétállapotú) modulációs rendszerek

A gyakorlat digitális jelforrásai kétállapotú jeleket állítanak elő (M=2), sőt ezek a jelalakok is igen egyszerűek: NRZ (Non-Return-to-Zero) impulzusok. Így egy digitális jelfolyam egy “alacsony” és egy “magas” feszültségszint váltakozó sorozatából áll; szintváltás csak $t=kT$ időpontokban következhet be. Ilyen jelsorozattal való moduláció esetén indokolt ezt analóg jelnek tekinteni és az előző szakaszban megismert modulációs eljárások egyikét vagy másikát alkalmazni. Így az alábbi modulációkhoz jutunk.

i.Bináris amplitúdómoduláció vagy “amplitúdóbillentyűzés” (Angolul: Amplitude Shift Keying, rövidítve ASK)

A két jelalak:

$$s_1(t) = \sqrt{2}A\cos\omega_c t; s_2(t) = 0; t \in (0,T) \qquad (1.5.27)$$

Ez egy egydimenziós (*D*=1) jelkészlet Az egyetlen bázisfüggvény:

$$\varphi(t) = \sqrt{2/T}\cos\omega_c t \qquad (1.5.28)$$

Az AM-DSB-NSC moduláció digitális alakját ismerjük fel ebben, melyben *h*=1.

ii.Bináris frekvenciamoduláció vagy “frekvenciabillentyűzés” (Angolul: Frequency Shift Keying, rövidítve FSK)

A két jelalak:

$$\begin{aligned} s_1(t) = \sqrt{2}A\cos\omega_c t;\ s_2(t) = \sqrt{2}A\cos(\omega_c + \delta\omega)t;\ t \in (0,T) \\ \varphi_1(t) = \sqrt{2/T}\cos\omega_c t;\ \varphi_2(t) = \sqrt{2/T}\cos(\omega_c + 2\pi/T)t;\ \delta\omega \geq 2\pi/T \end{aligned} \qquad (1.5.29)$$

Ebben a frekvenciamoduláció digitális változatát ismerjük fel. Mint látjuk, ez kétdimenziós jelkészlet. Amennyiben $\delta\omega$=2π/*T*, a két jelvektor merőleges egymásra; az ilyen jelkészletet ortogonális jelkészletnek nevezik.

iii.Bináris fázismoduláció vagy “fázisbillentyűzés” (Angolul: Phase Shift Keying, PSK)

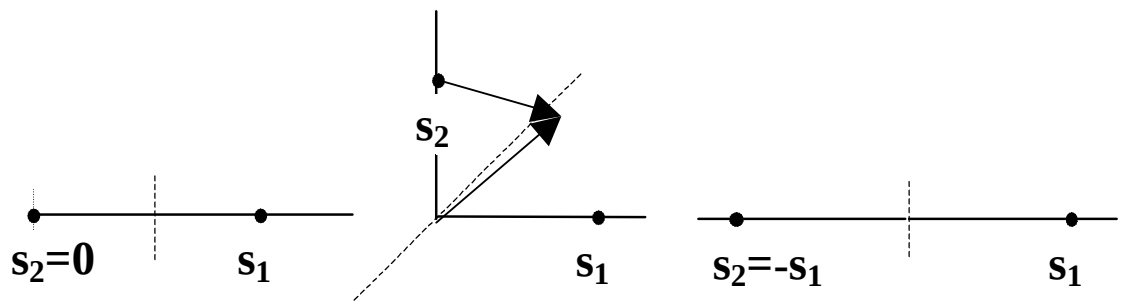
A két jelalak és a bázisfüggvények:

$$\begin{aligned} s_1(t) = \sqrt{2}A\cos\omega_c t;\ s_2(t) = \sqrt{2}A\cos(\omega_c t + \Phi);\ t \in (0,T) \\ \varphi_1 = \sqrt{2/T}\cos\omega_c t;\ \varphi_2 = \sqrt{2/T}\sin\omega_c t \end{aligned} \qquad (1.5.30)$$

Általános esetben ez is egy kétdimenziós jelkészlet és a fázismoduláció megfelelője; ha azonban $\Phi=\pi$, s_1=-s_2 és *D*=1; ekkor tekinthetjük az AM-DSB-SC digitális megfelelőjének is.

Az olyan jelkészletet, melyben s_1=-s_2, antipodális jelkészletnek nevezik; kimutatható, hogy a kétállapotú jelek között ennek a legkisebb a hibaaránya.

Az 1.5.6. ábrán feltüntettük e három modulációs formának megfelelő jelvektor-elrendezést.



1.5.6. ábra. ASK, ortogonális FSK és antipodális PSK jelkészlet; a döntési tartományhatárokat is feltüntettük(a szaggatott vonalak), valamint FSK-nál egy hibás döntést okozó zajvektort (az s_2-ből kiinduló) és a jel+zaj vektort.

Digitális jelek demodulálása, döntése és hibavalószínűsége

Digitális jeleket átvivő rendszer feladata nem az, hogy az adott jelalakot minél hűségesebben (minél kisebb torzítással) jelenítse meg, hanem hogy eldöntse: az *M* lehetséges közül éppen melyik jelalakot adták. E *döntési* művelet végrehajtásához a vevőnek a jelalakokat ismernie kell, azoknak a vevőben tárolva kell lenni. A döntő áramkör azt állapítja meg, hogy a vett (torzítással, zajjal, interferenciával elcsúfított) jel melyik "lehetségeshez" hasonlít a legjobban. Kicsit precízebben: ha a vett jelalak *r*(*t*), a hibaarány akkor lesz a legkisebb, ha a annak a javára döntünk, melyre

$$\Pr\{s_i(t)|r(t)\} = \max; i = 1,2,\ldots M \qquad (1.5.31)$$

Azonos a-priori valószínűségek esetén a vett jelhöz legközelebbi jelvektor javára kell dönteni – ezt tüntettük fel az 1.5.6. ábrán

Az imént azt mondtuk, hogy a döntéshez szükség van a jelalakok ismeretére a vevőben. A teljeskörű ismeret a jelen esetben magábafoglalja azt is, hogy a vevőben a jel fázisát is ismerni kell. Ehhez a fázisra vonatkozó információt a vevőben (a vett jelből) elő kell állítani, miután annak eleve-tárolására nincs lehetőség. Az olyan demodulátort, amely ezt el is végzi *koherens demodulátornak* nevezik. Koherensen demodulált jelek hibaaránya a három ismertetett modulációs rendszerben.

$$P_{E,PSK} = \frac{1}{2}\operatorname{erfc}\sqrt{\frac{E}{N_0}}; P_{E,FSK} = \frac{1}{2}\operatorname{erfc}\sqrt{\frac{E}{2N_0}}; P_{E,ASK} = \frac{1}{2}\operatorname{erfc}\sqrt{\frac{E}{4N_0}}; \qquad (1.5.32)$$

ahol *E* egy bit energiája (ASK-nál a csúcs-energiája), N_0 a zaj spektrális sűrűsége.

Digitális jelek által elfoglalt frekvenciasáv; kettőnél több állapotú modulációk

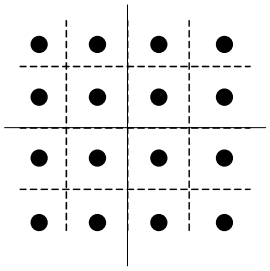
Rádióátvitel esetén az átviteli rendszer legköltségesebb eleme az elfoglalt frekvenciasáv. Elméletileg ez végtelen: mivel az időfüggvények tartója véges, Fourier-transzformáltjuk tartója szükségképpen végtelen. *Gyakorlatilag* elfoglalt frekvenciasávnak azt tekinthetjük, ahol a jel energiájának nagy része – 90%-a, 95%-a – elhelyezkedik. E gyakorlatilag elfoglalt sáv (Hz-ben) – pontosabban *egy* dimenzió által elfoglalt frekvenciasáv – a jelidő reciprokának nagyságrendjébe esik. Így pl. egy 155 Mbit/sec sebességű STM-1 jelfolyam 155 MHz-et foglal el; csökkentése feltétlenül kívánatos, ami a jel-idő növelését teszi szükségessé. Adott sebességű

forrás jel-idejének növelésére egyetlen mód van: ha a biteket nem egyenként visszük át, hanem több – *n* db – bitet összefogunk egy szimbolummá. Akkor egy-egy szimbolum időtartama *nT* lesz és így (amennyiben a dimenziók számát nem növeljük) az elfoglalt frekvenciasáv *n*-ed részére csökken. Azonban, míg 1 bitnek 2 lehetséges állapota van, *n* bitnek $M=2^n$ állapota, ezek átvitele *M* állapotú modulációt igényel.

Ha az 1 vagy 2 dimenziós (jel)térben nem 2 hanem *M* pontot – vektort – kell elhelyeznünk, ezek sokkal közelebb kerülnek egymáshoz; így már kisebb zaj is hibás döntést eredményez, megnő a hibavalószínűség. Vagy, ennek elkerülésére, meg kell növelnünk a jelek energiáját, hogy a vektorvégpontok távolabb kerüljenek egymástól. Így mondhatjuk: digitális átvitelben az elfoglalt frekvenciasáv (tetszőlegesen) csökkenthető, a szükséges energia növelése árán.

Kimutatható, hogy az M-PSK a legjobb modulációs rendszer, ha az állapotok száma nem nagyobb 6-nál. Ennél nagyobb állapotszám esetén a fázis *és* az amplitúdó változtatása jobb eredményt (kisebb hibaarányt) ér el. Gyakran használatos az optimaálishoz közel álló kvadratúra amplitúdó moduláció, QAM. Pl. 16QAM vektorelrendezését mutatja az1.5.7. ábra. MQAM hibaaránya gaussi csatornán

$$P_E = 2\left(1-1/\sqrt{M}\right)\operatorname{erfc}\frac{\sqrt{E_{peak}}}{\sqrt{2N_0}\left(\sqrt{M}-1\right)} \qquad (1.5.33)$$



1.5.7. ábra 16QAM jelkészlet; a szaggatott vonalak itt is a döntési tartományok határát jelölik

Folytonos fázisú modulációk

E modulációs rendszerek három előnyös tulajdonsággal rendelkeznek, ami egyre növekvő elterjedésükhöz vezetett:

- takarékosan bánnak az elfoglalt frekvenciasávval; (emiatt másik szokásos gyüjtőnevük: frekvencia-takarékos modulációs rendszerek);
- kódolási nyereséget biztosíthatnak;
- szemben más (pl. QAM) frekvencia-takarékos modulációval, meglehetősen érzéketlenek nemlineáris torzításokkal szemben.

Az alábbiakban röviden áttekintjük e tulajdonságokat.

Kimutatható, hogy egy folytonos függvénynek, melynek első q-1 deriváltja is folytonos, de q-adik deriváltja szakadásos, Fourier-transzformáltja $1/\omega^{q+1}$ szerint csökken; továbbá véletlenszerű bitfolyam spektrális sűrűségfüggvénye arányos az elemi jel Fourier-transzformáltjának négyzetével. (Például egy digitális jelfolyam melynek amplitúdója állandó és fázisa folytonos, spektrális sűrűsége legalább ω^{-4} szerint csökken.) Így ha a q kitevő elég nagy, szűrő nélkül is elérhetjük, hogy a szomszédos csatorna sávjába átnyúló zavaró spektrum-rész elfogadhatóan kicsi lesz. Keskenysávú átvitelben ez különösen fontos.

Továbbá a fázis folytonossága, memóriát és bizonyos fajta redundanciát visz az ilyen átvitelbe. E redundancia hibajavító dekódolást tesz lehetővé.

Végül emlékeztetünk arra, hogy nemlineáris torzítás akkor lép fel, ha nemlineáris karakterisztikájú eszközön olyan jel halad át, melynek burkolója nem állandó. A nemlineáris torzítás következménye lehet a hibaarány megnövekedése valamint a spektrum kiterjedése. A most tárgyalt rendszerekben

- az információt a fázis hordozza;
- az amplitúdó (a burkoló) állandó és a fázis folytonos.

Ilyenmódon a nemlineáris karakterisztika nem okoz torzítást, így hibaarány-növekedést sem (ugyanis a *fázis-függvény* csak *AM-PM konverzió* következtében torzulna el; állandó amplitúdójú jelen ilyesmi nem lép fel – nincs AM).

Kódolt modulációs rendszerek

Modulált átviteli rendszerben kézenfekvőnek tűnik (legalább is most, hogy már kitalálták) redundáns átviendő szimbolumok helyett az elfoglalt frekvenciasáv csökkentése érdekében, redundáns állapotok beiktatása. Így például 2/3 arányú kódolás esetén lehetségesnek tűnő megoldás a következő: 4-állapotú modulációból kiindulva (a gyakorlatban QPSK) minden 2-bites adat-szimbolumot egy 3-bites kódszimbolummá kódolunk; ezt átvihetjük 8PSK - modulációval amivel

célkitűzésünket elvileg megoldottuk. A gyakorlatban e megoldás nem nagyon előnyös: 8PSK átvitel kb. 4 dB-lel nagyobb teljesítményt igényel mint a QPSK, mivel kisebb a vektorok távolsága. A fentebb vázolt eljárásban ezért olyan nagy kódolási nyereségre volna szükség, mely nemcsak kompenzálja e hátrányt hanem még ezen felül is számottevő eredő nyereséget okoz.

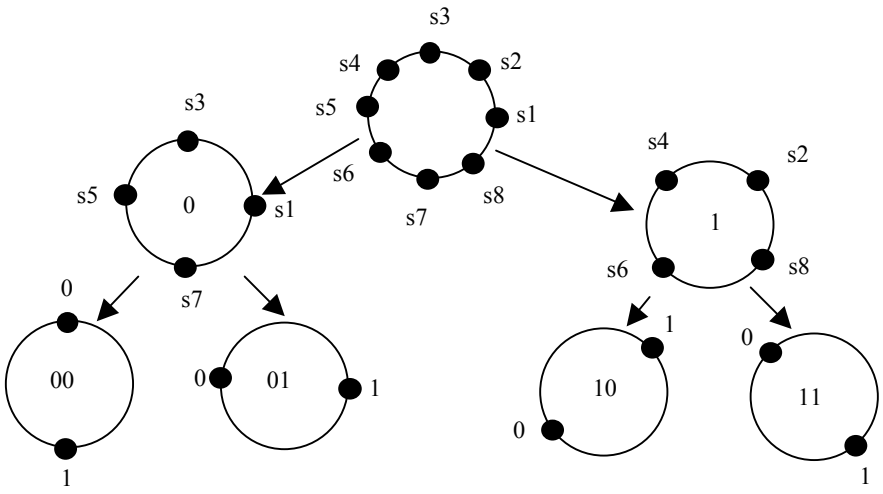
A probléma egy igen hatékony megoldása az Ungerböck által bevezetett jelkészlet-felosztás. Az 1.5.8. ábrán 8 állaputú PSK jelkészlet felosztása látható 4 db PSK-ra. A rész-jelkészletek illetve a jelvektorok melletti bináris számok azt mutatják, hogy milyen módon képeződnek le a (három bites) kódvektorok a nyolcállapotú jeltér egyes pontjaiba. Az ábrán látható, hogy a részjelkészletek vektorainak távolsága lényegesen nagyobb, mint a teljes - 8PSK - jeltér jeleié.

A 8PSK úgynevezett trellis-kódolt modulációs rendszer (TCM, Trellis-Coded-Modulation) működése ezek alapján úgy képzelhető el, hogy a bit-hármas első két bitje azt mutatja meg, hogy melyik antipodális rész-jelkészletről van szó, a harmadik pedig azt, hogy ennek melyik bitjéről. Továbbá, a BPSK rész-jelek távolsága a gyakorlatban eléggé nagy. A többi bit távolságának növelésére pedig megfelelően nagy kényszer-hosszú konvolúciós kódolást alkalmaznak; a 8PSK példában ennek kódaránya 1/2.

A fenti példa megfelel az általánosan alkalmazott eljárásnak, a következők szerint:

- a rendelkezésre álló frekvenciasáv ismeretében meghatározhatjuk n-et, a (kódolatlan) szimbolumok bitjeinek számát;
- k = n-1 biten k/(k+1) arányú konvolúciós kódolást hajtunk végre;
- a kódolatlanul hagyott (mondjuk utolsó) bittel együtt létrejött m=n+1 bites kódszavakat megfelelően leképezzük a 2-dimenziós, 2^{n+1} állapotú jeltérre;
- a legutóbbi lépés hatékony végrehajtása érdekében kiválasztjuk az adott jeltérbeli elrendezés legnagyob minimális távolságú kétállapotú rész jelkészleteket.

  Ha az eredő állapotszám 8-nál nagyobb, PSK helyett QAM-et alkalmazunk.



1.5.8. ábra 8PSK jelkészlet felosztása négy antipodális jelkészletre.

1.5.7. Moduláció az optikai tartományban

Az optikai frekvenciatartomány jelentősen eltér az elektromosnak nevezhető tartománytól. Az átviteli közeg sávszélessége sok nagyságrenddel nagyobb, mint a rádiófrekvenciáké. Csaknem kizárólag vezetékes összeköttetéseket alkalmaznak, így a különböző felhasználók nem interferálnak egymással, nem zavarják egymást. Elsősorban e két ok következtében gyakorlatilag mindig megelégszenek kétállapotú átvitellel.

Az optikai sáv egy további tulajdonsága plauzibilissé és így csaknem kizárólagossá tesz egy speciális, az elektromos tartományban nem alkalmazott modulációt, az *intenzitásmodulációt*. A fény intenzitása a teljesítmény vagy a teljesítménysűrűség szinonimája; így az elektromos moduláló jellel – feszültséggel, gyakrabban árammal – a fény *teljesítménye* arányos. Ezt az összerendelést az indokolja, hogy az elektromos áram áll elemi részecskékből, elektronokból, míg az optikai sávban a teljesítmény, fotonokból. Minden modulátorban, demodulátorban az elektronok száma és a fotonok száma arányos egymással; ebből következik az elektromos áram és az optikai teljesítmény közötti arányosság.

Az utóbbi ténynek van egy koncepcionális következménye, mely azonban nem jár különösebb gyakorlati következménnyel. Nevezetesen, miután az optikai *teljesítmény* az, ami az információt hordozza, az átvitel eleve nemlineáris: az

elektromos *jellel* (az adott esetben árammal) az optikai teljesítmény arányos, így az *optikai jel* (vagyis a térerősség) az elektromos jel *négyzetgyökével*. Ennek az ideális négyzetgyökvonásnak azután valóan nincs gyakorlati következménye: az optikai jel elektromossá konvertálása ugyancsak ideális *négyzetreemeléssel* egyenértékű. Jelentősége csak optikailag is szélessávú jeleknél lehetne, ilyen azonban még ma is csak a komolytalan futurológia körébe tartozik.

Az intenzitásmoduláció sajátságai igen hasonlatosak az AM-DSB-NSC rendszerhez: a jel – ebben az esetben az optikai teljesítmény – persze csak pozitív lehet; a modulált fény egy modulálatlan intenzitáshoz adódik, és a modulációs index határozottan kisebb 1-nél.

Noha csaknem kizárólagos az intenzitásmoduláció alkalmazása, kidolgoztak olyan optikai modulációs eljárásokat is, melyek analógak az elektromos modulációkkal: PSK, FSK; meg egy olyat is, melynek elektromos megfelelője nem használatos, a polarizáció-modulációt (PolSK). Ezek az intenzitásmodulációnál lényegesen bonyolultabbak másfelől nagyobb kapacitást biztosítanak az optikai csatornának. A mai technika még nem igényli alkalmazásukat

Irodalomjegyzék

[1.5.1] J. Proakis: Digital Communications, McGraw-Hill, 1998

[1.5.2] H.L. Van Trees: Detection, Estimation and Modulation Theory I-II, Wiley

[1.5.3] D. Ventre: Communications Analogiques, Ellipses

[1.5.4] Frigyes I.: Hírközlő rendszerek, Műegyetemi Kiadó, 1998

1.6. Hullámtan

Szerző: dr. Nagy Lajos

Lektor: dr. Veszely Gyula

1.6.1. A rádióspektrum

Az elektromágneses hullámok eddig megismert frekvenciatartománya közel zérustól mintegy 10^{23} Hz-ig terjed. Ebben az igen széles tartományban helyezkednek el többek között a rádióhullámok és a fény is.

A rádióspektrum az elektromágneses spektrumnak az a része, amely mesterséges úton viszonylag jó hatásfokkal előállítható, kisugározható és felfogható, és ezáltal különféle rádiószolgálatok számára felhasználható.

Jelenlegi ismereteink szerint ez a tartomány 9 kHz és mintegy 3000 GHz között helyezkedik el. A rádióhullámokra az jellemző, hogy mesterséges vezetés nélkül terjednek.

1.6.2. Terjedési módok

A földfelszini műhold-műhold és föld-műhold rádióösszeköttetések alapvető terjedési módja az adó-, és vevőantenna között közel szabad térben terjedő hullám. Ezen terjedési mód főleg a mikrohullámú frekvenciasávra jellemző.

Alacsony antennamagasság és kis irányítottságú antennák esetén a direkt hullám mellett megjelenik a - főképp - talajról reflektálódott hullám, így alakul ki a kétutas terjedési mód.

Alacsony frekvenciákon jó vezetőképességű talaj mellett alakulnak ki a föld görbületét követő felületi hullámok.

Az ionoszféra ionizáltságától és a frekvenciától függően különböző mértékben téríti el a rádióhulámokat az egyenes vonalú terjedéstől, az ilymódon a földre visszahajlított hullámok az ionoszférikus terjedési mód hordozói.

A földi légkör troposzférikus rétegének magassággal gyorsan változó törésmutatójának következtében fellépő hullámszóródással valósul meg a troposzférikus szórású terjedési mód.

1.6.3. A frekvenciasávok terjedési módjai és rendszerjellemzői

Az ELF (3 kHz alatt) és VLF (3-30 kHz) sávokban az ionoszféra a tápvonal módusú terjedés felső, a föld felszíne pedig a tápvonal módusú terjedés alsó határa. Nehéz irányított ill. jó hatásfokú adóantennákat készíteni.

Az ELF sáv rendszerjellemzői az alacsony információs sebesség, tipikus szolgálatok a rövid távolságú - víz alatti, búvárok között és a nagy távolságú - tengeralattjárók közötti kommunikáció. Az elérhető adatsebesség 1 bit/s körül van.

A VLF sáv tipikus szolgálatai a világméretű távíróösszeköttetés a hajókkal, nagy távolságú állandóhelyű összeköttetések, navigációs célok (Omega), viharjelző szolgálatok, idő etalonok.

Az LF (30-300 kHz) és MF (300-3000 kHz) sávokban 100 kHz-ig még továbbra is csak a föld görbületét követő felületi hullámok jellemzőek, fölötte megjelennek a térhullámok is. Az LF sáv tipikus szolgálatai a nagy távolságú összeköttetés hajókkal, nagy távolságú állandóhelyű összeköttetések, műsorszórás, rádiónavigáció.

Az MF sávon több elemű, irányított antennák, L, T elemek a jó hatásfokú adóantennák, vevőantennaként főleg ferritantennákat alkalmaznak. Az MF sáv tipikus szolgálatai a műsorszórás, rádiónavigáció, néhány földi, tengeri és légi mozgó szolgálat, néhány állandóhelyű szolgálat,

A HF (3-30 MHz) sávban a térhullámok csak az ugrástávolság után jelentkeznek, a felületi hullámok kisebb távolságon, főleg tengervíz felett jellemzőek.

A sáv jellemzően alkalmazott antennái a log-periódikus antennák (horizontális vagy vertikális), vertikális ostorantenna, horizontális dipólrendszer.

Tipikus szolgálatok: állandóhelyű pont-pont összeköttetések, földi (az ugrástávolságnál nagyobb távolságra), tengeri, légi mozgó szolgálat, nagy távolságú műsorszórás,

A VHF (30-300 MHz) UHF (300-3000 MHz) sávokban az atmoszféra hatása a VHF sávon refrakció és reflexió a törésmutató index irregularitásokon, szporadikus E reflexió, ionoszférikus szórás, Faraday forgatás és ionoszférikus szcintilláció a föld-műhold rádiósszakaszon. A terep hatása reflexió nagyobb hegyekről, diffrakció a völgyekbe. A felületi reflexió többutas terjedést okoz látóhatáron belüli összeköttetéseknél. Az összeköttetések jellemző antennái a több elemes Yagi antennák, helixek.

A VHF sáv tipikus szolgálatai a hang és kép műsorszórás; földi, légi és tengeri mozgó szolgálatok, mobil telefonok és vezeték nélküli telefonok, rádiónavigációs nyalábok.

Az UHF sávban az atmoszféra hatása refrakció, továbbá reflexió az alacsonyabb frekvenciákon, és duct a magasabb frekvenciákon.

A törésmutató index fluktuáció miatt horizonton túli szórás lép fel 500 MHz felett.

A terep hatására legjellemzőbb a hegyek, épületek által okozott árnyékolás.

A jellemző antennatípusok alacsonyabb frekvenciákon a Yagi antennák, magasabb frekvenciákon a paraboloid reflektor antennák.

Az UHF sáv tipikus szolgálatai a TV műsorszórás, légi navigáció, leszállító rendszer, rádiólokáció, mobil szolgálatok, cellás rádiótelefon rendszerek.

Az SHF (centiméteres hullámok, 3-30 GHz) tartományában az előző sávokhoz képest a csapadék már számításba veendő, változó csillapítást okoz. Nagy nyereségű, forgásparaboloid és tölcsérantennákat alkalmaznak. A tipikus rendszerek a fix telepítésű földi pont-pont, pont-multipont, műholdas mobil hírközlési alkalmazások és rádiólokáció.

Az EHF (30-300 GHz, milliméteres hullámok) és (300-3000 GHz, szubmilliméteres hullámok) sávokban a csapadék csillapítása mellett az atmoszférikus gázok csillapítása is jelentős. Nagynyereségű antennaként az EHF sávon paraboloid reflektor, a szubmilliméteres sávon lencseantennákat alkalmaznak. A sávokban megvalósítható rendszerek kis távolságú látóhatáron belüli összeköttetések és távérzékelés.

1.6.4. Elektromágneses hullámok

A rádiózásnál alkalmazott terjedési módokat a terjedés fizikai elve szerint két csoportra oszthatjuk, ezek a térhullámú és a felületi hullámú terjedési módok.

A térhullámú terjedési módra jellemző, hogy az elektromágneses hullámok a Föld felületétől elszakadva, az elektromos és mágneses térerősség vektorszorzataként kifejezhető Poynting vektor irányába terjednek. A forrásoktól távol ezen hullámok síkhullámok, melyeket a forrásmentes térre felírt Maxwell egyenletek megoldásaként kapunk.

Síkhullámú terjedés

Az I és II Maxwell-egyenletek szinuszos időfüggésre, árammentes térre, levegőben (ε_0,μ_0):

$$rot\mathbf{H} = j\omega\varepsilon_0\mathbf{E} \qquad (1.6.1)$$

$$rot\mathbf{E} = -j\omega\mu_0\mathbf{H} \qquad (1.6.2)$$

Néhány átalakítást elvégezve kapjuk a mágneses térerősségre az alábbi homogén Helmholtz-egyenletet:

$$\frac{\partial^2 H_{x,y,z}}{\partial x^2} + \frac{\partial^2 H_{x,y,z}}{\partial y^2} + \frac{\partial^2 H_{x,y,z}}{\partial z^2} + \omega^2\varepsilon_0\mu_0 H_{x,y,z} = 0 \qquad (1.6.3)$$

A (1.6.3) egyenletet elégítsük ki az alábbi megoldással:

$$H_x = H_z = 0 \qquad (1.6.4)$$

$$H_y = H_y^0 \cdot e^{-j\beta z} \qquad (1.6.5)$$

A (1.6.5) kifejezést helyettesítsük a (1.6.3) egyenletbe, melyből megállapíthatjuk, hogy az teljesül, ha:

$$\beta^2 = \omega^2\varepsilon_0\mu_0 \qquad (1.6.6)$$

Az I Maxwell egyenletből pedig az elektromos térerősséget fejezzük ki:

$$\mathbf{E} = \frac{1}{j\omega\varepsilon_0} rot\mathbf{H} = \frac{-1}{j\omega\varepsilon_0}\mathbf{e}_x\frac{\partial}{\partial z}H_y = \frac{-\mathbf{e}_x}{j\omega\varepsilon_0}H_y^0(-j\beta)e^{-j\beta z} = E_x\mathbf{e}_x \qquad (1.6.7)$$

Az elektromos és mágneses térerősségek kifejezéséből látszik, hogy azok egymásra merőlegesek, azonos fázisúak és hányadosuk

$$\frac{E_x}{H_y} = \frac{\beta}{\omega\varepsilon_0} = \sqrt{\frac{\mu_0}{\varepsilon_0}} = 120\pi \qquad (1.6.8)$$

ami a szabad tér hullámimpedanciája.

A polarizáció

A térhullám polarizációját síkhullámokra értelmezzük és a térerősségvektor által a terjedési irányra merőleges síkban leírt alakot értjük. Az előzőekben megmutattuk, hogy a szabadtérben terjedő elektromágneses hullámok elektromos és mágneses térerőssége között állandó kapcsolat van, így a polarizációt az elektromos térerősségvektor által leírt görbével egyértelműen jellemezhetjük.

Általános esetben a síkhullám polarizációja elliptikus, melynek speciális eseteként adódik a körös ill. lineáris polarizáció. A térerősségvektor forgásának iránya a hullámok terjedési irányából megfigyelve óramutató járása szerinti, vagy azzal ellentétes lehet, ekkor beszélünk jobbforgású (clockwise – CW) ill. balforgású (counterclockwise – CCW) polarizációkról.

Reflexió

A reflexiót homogén, végtelen kiterjedésű, síkkal határolt térrészek határfelületén értelmezzük. A reflektált hullám amplitudóját, fázisát és polarizációját a térrészek anyaga és felületük egyenetlensége határozza meg. Ha a felszín sík és tökéletesen síma, akkor spekuláris reflexió alakul ki. Ha a beeső hullám síkhullám, akkor a visszavert hullám is az lesz és az energia egyetlen diszkrét irányba terjed. Ez az ideális eset elméletileg jól leírható, ha a veszteségmentes dielektrikumra vonatkozó Snell-Descartes törvényt a komplex ε és komplex μ bevezetésével veszteséges dielektrikumokra általánosítjuk.

A továbbiakban a reflexiós tényezőt vizsgáljuk meg két térrészre az alábbi két polarizációra. (1.6.1. ábra)

A reflexiós tényezőt mint a reflektált és beeső hullám elektromos térerőssége amplitudóaránya definiáljuk.

$$\Gamma = \frac{E_r}{E_i} \qquad (1.6.9)$$

A reflexiós tényező horizontális polarizációra

$$\Gamma_h = \frac{\sin\vartheta - \sqrt{\varepsilon^* - \cos^2\vartheta}}{\sin\vartheta + \sqrt{\varepsilon^* - \cos^2\vartheta}} \qquad (1.6.10)$$

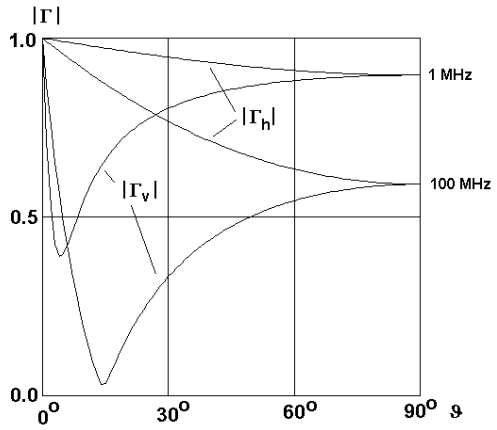
A reflexiós tényező vertikális polarizációra

$$\Gamma_v = \frac{\varepsilon^* \sin\vartheta - \sqrt{\varepsilon^* - \cos^2\vartheta}}{\varepsilon^* \sin\vartheta + \sqrt{\varepsilon^* - \cos^2\vartheta}} \qquad (1.6.11)$$

ahol

$\varepsilon^* = \varepsilon' + \frac{\sigma}{j\omega\varepsilon_o} = \varepsilon' - j\,60\lambda\,\sigma = \varepsilon' - j\,\varepsilon''$ a komplex dielektromos állandó.

A rádióhullámok terjedésének vizsgálatakor a talajreflexió vizsgálata a legfontosabb reflexiós feladat. Ábrázoljuk a talajreflexió abszolut értékét két frekvenciára ϑ_B beesési szögnél vertikális polarizációra a $|\Gamma|$ minimumot ér el.



1.6.2. ábra. A talajreflexiós tényező abszolút értéke

Ha σ=0, akkor $tg\vartheta_B = 1/\sqrt{\varepsilon'}$. Ennél a szögnél $\Gamma_v = 0$ és ϑ_B a Brewster szög, ha $\sigma \neq 0$ akkor $tg\vartheta_B \cong 1/\sqrt{\varepsilon'}$ és ϑ_B a pszeudo Brewster szög.

Diffrakció

A hullámokat a szabad terjedésben a talajon kívül tereptárgyak is akadályozhatják.

Az akadály modellezése késél, parabolikus henger vagy dielektromos ékkel történik. A gyakorlatban elterjedten a késél diffrakciós modellt alkalmazzák.

A Huygens elv értelmében a terjedő hullám frontja új hullámforrásként viselkedik és így az elektromágneses hullámok ezen új forrásokból származó hullámok szuperpozíciójaként állítható elő.

Fresnel geometriai diffrakcióelmélete értelmében a 1.6.3. ábrán látható geometriára a vételi térerősség a következő Fresnel integrállal írható fel.

$$\frac{E}{E_o} = \frac{1}{1-j}\int_{\nu_o}^{\infty} \exp(-j\frac{\pi}{2}\nu^2)d\nu \quad (1.6.12)$$

ahol az 1. Fresnel zóna sugara r_1 és az akadály relatív benyúlása ν_o:

$$r_1 = \sqrt{\frac{\lambda d_1 d_2}{d_1 + d_2}} \quad \rightarrow \quad \nu_o = \sqrt{2}\frac{h}{r_1} = h\sqrt{\frac{2(d_1+d_2)}{\lambda d_1 d_2}} \qquad (1.6.13)$$

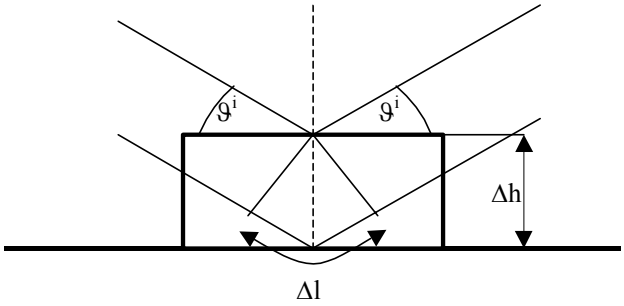
A késélként modellezett diffrakciós csúcs által okozott többletcsillapítás (szabadtéri térerősséghez képest) a késél ν_0 relatív magasságának függvényében a 1.6.4. ábrán látható. A ν_0 paraméter a késélnek a Fresnel ellipszisekbe való relatív benyúlását adja meg, mely pozitív, ha a késél a látóvonal fölé nyúlik, negatív, ha a késél a látóvonal alatt helyezkedik el.

Szóródás

A szóródást lényegében egyenetlen felületen történő rendezetlen reflexiók együtteseként kezelhetjük. A vizsgálataink föleg a felületi egyenetlenség

jellemzésével foglalkoznak és a Rayleigh kritériumot alkalmazzuk a felület síma ill. egyenetlen voltának eldöntésére. Ha a felület egyes pontjaiból reflexióval származó hullámösszetevők közötti maximális fáziseltérés $\pi/2$-nél kisebb, akkor a felület síknak tekinthető, ellenkező esetben egyenetlen. A fáziseltérésből az úthosszkülönbségekre $\lambda/4$ adódik.

A talajegyenetlenségből következő hullámísszetevők úthosszkülönbsége az 1.6.5. ábrából $\Delta l = 2 \cdot \Delta h \cdot \sin \vartheta^i$, így a Rayleigh kritériumból következő maximális megengedett talajegyenetlenség $\Delta h = \lambda / (8 \sin \vartheta^i)$.



1.6.5. ábra. Talajegyenetlenség modellje

Egyenetlen felületekre a felület magassági eloszlását Gauss eloszlásként modellezik, a szórási veszteség ρ_s megadható

$$\rho_s = \exp\left[-8 \left(\frac{\pi \sigma_s \sin \vartheta^i}{\lambda} \right)^2 \right] \quad (1.6.14)$$

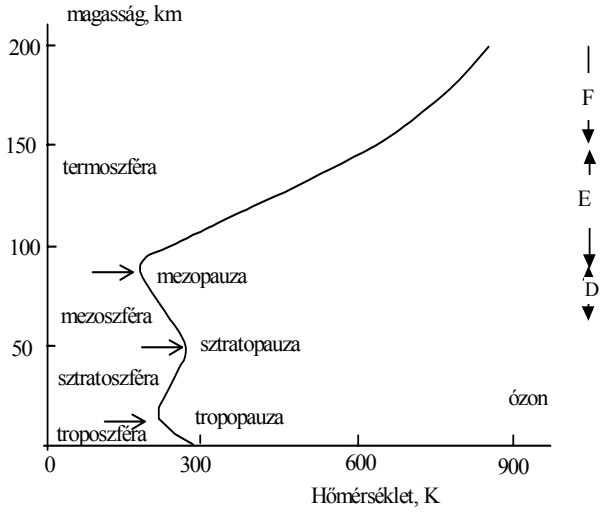
ahol σ_s a felület magasságának szórása. Így az egyenetlen felületről történő szórás reflexiós tényezője

$$\Gamma_{egyenetlen} = \rho_s \cdot \Gamma_{sík} \qquad (1.6.15)$$

1.6.5. A földi atmoszféra hatása

Az antenna által létrehozott teljesítménysűrűség két különböző fizikai hatás következtében csökken a hullám atmoszférában történő terjedése közben:

- az antenna által a térbe kisugárzott hullám divergál
- a terjedést biztosító közeg elnyeli vagy szétszórja a hullámokat, melynek eredete



- -az atmoszférikus gázok molekuláris abszorpciója;
- -az atmoszférában lévő folyadék vagy szilárd részecskék álta okozott abszorpció vagy szóródás (esőcseppek, hó, jég részecskék).

Ezek a hatások néhány GHz feletti frekvencián kezdenek jelentkezni és hatásuk nagyon gyorsan növekszik a növekvő frekvenciával. Az előzőeken túl az atmoszférában lebegő részecskék az atmoszférán keresztülhaladó hullám polarizációjának megváltozását is okozhatják.

A légköri abszorpció

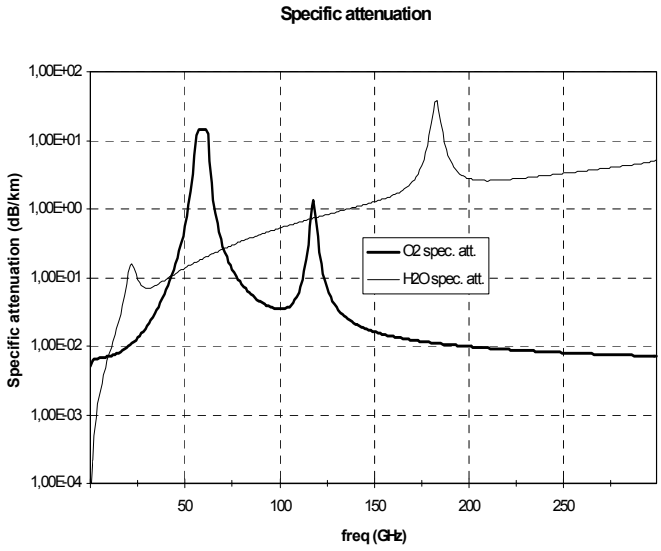
Mivel a nitrogénnek nincsen elnyelési sávja a rádiófrekvenciás tartományban, ezért a molekuláris abszorpciót főképpen az oxigén és vízgőz molekulák elnyelő hatása okozza.

A 350 GHz alatti frekvenciákon az *oxigénnek* egy izolált elnyelési frekvenciavonala van 118,74 GHz-en és nagyon sok egymáshoz közeli elnyelési vonala 50 és 70 GHz között. Az atmoszféra alsó részében ezek a vonalak folytonos sávvá szélesednek.

A 350 GHz alatti frekvenciatartományon a *vízgőznek* három elnyelési vonala van, 22.3 GHz, 183.3 GHz és 323,8 GHz-en. Magasabb frekvencián, a szubmilliméteres és infravörös sávban további intenzív elnyelési vonal jelentkezik.

A vízgőz alacsony koncentrációja mellett a vízgőz csillapítása a koncentrációjával arányosnak tekinthető.

A 1.6.7. ábrán a fajlagos csillapítást mutatjuk be. A vízgőz koncentráció egyenlő 7.5 g/m^3-rel, mely megfelel 1% vízgőz molekula és 99% száraz levegő molekula keverékének. Ezen érték egy átlagos, talajszint magasságában, 50% relatív páratartalmat jelent 16.5^{o}C levegő hőmérséklet mellett, vagy 75% relatív páratartalmat 10^{o}C levegő hőmérséklet mellett.



1.6.7. ábra. Az atmoszférikus gázok által okozott csillapítás

A csapadék csillapítása

Általában az eső által okozott csillapítás az elsődlegesen vizsgált jelenség. A gyakorlatban rendszerint az esőintenzitás R (csapadék milliméterben óránként) mérhető egyszerűen. A csendes szemerkélő eső megfelel R=0.25 mm/óra intenzitásnak, könnyű zápor megfelel 1 mm/óra , közepes eső 4 mm/óra , erős zápor 16 mm/óra, és felhőszakadás több cm/óra esőintenzitásnak. A cseppméret eloszlás az esőintenzitás függvénye, nagyobb esőcseppekkel a nagyobb esőintenzitásokkor. Marshal és Palmer a következő empirikus formulát állapította meg:

$$N(a) = N_0 e^{-\Lambda a} \qquad (1.6.16)$$

ahol $N_0 = 1.6 \times 10^4 \; mm^{-1}/m^3$ és $\Lambda = 8.2R^{-0.21} mm^{-1}$, ***a*** a cseppek sugara mm-ben.

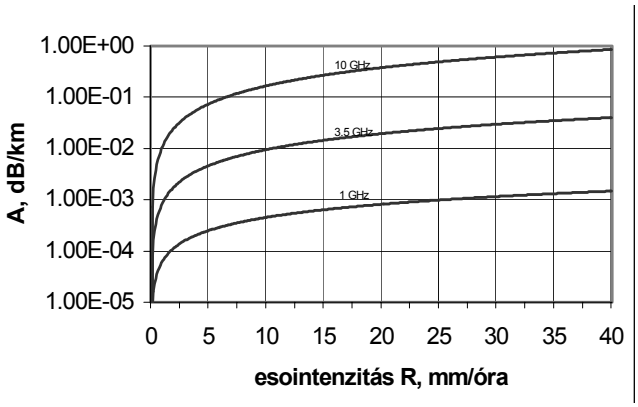
Ezt a modellt használják a legtöbb elméleti esőcsillapítás számításnál. A kifejezés jó egyezést mutat a Laws és Parsons által mért eloszlásokkal.

A rádióösszeköttetések méretezéséhez egyszerű csillapításképletek a kedveltek, melyek az esőintenzitás, frekvencia és hőmérséklet függvényében megadják a fajlagos csillapítást. Ilyen a mérésekkel jól egyező kifejezés a következő:

$$A = cR^b \quad [dB/km] \quad (1.6.17)$$

ahol *c* és *b* frekvenciától és az eső hőmérsékletétől függő konstansok. A hőmérséklettől való csillapításfüggés a víz dielektromos állandójának hőmérsékletfüggésével magyarázható.

Az (1.6.17) kifejezés felhasználásával az 1.6.8. ábrán néhány fajlagos csillapítás eredményt mutatunk be 1, 3.5 és 10 GHz-re az esőintenzitás függvényében.



1.6.8. ábra. Fajlagos esőcsillapítás 1, 3.5 és 10 GHz-en az esőintenzitás függvényében

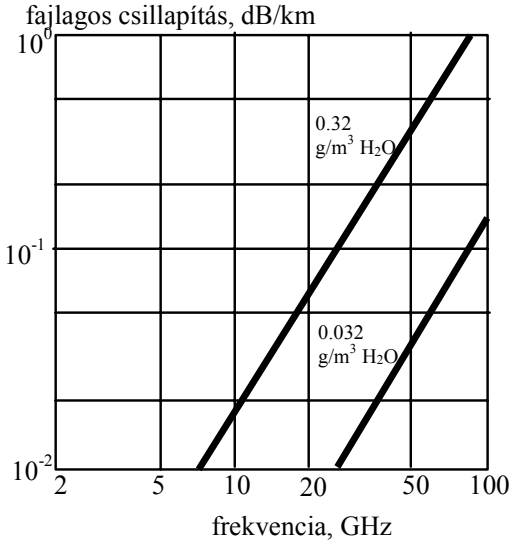
Az eső további hatása a kettős polarizációval működő rádiórendszereknél jelentkezik, ez *a rádióhullámok depolarizációja*. A jelenség lényegében a névleges polarizócióból az ortogonális polarizációba történő energia transzformáció.

A radarelmélet szerint az esőcseppek bisztatikus szórást is okoznak. Ez a hatás a térosztásos multiplex rendszereknél jelentkezik, ahol szomszédos csatornás interferenciát okozhat, ha az egyik nyalábból szórt jel a másik szektorban elhelyezkedő vevőantennára jut.

Mikrohullámú és milliméter hullámú sávban a *köd által okozott csillapítás* hasonló törvényszerűségekkel és egyenletekkel írható le, mint az eső által okozott csillapítás. A lényeges különbség az, hogy a köd jóval kisebb részecskékből tevődik össze, ezen részecskék mérettartománya 0.01 to 0.05 mm sugarat jelent. 300 GHz

alatti frekvenciákon a köd által okozott csillapítás a vízgőztartalom függvényében lényegében lineárisnak tekinthető egy adott frekvencián. A vízgőz tartalom felső határa 1 g/m^3 körülire tehető, a legtöbb természetben előforduló köd vízgőztartalma ennél lényegesen kisebb. 0.032 g/m^3 vízgőztartalom megfelel a 600 m látótávolsággal jellemezhető ködnek, 0.32 g/m^3 vízgőztartalom pedig körülbelül 120 m látótávolságúnak.

A köd által okozott fajlagos csillapítást a frekvencia függvényében a 1.6.9. ábrán mutatjuk be az előző két vízgőztartalomra.



1.6.9. ábra. A köd csillapítása a frekvencia függvényében két koncentrációra

300 GHz frekvencián a nagy sűrűségű köd csillapítása is legfeljebb 1 dB/km, emiatt a rádióösszeköttetések tervezésekor az esőcsillapítás kompenzálására beállított csillapítás tartalék a köd csillapítást mindig ellensúlyozza.

A mikrohullámú frekvenciákon a száraz hó csillapítása legalább egy nagyságrenddel kisebb, mint az eső csillapítása azonos csapadék intenzitás esetére. A nedves hó csillapítása ezzel szemben összemérhető az eső csillapításával, a milliméteres hullámsávon azt meg is haladhatja. Egyes mérések szerint száraz hóra is 0.96 mm-en az esőénél nagyobb csillapítást kapunk azonos csapadék intenzitás esetén.

A légkör törésmutatója-refrakció

Az előző fejezetekben feltételezett egyenes vonalú hullámterjedés csak speciális esetekben valósul meg, így például a műholdak közötti összeköttetéseknél. Minden más rádióösszeköttetésnél az elektromágneses hullámok az atmoszférán haladnak keresztül. A rádióhullámok refrakciója szempontjából az atmoszféra legalsó - a Földfelszínhez legközelebbi rétegének - a troposzférának van szerepe, mert a légköri gázoknak itt még akkora a koncentrációja, hogy a rádióhullámokat jelentősen eltérítik az egyenes vonaltól.

Az egyenes vonalú terjedéstől való eltérést a troposzféra törésmutatójának hely szerinti változása okozza, amit közvetlenül a légkör molekuláris felépítése okoz.

A törésmutató értéke két fő tényező miatt tér el 1-től:

- a légköri gázok molekuláit a beérkező elektromos tér polarizálja;
- molekuláris rezonancia, mely csak keskeny frekvenciasávban, lényegében csak 22 ill. 60 GHz környékén jelentkezik.

Az első hatás gyakorlatilag frekvenciafüggetlen a milliméteres hullámok tartományáig.

A Föld felszínéhez közel a levegő törésmutatója $n \cong 1.0003$.

A számítások egyszerűsítésére vezessük be a törésmutató indezet, a következő összefüggéssel:

$$n=1+10^{-6}N \quad (1.6.18)$$

Az ITU-R a troposzféra törésmutató indexére a következő kifejezést közli.

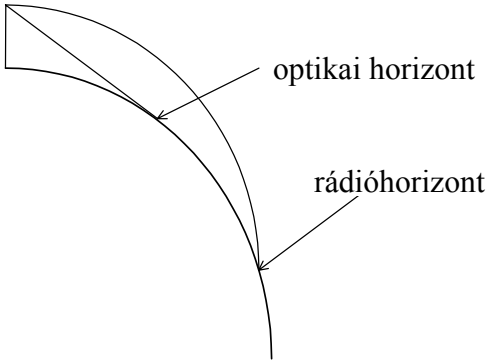
$$N = (1-n)10^6 = 77.6\frac{\cdot p}{T} + 3.73 \cdot 10^5 \frac{e}{T^2} \qquad (1.6.19)$$

p a légnyomás, *e* a vízgőz parciális nyomása, T az abszolút hőmérséklet.

Ezen meteorológiai tényezők a magasság függvényében változnak

Standard atmoszférában n a magassággal csökken, emiatt a rádióhullámok a Föld felé hajlanak el.

N értéke a standard atmoszférában közelítőleg 300.



1.6.10. ábra. Hullámok refrakciója

A gyakorlati számításokban bevezetjük a *K* Földsugár tényezőt, ami lehetővé teszi a $K \cdot R_0$ módosított Földsugár alkalmazásával a hullámterjedési feladatok egyenes vonalú terjedési feladatokra visszavezetését.

1.6.6. Szabadtéri rádióösszeköttetés

A rádiós hullámterjedés determinisztikus analízis módszerei lényegében csak néhány egyszerű esetben alkalmazhatók. Legtöbbször ezen egyszerű modellek kiterjesztése szükséges, de ezen analízis eredményei általában megadják az alapvető hullámterjedési mód csillapítását. Az 1.6.4.1. fejezetben bemutatott síkhullámú terjedési mechanizmust az antennák távolterében alkalmazhatjuk. Most pontszerű forrást feltételezve kiszámítjuk két – a szabad térben elhelyezett – antennából álló szabadtéri rádióösszeköttetés csillapítását.

Ha az adóantenna a szabad térbe sugároz, azaz távol helyezkedik el a föld felszínétől és egyéb objektumoktól, akkor G_T nyereségű adóantenna és P_T adóantennába betáplált teljesítmény esetén a vevőantennát az adóantenna fő sugárzási irányában elhelyezve a teljesítmény sűrűség a *d* távolságban elhelyezett vevőantenna helyén

$$S = \frac{P_T G_T}{4\pi d^2} \qquad (1.6.20)$$

A vevőantennából kivehető maximális hatásos teljesítmény a vevőantenna A_e hatásos felületével kifejezve így:

$$P_R = \frac{P_T G_T}{4\pi d^2} A_e = \frac{P_T G_T}{4\pi d^2} \cdot \frac{\lambda^2 G_R}{4\pi} \quad (1.6.21)$$

ahol G_R a vevőantenna nyeresége, λ üzemi hullámhossz.

Az előzőek alapján a szabadtéri rádiócsatorna csillapítása (szabadtéri csillapítás) a következőképpen írható fel

$$L_F = \frac{P_R}{P_T} = G_T G_R \left[\frac{\lambda}{4\pi d} \right]^2 \qquad (1.6.22)$$

A szabadtéri csillapítást általában dB-ben fejezzük ki és így az (1.6.22) kifejezést az *f* üzemi frekvenciával a következőképpen írjuk fel

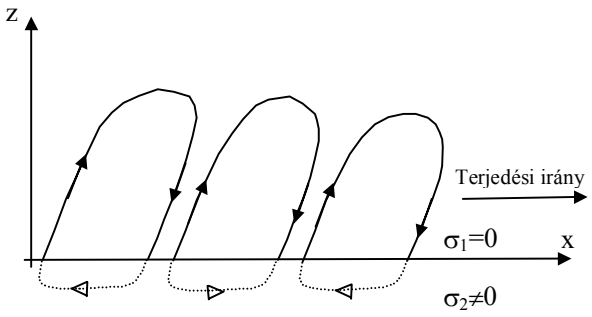
$$L_F = \left(G_T^{dB} + G_R^{dB} \right) - 20\log f - 20\log d + 147.6 \qquad (1.6.23)$$

Gyakran használjuk a szakaszcsillapítás mellett az L_B izotróp antennák között értelmezett szakaszcsillapítást, mely

$$L_B = -32.44 - 20\log f^{MHz} - 20\log d^{km} \qquad (1.6.24)$$

1.6.7. Felületi hullámú terjedés

A felületi hullám a jól vezető Föld és a levegő határfelülete mentén alakul ki a hullámhosszhoz képest kis antennamagasságok esetén, mivel ekkor a közvetlen és reflektált hullám kioltja egymást. A felületi hullám elektromos erővonalai az 1.6.11. ábrán látható alakúak.



1.6.11. ábra. Felületi hullám terjedése

A levegőben folyó eltolási és a talajban folyó vezetési áram zárt hurkot alkot. Az elektromos erővonalak a talaj véges vezetőképessége miatt a haladás irányába megdőlnek.

Sommerfeld 1909-ben publikálta a vertikális felületi hullámok csillapítási tényezőjét.

$$E = E_0 \cdot A(p) \quad (1.6.25)$$

ahol

A(p) a vertikális felületi hullámok csillapítási tényezője

E_0 a szabadtéri térerősség

A felületi hullámok térerőssége az adóantennától nagy távolságra a távolság négyzetével arányosan csökken. Az előzőekben bemutatott sík földre vonatkozó modell $d[km] = 80/\sqrt[3]{f[MHz]}$ határig használható, effölött Sommerfeld a görbült földre érvényes korrekciós tényezőt vezetett be.

1.6.8. Troposzférikus szórás

A földi légkör törésmutatója (ahogy a refrakció jelenségénél bemutattuk) hosszú idő átlagában jól leírható módon, szabályosan változik. Emellett a levegő törésmutatójában mindíg előfordulnak diszkontinuitások is, amelyeknek az oka a levegő páratartalmának, hőmérsékletének vagy nyomásának hely szerinti gyors megváltozása. Ezek a változások csekélyek, de nagy adóteljesítmény mellett jelentős szórt teljesítménysűrűséget hoznak létre.

A troposzférikus összeköttetéseket általában 200 MHz-től 10 GHz-ig alkalmazzák. Az alsó frekvenciahatárt a szükséges nagy nyereségű antennák jelentős mérete korlátozza, a felső frekvenciahatár felett pedig az atmoszférikus gázok és csapadékcsillapítás miatti szakaszcsillapítás válik túl naggyá. A troposzférikus összeköttetések tipikus távolsága néhány száz km, és az összeköttetések megvalósülásához szükséges nagyságú szórás a troposzféra 10 km alatti részében alakul ki.

1.6.9. Ionoszférikus terjedés

Az ionoszféra ionizáltságának fő forrása a Nap ibolyántúli és részecskesugárzása ill. a földi légkörbe jutó meteoritok ionizáló hatása. Mivel az ionizációt főképp a Nap okozza, ezért az ionoszféra állapota szorosan összefügg a naptevékenységgel, így jól jellemezhető és előrejelezhető a napfoltszámmal.

A rádióhullámok terjedésére az ionoszféra komplex dielektromos állandójának magassággal történő változása miatt az ionoszférikus rétegekben (1.6.6.ábra) történő refrakció jellemző. Minden réteghez tartozik egy maximális frekvencia, mely az adott rétegről merőlegesen még visszatörik a föld felé, ezt a réteg kritikus frekvenciájának nevezzük.

Az ionoszférikus összeköttetések egy vagy több ionoszférikus ugrással valósíthatók meg, akár több ezer km szakasztávolsággal.

Irodalomjegyzék

[1.6.1.] M.P.M. Hall, L.W.Barclay, M.T.Hewitt: Propagation of Radiowaves, The IEE, London, 1996.

[1.6.2.] R.E. Collin: Antennas and Radiowave Propagation, McGraw-Hill Book Company, New York, 1985.

[1.6.3.] L. Boithias: Radio Wave Propagation, North Oxford Academic, Publ. Ltd., London, 1987.

1.7. Forgalomelmélet

Szerző: dr. Molnár Sándor

Lektor: dr.Jereb László

1.7.1. Bevezetés

A forgalomelmélet [1.7.1], [1.7.7] a távközlési hálózatok teljesítményanalízisének és tervezésének az alaptudománya. Agner Krarup Erlang dán matematikus (1878-1929) [1.7.6], volt az első, aki a XX. század elején megfogalmazta a matematikai problémát. Az elmélet a telefonhálózatok fejlesztésével párhuzamosan fejlődött, és lényegi elemévé vált a klasszikus távközlési hálózatok tervezésének [1.7.7].

A forgalomelmélet, a kapcsolási és hálózati technikák átalakulása következtében lényegi változáson ment át az elmúlt évtizedben, és magabá olvasztotta az operációkutatás és a sorbanálláselmélet legújabb eredményeit is. A forgalomelmélet folyamatos evolúciója figyelhető meg, ahogyan a különféle tudományágak erdeményeit magába integrálja és alkalmazza. A forgalomelmélet tárgya a forgalmi igények, hálózati erőforrások és a teljesítményjellemzők közötti kapcsolat matematikai modellezése. A forgalmi igények a természetük szerint statisztikusak, így a sztochasztikus folyamatok elméletéből származtathatóak.

Ebben a fejezetben először áttekintjük a hálózati forgalom legfontosabb jellemzőit. A forgalom természete alpavetően meghatározza azt a forgalomelméletet ami ma a rendelkezésünkre áll. Ezután áttekintjük a forgalomelmélet elemeit, beleértve a jelöléseket, a rendszerek típusait és a legfontosabb forgalomelméleti összefüggéseket. Az ismertetett alapok alkalmazása a 3.3 fejezetben található, ahol a forgalmi modelleket és tervezési elveket ismertetjük.

1.7.2. A hálózati forgalom jellemzői

A hálózati forgalom jellemzői a mai adathálózatokban (pl. Internet) teljesen eltérnek a klasszikus telefonhálózatok forgalmának jellemzőitől, és egy bonyolultabb

problémával kell szembenéznünk [1.7.8], [1.7.9]. Az eltérés lényege, hogy a telefonforgalom lényegileg *statikus* jellegű. Ezért lehetett tipikus felhasználói viselkedést találni. A forgalmi jellemzők *korlátozott változékonysága* pedig lehetővé tette, hogy átlagértékekkel számoljunk, melyek jól leírták a forgalom jellegét.

A telefonhálózatok forgalmának statikus jellege tette lehetővé, hogy olyan "univerzális törvényeket" találjunk, mint a hívások keletkezésének *Poisson természete* [1.7.8], [1.7.9]. Ez a "szabály" azt mondja, hogy a hívások függetlenek és a hívások közötti idő exponenciális eloszlású. A Poisson hívásérkezési modellnek rendkívül nagy volt a népszerűsége az elmúlt ötven évben. A Poisson modell sikere az egyszerűségében rejlik, ami a gyakorlat számára egy fontos kritérium.

Egy hasonló "univerzális törvény" az, hogy a POTS forgalomban a hívások tartásideje közelítőleg *exponenciális eloszlású*. Ez a modell szintén a legtöbb esetben igen jól közelíti a valóságos folyamatokat. Ezenkívül értéke az egyszerűség és analitikus kezelhetőség. Annak ellenére használták ezt a modellt, hogy sokszor a telefonhívások tartásidejének eloszlása eltért az exponenciális jellegtől. Ez az eltérés azonban nem okozott jelentős hibákat az analízisben, ami szintén a Poisson hívásérkezési modellnek köszönhető. Ugyanis számos teljesítményjellemző Poisson érkezési folyamat esetén nem függ a tartásidő eloszlásától, csak annak átlagértékétől.

Jelentős változás következett be ezen "univerzális törvények" érvényességét illetően amikor a telefonhálózatokat nem csak beszéd, hanem FAX és Internet hozzáférésre is használni kezdték. Ezen szolgáltatások statisztikus természete ugyanis jelentősen eltér a beszédforgalométól. Például a hívások ídőtartama sokkal hosszabb és sokkal változékonyabb, mint a beszédhívások. A web népszerűségének növekedésével pedig egyre több és több ember kezdte el használni a klasszikus telefonhálózatot az Internet elérésére. Ezek a változások késztették arra a kutatókat, hogy felülvizsgálják a régi modelleket.

Az adathálózatok forgalmának természete jelentősen eltér a telefonhálózatokétól, de minden olyan próbálkozás, hogy a telefonhálózatoknál jól bevált "univerzális törvényeket" itt is találjunk, eddig kudarcot vallott [1.7.8]. Ennek a legfőbb oka, hogy az adathálózatok forgalma sokkal változékonyabb mint a beszédforgalom. Leegyszerűsítve azt is állíthatjuk, hogy lehetetlen találni általános modellt, mert az adatkommunikációban minden egyes kapcsolat ideje a nagyon

rövidtől az extrém hosszúig változhat és az adatsebesség is széles tartományban bármi lehet. Az adatforgalomnak nincs meg az a homogén természte, mint a beszédforgalomnak. Az adatforgalom börsztösségének az egyik fő oka az, hogy itt többnyire gépek kommunikálnak egymással és nem emberek.

Az adatforgalom nagy változékonyságát megtaláljuk mind az *idő* tartományában (a forgalom összefüggőségi kapcsolatai nem csengenek le exponenciális gyorsasággal, mint a beszédforgalom esetében hanem annál sokkal lassabban, hosszú idejű összefüggések is jelen vannak) mind a *méret* tartományában (a forgalmi egységek méretének eloszlása nem exponenciális, mint beszédforgalomnál, hanem a hosszú farkú eloszlások a tipikusak, pl. a web által letöltött objektumok mérete). Ezen tényezők miatt új modellek és technikák kifejlesztése fontossá vált. A hosszabb időskálájú összefüggőségi struktúrát statisztikusan leírhatjuk a *hosszú idejű összefüggőséggel* (long-range dependence, LRD) ahol az autokorrelációs függvénynek hatványalakú lecsengése van. Az extrém méretbeli változékonyságot pedig a végtelen szórású *hosszú farkú eloszlások* segítségével (heavy-tailed distributions*)* írhatjuk le, ami pl. a Pareto eloszlással való modellezést jelentheti. A hatványalakú lecsengési tulajdonságok mind térben és méretben gyakran okozzák azt, hogy a forgalomnak *fraktális* természete van [1.7.8].

A fraktáltulajdonsgok egyik fontos megjelenése az *önhasonlóság*. Ez azt jelenti, hogy a forgalom számos statisztikus jellemzője azonos marad több időskálán keresztül. Az egyszerű önhasonló modellek, amelyeket az elmúlt évtizedben széleskörben alkalmaztak úgy tűnik sikeresen tudják modellezni a fraktális adatforgalmat. A jelenlegi kutatások azonban azt mutatják, hogy a forgalomnak sokkal kifinomultabb börsztszerkezete van, amelyet inkább a *multifraktál* modellek segítségével lehet leírni. Ahol a monofraktál jellegű önhasonló modellek nem megfelelőek, ilyen multifraktál modellek alkalmazása válik szükségszerűvé [1.7.8].

Az adatforgalom jellegének nagy változékoysága mellett egyéb tényezők is megnehezítik az adatforgalom modellezését. Az Internet forgalma megduplázódik minden évben. Ez a gyors forgalom növekedés és az előre nem látható új, esetleg tömegesen jelentkező alkalmazások (“killer applications”) felboríthat minden jövőre vonatkozó forgalmi predikciót. Mindamellet eddig az Internet történetében csak három ilyen alkalmazás volt, ami drasztikusan megváltoztatta az Internet forgalom természetét (az e-mail, a web és a manapság terjedő Napster (zenei műsorok

cseréje) tipusú alkalmazások), de senki sem tudja, hogy mikor jelenik meg esetleg egy olyan népszerű alkalmazás ami alapvetően megváltoztatja az Internet forgalmának összetételét. A helyzet még bonyolultabbá válik, ha a különböző alkalmazások eltérő minőségi követelményeit (Quality of Service, QoS) is figyelembe vesszük, amelyek eltérő forgalmi jelleget is vonnak maguk után.

Annak érdekében, hogy leírhassuk a forgalom jellemzőit mind a streaming mind az elasztikus forgalom esetében számos modellt fejlesztettek ki. A megfelelő modell alapján pedig kidolgozható az alkalmas méretezési technika. A 3.3. fejezetben a fontosabb modelleket és méretezési elveket ismertetjük.

1.7.3. A forgalomelmélet alapfogalmai

Ebben a fejezeten a legfontosabb forgalomelméleti alapelveket ismertetjük [1.7.1].

1.7.3.1. Fogalmak és jelölésrendszer

A hálózatban egy kapcsolat felépítési igényt *hívásnak* nevezünk, amelyet egy *előfizető* kezdeményez. A hívás időtartama a *tartási idő* vagy *kiszolgálási idő*. A *forgalom* az egyégnyi időre eső teljes tatásidő. A forgalom egysége az *erlang* (erl vagy E) amelyet a forgalomelmélet megalapítójáról neveztek el.

A forgalomnak a következő fontos tulajdonságai vannak:

1. A forgalom (felajánlott forgalom) $a=ch$ (erl) ahol c az egységnyi idő alatt kezdeményezett hívások száma és h az átlagos tartásidő.
2. A forgalom (felajánlott forgalom) egyenlő az átlagos tartásidő alatt kezdeményezett hívások számával.
3. A forgalom (átvitt forgalom) amit egy trönk továbbít az egyenlő a trönk foglaltságának valószínűségével.
4. A forgalom (átvitt forgalom) amit egy trönkcsoport továbbít az egyenlő a csoportban levő foglalt trönkök számának átlagával.

1.7.3.2. Forgalomelméleti rendszerek osztályozása

Kapcsolórendszer a bemeneti portokat köti össze a megfelelő (kijelölt, kiválasztott) kimeneti potrtokkal. Egy rendszert *teljesen elérhetőnek* nevezünk, ha bármelyik bemeneti port bármelyik kimeneti porthoz csatlakoztatható. *Torlódásnak*

nevezzük a rendszer azon állapotát amikor valamilyen kapcsolat nem létesíthető a foglalt kimeneti portok vagy foglalt belső utak miatt. A rendszert *várakozásos rendszernek* nevezzük, ha torlódás esetén a bejövő hívás várakozni tud. Amennyiben nincs várakozási lehetőség torlódás esetén, a rendszert *veszteséges rendszernek* nevezzük.

A teljes elérhetőségű rendszereket a kövekezőképpen írhatjuk le [1.7.1]:

1. Bemeneti folyamat: Ez a hívások érkezésének folyamatát írja le.

2. Kiszolgáló mechanizmus: Ez leírja a kimenetek számát, a kiszolgálasi idő eloszlását, stb.

3. Sorbanállási diszciplína: Ez specifikálja a hívás kezelésének módját torlódás esetén. Késleltetéses rendszerekben a legegyszerűbb sorbanállási szabály a “first-in first-out” (FIFO), “last-in first-out” (LIFO), prioritásos rendszerek, processzor megosztás, stb.

A teljes elérhetőségű rendszerek osztályozására a *Kendall jelölésrendszert* használjuk [1.7.1], [1.7.3], [1.7.4] David A. Kendall brit statisztikus tiszteletére:

A/B/C/D/E-F

ahol

- *A* jelöli az érkezések közötti idő eloszlását,
- *B* jelöli a kiszolgálási idő eloszlását,
- *C* jelöli a párhuzamos kiszolgálók számát,
- *D* jelöli a rendszer kapacitását,
- *E* jelöli a véges felhasználó populáció nagyságát
- *F* jelöli a sorbanállási szabályt.

A következő jelölések használatosak:

- M: exponenciális (markovi)
- E_k: k-állapotú Erlang
- H_n: n-rendű hiperexponenciális
- D: determinisztikus
- G: általános
- GI: általános független
- MMPP: markov modulált Poisson folyamat
- MAP: Markov érkezési folyamat

Például az *M/M/1/∞/∞*-FCFS jelölés egy olyan sorbanállási rendszert reprezentál, melyben Poisson érkezések vannak és a kiszolgálási idő exponenciális eloszlású. A rendszernek egy kiszolgálója van és végtelen méretű tárolója. A felhasználók populációja végtelen és “először be először szolgál” a kiszolgálási szabály.

1.7.3.3. Alapösszefüggések

PASTA

Poisson érkezési folyamat estén (exponenciális az érkezések közötti idők eloszlása) állandósult állapotban egy tetszőleges pillanatban az aktív kapcsolatok számának eloszlása egyenlő az érkezés pillanatában levő kapcsolatok számának eloszlásával. Ennek az összefüggésnek *PASTA* (Poisson arrivals see time averages) a neve [1.7.1] mert ez a valószínűség egyenlő az aktív kapcsolatok átlagos időhányadával ha hosszú időn keresztűl vizsgáljuk a rendszert.

Markov tulajdonság

Exponenciális érkezések közötti idők eloszlása esetén egy tetszőleges időpillanat után a hátralévő idő eloszlása is exponenciális ugyanazzal a paraméterrel, az a modell, melynél az érkezések közötti idő és a kiszolgálási idő is exponenciális *Markovi modell* [1.7.1]. Ettől eltérő esetekben a modellt *nem Markovi modellnek* nevezzük.

Little Formula

Az $N=\lambda W$ összefüggést *Little formulának* [1.7.1], [1.7.3], [1.7.4] nevezzük, ahol N a rendszerben tartozkodó igények átlagos száma, λ az igények átlagos érkezési intenzitása és W az átlagos várakozási idő a rendszerben. A Little formula érvényes minden stacionárius rendszerre ahol a rendszerben igények nem születnek és nem vesznek el.

Veszteségi Formula

Annak a valószínűsége, hogy egy tetszőleges igény elvész a rendszerben [1.7.5]

$$P_{loss} = 1 - \frac{1-\phi}{\rho},$$

A veszteségi formula többkiszolgálós rendszerekre is érvényes, ahol ρ az egy kiszolgálóra jutó átlagos kihasználtság és ϕ a valószínűsége annak, hogy egy tetszőleges kiszolgáló szabad.

Hátralévő munka és a rendszerben levő igények száma

Állandó kiszolgálási idejű egy kiszolgálós rendszerben a következő az összefüggés a hétralévő munka (unfinished work) V_t és a rendszerben levő igények száma X_t között: $X_t = \lceil V_t \rceil$. Ez alapján a következő komplementer eloszlásfüggvény írható fel [1.7.5]:

$P(X_t > n) = P(V_t > n)$, ahol *n* egész.

Csomagvesztési valószínűség és sorhossz farokeloszlás

Tekintsünk egy diszkrét idejű *G/D/1* rendszert állandó csomaghosszal (cellák). A cellavesztési valószínűségre a következő felső korlátot írhatjuk fel [1.7.5]

$$\rho P_{loss} \le P(X_t^{\infty} > K),$$

ahol ρ a kihasználtság, X_t^{∞} a sorhossz egy hipotetikus végtelen kapacitású sorban.

Az általánosított Beneŝ formula

Tekintsünk egy kiszolgálási rendszert végtelen kapacitású tárolóval. Tételezzük fel, hogy a rendszer stacionárius, tehát a 0 időpont egy tetszőleges időpontot reprezentálhat. A kiszolgálási kapacitás 1 egységnyi munka egységnyi idő alatt. A rendszerben levő munka komplementereloszlás függvénye a 0 időpntban [1.7.5]

$$P(V_0 > x) = \int_{u>0} P(\xi(u) \ge x > \xi(u+du) \qquad \acute{e}s \qquad V_{-u} = 0),$$

ahol, u egy idő, $\xi(t) = A(t) - t$, $t \ge 0$, és *A(t)* a rendszerbe érkező munka idő egységben kifejezve a [-*t*,0) intervallumban. Ez az eredmény érvényes a legtöbb

számunkra érdekes sorbanállási rendszerre és rendkívűl hasznosnak bizonyult a forgalomelméletben.

1.7.4. Az M/G/1 sorbanállási rendszer

Az az egykiszolgálós sorbanállási rendszer melyben az igények Poisson folyamat szerint érkeznek és a kiszolgálási idő tetszőleges eloszlású (*M/G/1*) egy fontos kategória. A következőkben áttekintjük ezen rendszer lényeges jellemzőit [1.7.1], [1.7.2], [1.7.3], [1.7.4].

A következő jelöléseket használjuk:

- *W*: várakozási idő a sorban
- *T*: a rendszer válaszadási ideje
- N_q: igények száma a sorban
- *N*: igények száma a rendszerben
- *S*: kiszolgálási idő

Az átlagos várakozási idő és az igények átlagos száma a sorban az *M/G/1* rendszerben a következőképpen számítható [1.7.2]:

$$\overline{W} = \frac{\rho E(S)(1+c_S^2)}{2(1-\rho)},\ \overline{N}_q = \frac{\rho^2(1+c_S^2)}{2(1-\rho)}$$

ahol ρ a kihasználtság és c_S^2 a kiszolgálási idő relatív szórásnégyzete: $c_s^2 = \frac{E^2(S)}{S^2}$.

A rendszerben levő igények számának eloszlása a Pollaczek-Khinchin egyenlet segítségével számolható [1.7.2]:

$$G_N(z) = L_S(\lambda(1-z))\frac{(1-\rho)(1-z)}{L_S(\lambda(1-z))-z},$$

ahol $G_N(z)$ az *N* generátorfüggvénye, $L_X(s)$ az *X* Laplace transzformáltja és λ a Poisson folyamat érkezési intenzitása. Ez alapján a legfontosabb *M/G/1* rendszerekre a következő eredmények adódnak.

Sorbanállási rendszer	Sorhossz eloszlás P(N=n)
M/M/1	$(1-\rho)\rho^n$
M/H_2/1	$q(1-\alpha_1)\alpha_1^{\ n} + (1-q)(1-\alpha_2)\alpha_2^n$
M/D/1	$(1-\rho)\sum_{k=0}^{n} e^{k\rho}(-1)^{n-k}\frac{(k\rho+n-k)(k\rho)^{n-k-1}}{(n-k)!}$
M/E_k/1	$(1-\rho)\sum_{j=0}^{n}(-1)^{n-j}\frac{\alpha^{n-j-1}}{(1-\alpha)^{kj}}\left[\binom{kj}{n-j}\alpha+\binom{kj}{n-j-1}\right]$

A táblázatban H_2 a hiperexponeniális eloszlást jelenti α_1, α_2 és q paraméterekkel. E_k jelöli a k állapotú Erlang eloszlást α és k paraméterekkel. Ez a két eloszlás azért fontos, mert segítségükel tetszőleges *M/G/1* rendszert közelíthetünk. Ahol a kiszolgálási idő relatív szórásnégyzete 1 vagy az alatt van az *M/E_k/1* rendszert, ahol ez az érték 1 vagy annál nagyobb az *M/H_2/1* rendszert használhatjuk az approximációhoz.

1.7.5. Általános sorbanállási rendszerek

Az általános sorbanállási rendszerek (*G/G/n* sorok) rendszerint nehezen kezelhetőek, de van néhány olyan alosztály amely analitikusan is jól kezelhető. Például az *G/M/1* tipusú sorbanállási rendszerek kevésbé fontosak mint az *M/G/1* duálpárjuk, de az analízisük mégis egyszerűbb. Egy fontos eredménye a *G/G/1* rendszerek analízisének a *Lindley's integrál egyenlet* [1.7.1], [1.7.3] amivel a stacionárius várakozási idő eloszlása számítható:

$$F_W(t) = \int_{-\infty}^{t} F_W(t-v)dF_U(v)\,,$$

ahol *U =S-A* és *A* jelöli az időt két egymás után érkező igény között.

1.7.6. Forgalomelméleti technikák

A klasszikus sorbanállási módszereken kívül számos approximációt, korlátot és technikát dolgoztak ki a forgalomelméleti rendszerek kezelésére. Ebben a fejezetben áttekintünk néhányat a legfontosabbak közül.

A *folyadékmodell approximáció* [1.7.3] egy nagyon hasznos approximáció, amikor a vizsgált időskálán rengeteg forgalmi egység talalható (pl. csomagok).

Ebben az esetben a forgalmat tekinthetjük egy folytonos folyamnak, hasonlóan ahogy a folyadék áramlik egy csőrendszerben. Legyen *A(t)* és *D(t)* a (0,*t*) intervallumban érkező, illetve távozó igények száma. Az igények száma a rendszerben adott *t* időpontban *N(t)=A(t)-D(t)*, ahol feltételezzük, hogy a rendszer eredetileg üres volt. A nagy számok gyenge törvénye alapján amint *A(t)* növekszik egyre közelebb kerül annak átlagértékéhez, és ugyanez érvényes *D(t)* értékére is. A folyadékmodell approximációban egyszerűen az *A(t)* és a *D(t)* véletlen változókat helyettesítjük azok átlagával. Így végeredményben egy folytonos determinisztikus folyamatot kapunk. A folyadékmodelleket nagyon gyakran használják a forgalomelméletben.

A folyadékmodell approximáció az átlagértékeket használja, de az érkezési folyamat változékonysága ezen értékek körül nincs számításba véve. A *diffúziós approximáció* [1.7.3] kiterjeszti ezt a modellt és a központ határeloszlás tétel alapján az átlag körüli változásokat normális eloszlással modellezi. A diffúziós approximáció segítségével bonyolult sorbanállási rendszereket analizálhatunk. Például a complex *G/G/1* rendszer sorhossz elszlása számítható diffúziós approximáció segítségével.

Egy érdekes módszer amit a *maximum entrópia módszerének* [1.7.3] hívnak az információelméletből származik. Az alapelv a Bernoulli féle elégtelen információk elve, amely azt állítja, hogy ha semmilyen információnk nincs egy valószínűségi változóról, akkor egyenletes eloszlásúnak feltételezhetjük. Egy valószínűségi változó entrópiája minimum (zéro) amikor értéke ismert, bizonyos. Az entrópia maximális amikor a változó egyenletes eloszlású, mert egy esemény kimenetelének maximális a bizonytalansága. Az ötlet az, hogy az entrópiát maximalizálhatjuk külső feltételek ismeretében. A módszert sikerrel alkalmazzák a sorbanálláselméletben.

Egyéb módszereket, mint pl. a sorbanállási hálózatok számos megoldási technikával (fixpontos módszer, dekompozíciós technikák) szintén szép számmal fejlesztettek ki. Az érdeklődő olvasók az irodalomjegyzékben találnak referenciákat ezekre.

Irodalomjegyzék

[1.7.1] H. Akimaru. K. Kawashima: Teletraffic, Theory and Applications, Springer-Verlag, 1999.

[1.7.2] R. Nelson: Probability, Stochastic Processes, and queueing Theory, Springer-Verlag, 1995.

[1.7.3] P. G. Harrison, N. M. Patel: Performance Modelling of Communication Networks and Computer Architectures, Addison-Wesley, 1993.

[1.7.4] R. Jain: The Art of Computer Systems Performance Anaysis, Wiley, 1991.

[1.7.5] J. Roberts, U. Mocci, J. Virtamo (eds.), Broadband Network teletraffic, Springer-Verlag, 1996.

[1.7.6] E. Brockmeyer, F. L. Halstrom, A. Jensen: The Life and Works of A. K. Erlang, Acta Polytechnica Scandinavica, 1960.

[1.7.7] R. Syski, Introduction to Congestion Theory in Telephone Systems, Oliver and Boyd Ltd. 1960.

[1.7.8] W. Willinger, V. Paxson: Where Mathematics Meets the Internet, Notices of the American Mathematical Society, vol.45, no.8, Aug. 1998, pp. 961-970.

[1.7.9] J. Roberts, Traffic Theory and the Internet, IEEE Communications Magazine, January 2000.

1.8. Adatvédelem és nyilvánosság

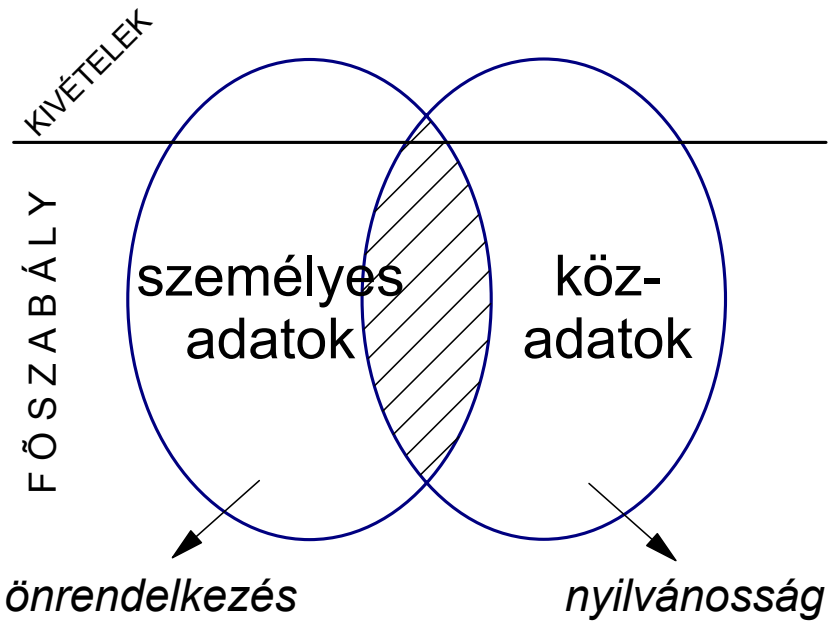
Szerző: Székely Iván

Lektor: dr. Vajda István

1.8.1. Alapmodellek

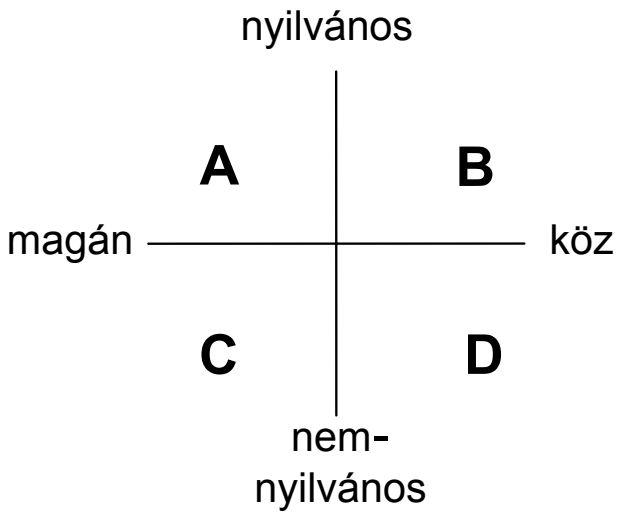
Az információktól a strukturált, visszakereshető formában rögzített adatokig és az adatoktól a kontextusba ágyazott, újraértelmezett információig terjedő teljes vertikum sajátos szempontú osztályozásának alapja az adatok és információk személyes, illetve közérdekű volta. E két alapkategória lefedi az információs rendszerekben kezelt és a távközlő hálózatokon továbbított adatok teljes körét (1.8.1. ábra). Noha ez a kategorizálás eltér a távközlés és az informatika területén szokásos műszaki-tudományos osztályozás logikájától, mégis olyan alapvető szabályokat vezet be, amelyek meghatározzák az adatkezelés kereteit és befolyásolják annak műszaki megvalósítását. E szabályok nem csupán jogi vagy társadalomtudományi jellegűek: az adatkezelés filozófiájától annak tételes alapelvein, a jog, a szabályozás és önszabályozás eszközein át az információs-kommunikációs technológiáig terjednek, s végső soron a korszerű adatkezelés közegében a nyilvánosság és titkosság alapvető kereteit határozzák meg.

Az ábrán látható *Székely-féle* (filozófiai-jogelméleti indíttatású) *modellben* mindkét alapvető adatkategóriára egy-egy főszabály vonatkozik: a személyes adatokra az önrendelkezés, a közadatokra a nyilvánosság. (E két fogalom jogi terminológiában használt megfelelője az információs önrendelkezés, illetve az információszabadság.) A vonal alatti területeken érvényesülnek a főszabályok, a vonal felett a kivételek. Az átfedő (satírozott) terület azon személyes adatokat tartalmazza, például a közfunkciót betöltő személyek e tevékenységével összefüggő személyes adatait, amelyekre nem az önrendelkezés, hanem a nyilvánosság főszabálya vonatkozik. A Székely-féle alapmodell kritikája, hogy nem jeleníthetők meg rajta a nem állami (üzleti, társadalmi) szervezetek adatkezelési viszonyai. Továbbfejlesztett változata ezért nem három, egymást részben átfedő körből vagy ellipszisből áll, hanem három hosszúkás, ívelt idom gyűrűjéből, ahol a szomszédos



idomok végei átfedik egymást. *Heller* és *Rényi* (szociológiai-tömegkommunikációs indíttatású) modelljében az információ magán–köz, és nyilvános–nem nyilvános attribútumai egymástól függetlenül, egy kétdimenziós koordinátarendszerben jelennek meg, s az így létrejövő négy tartományból kettőt "természetesnek", kettőt pedig magyarázatot igénylőnek, azaz kivételesnek tekinthetünk (1.8.1./b ábra).

Az adatok bárki számára, vagy csakis meghatározott személyek számára hozzáférhetővé tételét, és mások számára hozzáférhetetlenné tételét számos jog és érdek határozza meg. E jogok és érdekek testesülnek meg a hagyományos, nevesített titokkategóriákban, s érvényre juttatásukhoz az információs technológia mindenkori állásának megfelelő eszközök és eljárások tartoznak. Fontosabb nevesített titokkategóriák: üzleti titok, államtitok, szolgálati titok, magántitok, ügyvédi titok, orvosi titok, banktitok, gyónási titok stb. E kategóriák egy része (pl. orvosi titok) ráerősít a főszabályokra, más része (pl. államtitok) azok jelentős kivételeit képezik, s ennek megfelelően a modell különböző tartományaiban helyezhetők el. Általános jellemzőiket többnyire törvények és más jogszabályok, etikai kódexek tartalmazzák; konkrét alkalmazásukat e kereteken belül az információs rendszerek felelős kezelői határozzák meg. A személyes/közérdekű dichotómia azonban e titokkategóriák alkalmazásakor is érvényesül.

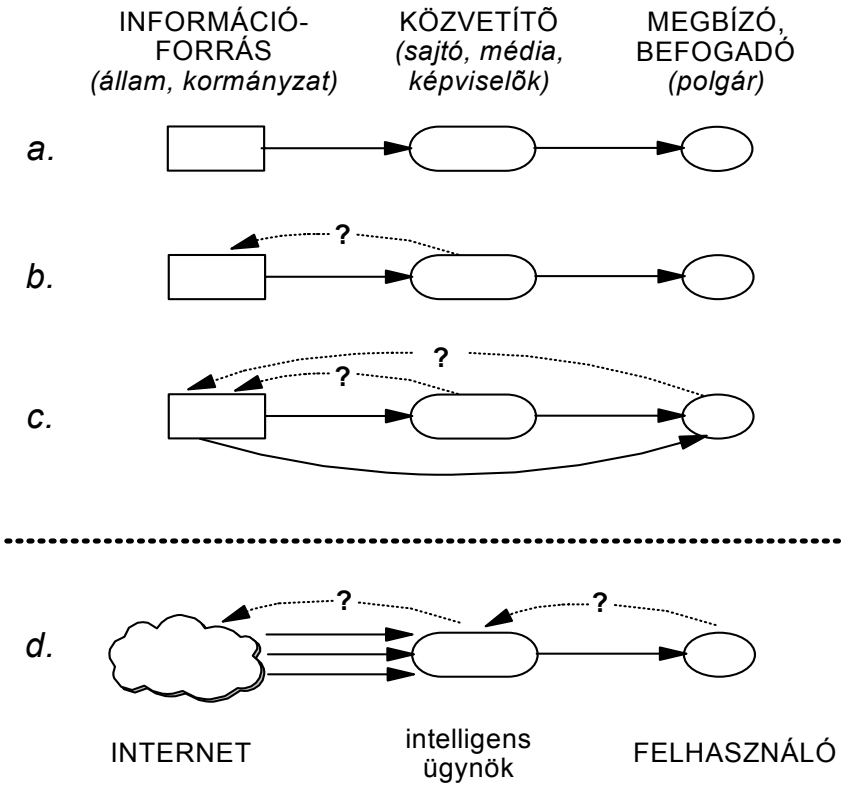


1.8.1./b ábra

Érdemes megjegyezni, hogy a fenti adatkezelési kategóriák és főszabályok már jóval a gépi adatfeldolgozás és a korszerű távközlés megjelenése előtt kialakultak, aktualitásukat azonban ezek elterjedése jelentősen erősítette. A mindenkori korszerű informatika és távközlés alkalmazása ugyanis számos elméleti és gyakorlati problémát vet fel az egyén és az információs hatalom, vagy más kapcsolatrendszerben az állam és az állampolgár, az üzleti szféra és az ügyfél, vagy általánosságban az információs szempontból erősebb és gyengébb fél viszonyában. E problémák egyik fő ága a személyes magánszféra információs határainak megváltozásából, az információs hatalom *mint az egyént ellenőrző és befolyásoló tényező* koncentrálódásából, a másik fő ága az egyén társadalmi részvételét meghatározó információs státusának megváltozásából, az információs hatalom *mint közinformációkat kezelő monopólium* koncentrálódásából ered.

Míg az információs önrendelkezés biztosításának történelmi folyamatát egyensúlyi állapotok és azoknak az új ICT által indukált felbomlási szakaszai sorozataként értelmezhetjük, az információszabadság elvi és gyakorlati megvalósulását evolúciós folyamatként. [1.8.2.]

A közinformációhoz való hozzáférés evolúciós modellje



1.8.2. ábra

Az *a.* szakaszt, leegyszerűsítve, a *képviseleti demokrácia* információs modelljének tekinthetjük: a képviselő (és az átvitt értelemben a polgárt képviselő sajtó) eljuttathatja a közinformációkat azok forrásától a befogadókig. A *b.* szakasz a *sajtószabadság* modellje: a "képviselő" nemcsak közvetíthet, hanem privilégiumait kihasználva követelhet is információkat, s azokat eljuttathatja a befogadókig. A *c.* szakasz a ma aktuális *információszabadság* modellje: a befogadó maga is közvetlenül követelhet információkat és azokat a közvetítő kiiktatásával kaphatja meg. (Az információk közvetítése a korszerű ICT alkalmazásával vagy anélkül is történhet.)

Az információszabadság joga az internet használatának általánossá válásával a *dezintermediáció* illúzióját veti fel. A közvetítők kihagyása azonban nemcsak azért illúzió, mert a globális hálózaton (jogszerűen) eleve csak az az információ található meg, amit oda valaki a nyilvános hozzáférés biztosítása céljából feltett, hanem azért

is, mert a felhasználó két alapvető problémával: a mennyiségi és a minőségi problémával szembesül. E problémák részleges orvoslására született reintermediációs megoldások egyike a *d.* szakasznál ábrázolt *intelligens ügynök* alkalmazása, ahol a személyre szóló tudásbázist tartalmazó ügynök előszelektálja, mintegy előemészti a befogadó által kért információt. (Meg kell jegyezni azonban, hogy a közinformációkhoz való hozzáférés megkönnyítésére alkalmazott ügynök egyúttal a felhasználó manipulálásának egyik leghatékonyabb, ma még kevéssé ellenőrizhető eszköze.) Itt az "internet" természetesen csak virtuális információforrás, mögötte valódi forrás áll; a szereplők megváltozását a szaggatott vonal jelzi.

1.8.2. Meghatározások

Az adatkezelés fenti attribútumaihoz kapcsolódó fogalmak egységes használata ma már általános követelmény a korszerű információs és távközlő szolgáltatásokban. (Magyarországon még ma is születnek színvonalas művek téves/elavult fogalomhasználattal, s a műszaki értelmiség körében az adatvédelem/adatbiztonság fogalom-páros ma is sok esetben félreértések forrása.) Az alábbiakban a legfontosabb fogalmak rövid meghatározását adjuk a nemzetközileg elfogadott terminológia alapján.

Adatvédelem (data protection): a személyes adatok gyűjtésének, feldolgozásának és felhasználásának korlátozását, az érintett személyek védelmét biztosító alapelvek, szabályok, eljárások, adatkezelési eszközök és módszerek összessége. (Az adatvédelem fogalmilag csak személyes adatok esetében értelmezhető.)

Adatbiztonság (data security): itt használt értelmében az adatok jogosulatlan megszerzése, módosulása és tönkremenetele elleni műszaki és szervezési megoldások rendszere. (Az adatbiztonság személyes és nem személyes adatok esetében egyaránt értelmezhető.)

Röviden: az adatvédelem az *adatalanyok* védelme, az adatbiztonság maguké az adatoké.

Személyes adat: a meghatározható természetes személlyel kapcsolatba hozható adat, az adatból levonható, az érintett személlyel kapcsolatba hozható

következtetés. Személyes minőségét az adat mindaddig megőrzi, amíg kapcsolata az érintettel helyreállítható.

Közérdekű adat: az állami vagy helyi önkormányzati feladatot vagy egyéb közfeladatot ellátó szerve vagy személy kezelésében lévő, a személyes adat fogalmába nem tartozó adat. (Az adat közérdekű mivoltában tehát nem annak deklarálása, nem az adatkezelő tulajdoni formája, hanem a közfunkció a döntő elem.)

Adatalany: az érintett személy, akivel az adat kapcsolatba hozható.

Adatkezelés: az alkalmazott eljárástól függetlenül a személyes adatok felvétele, tárolása, feldolgozása, hasznosítása, megváltoztatása, továbbítása, nyilvánosságra hozatala.

Adatfeldolgozás: az adatkezelési műveletek, technikai feladatok elvégzése, függetlenül az alkalmazott eszköztől és módszertől.

Adatkezelő: az a természetes vagy jogi személy, vagy jogi személyiséggel nem rendelkező szervezet, aki (amely) az adatkezelés célját meghatározza, a rá vonatkozó döntéseket meghozza és végrehajtja, illetőleg a végrehajtással adatfeldolgozót bízhat meg.

Adatfeldolgozó: az a természetes vagy jogi személy, vagy jogi személyiséggel nem rendelkező szervezet, aki (amely) az adatkezelő megbízásából személyes adatok feldolgozását végzi. (Adatkezelő és adatfeldolgozó tehát önálló szerv vagy személy; kapcsolatuk ma jellemző példája az *outsourcing.* Az adatkezelés jogszerűségéért az adatkezelő felel.)

1.8.3. Az adatvédelem alapelvei

Az információs önrendelkezés főszabálya és kivételei érvényre juttatásának következő szintjét a nemzetközileg elfogadott tartalmú, tételesen megfogalmazott adatvédelmi alapelvek képezik. Az alábbiakban az alapelveket az OECD Adatvédelmi Irányelveinek [1.8.2.] csoportosítását követve, kivonatosan ismertetjük.

1. *Az adatgyűjtés korlátozásának elve*

Személyes adatok gyűjtése csak törvényes és tisztességes eszközökkel, az adatalany tudtával és belegyezésével történhet.

2. Az *adatminőség elve*

Az adatoknak az adatkezelés céljával összhangban pontosnak, teljesnek és aktuálisnak kell lenniük.

3. A célhozkötöttség elve

Személyes adatokat csak előre meghatározott célból, csak a cél megvalósulásához szükséges mértékben és ideig lehet kezelni.

4. A *korlátozott felhasználás elve*

Az adatokat csak az adatalany hozzájárulásával vagy törvényi felhatalmazással lehet felhasználni.

5. A *biztonság elve*

Az adatokat a technika mindenkori állásának megfelelő ésszerű intézkedésekkel védeni kell a jogosulatlan hozzáférés, megváltoztatás, nyilvánosságra hozás, sérülés és megsemmisülés ellen.

6. A *nyíltság elve*

Az adatkezelés tényének, helyének és céljának, az adatkezelő személyének, valamint az adatkezelési politikának nyilvánosnak kell lennie.

7. A *személyes részvétel elve*

Az adatalany megismerheti a rá vonatkozó adatokat, azokat (ha helyénvaló) helyesbítheti, kiegészítheti vagy töröltetheti.

8. A *felelősség elve*

Az adatkezelő a felelős a fenti elvek betartásáért, s bizonyítani kell tudnia az adatkezelés jogszerűségét.

1.8.4. Nemzeti és nemzetközi szabályozás

Az adatvédelem következő szintjét a nemzetközi szerződésekben, irányelvekben és más dokumentumokban meghatározott feltételek és követelmények rendszere alkotja. A legfontosabb nemzetközi dokumentumok: az OECD Adatvédelmi Irányelvei, az Európa Tanács Adatvédelmi Egyezménye [1.8.3.] és az Európai Unió Adatvédelmi Direktívája. [1.8.4.] Megjegyzendő, hogy a magyar jogi szakirodalom a direktívát is általában "irányelv"-nek fordítja; az eredeti kifejezés megtartását az indokolja, hogy az irányelv (guidelines) követése nem kötelező, míg a direktívában (directive) foglaltak bevezetése a belső jogba igen.

Az 1980-ban született OECD irányelvek és az 1981-es ET egyezmény párhuzamosan készült, mindkettő tartalmazza az adatvédelmi alapelveket, de amíg az OECD irányelvek a határátlépő adatáramlás szükségességét (és annak garanciáit) hangsúlyozza, addig az ET egyezmény célja az információs jogok biztosítása a határátlépő adatáramlásban. További különbség, hogy az irányelvek követése csak ajánlott, az egyezmény betartása pedig kötelező a csatlakozó országok számára (Magyarország 1993-ban aláírta, 1997-ben ratifikálta és 1998-ban kihirdette az egyezményt).

Az 1995-ben, ötéves vita után elfogadott EU direktíva az adatkezelés azon közös, részletes szabályait határozza meg, amelyeket az EU tagállamoknak kötelező belső jogukba emelniük; ennek határideje 1998 októbere volt. Amelyik országnak volt korábbi adatvédelmi törvénye és gyakorlata, azt szükség esetén hozzá kellett igazítani a direktíva előírásaihoz, az újonnan hozott jogszabályok, tagfelvételre váró országokban is, már a direktíva szellemében készültek.

Ahogy Colin Bennett kanadai politológus már a nyolcvanas években felismerte, az adatkezelés terén nemcsak technológiai konvergencia, hanem "policy konvergencia" is tapasztalható. Fontos gyakorlati következményekkel jár azonban, hogy amíg az ET egyezmény a szerződő országok számára *azonos (ekvivalens) védelmet* ír elő, addig az EU direktíva csak *megfelelő (adekvát) védelmet*. Az egyezményen, és az EU tagországain kívüli, harmadik országba ugyanis csak akkor lehet korlátozások nélkül, például automatikus távközlő és információs rendszerek segítségével személyes adatokat továbbítani, ha a harmadik ország a megkövetelt adatvédelmi szintet képviseli. Az ekvivalens védelem azonos szabályozást és gyakorlatot követel, az adekvát védelem viszont eltérő szabályozási környezetben és alternatív eszközök és módszerek alkalmazásával is elképzelhető.

Az adekvát védelem problémája kiélezetten jelentkezik az EU és az Egyesült Államok vitájában. Az európai és az amerikai modell eltéréseit az *1.8.1.. táblázat* foglalja össze. Az európai modellt a nyugat-európai országok és a fejlettebb új demokratikus országok képviselik, az amerikai modellt (kisebb eltérésekkel) az USA, Ausztrália, Új-Zéland, valamint – korábban – Kanada.

Egy nem-adekvát védelmi kategóriába tartozó országgal szemben a személyes adatok határátlépő áramlásában korlátozásokat kell alkalmazni, s ez a globalizálódó információáramlás korszakában jelentős gazdasági és politikai

	Európai modell	Amerikai modell
Szektor:	magán + köz	köz
feldolgozás:	automatikus + manuális	automatikus
Lefedés:	Általános	mozaikszerű
ellenőr:	Van	nincs

1.8.1. táblázat. Adatvédelmi szabályozás

következményekkel járhat. Jelenleg az EU illetékes testületei, ellentmondó határozatok sora után, ideiglenes jelleggel adekvát adatvédelmi szintűnek fogadja el a Safe Harbor Principles elnevezésű, az adatkezelők önkéntes csatlakozásán és önkorlátozásán alapuló elvek követőinek tevékenységét, a csatlakozók száma azonban alacsony, az elvek megsértésének pedig nincs szankciója. Magyarország, Svájc után második EU-n kívüli országként 2000-ben hivatalosan megkapta az adekvát státust.

A nemzetközi szinten meghatározott adatkezelési kritériumok azonban közvetlenül is alkalmazandók az adatkezelési rendszerek tervezésében és működtetésében: az adatalany hozzájárulásának három, az EU által meghatározott kritériumának (önkéntesség, határozottság, tájékozottság) például a webes felületű online adatkezelési rendszereknek is eleget kell tenniük. Hasonlóképpen, a személyes részvétel elvének érvényesíthetősége (például az egyénekkel kapcsolatba hozható tranzakciók elkülöníthetősége és visszakereshetősége) az adatalanyok átgondolt azonosítási rendszerének kialakítását igényli az adatbázisokban.

A *belső (nemzeti) jog* az adatvédelem soron következő szintje. A magyar szabályozás korszerű, az európai hagyományokat követi. Sajátossága, hogy az adatvédelmet és az információszabadságot egy közös törvény szabályozza. A szabályozás vertikuma az Alkotmánytól a keretjellegű Adatvédelmi törvényen [1.8.5.] át a szektorális adatvédelmi törvényekig, és rendeletekig terjed.

A nemzeti adatvédelmi jog érvényesülését független ellenőri intézmények ellenőrzik. Ezek hatáskörüket és legitimitásukat tekintve többfélék: lehetnek testületek, mint a francia CNIL (Commission Nationale de l'Informatique et des Libertés), egyszemélyi tisztségek, mint a brit Information Commissioner (korábban Registrar); választhatja őket parlament (Németország) vagy kinevezheti kormány (Hollandia); jogosítványaik lehetnek hatósági, bírói vagy ombudsmani jellegűek. A

magyar adatvédelmi biztost a másik két országgyűlési biztossal együtt 1995-ben választotta meg először a Parlament; jogosítványai ombudsmani jellegűek és mind az adatvédelem, mind az információszabadság területére kiterjednek. Tevékenységének súlypontját a panaszügyek kivizsgálása képezi, de hivatalból is indíthat vizsgálatot állami és magánszektorbeli adatkezelőknél egyaránt. Vizsgálatának eredményei ajánlások, amelyek követése jogi értelemben nem kötelező, elfogadottságuk aránya azonban magas. Véleményezi az adatkezelést érintő jogszabályok tervezetét és ellenőrzi az állam- és szolgálati titokká minősített adatok minősítésének indokoltságát. Titokfelügyeleti jogosítványa hatósági jellegű, visszaminősítésre vonatkozó felhívásait végre kell hajtani. [1.8.6.]

1.8.5. PET technológiák

Az adatvédelmi elvek és rendelkezések megvalósításának technológiai szintjét képviselő PET (Privacy Enhancing Technologies) összefoglaló néven ismert változatos információs és kommunikációs technológiákat abból a célból fejlesztették ki, hogy ne csak az adatokat, hanem az adatok *alanyait* is védjék a visszaélések ellen. A PET technológiák célja a tágabb technológiai környezet által okozott magánéleti sérelmek kifejezett csökkentése; közelebbről az, hogy az új technológiák által nyújtott előnyöket a személyes magánszféra további sérelme nélkül (vagy e sérelmeket legalább is mérsékelve) lehessen igénybe venni. A rendeltetésszerűen használt PET eszközök és rendszerek mindig a *gyengébb felet* (jellemzően az adatalanyt) védik az információs túlhatalommal rendelkező erősebb féllel szemben.

A PET technológiák többféle szempont szerint csoportosíthatók.

(a) A *Burkert-féle csoportosítás* [1.8.7.] szerint vannak

- szubjektum-orientált,
- objektum-orientált,
- tranzakció-orientált, és
- rendszer-orientált technológiák.

Az első esetben a technológia az adatalanyra irányul (pl. kártyabirtokosok anonimitásának biztosítása), a második esetben az eszközre (anonim fizetőeszközök), a harmadik esetben a hálózati tranzakció nyomainak eltüntetése a

cél, a negyedik esetben pedig több technológia rendszerszerű közös alkalmazásáról van szó.

(b) Más szempont szerint megkülönböztethetünk

- meglévő rendszerek biztonságát növelő technológiákat,
- új adattárolási és -hozzáférési technológiákat, és
- tranzakció-alapú technológiákat.

(c) Csoportosíthatjuk a PET technológiákat aszerint, hogy melyik adatvédelmi alapelv érvényesülését segítik elő.

(d) Végül megkülönböztethetünk

- technológia-alapú, és
- humán interakció alapú
- PET-eket.

Fontos megjegyezni, hogy ugyanannak a halmaznak többféle szempont szerinti felosztásáról van szó. Néhány példa PET alkalmazásokra:

Bioscrypt. A biometrikus rejtjelezés (biometric encryption) két kiinduló elemből, egy biometrikus elemből (pl. digitalizált ujjlenyomat) és egy nem-biometrikus elemből (pl. kulcs, PIN vagy pointer) hozza létre a bioscryptet. A bioscryptes alkalmazások működtetésének feltétele a biometrikus elem forrásának produkálása, vagyis az adatalany jelenléte. A biometrikus rejtjelezés sajátos alkalmazásai közé tartozik az "anonim adatbázis", ahol a személyazonosító adatok és a lényegi adatok egy adatbázis mezőiben kvázi-véletlenszerűen vannak elszórva, s a köztük lévő kapcsolatot a bioscryptben lévő pointer tartalmazza. A (b) csoportosítás szerint új adattárolási és hozzáférési technológiáról van szó, amely a (c) szerint a célhozkötöttség elvének érvényesülését segíti elő. Felhasználható hálózaton történő személyközi kommunikáció résztvevőinek és az üzenet tartalmának illetéktelenek előli elrejtésére; ehhez a partnereknek közös bioscryptet kell előállítaniuk, amelyet bármelyikük biometrikus elemével (ujjlenyomatával) fel lehet bontani, s felszabadíthatják a bioscryptbe csomagolt szimmetrikus kulcsot. Felhasználható továbbá elektronikus kereskedelmi alkalmazásokban: itt három résztvevős adatkapcsolatról van szó (Bank, Ügyfél, Bolt), ugyancsak közös bioscrypttel, de a Bank itt ujjlenyomat vagy hasonló biometrikus elem helyett egy megfelelő bonyolultságú egyedi mesterséges mintázatot használ a bioscrypt előállításához.

Platform for Privacy Preferences (P3P). Internetes, jellemzően angol nyelvű elektronikus kiskereskedelmi alkalmazásokra kifejlesztett technológia, amely a távoli adatkezelő és az adatalany közti személyesadat-áramlást egy standardizált alkufolyamattá alakítja. Humán interakció alapú technológia, amely egyúttal a személyes részvétel elvének érvényesülését segíti elő (az adatalany tudtával és ellenőrzésével kerülnek adatai a távoli adatkezelő birtokába).

Anonim remailer. Elektronikus levelező szolgáltatások, amelyek nemcsak az üzenet tartalmát, hanem annak tényét, időpontját, gyakoriságát, résztvevőinek kilétét is elrejtik az illetéktelen megfigyelő elől. Fejlett (Mixmaster típusú) változataik az üzenetek késleltetését, véletlenszerű továbbítását, az azonos üzenetek kiszűrését, a remailerek láncbafűzését, valamint szabványos üzenetméretet és rétegesen kódolt formátumot alkalmaznak.

Digitális fedőnevek. Számos PET alkalmazás kínál internetes szolgáltatások felhasználóinak ún. *nym*-eket. Ezek az egyedi digitális fedőnevek az egyes adatkezelőkkel vagy bizonyos típusú szolgáltatókkal való kapcsolat állandó *alias*-aiként, vagy egyszer használatos azonosítókként használhatók fel.

Hitelesítés azonosítás nélkül. A pénzinformatikai alkalmazások céljára kifejlesztett PET technológiák egyik alapkérdése a "hitelesítés kontra azonosítás" problémája. Az elméleti kutatások és a létező rendszerek tapasztalatai egyaránt igazolják, hogy a bank mint adatkezelő lemondhat az ügyfelek tranzakcióinak teljes körű ismeretéről, s ez paradox módon erősíti a rendszer adatbiztonsági szintjét, egyúttal biztosítja a tranzakciók hitelességét. A két elvi megoldás egyike a tranzakció részekre bontása oly módon, hogy a hitelesítés a tranzakció egészén végigkövethető legyen, az azonosítás azonban mindig csak két-két pont között jöjjön létre, s az egyes pontokon az egyedi ügyfélazonosítón egyirányú (nehezen visszafejthető) átalakítás történjen. A másik megoldás az ismétlődő, triviális tranzakciók egyszer használatos digitális fedőneveken történő bonyolítása.

1.8.6. . Adatvédelmi szakértelmek, teendők

A technológiai konvergencia és az online szolgáltatások terjedése szükségessé teszi az adatvédelmi követelmények gyakorlati érvényesítésének hozzáigazítását a változó technológiai, jogi és szervezeti környezethez. Különös

figyelmet igényel az adat és az adatalany közötti kapcsolatba hozhatóság kritériumainak rögzítése, a kapcsolat helyreállíthatóságának értelmezése, továbbá az adatkezelések összekapcsolhatósági feltételeinek, az adatalany online hozzájárulásának adatvédelmi szempontú biztosítása, valamint az adattovábbítás (különösen az automatikus adattovábbítás) feltételeinek meghatározása, a betekintési és törlési jog érvényesíthetőségének biztosítása, az adatkezelő és adatfeldolgozó viszonyának és funkcióinak meghatározása.

Az adatvédelmi követelmények érvényesítéséhez mind statikus, mind változó környezetben rendszerszemléletű, a tágan értelmezett informatikai, jogi és társadalomtudományi ismereteket magában foglaló szakértelem szükséges. Az adatkezelések adatvédelmi követelményeit, szervezeti szintű eljárási szabályait és felelősségi viszonyait célszerűen belső *adatvédelmi szabályzat* tartalmazza. Az adatkezelések adatvédelmi szempontú ellenőrzésének célszerű módja az *adatvédelmi auditálás,* amely egységes rendszerben vizsgálja a szervezet személyesadat-kezelésének folyamatát, az alkalmazott eljárásokat és technológiákat, a személyesadat-tartalmú termékeket és szolgáltatásokat. (Az adatvédelmi audit nem azonos az *adatbiztonsági* audittal, az utóbbinak egyes eredményei azonban részét képezik az adatvédelmi auditálás során vizsgált területeknek.) Előzetes adatvédelmi szempontú vizsgálatot indokolt végezni új személyesadat-kezelési rendszerek tervezésénél, felépítésénél. Az adatvédelmi státust rendszeresen ellenőrizni szükséges.

Irodalomjegyzék

[1.8.1.] Székely Iván: Az adatvédelem és az információszabadság filozófiai, jogi, szociológiai és informatikai aspektusai. Kandidátusi értekezés. Budapest, 1994.

[1.8.2.] Guidelines on the Protection of Privacy and Transborder Flow of Personal Data. Organisation for Economic Co-operation and Development (OECD), Paris 1980.

(Magyarul: Az OECD Tanács ajánlása a magánélet védelmét és a személyes adatok határátlépo áramlását szabályozó irányelvekre. In: INFORMATIKA ? JOG ? KÖZIGAZGATÁS, Nemzetközi dokumentumok I., InfoFilia, Budapest 1991., 5.1–5.39 old.)

[1.8.3.] Convention for the protection of individuals with regard to automatic processing of personal data. Council of Europe, Strasbourg, 28 January, 1981. European Treaty Series No. 108. (Magyarul: Egyezmény a személyiségnek a személyes adatok automatikus kezelésével kapcsolatos védelmérol. In: INFORMATIKA ? JOG ? KÖZIGAZGATÁS, Nemzetközi dokumentumok I., InfoFilia, Budapest 1991., 3.1–3.55 old.)

[1.8.4.] Directive 95/46/EC of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data. Official Journal of the European Communities No. L 281/31, 23.11.1995

(Magyarul: Az Európai Parlament és a Tanács 95/46/EC Irányelve az egyénnek a

személyes adatok feldolgozásával kapcsolatos védelmérol és ezeknek az adatoknak

a szabad áramlásáról. Adatvédelmi Biztos Irodája, Budapest 1995.)

[1.8.5.] 1992. évi LXIII. törvény a személyes adatok védelméről és a közérdekű adatok nyilvánosságáról.

[1.8.6.] Az adatvédelmi biztos beszámolója 1995–96, 1997, 1998, 1999, 2000 Budapest.

[1.8.7.] Herbert Burkert: Privacy-Enhancing Technologies: Typology, Critique, Vision.

In: Agre, P.E. – Rotenberg, M. (eds.): Technology and Privacy: The New Landscape.

MIT Press, 1997.

1.9. Adatbiztonság

Szerző: dr. Nemetz Tibor

1.9.1. Alapfogalmak

Ismeretek, gondolatok rögzítésének igénye az emberiség fejlődésének korai szakaszában megjelent: ábrák, képek, jelek, majd később betűk segítségével írásban rögzítették ezeket. Az írás célja kettős volt: valaki a későbbi időben való reprodukálhatóság érdekében akarta rögzíteni gondolatait, vagy egyszerűen közölni kívánta másokkal (általában térben, esetleg időben távol levőkkel) ezeket a gondolatokat. Az írásbeliség kialakuláával gyakorlatilag egyidejű az az igény, hogy a rögzített ismereteket ne szerezhesse meg illetéktelen személy, még akkor sem, ha a teljes írott szöveg birtokába jút. Az igény elsősorban a hatalom világi és egyházi urai részéről jelentkezett. Hadvezérek, diplomaták részéről az igény természetes és alapvető jelentőségű volt. Gyakorlatilag egyaránt kisajátították maguknak az elméletet és a gyakorlatot

Kezdetben az információ titkosságának biztosítására fizikai megoldásokat használtak: értelmes szöveget tartalmazó információ hordozókba, vagy mesterségesen konstruált szövegekbe rejtették el az információt. A bizalmas közléseket teljesen elrejtő írásos formákra (például láthatalan írásra) szolgál a görög eredetű **szteganográfia** elnevezés. A betűírás kialakulásával azonban az információ rejtésnek hamarosan kialakult egy egyszerűbben megvalósítható, ám mégis biztonságosnak tartott módja, a helyettesítés. Az egyik első betű-kódot a zsidók már krisztus előtt használták. Ábécéjük első betűjét az utolsóval, a másodikat az utolsó előttivel, és így tovább, cserélték fel. Ismertebb lett azonban Caesar római császár eljárása. Ő egy írott szövegben minden betűt az alfabetikus sorrendbe írt 25 betűs latin ábécében a rá következő negyedik betűvel helyettesített, miközben a szóközt mindig megtartotta. Ma **Caesar eljárás** alatt értjük az output ábécé rögzített sorrendje mellett a betűk helyettesítését az alfabetikusan rájuk következő **c**-edik betűvel. Ekkor tehát **c** megválasztása egy kulcs rögzítését jelenti. A titkosítás, és a

titkosított szövegből az eredeti visszaállítása egyaránt könnyű feladat a **c** kulcs ismeretében.

Ezen a ponton célszerűnek látszik néhány alapvető elnevezést bevezetni, és fogalmat meghatározni.

Nyílt szöveg: az az írott szöveg, amelyet titkosítani akarnak.

Rejtjelezés: az a folyamat, amelynek a segítségével az írott szöveget illetéktelenk számára értelmezhetetlenné alakítják.

Rejtjeles szöveg: a nyílt szövegből a rejtjelzés eredményeként előálló új szöveg.

A **kriptográfia** általános értelemben mindazoknak az eljárásoknak, algoritmusoknak, biztonsági rendszabályoknak kutatását, alkalmazását jelenti, amelyek írásbeli információknak illetéktelenek előli elrejtését hivatottak megvalósítani. Szűkebb értelemben szokás a kriptográfiát a rejtjelző algoritmusok világára korlátozni. Ez a szűkebb terület lényegében véve matematika, amely a matematika számos részterületét öleli fel.

A kriptográfiai algoritmusok a rejtjelzés konkrét transzformációi. Ezek egy véges X input ábécé véges $\underline{x} \in X^{\infty}$ sorozataihoz rendelnek hozzá egy véges Y output ábécé betűiből álló $\underline{y} \in Y^{\infty}$ sorozatot. A hozzárendelést az információelméletben szokásos módon értelmezett **kódok egy paraméteres családja** valósítja meg. Minden konkrét rejtjelzés során ezen paraméterek egyikét kell kiválasztani, és a neki megfelelő kódot kell alkalmazni. A kiválasztott paramétert **kulcs**nak nevezik, a lehetséges paraméterek halmazát **kulcstér**nek.

Kulcsgondozás az a végrehajtandó terv, amely szerint az alkalmazásra kerülő kulcsot kiválaszják, és hírközlés esetén a partnerhez abszolút megbizhatóan eljuttatják. Tartalmazza továbbá a kulcsok biztonságos őrzésének módjait is.

Kriptoanalizis: mindaz a tevékenység, amely a rejtjeles szöveg alapján a nyílt szövegre visszaállítására irányul. Szokásos magyar elnevezés még a **rejtjelfejtés** is.

Kriptológia egyesíti a kriptográfia és a kriptoanalizis területeit. Rajtuk kívül magában foglalja az illetéktelen megfejtés fizikai segédeszközeit és hírszerzési vonatkozásait is.

Az általános értelemben használt Caesar eljárás annyi különböző titkosítási lehetőséget ad meg, ahány betű az ábécében van (amelyek közül egy "triviális" is van, nevezetesen a **c=0** mértékű eltolás). Ha egy illetéktelen ismeri, hogy a titkosítás egy Caesar típusú eljárással történt, akkor teljes kipróbálással meg is fejtheti a titkosított szöveget. Ehhez pusztán fel kell írnia valamennyi lehetséges **c** értéknek megfelelő "eltolt" szöveget, és kiválasztani **az** értelmeset. Itt a határozott névelő azt sugallja, hogy egy értelemes szöveg adódik, és ez így is van, ha a szöveg elegendően hosszú.

Az "*elegendően hosszú*" kifejezés értelmezése lehetővé teszi, hogy az algoritmikus rejtjelzés egyik alapvető fogalmát, az **egyértelműségi pontot** megvilágítsuk. Egy rejtjelző algoritmus egyértelműségi pontja az a minimális távolság, amely lehető teszi a nyílt szöveg megfejtését, vagy magának a kulcsnak az azonosítását. Ha ilyen nincs, akkor a távolságot végtelennek tekintjük. Maga a távolság függ anyílt szövegtől, tehát ez egy valószínűségi változó. Szokás ennek a várható értékét tekinteni egyértelműségi pontnak, bár maga a fogalom nehezen precizírozható.

Európai nyelvek esetén a Caesar eljárás egyértelműségi pontja 4-5 betű. Állításunk szabatos bizonyítása helyett javasoljuk, hogy az olvasó erről "empirikusan" győződjön meg. Válasszon egy könyvből (napilapból) egy 4-5 betűs blokkot, és próbálja azt előlről-hátulról hosszabb értelmes szöveggé kiegészíteni, például teljes kipróbálással. Tapasztalataink szerint az esetek döntő többségében ez csak egyféleképpen tehető meg.

1.9.2. Klasszikus rejtjelző eljárások.

A Caesar eljárás egy kevésbé megfejthető változata az **egyszerű helyettesítés**. Ennek alkalmazása esetén a nyílt szövegben az input ábécé valamennyi betűjét helyettesítjük az output ábécé valamelyik betűjével, ugyanazt a betűt mindig ugyanazzal, különböző input betűket különböző output betűkkel. Mivel a fantázia rendkívül sokféle output ábécét tud elővarázsolni, így a kulcstér is végtelen nagynak tűnhet. Valójában azonban ezek nagy része egymással ekvivalens. Nevezetesen igaz a következő

Tétel: Tetszőleges output ábécé melletti tetszőleges helyettesítés ekvivalens az output ábécé = input ábécé melletti valamelyik helyettesítéssel abban az értelemben, hogy az egyik akkor és csak akkor fejthető meg, ha a másik is megfejthető.

Ez a tétel segít az egyszerű helyettesítést alkalmazó rejtjelzés **biztonsági vizsgálat**ában. Így meg lehet számolni, hogy egy adott ábécé méret mellett hány különböző kód létezik: annyi, ahány féle hozzárendelés van. Ezek a hozzárendelések az ábécé permutációi, tehát ha az ábécé mérete **k**, akkor **k!** különböző kód van. Ez már a k=25 betűs latin ábécé esetén is túlságosan nagy ahhoz, hogy teljes kipróbálással meg lehessen találni a konkrétan alkalmazott kódot. Ez volt az alapja nagyon sokáig annak a tévhitnek, hogy az egyszerű helyettesítés a Caesar eljárással szemben megfejthetetlen. A tapasztalat azonban azt mutatja, hogy 9-10 éves iskolások is meg tudnak fejteni néhány száz betűből álló írott szöveget. Fejtéshez az *adott nyelv statisztikai tulajdonságai*t lehet kihasználni. Minden nyelvben vannak gyakoribb és ritkább betűk. Ennek megfelelően a rejtjeles szövegben a gyakoribb betűket helyettesítő betűk is gyakoribbak lesznek, a ritkább betűk képei pedig ritkábban fordulnak elő. Hasonló igaz a betűpárokra, betűhármasokra is. A gyakorlott fejtők egészen rövid rejtjeles szövegeket is meg tud fejteni. Empirikus tény, hogy angol nyelv esetén az egyszerű helyettesítés egyértelműségi pontja 20-25 betű. (Ennyiből persze maga a transzformáció nem állítható vissza, pusztán a nyílt szöveg.)

A megfejthetőség tényét már az ókorban felismerték. Ezért újabb csodaszerként a "betűcsere" helyett a "helycserés" eljárást kezdték alkalmazni. Ennek során a nyílt szöveget egymáshoz csatlakozó adott hosszúságú blokkokba osztották, majd a blokkokon belül az egyes pozíciokban álló betűket cserélték ki egymással, mégpedig valamennyi blokkra ugyanazt a cserét alkalmazva. Az eljárás neve **pozíciócserés eljárás**, bár magyarul *keverés*nek is nevezik.

Ebben az esetben az egyszerű betűstatisztika nem nyújt segítséget az illetéktelen fejtőnek, hiszen az a keverés után nem változik meg. Eggyel magasabb rendű, betűpárokra vonatkozó statisztika azonban most is alkalmazható. Ez az alkalmazás a *statisztikai hipotézis vizsgálatok* területére vezet el bennünket. A kis lépések stratégiáját követi. Ahelyett, hogy a keverést megvalósító permutációt a maga teljességében meg akarnánk határozni, csak arra a kérdésre keressük aa

választ, hogy a rejtjeles szöveg adott pozicíójában elhelyezkedő betüket a nyílt szövegben melyik pozíció betűi követik.

A középkor végére az is nyilvánvalóvá vált, hogy az egyszerű helyettesítés megfejtése amiatt volt lehetséges, hogy mindössze egyetlen helyettesítő ábécét használtak. Így többé-kevésbé szükségszerű lett **többábécés eljárások** alkalmazása. Egy ilyen eljárás kidolgozása Blaise de Vigenère (1523-1596) tevékenységéhez füződik. Kimerítő összefoglalást adott kora rejtjeltudományának állásáról. Számos többábécés eljárást dolgozott ki, amelyek az ábécé betűi között értelmezett összeadást hasznosítottak, így munkássága a véges csoportok előfutárjának is tekinthető. Életében módszerei nem kerültek felhasznllásra, jelentőségüket csak mintegy 200 évvel ismerték fel.

A róla elnevezett Vigenère-eljárás nemcsak történelmi jelentőségű: ez vezetett el az elméletileg bizonyítottan fejthetetlen rejjelzési eljárás feltalálásához. Az eljárásban használt összeadás rendkívül egyszerű. Ehhez rendezzük (alfabetikus) sorrendbe az ábécé betűit, és számozzuk meg őket 0-val kezdve a számozást. Amikor két betűt össze akarunk adni, akkor adjuk össze a sorszámukat. Ha a sorszámok összege kisebb az ábécé betűinek számánál, akkor az ennyiedik betű lesz a kívánt összeg. Ha a sorszámok összege nagyobb, akkor vonjuk le az összegből az ábécé betűinek számát. A maradék megint kisebb lesz a betűk számánál, így megint lesz ilyen sorszámú betű az ábécében. Most a betűk összege az ilyen sorszámú betű lesz. Ez a művelet nyilvánvalóan

bármelyik két betú esetén elvégezhető. Az is igaz, hogy bármely két betű esetén az

Egy adott betű + ismeretlen betű = egy másik adott betű

típusú egyenlet mindig egyértelműen megoldható. Ezek előrebocsátása után Vigenère eljárása:

- Választunk egy kulcsszót
- Leírjuk a nyílt szöveget
- A kulcsszót állandóan ismételve, betűről-betűre a nyílt szöveg fölé írjuk
- Az egymás alatt álló betűket összeadjuk. Az összegsorozat adja a rejtjeles szöveget.

Ez a rejtjelzés nyilván a kivánás segítségével mindig megoldható az

Ismeretlen nyílt betű = rejtjeles betű - kulcsbetű

összefüggés alapján.

Ez az eljárás is magában tartalmazza az illetéktelen fejtés lehetőségét. Mindenekelőtt *egyszerű statisztikai teszttel meghatározható a kulcsszó hossza*. Ha a kulcsszó hossza ismert, akkor az azonos kulcsbetűnek megfelelő rejtjeles betűk egy-egy oszlopba gyűjthetők, és ők már mindössze egy Caesar típusú egyszerű helyettestéssel lettek rejtjelezve. Ennek a kulcsa éppen a megfelelő kulcsbetű. (csak a megfejtés egy lehetséges ötletét írtuk le!)

A Vigenère eljárás megfejthetőségét az teszi lehetővé, hogy a kulcsszó véges, ezért a rejtjeles szöveg azonos kódábécével rejtjelezett részekre bontható szét. Ez nem lehetséges, ha a kulcsszó legalább ugyanakkora, mint maga a nyílt szöveg. Ha még azt is lehetne biztosítani, hogy a rejtjeles szöveg és a nyílt szöveg statisztikailag független, tehát a rejtjeles szöveg teljes ismeretében sem lehet semmivel sem többet mondani a nyílt szövegről, mint ennek ismerete nélkül, akkor az illetéktelen fejtés is reménytelen lenne. Ezt a gondolatot hasznosította Vernam amikor a **One time pad** néven ismertté vált szabadalmát bejelentette. A Vigenère eljárásban azt a változtatást hajtotta végre, hogy a kulcsszó méretét nagyon nagyra választotta, magát a kulcsszót pedig betűről betűre haladva véletlenszerűen, egyenletes eloszlással sorsolta ki. Ezt a kulcssorozatot egy-egy tépő-blokk-füzetbe rögzítették kettő példányban. Egyik példány a rejtjelzőé lett, amásik a megoldóé. A két füzet kiosztását, őrzését a legszigorúbb biztonsági rendszabályok mellett kellett végrehajtani. Minden lapot csak egyszer lehetett felhasználni, erre utal a névben szereplő *one time*. A második világháborúban a nagyhatalmak fontosabb üzenetek továbbítására ennek az eljárásnak valamelyik változatát használták.

A véletlen választással kapcsolatban itt fogalmazzuk meg a kriptogáfában már a múlt században általánosan elfogadott elvet.

Egy *kriptográfiai eljárás bevizsgálása során alapvető feltétel*, hogy az ellenfél (potencális illetéktelen megfejtő) a paraméteres kód-családot teljesen ismeri (például megvette egy alkalmazottól). Ilyenkor a titkosságot csak a kulcs megválasztása jelenti, így azt úgy kell megválasztani, hogy illetéktelenek számára maximális bizonytalanságot tartalmazzon. Ezt úgy lehet biztosítani, hogy minden új nyílt szöveghez egy új kulcsot választunk, mégpedig minden előzőtől függetlenül, egyenlő

valószínűséggel. Ennek módszereivel a matematika egy nagy részterülete, a *véletlenszám generálás* témaköre foglalkozik.

Az újkor hajnalán az egyszerű helyettesítés egy más típusú általánosítása is megtörtént. Ennek során nem betűket, hanem betűcsopostokat helyettesítettek, mégpedig általában nem ugyanakkora, hanem rövidebb sorozatokkal. Ez az általánosítás Antoine Rossignol felfedezése. Összegyütött a királyi udvar rejtjeles forgalmát jellemző tipikus szófordulatokat, neveket, és ezeket **egy kódkönyv**be rendezte. A kódkönyv valamennyi tételét egy rövid betűcsoporttal azonosította. Az 1600-as évek második felében így konstruált Nagy Chiffrè nevű kódkönyve hosszú ideig megfejthetetlen maradt, csak 1890-ben sikerült megfejteni. Magyar vonatkozása a kódkönyvek korai történetének, hogy a Napkirály Habsburgok elleni külpolitikai tevékenysége során II. Rákóczi Ferenchez is eljuttatott kódkönyveket, ilyenekkel folytatták titkos levelezésüket.

A kódkönyvek jelentették az első lépéseket a rejtjelzés nyilvánossága felé. Röviddel Morse találmánya után egy Smith nevű ügyvéd a nyilvános távirati kereskedelmi forgalom számára szerkesztett egy titkos üzleti levelezést szolgáló "beststeller" kódkönyvet, amely egyúttal a hatékony adattömörítés és a rejtjelzés kapcsolatára is rámutatott.

Az ipari forradalmat követően a titkos üzenetek száma és terjedelme annyira megnövekedett, hogy a rejtjelzés gépesítése mindennapi szükségletté vált. Már a múlt század második felében megjelentek azok az eszközök, amelyek mai **rejtjelező gépek** elődeinek tekinthetők. Valamennyien a többábécés eljárások technikai kivitelezését segítették. Az első rejtjező gépek a Vigenère eljárás valamely változatát igyekeztek gépesíteni. Ilyen volt az 1867-ben konstruált Wheatstone korong, majd az 1891-ből származó Baseries féle rejtjelző korong.

A rejtjelzési eljárások két utóbbi köre uralkodó szerepet töltött be a II. világháborúig, illetve annak elején. A két legfeljettebb eljárás egyike kódkönyveket alkalmazott a másik pedig tárcsás rejtjelző gépeket, ahol az egymás utáni rejtjelző kódokat a tárcsák léptetése automatikusan generálta. Meglepő módon csak lassan ismerték fel, hogy ezek nem biztosítanak hatékony titokvédelmet.

A megfejthetetlenségbe vetett tévhit tipikus példája a németek által használt *ENIGMA gép*. Ez a gép valóban nagyon bonyolult matematikai algoritmust használt.

A gép működését két lengyel matematikusnak sikerült rekonstruálni, majd a rejtjelfejtési munka Lengyelország lerohanása után Angliában folytatódott. 1942 végére az angolok már "on-line" tudták olvasni a német Enigma üzeneteket.

Az Enigma mellett természetesen más mechanikus és elektromos rejtjelzőgépek is forgalomba kerültek. Közülük a legelterjedtek (bár megváltozott formában) ma is forgalomban levő *Hagelin gépek*. Ezek megfejtési algoritmusai is publikusak.

A kriptográfia történetéről részletes leírást ad Kahn [1.9.4.]

1.9.3. A Shannon elmélet

A II. világháború alatt már a rádiós hírközlés vált általánossá. Ennek lehallgatása rutin feladat lett. Így a rejtjelfejtők hatalmas mennyiségű olyan rejtjeles szöveg birtokába jutottak, amelyek ugyanazon algoritmussal, bár többnyire táviratonként eltérő kulccsal készültek. Ez volt az egyik oka, hogy a világháború elején a rejtjelzők – rejtjelfejtők párharcában az utóbbiak voltak az eredményesebbek. A rejtjelzőgépek és más, sokáig megfejthetetlennek tartott eljárások sorozatos megfejtését kísérő csalódások kikövetelték egy olyan elmélet létrejöttét, amelyik *tudományos szigorúsággal* képes *megvizsgálni egy adott kriptorendszer megbizhatóságát*. Ehhez először azt a környezetet kellett modelezni, amelyben a kriptográfia működött, tehát magát a hírközlést. Ezt a szükségszerűség szülte feladatot **Claude E. Shannon** oldotta meg még a háború alatt, majd 1948-49-ben publikálta [1.9.8.], [1.9.9.]. Shannon kifejlesztette a **hírközlés** máig használt **matematikai elméletét**, és ehhez kapcsolódva **a titkos rendszerek hírközlési elméletét**. Az elméleten belül lehetőség nyílt annak szabatos bizonyítására, hogy a Vernam által szabadalmaztatott egyszeri véletlen átkulcsolás elméletileg is fejthetetlen rejtjelzési eljárás, ha pusztán a rejtjles szöveg alapján kell a fejtést elvégezni.

C.E. Shannon II. világháborús rejtjelfejtői tevékenységének tapasztalatait felhasználva alkotott hírközlési és kriptográfiai matematikai modelljét az **információelmélet** megalapozásának tekintik.

A modell blokkdiagrammját az 1.9.1. ábra mutatja. Ezen az ábrán az illetéktelen fejtő bekapcsolódásának a lehetősége is fel van tüntetve, utalva az

illetéktelen információszerzés lehetőségére is. Már ebben a modellben felhívjuk a figyelmet arra a tényre, hogy a legbiztonságosabb algoritmus esetén is szükséges egy **jól megtervezett üzemeltetési szabályzat** maradéktalan betartása, mert e nélkül az elméletileg megfejthetetlen rejtjelzési algoritmus is könnyen fejthetővé válhat.

Erre vonatkozóan ismertetünk egy negatív példát az elméletileg fejthetetlen a véletlen egyszeri átkulcsoláshoz kapcsolódóan. Egy rendkívül durva hiba ugyanannak az átkulcsoló sornak ismételt felhasználása. A *kulcsismétlés* egyszerű (kézzel is végrehajtható) statisztikai teszttel felismerhető. Felismerés után a két rejtjeles szöveget egymásból kivonva megszabadulunk a biztonságot jelentő kulcstól, amit a különbségszöveg már nem tartalmaz. Ez a különbségszöveg a két nyílt szöveg különbsége lesz. Ebből a különbségből a két nyílt szöveg már a múlt századból ismert módon, *Kerchkoff* által kdolgozott módszerrel visszaállítható (esetleg minimális hibával). A kulcsismétlés hibáját számos ország elkövette, köztük a Szovjetunió. Ennek oka első sorban a "tiszta" használat esetén felépő jelentős pénzügyi kiadás volt. Valódi véletlenszám generátort kellett alkalmazni. A generált sorozatokat biztonságosan kellett tárolni, és abszolút biztonságos csatornán eljuttatni a felhasználókhoz, mégpedig a háborús körülmények között. Ennek a kemény feladatnak az egyszerűsítését jelentette, hogy ugyanazt a véletlenszám sorozatot bizonyos eltolással ismételten felhasználták. A *Venona-papers* címmel idézett, a 90-es évek végén nyilvánosságra került tanulmányok szerint az USA hírszerzése külön csoportot szervezett a kulcsismétlést használó távíratok összegyüjtésére és megfejtésére. Ebben az esetben a modell ábrájában a "Segédinformáció" a hírszerzés által beszerzett információ volt, amely felderítette az ismételt kulcsfelhasználás gyakorlatát.

A segédinformációk nemcsak megkönnyíthetik a rejtjelfejtést, de esetleg feleslegessé is tehetik azt. A számítógép képernyőjét elektronikusan leolvasva például közvetlenül is meg lehet ismerni a nyílt szöveget. A rejtjelzési környezet ilyen jellemzőinek feltérképezése elengedhetetlen egy "jó" biztonsági rendszabályzat elkészítéséhez. A kriptográfiai módszerek alkalmazása az algoritmikus meggondolásoknál sokkal szélesebb skálán mozog.

1.9.4. Nyilvános rejtjelzési eljárások: DES, AES, IDEA

A rejtjelzés klasszikus, hagyományos alkalmazási területe a katonaság, hírszerzés és diplomácia. Ezekre a központi irányítás jellemző. Általában létrehoznak egy rejtjelző központot. Ennek a feladata a rejtjelző algoritmus kiválasztása, a biztonsági rendszabályzat kialakítása és betartatása, a kulcsgenerálás folyamatos megoldása és kulcskiosztás megszervezése. Az alkalmazás résztvevői jól meghatározott hírközlő végpontok, amelyek egy zárt hálózatot képeznek.

Zárt hálózat: olyan hírközlési hálózat, amelyben minden felhasználó előre ismert, nyilvántartott. Senki sem léphet be új felhasználóként önkényesen a rendszerbe. A rendszeren belüli forgalmazás csak a hálózat által elfogadott azonosítás megtörténte után lehetséges.

A hírközlés robbanásszerű fejlődése azt az igényt is megteremtette, hogy akár ismeretlen emberek is biztonságos, mások számára érthetetlen levelezést folytathassanak egymással. Ennek megfelelően a korábbi *zárt hálózatok* helyett egyre nagyobb jelentőséggel bírnak a *nyílt hálózatok.*

Nyílt hálózat: olyan hírközlési hálózat, amelyhez nyilvánosan közzétett szabályok szerint bárki csatlakozhat.

Első lépésként egymást ismerő üzletemberek csoportjai kívántak egymással rejtjeles kapcsolatot létesíteni. Ez nem jelentett a klasszikus alkalmazásoktól lényegesen eltérő feladatot. Közösen megválaszthatták az alkalmazni kívánt rejtjelző algoritmust, megválaszthatták egy rövidebb időszakra érvényes rejtjelkulcsaikat és azokat személyesen kicserélhették egymással. Ők is egy zárt hálózatot alkottak azzal a különbséggel hogy a hálózat bővítése adminisztratilag egyszerűbben, közös akarattal történt. Egy ilyen nyított hálózat is lehet központi szervezésű, például egy cég és leanyvállalatai (vagy ügyfelei) között. A rejtjelzési algoritmusok ebben az esetben nem igényeltek semmi újabb elemet, a hagyományos módszereket továbbra is lehetett alkalmazni. Elérhető szakismeretük hiányosságai, anyagi lehetőségeik korlátai azonban megkövetelték a szakmai segítséget az algoritmusok megválasztásában és a biztonságos működés rendszabályzatának a létrehozásában. Ennek megvalósítása azonban már államérdek volt, ahogyan azt először az USA-ban felismerték. Rejtjelzési szabványra volt szükség, amely

garantálni tudta felhasználóinak a megbízható üzenetváltást. Az első ilyen szabvány a DES (Data Encrytion Standard).

Ezt 1976 decemberében a *National Bureau of Standards*, USA, jelentette be, mint egy új "Nemzeti Adatfeldolgozási Szabványt" (FIPS No. 46. Leírása megtalálható az USA szabványok gyüjteményében, lásd National Bureau of Standards [1.9.6.]

A DES 64 bites (8 byte-os) nyílt üzenet blokkokat képez le ugyancsak 64-bites rejtjeles blokkokba 56 bit nagyságú kulcsméret mellett. Az IBM által kifejlesztett block-algoritmus Shannon egy keverő transzformációját használva biztosítja, hogy a blokkon belül a kimenet minden bitje függ a bemenet minden bitjétől. A szabvány tartalmazza azt a kikötést is, hogy csak hardware implementált változata használható az USA-n belül, és az USA kormányzat megtiltotta, hogy ezt a hardware kivitelezést exportálják. Magát az algoritmust 5 évenként biztonsági vizsgálatnak vetették alá. Ez utoljára 1994-ben történt meg, amikor 1998-at jelölték meg a felhasználhatóság utolsó határának. Ennek ellenére kissé módosított változatban ma is általánosan használják.

A DES megfejthetőségére utaló publikációk sorát Martin Hellman egy 1977-ben tartott előadása nyította meg, aki a teljes kipróbálást is kivitelezhetőnek tartotta megfelelően épített hardware segítségével. A DES halálához a döntő csapást Biham és Shamir munkássága adta meg, akik 1993-ban egy újonnan kifejlesztett módszer, a "Differential Cryptanalysis" segítségével adtak meg egy fejtési eljárást. 1994-ben Matsui egy más típusú, u.n. lineáris kriptanalízis módszert dolgozott ki, amely már gyakorlatilag kivitelezett fejtésről számolt be.

Javításként először a fejtést gyakorlatilag kivitelezhetetlennek beállító eljárásként a DES kétszeres alkalmazását javasolták. Ehhez az üzenetet kétszer kellett egymás után rejtjelezni két, egymástól függetlenül választott kulccsal, ami a kulcs méretét immár 108 bit hosszúságúvá tette. Ezzel kapcsolatban már 1992-ben Merkle és Hellman nyilvánvalóvá tette azonban, hogy ha a "simple DES" fejthető, akkor egy "Középen találkozunk" elnevezésű módszer lehetővé teszi a "Double DES" fejtését is.

A DES jelenlegi legfejlettebb változata a "Triple Des, 3-DES". Ez vagy kettő, vagy három 58 bites kulccsal dolgozik. Az üzenetet először az első kulccsal

rejtjelezik normál DES módban, majd a második kulccsal a megoldó algoritmust alkalmazzák rejtjelezőeljárásként. Az így nyert közbülső szövegre alkalmazzák ismét az első, három kulcsos rendszerben a harmadik kulcsot. Ismertnek tekintik, hogy több helyen építettek olyan célgépet, amellyel a 3xDES-t fejteni lehet..

Az export-tilalom miatt számos konkrét chip-megvalósítás található a piacon, és a pénzügyi szféra több nemzetközi szabványában is található 3-DES elem. A Triple-DES chipek vásárlóinak ajánlatos erősen kritikus szemmel megvizsgálni az alkalmazni kívánt megoldást. Az Internetről letölthető szoftver alkalmazása a leghatározottabban kerülendő.

Mivel az algoritmus publikus, így jogtisztán programmozható. Programmozás során számos matematikai trükk tudja az időfaktort csökkenteni.

2001 őszén a NIST új amerikai rejtjelzőszabványt hírdetett ki, az **AES: Advanced Encryption Standard-t.** A **DES-**sel kapcsolatosan felmerült hiányosságok együtt kikényszerítették a National Institute of Standards and Technology (NIST) azon döntését, hogy ki kell fejleszteni a DES utódját, amely az Advanced Encryption Standard (AES) nevet kapta. A pályázatot 1997 szeptemberében írták ki, közzétéve azon elvárásoknak a listáját, amelynek az AES algoritmusnak meg kell felelni. Ugyanakkor deklarálták, hogy a benyújtott rejtjelzési algoritmusok nyilvánosak, szabadon felhasználhatók lesznek. A kiírás szerinti elvárások:

- Legyen blokkos algoritmus, 128 bites blokkmérettel
- A 128, 196 és 256 bites kulcsméret opcionálisan egyaránt megválasztható
- Legyen az algoritmus nyilvános, jogdíj nélkül használható
- Álljon ellen valamennyi ismert rejtjelfejtési módszernek
- Legyen világos, logikus szerkezetű, áttekinthető
- Mind a kódolás, mind a dekódolás gyors legyen
- Kevés memóriát foglaljon el
- Többféle processzoron is hatékonyan implementálható legyen

Ezeknek az elvárásoknak megfelelő rejtjelzési algoritmus várhatóan hosszú távra, akár 20-25 évre is megoldja a polgári életben keletkező adatok biztonságos védelmének a kérdését.

A beérkezett pályaművek közül 15 felelt meg a formai elvárásoknak. A NIST *1999. március 22-23 között Rómában* megrendezett második szakértői konferencia eredményeként már csak öt algoritmus maradt versenyben:

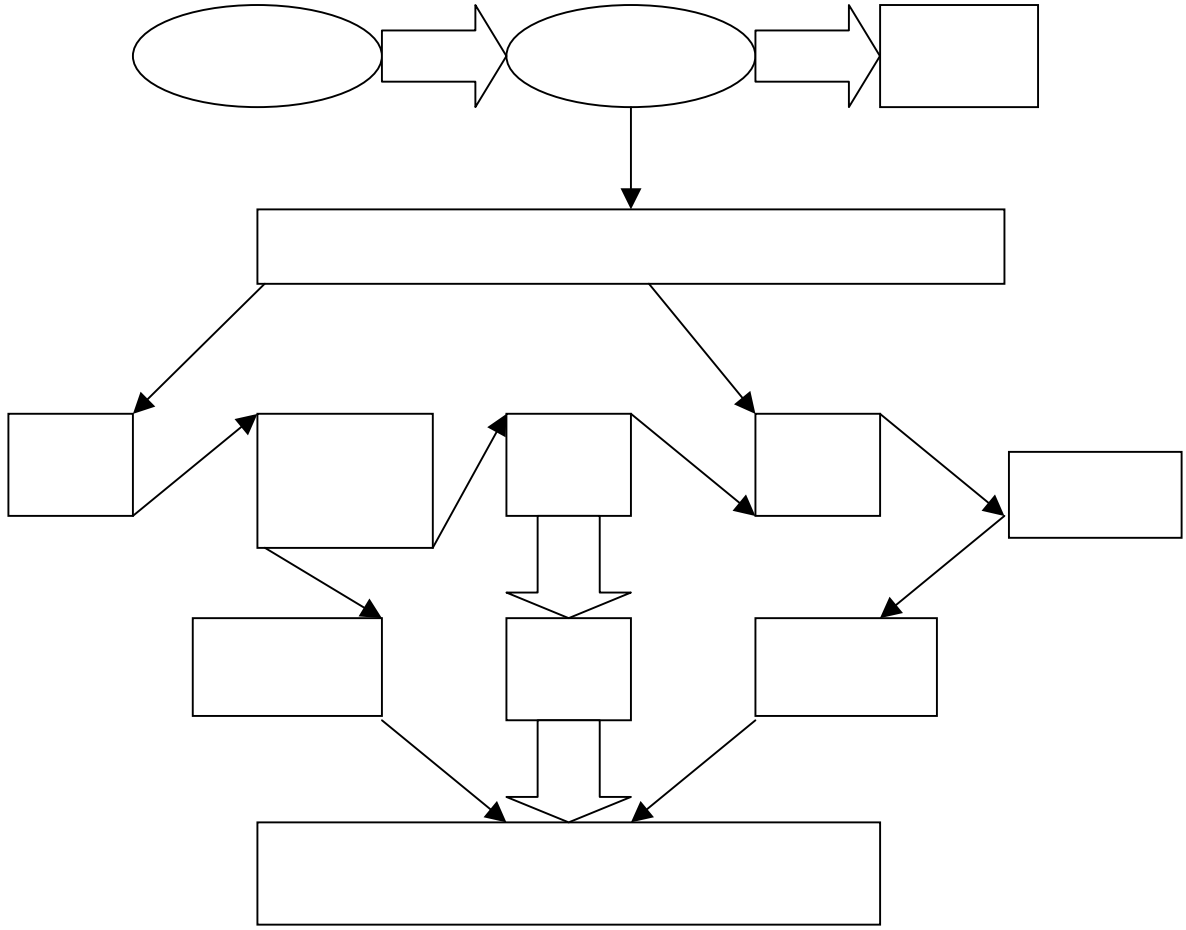
A győztes algoritmust 2000. október másodikán jelentették be. A versenyt **a RIJNDAEL algoritmus** nyerte meg, melynek szerzői**, Rijmen, Daemen belga kriptográfusok** A Rijndael algooooritmus teljes mértékben megfelel a fent leírt feltételeknek. Az algoritmus több érdekes technikát alkalmaz (pl. byte-szintű operációk a 256 elemű véges test fölött.). Szerkezete nem követi a DES struktúráját, bár ez is számos iterációs lépésben valósul meg. Minden iteráció három rétegből áll (lineáris keverés réteg, nemlineáris réteg, kulcs-addíciós réteg), amelyek szerepe különböző. (lásd: [1.9.1.])

A Rijndael algoritmus egyszerű, világos, bármely programozási nyelven gyorsan programozható. Az Internetről számos változat szívható le, bár ezt a megoldást senkinek sem ajánljuk.

Az AES pályázat kiírása elótt is számos megoldást javasoltak a DES kiváltására, helyettesítésére. Ezek közül a legelterjedtebb az **IDEA Cipher**. Ezt Lai és Massey javasolták 1990-ben, mint "Proposed Encryption Standard"-t. Lai 1992-ben fejlesztette tovább az IDEA (= Ideal Data Encryption Algorithm) néven ismertté vált eljárássá.

Az IDEA gondosan választott alapvető, de kielégítő bonyolultságú matematikai műveletek speciális kombinációit használja fel. Ezeket a műveleteket 16 bites blokkonként alkalmazza 64 bites nyílt szöveg blokkokra, 128 bites kulcs felhasználása mellett. A blokkon belüli kimenet bitek mindegyike minden bemeneti bittől függ. Bizonyítottan rendelkezik a Shannon által megkövetelt keverési és szétterjesztési tulajdonságokkal. A matematikai műveletek egyszerűsége gyors és egyszerű technikai megoldásokat tesz lehetővé mind software, mind pedig hardware szinten. Ezeknek a megoldásoknak némelyike eredményes fejtést enged meg, így ezekre megvásárlásuk során különös figyelmet kell fordítani.

Az IDEA eljárás számos országban szabadalom védelem alatt áll.Az Európára érvényes szabadalom bejegyzésének ideje 1993. június 30, száma EP 0 482 154 B1. Termékbe beépítve használatához az ASCOM engedélye szükséges. A



http://www.szcom.ch/Web/systec/security/license/htm címen minderről bővebb felvilágosítás olvasható.

A klasszikus rejtjelző eljárások egyetlen kulcsot használnak rejtjelzésre és megoldásra (miközben a megoldó algoritmus nem feltétlenül egy fordított sorrendben végrehajtott rejtjelzés). Ezért az utóbbbi időben rájuk **szimmetrikus kulcsú eljárások** néven szokás hivatkozni.

A klasszikus kriptográfia módszereit alkalmazták az analóg jeleket használó első **beszédtitkosítók** is. Ezek az időtengely egységintervallumát sok kis egyenlő hosszúságú részintervallumra osztva a rész-intervallumokat permutálták. Ugyancsak permutálták az amplitudótengely hasonlóan képzett kis intervallumait is. Napjaink gyakorlatában az analóg jeleket először *digitalizálják*, majd az így adódó digitális sorozatokat nyílt szövegnek tekintve rejtjelzik.

1.9.5. Nyilvános kulcsú kriptográfia.

Napjainkban a digitalizált hírközlés forradalma zajlik. Ahogyan a korábbi nagy társadalmi változások mindig kialakították a maguk sajátos kriptográfiai eljárásait, ugyanúgy az elektronikus társadalom is kialakítja a maga saját formáit. A digitális hírközlés egyik elvárása, hogy ismeretlen emberek abszolút biztonságosan tudnak egymással titkos üzeneteket váltani. Napjainkra ennek az elvárásnak a technikai biztosítékai is rendelkezésre állnak.

A nyílt hálózatok digitális világában a biztonság két különböző típusú követelmény kielégítését jelenti. A személyiséggel kapcsolatos jogok, valamint az adatok biztonsága érdekében másféle intézkedésekre van szükség. Két szokásos meghatározás:

Adatbiztonság: digitális adatok sérthetetlenségét, illetéktelen személyek által történő megismerését megakadályozó módszerek összesége.

Adatvédelem: azoknak a módszereknek az összesége, amelyek megakadályozzák az adatok alapján a személyiségi jogok megsértését, például hozzáférésvédelem biztosításával. Ennek eszköze a nyilvános kulcsú kriptográfia.

A **nyilvános kulcsú kriptográfia** olyan kriptográfiai rendszer, amelynek a résztvevői közös algoritmust használnak rejtjelezésre. Nyílt hálózatokban használják, amelybe bárki beléphet résztvevőként. Az algoritmusnak két - a használóktól függő egyedi - kulcsa van. Ezek egyikét (a nyilvános kulcsot) a nyílt hálózat résztvevői nevükkel együtt nyilvánosságra hozzák, a másikat titokban tartják. (Ezt titkos kulcsnak nevezik.) A kulcsok egyikét a rejtjelzésre, a másikat a (jogosult) megoldásra használják.

Olyan közös rejtjelzési algoritmust használnak, amelyben a rejtjelzést a nyilvános kulcs birtokában könnyű elvégezni, de pusztán ezzel a nyilvános kulccsal a dekódolás gyakorlatilag nem kivitelezhető. A titkos kulcs segítségével azonban a dekódolás is gyors művelet. Ezt a filozófiát megvalósító rendszerek gyűjtőneve: *Nyilvános kulcsú rendszerek* (public key cryptosystems).

Maga az elképzelés a *Hellman-Diffie* szerzőpárostól származik. Az első és máig az egyetlen megbízhatónak látszó, technikailag kivitelezhető eljárást Rivest, Shamir, Adleman adta meg 1978-ban, lásd [1.9.7.]. Matematikai alapja a

számelmélet Fermat tétele, és az a tény, hogy nagy számok osztóinak meghatározása rendkívül bonyolult feladat

Az **RSA algoritmus** a modulo aritmetikában az ismeretlent hatványban tartalmazó egyenletek megoldásának nagyfokú bonyolultságát használja ki, így megfelelő nagyságú modulus esetén a megoldás technikai kivitelezhetetlensége szolgáltatja a biztonságot. A "megfelelő nagyság" itt kritikus szerepet játszik: a kezdetben biztonságosnak ítélt 40 bit hosszú kulcsok helyett ma nem nevezhető biztonságosnak 1024-nél rövidebb kulcs.

A nyilvánosságra hozott kulcs egy (E,M) egészekből álló számpár. A titkosítás ezek segítségével történik. Először a nyíltszöveg adott hosszúságú blokkjait az M modulusnál kisebb egész számmá alakítják, majd ezt a számot M modulusban felemelik az E-edik hatványra. Ez a szám, illetve ennek az átviteli csatornára befogadható sorozattá kódolt változata lesz a titkosított üzenet. A titkos kulcs a nyilvános kulcshoz hasonlóan egy (D,M) számpár, ahol M azonos az előzővel, míg a D dekódoló exponens úgy van megválasztva, hogy a titkosított üzenetnek megfelelő modulo-M számot ennyiedik hatványra emelve az eredeti üzenet adódik. Megbízható algoritmushoz M-et két nagyon-nagy prímszám szorzatának, E-t véletlenszerűen választják.

Mivel a modulus aritmetikában a hatványozás idő- és költségigényes, ezért szinte kizárólag hardware megoldások jönnek számításba. A piacon számos ilyen hardware (Chip) megoldás konkurál. Ezek főként a véletlen prímszám, vagy a véletlen exponens megválasztásban térnek el egymástól, de köztük a döntő különbséget a nagyszám aritmetika megválasztása jelenti. Az elsőre a "hibás" választások miatt elvégezhető fejtések lehetősége miatt fontos odafigyelni, míg a második a gyorsaság (időigényesség) miatt jelentős. Az RSA első alkalmazásai között szerepel a digitális aláírás, a hozzáférés védelem, és az üzenethitelesítés.

Digitális aláírással lehet garantálni digitális dokumentumok eredetét, eredetiségét, letagadhatatlanságát,. A digitális aláírástól ugyanazt az aláírtság-funkciót várjuk el, mint amit a kézírásos aláírás a hagyományos dokumentumok esetén biztosít.

A **digitális aláírás** olyan digitális (titkosított) karaktersorozat, melyet igen nagy valószínűséggel csak az aláíró kódolhatott, s ez magából a kódolásból következik és amely

- elektronikus adathordozón tárolható
- elektronikus csatornán továbbítható
- függ az aláíró személyétől,
- függ az aláírt szövegtől
- a küldött üzenetből képezett sűrítményt, "lenyomatot" tartalmazhat.
- tartalmazhat az aláírás megtörténtjének időpontját rögzítő időpecsétet.

A *digitális aláírás* során a hitelesítendő hosszú dokumentumokhoz egy rövid *STRINGet* rendelünnk hozzá úgy, hogy különböző dokumentumokhoz különböző stringek tartozzanak, miközben a stringek előállítása könnyű feladat. Ezeket a sztringeket *digitális lenyomatnak (sűrítmény nevezik*). Őket akarjuk azonosításra használni. Azt a leképezést, amely ezt a feladatot megoldja, **hash függvény**nek nevezik. A hash függvénnyel szembeni elvárásokat a következő követelmények összegezik:

- Gyakorlatilag lehetetlen egy adott outputhoz olyan új dokumentumot konstruálni, amelyiknek ugyanaz a lenyomata.
- Gyakorlatilag lehetetlen két olyan dokumentumot konstruálni, amelyeknek azonos a lenyomata.
- Ha a dokumentumban legalább egy bitet megváltoztatunk, akkor a megfelelő lenyomat sok bitben különbözik.Gyorsan számolható.
- Elegendő hosszúságú bementi dokumentum esetén a lenyomat véletlenszerűen viselkedik. 5-6 32 bites szóból álló bemenet már "elegendően" hosszúnak tekinthető.

A fenti követelményeknek megfelelő eljárások közül a **Secure Hash Algoritmus**, az SHA hash függvény a legáltalánosabban elterjedt. Az algoritmus inputja egy tetszőleges hosszúságú (de maximum 2^{64} bitből álló) tetszőleges dokumentum, az *outputja* pedig egy *160 bit hosszuságú string* (hash érték=message digest).

Az SHA függvényt az amerikai *Federal Information Processing Standard* sorozat 180-as számú dokumentuma szabványosítja. A szabványon 1996-ban apróbb változásokat hajtottak végre, az új változatot SHA-1 jelöli. Az új szabvány a **FIPS PUB 180-1** jelet kapta. Az algoritmus számítás-technikailag sokféle módon valósítható meg, amelyeknek azonban ugyanazon bementi sorozat esetén ugyanazt

a lenyomatot kell eredményezni. A szabvány az alábbi három stringet adja meg a szabvánnyal való megfelelés ellenőrzésére.

Megjegyezzük, hogy elfogadhatónak tartják az **MD4** és annak módosított **MD5 eljárás** használatát is. Ugyancsak használatos a CCITT által javasolt **32 bites CRC** algoritmus (ANSI X3.66 szabvány).

Az MD5 eljárás alkalmazása jogdíjköteles, az SHA-1 és a CRC jogtisztán is használható saját készítésű szoftver esetén. Valamennyire létezik több INTERNETes konkrét megvalósítás is, bár azok helyességét, vírusmentességét semmi sem garantálja. Helyénvaló itt idézni egy tipikus figyelmeztetést az Internetről: "Ilyen szabványok a Szövetségi Kormány Export Ellenőrzési rendeletének a hatálya alá esnek, Szövetségi Törvénykönyv, 15. címszó, 768-799. bekezdések."

Az elektronikus üzenetküldés egyre tömegesebbé válásának egyik következményeként igény merült fel arra, hogy elektronikus formában joghatást kiváltó üzenetet is lehessen küldeni (pl. szerződéses ajánlat, illetve annak elfogadása, teljesítés igazolása). Ehhez azonban szükséges, hogy az így kapcsolatba kerülő felek hitelesen megállapíthassák, kitől származik az üzenet és annak tartalma nem változott-e a feladás óta. Ezt az igényt hivatott a digitális aláírás kielégíteni. A digitális aláíírás jogi elismerhetősége érdekében számos ország **digitális aláírás-törvény**t alkotott, Európában elsőként a Német Szövetségi Köztársaság 1997 őszén. Hazánkban az Országgyűlés – felismerve és követve az egyetemes fejlődésnek az információs társadalom felé mutató irányát, az új évezred egyik legfontosabb kihívásának eleget téve – 2001 júnusában alkotott **törvény**t **az elektronikus aláírásról** annak érdekében, hogy megteremtse a hiteles elektronikus nyilatkozattétel, illetőleg adattovábbítás jogszabályi feltételeit az üzleti életben, a közigazgatásban és az információs társadalom által érintett más életviszonyokban.

A törvény értelmezése szerint

Elektronikus dokumentum: elektronikus eszköz útján értelmezhető bármilyen adat. **Elektronikus aláírás**: elektronikus dokumentumhoz azonosítás céljából végérvényesen hozzárendelt vagy azzal logikailag összekapcsolt elektronikus adat, illetőleg dokumentum.

Elektronikus okirat: olyan elektronikus irat, mely nyilatkozattételt, illetőleg nyilatkozat elfogadását, vagy nyilatkozat kötelezőnek elismerését foglalja magában.

Időbélyegző: elektronikus irathoz, illetve dokumentumhoz végérvényesen hozzárendelt, illetőleg az irattal vagy dokumentummal logikailag összekapcsolt igazolás, amely tartalmazza a bélyegzés időpontját, és amely a dokumentum tartalmához technikailag olyan módon kapcsolódik, hogy minden – az igazolás kiadását követő – módosítás érzékelhető.

Az időpecsétet a sürítmény elkészítése előtt helyezik el szokványosan a digitális dokumentum elejére. Ugyancsak ide helyezhetnek el az *aláírót azonosító adatokat*, például útlevélszámot, születési helyet, dátumot. Ezután következhet a sűrítmény elkészítése, amely tehát ezektől az adatoktól is függ. Az aláírást egy nyilvános kulcsú rejtjelzés teheti teljessé. Ezt megadott eljárás mellett az aláíró titkos kulcsával végzi el.

Az *aláírás ellenőrzése során* az ellenőrző először dekódolja a rejtjelzett sűrítményt a az aláíró nyílvános kulcsával, majd a rendelkezésre álló nyílt dokumentumhoz ő is meghatározza a sűrítményt. Ha ez és a dekódolt aláírás megegyezik, akkor az aláírást hitelesnek tekintik. A valóságban azonban ez az ellenőrzés mindössze azt bizonyítja, hogy az aláió megegyezik azzal, aki a nyilvános kulcsot a központi nyilvántartóba elhelyezte. Szükséges tehát annak az tanúsítása is, hgy az aláíró valóban az, akinek vallja magát. Ezt a *hagyományos közjegyzői szerephez hasonlóan* egy megbízható harmadik fél, a **hitelesítés szolgáltató** végzi. A Hitelesítés Szogáltató működését hazánkban a Hiradástechnikai Felügyelet engedélyezi a törvényben meghatározott szigorú feltételek szerint. Ezek a feltételek magukban foglalják azt a követelményt, hogy a Hitelesítés Szolgáltatónak feladata ellátásához rendkívül szigorú biztonsági feltételeket kielégítő infrastruktúrával kell rendelkezzen. Ezek a biztonságos titkosítási módszerek mellett magukban foglalnak számítástechnikai követelményeket, mint például *megbízható tűzfalak* alkalmazása, de kiterjednek a fizikai környezetre, katasztrófa tervekre, és a személyzetre is. A nemzetközi feltételeket, szabványokat kielégítő infrasruktúrát magyarul is az angol **Public Key Infrasrtuture** (*Nyilvános Kulcsú Infrastruktúra*) kifejezésből származó **PKI** rövidítés jelöli.

A törvény meghatároz néhány alapvető jogi fogalmat is:

1. **Aláírás-létrehozó adat**: olyan egyedi adat (jellemzően kriptográfiai magánkulcs), melyet az aláíró az elektronikus aláírás létrehozásához használ.

2. **Aláírás-ellenőrző adat**: olyan egyedi adat, (jellemzően kriptográfiai nyilvános kulcs), melyet az elektronikus iratot vagy dokumentumot megismerő személy az elektronikus aláírás ellenőrzésére használ.

3. **Aláírás-létrehozó eszköz**: szoftver vagy hardver, melynek segítségével az aláíró az aláírás-létrehozó adatok felhasználásával az elektronikus aláírást létrehozza.

4. **Aláíró**: az a természetes személy, akihez az elektronikus aláírás hitelesítés-szolgáltató (a továbbiakban: hitelesítés-szolgáltató) által közzétett aláírás-ellenőrző adatok jegyzéke szerint az aláírás-ellenőrző adat kapcsolódik.

Bár 2002-re a legtbb európai ország megalkotta a maga elektronikus aláírás törvényét, ezek összehangolása, pontosítása továbbra is napirenden maradt. Az Information and Communication Technology Standards Board égisze alatt az európai ipari és szabványosítási szervezetek kibocsátották az Európai Elektronikus Aláírás Szabványosítási Kezdeményezést (**EESSI=** European Electronic Signature Standardization Initiative). Ennek az a célkitűzése, hogy átfogóan elemezze a szabványosítási tevékenység jövőbeli igényeit az elektronikus aláírásokról szóló Európai Direktíva megvalósítása érdekében, különös tekintettel az üzleti környezetre.

A törvény lehetőségeet teremt arra, hogy igazolni lehessen: egy jogügylet valóban létrejött,

- meghatározott tartalommal,
- helyen és időben.
- Az aláírás azonosítható és
- egyetlen személyhez kapcsolódik

Ennek megfelelően egy digitális okírat akkor érvényes, ha teljesül a

- **Letagadhatatlanság:** Szerzője és tartalma letagadhatatlan.
- **Keltezés igazolás:** Megállapítható, hogy mikor készült.
- **Sértetlenség:** Sem szerzője, sem más nem változtathatja meg.
- **Bizalmasság:** Jogtalan hozzáférő nem tudja elolvasni.
- **Hitelesség:** Meg lehet győződni arról, hogy az készítette, akinek vallja magát.

1.9.6. Kriptográfiai protokollok

A PKI egyik fonots jellenzője, hogy a rendelkezésre állnak két végpont közötti gyors és automatikus kapcsolatteremtéshez szükséges feltételek. Részben ezt a feladatot végzik el a **kriptográfiai protokollok.** Kapcsolat létrehozó funkciójuk

mellett olyan rendszabályokat is tartalmaznak, amelyek biztosítják, hogy egy adott alkalmazásban a felhasznált algoritmusok a megkívánt titkosságot, vagy hitelességet nyújtsák. Nyílt hálózati környezetben két alapvetően különböző megközelítés bizonyult hatékonynak; a hálózati színtű és az alkalmazás-szintű hitelesítés. A különböző hálózati színt nagy mértékben meghatározza a tulajdonságokat és behatárolja a lehetőségeket.

A gyakorlatban alkalmazott protokollok illusztrálására röviden bemutatjuk a **Secure Socket Layer** (SSL) protokollt. Ennek alkalmazásakor a szerver bizonyítja saját azonosságát a kliens előtt. Tipikusan a kliens "Ki vagy?" kérdésére a szerver válasza: "X-Y vagyok, itt a nyilvános kulcsomhoz tartozó tanúsítvány, és a következő algoritmusokkal tudok rejtjelezni: (lista). Ezután a kliens generál egy véletlen mesterkulcsot, majd rejtjelzi a szerver nyilvános kulcsával (RSA alkalmazás). A szerver a titkos kulcsával dekódolja a mesterkulcsot (RSA), majd a mesterkulccsal rejtjelezve elküld egy üzenetet, ezzel hitelesíti magát. A kliens a mesterkulcsból származtatja a kapcsolati kulcsot (MD5 alkalmazás), amit a szerver is elvégez. Végül a kliens bizonyítja a saját azonosságát. (Az SSL ujabb változatát TLS-nek nevezik.)

Az **S-HTTP** protokollt ellátták az üzenethitelesítéshez, digitális aláíráshoz, kulcs kialakításhoz, rejtjelzéshez szükséges lehetőségekkel. A http nyelvhez hasonlóan az elvégzendő kriptográfiai funkciókat zárójeles parancsok hordozzák, ahol a parancs mellett a szükséges paraméterek is szerepelnek. A két oldal megtárgyalja a lehetséges kapcsolati paramétereket (kulcskialakítás módja, rejtjelző algoritmus, annak üzemmódja, hash függvény, digitális aláíráshoz használt módszer, stb), majd a http file-ba adott konvenciók szerint beágyazza a válaszmezőket.

A nyilvános kulcsú kriptográáfiai protokollok egy érdekes formáját arra használják, hogy valaki bebizonyíthassa társának, hogy ismer egy titkot, mégpedig úgy, hogy maából a titokból egyetlen bitet sem árul el. Ezeket a magyar gyakorlat is **zero-knowledge protokoll**oknak nevezi.

A kriptográfiai szolgáltatást nyújtó hálózatok, rendszerek *rendszergazádinak számos teendőt kell szem előtt tartania* munkája során. Ezekből csak néhányat említünk meg.

A védelem komplex rendszerként történő megszervezése során alapos környezettanulmányra van szükség, ami magában foglalja

- a beszerezhető rejtjelző eszközöket (technikát),
- a potenciális ellenfél érdekeit, anyagi és technikai lehetőségeit,
- a (saját) kiszolgáló személyzet képzettségét.

Meg kell fogalmazni azokat a vezérelveket (azt az adatvédelmi filozófiát), amelyek alapján az egyes részterületek szakemberei kidolgozhatják a konkrét kivitelezésre vonatkozó javaslataikat.

A rendszertervező alapvető feladata annak vizsgálata, hogy a biztonság különböző fokozatainak megvalósítása mibe kerül, és mekkora (anyagi) kockázattal jár, illetve az ellenfélnek mennyibe kerül az illetéktelen betörés a rendszerbe. Az ebből adódó döntés meghozatala viszont már nem az adatvédelmi szakember feladata.

A rendszergazda feladata a megfelelően kialakított adatkezelési rendszabályok betartatása. Ezek egyik legfontosabb momentuma a saját személyzettel szemben is védelem biztosítása. A folyamatos ellenőrzés érvényesítéséből számos rendszerszervezési feladat adódik, mint például a naplózás, véletlen feltöltés, munkatársak helyettesítése, illetéktelen segítségnyújtás megakadályozása.

Adatbankok esetén el kell végezni az adatvédelmi törvény lokális viszonyokra történő adaptálását, a titkos ügykezelés helyi szabályzatának kialakítását. Ehhez szükség van TÜK típusú szabályokra.

Egy új titkosító algoritmus elfogadásakor alapvető elv, hogy a bevizsgálást azon feltétel mellett kell elvégezni, hogy a potenciális ellenfél tökéletesen ismeri a rejtjelzési eljárást, képes kompromisszumot szerezni, de nem tudja megismerni az egyes rejtjelezésekhez alkalmazott kulcsot.

A sugárzásvédelem különösen fontos számítógépes adatforgalom esetén. Gondoskodni kell az elektromos sugárzás lehallgatása elleni védelemről.

Irodalomjegyzék:

[1.9.1.] J.Daemen, V.Rijmen: "The design of Rijndael", Springer, 2002

[1.9.2.] Diffie, W. (1988): : "*The first ten years of public-key cryptography*". Proc. of the IEEE, 76. 560-577.

[1.9.3.] Diffie, W. - Hellman.E. (1977): "*New directions in cryptography*" IEEE Trans. on Info. Theory, IT-22, 644-654.

[1.9.4.] Kahn, D. (1967): "The codebreakers", MacMillan, New York

[1.9.5.] MEH (1996): "Informatikai rendszerek biztonsági követelményei", Miniszterelnőki Hivatal, Informatikai Koordinációs Iroda

[1.9.6.] National Bureau of Standards (1977): "Data Encryption Standard". Washington, D.C.

[1.9.7.] Rivest, R.L. - Shamir, A. - Adleman, L. (1978*): "A method for obtaining digital signatures and public-key cryptosystems*". Comm ACM 21,. 120-126.

[1.9.8.] Shannon, C.E.(1949): "*Communication Theory of Secrecy Systems*", Bell Syst. Techn. J., 28, 656-715.

[1.9.9.] Shannon, C.E.(1951*): "Prediction and Entropy of Printed English*", Bell Syst. Techn. J., 30, 50-64.

1.10. Gráfelmélet és alkalmazásai

Szerzők: Laborczi Péter, Recski András

Lektor: Lajtha György

A távközlő hálózatok jól kezelhető matematikai modelljei a gráfok: a kapcsoló berendezéseket, útvonalválasztókat megfeleltethetjük a gráf pontjainak, az ezeket összekötő kábeleket, rádiós és fényvezető szakaszokat pedig a gráf éleinek. Ebben a fejezetben a távközlő hálózatok modellezéséhez szükséges legfontosabb gráfelméleti fogalmakat és módszereket tekintjük át.

1.10.1. Bevezetés

Kezdetben közvetlenül a fizikai átviteli közegeket használták információ átviteli csatornának, így modellezésük egyszerű volt. Az információátvitel iránti növekvő igény megköveteli, hogy a hálózatok bonyolultabb struktúrával rendelkező adatot (beszédet és képet) hatékonyan továbbítsanak. Az Internet Protocol (IP) a legelterjedtebb alkalmazás, ebben az útvonalak csomagonként is változhatnak, ezért itt fontos egy áttekinthető forgalomirányítási modell alkalmazása. Az IP technológiához fejlesztették ki a többprotokollos címkekapcsolást (MultiProtocol Label Switching - MPLS), amely teljesíti a minőségi igényeket. Az MPLS hálózatok címkekapcsolt útvonalakat (Label Switched Path - LSP) használnak a csomagok továbbítására. A címkekapcsolt útvonalrendszer elve hasonlít az aszinkron átviteli módú (Asynchronous Transfer Mode - ATM) hálózatokban használt virtuális útvonalak (Virtual Path -VP) ötletéhez. Szinkron digitális hierarchiában (Synchronous Digital Hierarchy - SDH) a rendezőket kell úgy konfigurálni, hogy a virtuális konténereket (Virtual Container - VC) optimális útvonalon továbbítsák. Hullámhosszosztásos nyalábolást (Wavelength Division Multiplexing - WDM) vagy többprotokollos hullámhosszkapcsolást (Multiprotocol Lambda Switching -MPλS) használó hálózatokban a hullámhosszcsatornákat kell kiépíteni.

Ezen és más jövőben használatos hálózatok kézenfekvő közös modelljei a *gráfok*, a különböző technológiát rejtő virtuális útvonalak modellje pedig a gráfelméleti *út* fogalom. Ennek a modellnek megvan az az előnye, hogy sok korábbi

gráfelméleti módszert alkalmazhatunk a távközlési hálózatokra. A hálózatmenedzsment egyik érdekes kérdése, hogy mely útvonalat használjuk egy kapcsolat (LSP, VP, hullámhosszcsatorna) kiépítésekor. Az útvonalak lehetnek statikusak, amikor egy kapcsolat állandóan ki van építve, vagy dinamikusak, amikor a kapcsolatokat csak igény érkezése esetén építjük ki. A statikus útvonalak állandó bitsebesség esetén hatékonyak, míg a dinamikusak csomós (börsztös) forgalom esetén előnyösek.

A különböző típusú kapcsolatok kiépítését, karbantartását és lebontását a *hálózat-menedzsment* kezeli. A hálózat menedzsmentjét vagy központilag szokták megvalósítani a *hálózat-menedzsment központ* segítségével, vagy bizonyos funkciókat elosztottan kezelnek. Az utóbbi esetben minden hálózati egységnek van egy modulja, amely az adott elemet kezeli. Ezek a modulok kommunikálhatnak a hálózat-menedzsment központtal. Minden egyes menedzselendő hálózati elemhez tartozik egy menedzsment információs bázis, mely tartalmazza azokat az információkat reprezentáló változókat, melyeket a hálózati elem menedzser figyel és kezel, például az adott szál típusát, a maximális bitsebességet és a védelem típusát.

A gráfelméleti algoritmusokat a hálózattervezés és menedzselés több területén hasznosítják. Pont-pont összeköttetések létrehozásánál általában legrövidebb utat vagy minimális költségű folyamot keresünk a gráfban attól függően, hogy a folyamok a két pont között szétágazhatnak-e. A meghibásodások kivédésére bonyolultabb algoritmusok állnak rendelkezésre, mivel ekkor több utat szoktak definiálni egy kapcsolat számára, melyek egymástól él- vagy pont-függetlenek is. Minimális költségű feszítőfát keresünk egy hálózatban, amikor pont-többpont összeköttetést szeretnénk létrehozni, valamint a mobil hálózatok hozzáférési hálózatai is legtöbbször fák. A hullámhosszosztásos nyalábolást használó (WDM) hálózatok hullámhossz-kiosztásánál vagy mobil hálózatok frekvencia-kiosztásánál a gráfszínezési algoritmusokat alkalmazhatjuk.

Az egyes témáknál kisebb betűmérettel utalunk arra, hogy azokat hogy lehet a távközlő hálózatok tervezésénél alkalmazni.

1.10.2. Alapfogalmak

Egy **gráf** egy rendezett pár, $G=(V,E)$, ahol V egy nem üres halmaz, E pedig ebből a halmazból képezhető párok egy halmaza. V elemeit **pontoknak**, E elemeit **éleknek** nevezzük. Ha egy gráf két élének van közös pontja, azokat **szomszédos éleknek** nevezzük, ha pedig két pontot él köt össze, azokat **szomszédos pontoknak** nevezzük. Egy pontra illeszkedő élek száma a pont **fokszáma**.

Egy $(v_0, e_1, v_1, e_2, v_2, ..., v_{k-1}, e_k, v_k)$ sorozatot **útnak** nevezünk, ha e_i a v_{i-1}-t és v_i-t összekötő él, a pontok pedig mind különbözőek. Ha $v_0=v_k$, de egyébként a pontok különbözőek, akkor ez egy **kör** a gráfban. Ha a gráf bármely két pontja között vezet út, akkor a gráf **összefüggő**.

Hálózatokat gyakran irányított gráfokkal szoktak modellezni (pl. amikor egy átviteli szakasz kapacitása más A pontból B-be, mint B pontból A-ba). Egy **irányított gráf** élei nem $\{v_1, v_2\}$ alakú rendezetlen párok, hanem (v_1, v_2) alakú rendezett párok. Egy ilyen (v_1, v_2) élnek v_1 a **kezdőpontja**, v_2 a **végpontja**. Az út és a kör definíciójának analógiájára definiálható az **irányított út** és az **irányított kör**. Egy irányított gráf **erősen összefüggő**, ha bármely pontjából bármely más pontjába vezet irányított út.

> Számos alkalmazásnál a gráf éleihez vagy pontjaihoz számokat rendelünk. Ezek jelenthetik például egy szakasz fizikai hosszát, késleltetését, vagy a szakasz igénybevételének költségét; vagy egy berendezés késleltetését, költségét. Ilyenkor **élsúlyozott**, ill. **pontsúlyozott** gráfról beszélünk.

A gráfokat legtöbbször mátrixként tároljuk. Egy irányítatlan gráf **pont-pont szomszédsági mátrixa** egyeseket tartalmaz (vagy élsúlyozott gráf esetén az él súlyát) azokon a helyeken, ahol a két pontot él köti össze, egyébként nullákat. Irányított gráfok esetében a számok előjelével adjuk meg az él irányítását. **Pont-él illeszkedési mátrix** esetén a nullától különböző elemek azt jelzik, hogy az adott él az adott pontra illeszkedik. Az él irányítását itt is előjellel adjuk meg.

1.10.3. Fák

Az összefüggő körmentes gráfokat **fáknak** nevezzük. Egy n pontú fa éleinek száma n-1. Egy fában bármely két pont között pontosan egy út létezik.

Ennek következménye, hogy fa struktúrájú hálózatokban az útvonalválasztás egyértelmű.

Az *F* gráfot *G* gráf **feszítőfájának** nevezzük, ha *F* fa, pontjainak halmaza megegyezik *G* pontjainak halmazával, és *F* élei szerepelnek *G*-ben is. Egy élsúlyozott gráfban egy **fa súlya** a fa éleihez rendelt súlyok összegét jelenti. Minden összefüggő gráf tartalmaz feszítőfát, melyek közül szeretnénk a minimális súlyút megtalálni. Ennek megoldására a mohó algoritmus a legegyszerűbb, amely az alábbiak szerint működik. Az éleket egyesével választjuk ki a következőképpen. Először válasszuk ki a gráfból a legkisebb súlyú élek egyikét. Tegyük fel, hogy már kiválasztottunk néhány élt. Ekkor válasszuk ki a legkisebb súlyú olyan élek egyikét, amely nem alkot kört az eddig már kiválasztottakkal. Ha ilyen nincs, megállunk, ha van, akkor az eljárást ismételjük.

Minimális költségű feszítőfát keresünk egy hálózatban, amikor pont-többpont összeköttetést szeretnénk létrehozni, valamint mobil hálózatok hozzáférési hálózatai is legtöbbször fák. Ezekben az esetekben azonban két további korlát merülhet fel: az egyik az, hogy a berendezések fizikai korlátai miatt a fa pontjainak fokszáma nem léphet túl egy adott értéket; a másik pedig az, hogy a késleltetés, vagy más minőségrongáló tényezők miatt az utak hossza is korlátozva lehet a fában.

Sokszor nem kell a hálózat összes pontjának szerepelnie a fában, hanem csak a hálózat valamely kitüntetett pontjai között szeretnénk kapcsolatot létrehozni. Ekkor egy másfajta gráfot keresünk a hálózatban.

Jelöljünk ki a gráfban néhány pontot, melyeket nevezzünk **terminálóknak**. Ekkor a terminálokat tartalmazó fákat **Steiner-fáknak** hívjuk, ezek közül keressük a minimális súlyút (ez lehet kisebb súlyú, mint a mohó algoritmus által szolgáltatott kifeszítő fa). A fa tartalmazhat olyan pontokat is, melyek nem terminálok, ezeket **Steiner-pontoknak** hívjuk.

1.10.4. Utak

Ha egy távközlő hálózatban két pont között kapcsolatot akarunk létrehozni, az a technológiától függően kétféle lehet: az adott kapcsolatot vagy egyetlen út mentén kell megvalósítanunk (ezzel foglalkozunk ebben a részben), vagy pedig az átvitel szétágazhat (1.10.1. ábra). Az utóbbi esetről a következő részben lesz szó.

Legtöbbször a minimális súlyú (legrövidebb) utat keressük a gráfban, azaz az út által használt élek súlyainak összegét szeretnénk minimalizálni. A legismertebb legrövidebb utat kereső algoritmus Dijkstra algoritmusa, amely akár irányítatlan, akár irányított gráfban alkalmazható, és egy adott pont és az összes többi pont távolságát adja meg, ha minden él súlya pozitív.

Dijkstra algoritmusában nincs megengedve, hogy az élek súlya negatív legyen, ellentétben két másik nevezetes algoritmussal, melyek negatív élsúlyokat megengedő irányított gráfokra is működnek. A gráf azonban ezekben az esetekben sem tartalmazhat negatív összsúlyú irányított kört. Ford algoritmusa egy adott pont és az összes többi pont távolságát adja meg, Floyd algoritmusa pedig minden pontból minden pontba meghatározza a távolságot.

Két utat **élfüggetlennek** nevezünk, ha nincs közös élük, és **pontfüggetlennek**, ha a végpontok kivételével nincs közös pontjuk. Pontfüggetlen utak nyilván egyben élfüggetlenek is.

> Hibatűrő hálózatok esetén két vagy esetleg több utat szoktak definiálni egy kapcsolat számára, melyek él- vagy pontfüggetlenek a védelem típusától függően.

Dijkstra algoritmusának egy módosítása, mely Suurballe algoritmusaként is ismert, talál két vagy több minimális összsúlyú él- vagy pontfüggetlen útvonalat egy pozitív élsúlyokat tartalmazó irányítatlan vagy irányított gráfban, amennyiben léteznek ilyenek.

Egy gráfot ***k*-szorosan összefüggőnek** nevezünk, ha legalább $k+1$ pontja van, és akárhogy hagyunk el belőle k-nál kevesebb pontot, a maradék gráf összefüggő marad. A gráf ***k*-szorosan élösszefüggő**, ha akárhogy hagyunk el belőle

k-nál kevesebb élt, összefüggő gráfot kapunk. A gráf akkor és csak akkor *k*-szorosan összefüggő, ha legalább *k*+1 pontja van, és bármely két pontja között létezik *k* pontidegen út, hasonlóan akkor és csak akkor *k*-szorosan élösszefüggő, ha bármely két pontja között létezik *k* élidegen út.

Egy irányított gráfban az *s* pontból *t*-be vezető páronként élidegen irányított utak maximális száma megegyezik az összes irányított *s*-*t* utat lefogó élek minimális számával. Hasonló állítás érvényes a páronként pontidegen irányított utak maximális számára is, ha a gráfban nincs *(s,t)* él.

1.10.5. Folyamok és vágások

Legyen *G* egy irányított gráf. Rendeljünk minden *e* élhez egy *c(e)* nemnegatív számot, amit az él **kapacitásának** nevezünk. Jelöljünk ki továbbá két *s, t* pontot *G*-ben, melyeket **forrásnak,** illetve **nyelőnek** hívunk.

> Az élek kapacitása például egy átviteli szakasz maximális bitsebességének (áteresztőképességének) felelhet meg, és *s*-ből *t*-be kell egy bizonyos mennyiségű (m_f) adatot eljuttatni.

Legyen *f(e)* az az adatmennyiség, ami az *e* élen folyik át. Az *f* függvény **megengedett függvény**, ha minden élre $f(e) \leq c(e)$, és minden pontra igaz a folyammegmaradás, azaz a pontba mutató éleken az *f* értékek összege megegyezik a pontból kimutató éleken vett összeggel, kivéve az *s* és *t* pontokat. *s* pontra a kimutató, *t* pontra pedig az odamutató éleken vett összeg m_f. Ilyenkor az *f* függvényt **folyamnak**, az m_f mennyiséget a folyam **értékének** nevezzük. Egy élt **telítettnek** hívunk egy folyamban, ha $f(e)=c(e)$, és **telítetlennek**, ha $f(e)<c(e)$.

A maximális értékű folyamot az alábbi algoritmus segítségével határozhatjuk meg. Ha van egy olyan irányított út *s*-ből *t*-be, amelynek minden élén a folyam értéke kisebb, mint a kapacitása, vagyis minden él telítetlen, akkor ezen út mentén a folyam értékét minden élen megnövelhetjük annyival, hogy az egyik él telített legyen. Növelhetjük a folyam értékét úgy is, hogy egy ellentétes irányítású élen csökkentjük a folyamot. Az ilyen utakat **javító utaknak** hívjuk. Egy tétel kimondja, hogy egy folyam értéke akkor és csak akkor maximális, ha nincs javító út *s*-ből *t*-be.

Osszuk a gráf pontjainak *V* halmazát két részhalmazra: az egyik halmaz legyen *X*, a másik *V–X*. Feltesszük, hogy *X* halmaz tartalmazza az *s* pontot, *V–X*

pedig tartalmazza a *t* pontot. Az élek azon halmazát, amelyeknek egyik végpontja *X*-beli, a másik pedig *V—X*-beli, a hálózati folyam egy **(*s,t*)-vágásának** nevezzük. A vágás értéke azon éleken levő kapacitások összege, amelyek egy *X*-beli pontból egy *V—X*-beli pontba mutatnak, azaz a vágás értékében csak az előremutató élek játszanak szerepet. Ford és Fulkerson tétele kimondja, hogy a maximális folyam értéke egyenlő a minimális vágás értékével. A fent említett javító utas módszerrel mind a maximális folyamot, mind a minimális vágást megtaláljuk, ha a lehetséges javítóutak közül mindig egy legrövidebbet választunk. Ha a kapacitások egész számok, akkor van olyan maximális folyam is, mely minden élen egész értékű.

Egyes alkalmazások esetében nemcsak az élekhez, hanem a pontokhoz is rendelünk kapacitást, és előírjuk, hogy a ponton legfeljebb ennyi adatmennyiség mehet át. Ezt a problémát visszavezetjük az előzőre a következőképpen. Minden *c(v)* kapacitású *v* pontot helyettesítünk két *v′*, *v″* ponttal. Az eddig *v*-be mutató élek új végpontja legyen *v′*, az eddig *v*-ből kiinduló élek új kezdőpontja pedig legyen *v″*, ezenkívül *v′*-ből mutasson egy él *v″*-be és ennek kapacitása legyen *c(v)* (1.10.2. ábra). Így az új *(v′, v″)* él kapacitásával fejezzük ki *v* pont kapacitását.

Ha megengedünk irányítatlan éleket, akkor azt két egymással ellentétesen irányított éllel helyettesíthetjük, azaz egy *c* kapacitású *{u,v}* irányítatlan él helyett felveszünk két *c* kapacitású *(u,v)* és *(v,u)* irányított élt.

Eddig a hálózatban egyféle adatfolyamot, azaz egyetlen terméket feltételeztünk. Ekkor **egytermékes folyamproblémáról** beszélünk. Távközlő hálózatokban azonban általában több kapcsolatot szeretnénk létrehozni, melyeket meg kell különböztetnünk (általában a forrás-nyelő párok alapján). Ezeket a

folyamproblémákat nem oldhatjuk meg egyenként a közös élkapacitások miatt, hiszen az egyes éleken a folyamértékek összege nem lehet nagyobb, mint az adott él kapacitása. Ekkor **többtermékes folyamproblémáról** beszélünk, melynek megoldása gyakran csak közelítőleg lehetséges.

1.10.6. Gráfok színezése

Egy *G* gráf ***k* színnel színezhető**, hogyha minden pontját ki lehet színezni úgy, hogy bármely két szomszédos pont színe különböző legyen. *G* **kromatikus száma** $\chi(G) = k$, ha *G* *k* színnel kiszínezhető, de *k*-1 színnel nem. Egy ilyen színezésnél az azonos színt kapott pontok halmazát **színosztálynak** nevezzük.

Egy gráf **teljes**, ha bármely két pontja között van él. *G* egy teljes részgráfját **klikknek** nevezzük. A *G*-ben található maximális méretű klikk pontszámát $\omega(G)$-vel jelöljük és a gráf **klikkszámának** nevezzük. Nyilvánvaló, hogy ha egy gráfban van egy klikk, akkor ennek semelyik két pontja nem lehet azonos színű, azaz a kromatikus szám nem kisebb, mint a klikkszám: $\chi(G) \geq \omega(G)$. Jelöljük a gráf maximális fokszámú pontjának fokszámát $\Delta(G)$-val. Ha elkezdjük tetszőleges sorrendben színezni a gráf pontjait, és egy újabb pontot akarunk kiszínezni, akkor ennek legfeljebb Δ szomszédja van kiszínezve, így a $\Delta(G)+1$-dik színt felhasználhatjuk a színezésre. Azaz a kromatikus szám legfeljebb eggyel nagyobb, mint a maximális fokszám: $\chi(G) \leq \Delta(G)+1$. Ezzel a kromatikus számra alsó és felső korlátot adtunk. Meg kell azonban jegyezni, hogy ezek a korlátok sok esetben nagyon távol vannak a kromatikus számtól.

Ha egy gráf lerajzolható a síkba úgy, hogy pontjai a sík különböző pontjai legyenek, élei pedig olyan, a végpontok között vezető folytonos vonalak, melyeknek nincs közös belső pontja, akkor a gráf **síkbarajzolható**. Minden síkbarajzolható gráfot ki lehet színezni négy színnel: $\chi(G) \leq 4$.

Egy gráf élei *k* színnel kiszínezhetők, hogyha minden élét ki lehet színezni úgy, hogy bármely két szomszédos él színe különböző legyen. *G* **élkromatikus száma**, $\chi_e(G)=k$, ha *G* élei *k* színnel kiszínezhetők, de *k*-1 színnel nem. Az élkromatikus szám nem lehet kisebb a maximális fokszámnál, hiszen az egy pontra illeszkedő éleket mind különböző színre kell színezni. Viszont egyszerű (azaz

hurokéleket és többszörös éleket nem tartalmazó) gráfokra az élkromatikus szám ennél legfeljebb eggyel lehet nagyobb: $\Delta(G) \leq \chi_e(G) \leq \Delta(G)+1$.

A gráfszínezést például hullámhosszosztásos nyalábolást használó (WDM) hálózatokban hasznosíthatjuk, ahol az útvonalakhoz kell hullámhosszakat rendelni. Azonos élen áthaladó útvonalak nem használhatják ugyanazt a hullámhosszt. Tehát konstruálunk egy olyan gráfot, melynek pontjai a WDM hálózat útvonalai, és két pont akkor és csak akkor van összekötve, ha a két útnak van közös éle. Ekkor a gráf kromatikus száma megadja a szükséges hullámhosszak minimális számát, egy konkrét színezés pedig megad egy hullámhosszkiosztást.

Egy másik alkalmazásnál, mobil hálózatok kialakításánál arra kell ügyelni, hogy bizonyos (pl. szomszédos) adók nem használhatnak azonos frekvenciát. Ekkor konstruálunk egy gráfot, melynek pontjai az adók, és két pont akkor van összekötve, ha a megfelelő adók nem használhatnak azonos frekvenciát. Ekkor a gráf kromatikus száma megadja a szükséges frekvenciák minimális számát, egy konkrét színezés pedig megad egy frekvenciakiosztást.

1.10.7. Algoritmusok bonyolultsága

Az eddig ismeretett algoritmusok nagyrésze (minimális összsúlyú feszítő fa, legrövidebb út, maximális folyam vagy minimális vágás keresése) a gyakorlatban is jól használhatóak: lépésszámuk a legrosszabb esetben is felülről becsülhető az input méretének egy (általában nem túl magas fokszámú) polinomjával, így egy számítógépes implementálás még sokezer pontú gráfok esetén is ésszerű időn belül biztosítja az optimumot. Hasonlóképp polinom időben eldönthető például, hogy egy gráf összefüggő vagy kétszeresen összefüggő-e, hogy egy irányított gráf erősen összefüggő-e, vagy hogy egy gráf síkbarajzolható-e.

Ugyanakkor számos olyan probléma van, melyekre egzakt, polinom-idejű algoritmus nem ismeretes. Az eddig említettek közül ilyen a minimális összsúlyú Steiner-fa, a leghosszabb út vagy maximális vágás keresése, a többtermékes folyamprobléma általános megoldása, egy gráf klikkszámának, kromatikus számának vagy élkromatikus számának meghatározása. Ezek a problémák az ún. **NP-nehéz** [1.10.10] problémák közé tartoznak: ha bármelyikre sikerülne valakinek polinom idejű algoritmust találnia, akkor annak az algoritmusnak, mint szubrutinnak a hívásával az összes többire is adódnék polinom idejű algoritmus. NP-nehéz problémák futási ideje

- a hálózat méretétől és a probléma típusától függően - nagyon hosszú lehet. E problémákra valószínüleg a jövőben is csak olyan megoldások várhatóak, ahol

(1) vagy a lépésszám a feladat méretének nem polinomja,

(2) vagy nem az optimumot, csak egy közelítést találunk,

(3) vagy más elven alapuló számítási modellre van szükség (pl. véletlen számok generálását is igénylő algoritmusra).

Fontos azonban kihangsúlyozni, hogy

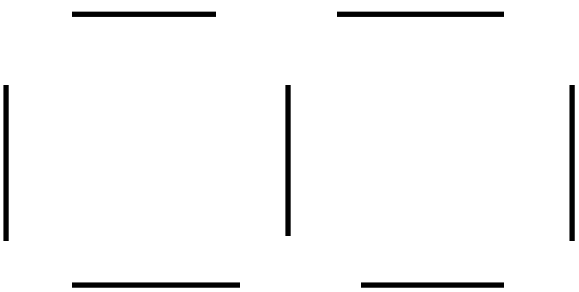
a. amennyiben a feladat mérete nem túl nagy, akkor ez a megkülönböztetés nem kritikus;

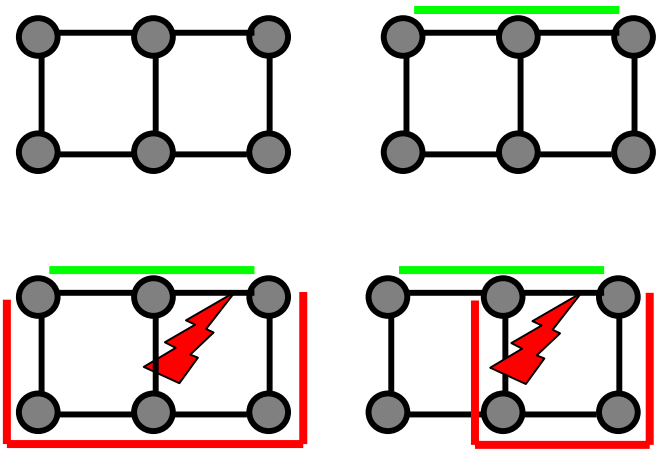
b. egy NP-nehéz problémának is lehetnek polinom időben megoldható speciális esetei (pl. élkromatikus szám meghatározása páros gráfokban, vagy leghosszabb irányított út keresése irányított kört nem tartalmazó gráfokban);

c. vannak olyan problémák is, melyekre nem ismeretes ugyan polinomrendű algoritmus, de - ismereteink szerint - nem tartoznak az NP-nehéz problémák közé sem (pl. gráfok izomorfiájának eldöntése).

1.10.8. Példák

Tekintsük az 1.10.3. ábrán látható hálózatot. Ezt a hálózatot modellezzük az 1.10.4.(a) ábrán látható gráffal. Ezek után egy útvonal kiépítéséhez legrövidebb útvonalat keresünk két pont között (1.10.4.(b) ábra). Ezt a forgalmat kétféleképpen lehet védeni: *útvonalvédelemmel* (1.10.4.(c) ábra) vagy *szakaszvédelemmel* (1.10.4.(d) ábra). Útvonalvédelem esetén minden egyes kapcsolat külön van védve, vagyis a forgalmat új útvonalon vezeti el a forrás és a nyelő csomópont között a hibás szakasz elkerülésével. Ez a megközelítés takarékosan bánik a védelmi kapacitással, de hosszú időbe telik a hiba észlelése. Az útvonalvédelem speciális





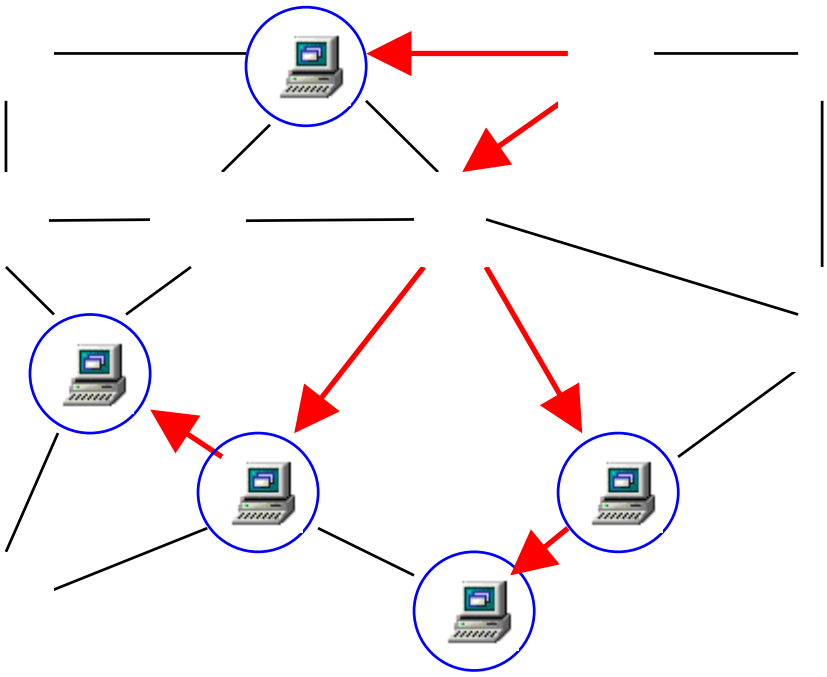
formája, ha az üzemi és a védelmi útvonal élfüggetlen egymástól. Ebben az esetben nem kell a hiba pontos helyét meghatározni, így a forgalmat azonnal helyreállíthatjuk. Szakaszvédelem esetén csak a meghibásodott szakasz forgalmát kell egy kerülő úton elvezetni.

Az 1.10.5. ábrán egy kiépített Steiner-fára láthatunk példát, ahol a bekarikázott csomópontok a terminálok. Ez a Steiner-fa egy Steiner-pontot tartalmaz, vagyis egy további klienst használunk ahhoz, hogy minden terminált elérjen a fa.

1.10.9. Irodalom

A gráfelmélet elemeit az [1.10.1] vagy [1.10.2] könyvekből és a [1.10.3] példatárból lehet alaposan megismerni. Az alkalmazások iránt érdeklődő könnyebb bevezetést talál [1.10.4]-ban. A gráfelméleti algoritmusokat nagyon igényesen tárgyalja [1.10.5]. A téma aktív kutatói számára a [1.10.6] kézikönyvet ajánljuk.

Magyar nyelven bevezető egyetemi jegyzet [1.10.7] vagy [1.10.8], míg [1.10.9] már középiskolás szakkörökben is jól feldolgozható. Az algoritmusokról pedig [1.10.10] ajánlható.



Irodalomjegyzék

[1.10.1] J. A. Bondy, U. S. R. Murty, "Graph Theory with Applications", The Macmillan Press Ltd, London, 1976

[1.10.2] R. Diestel, "Graph Theory", Springer, Berlin, 1997 and 2000

[1.10.3] L. Lovász, "Combinatorial Problems and Excercises", North-Holland, Amsterdam, 1979 and 1993
magyarul: Lovász László, "Kombinatorikus problémák és feladatok", Typotex, Budapest, 1999

[1.10.4] J. Gross, J. Yellen, "Graph Theory and its Applications", CRC Press, Boca Raton, 1999

[1.10.5] R. E. Tarjan, "Data Structures and Network Algorithms", SIAM, Philadelphia, 1983

[1.10.6] R. L. Graham, M. Grötschel, L. Lovász, "Handbook of Combinatorics", Elsevier, Amsterdam and The MIT Press, Cambridge, MA, 1995

[1.10.7] Hajnal Péter, "Gráfelmélet", Polygon, Szeged,1997

[1.10.8] Katona Gyula Y., Recski András, Szabó Csaba "A számítástudomány alapjai", Typotex, Budapest, 2002.

[1.10.9] Andrásfai Béla, "Ismerkedés a gráfelmélettel", Tankönyvkiadó, Budapest, 1973

[110.10] Rónyai Lajos, Ivanyos Gábor, Szabó Réka, "Algoritmusok", Typotex, Budapest, 1999

1.11. Hálózatok rétegmodellje

Szerző: Mazgon Sándor

Lektor: dr. Bartolits István

Az információközlés rendszabályai a technológia fejlődésével változtak, mind összetettebbé váltak és kellett, hogy váljanak az ember-ember kapcsolat mellett megjelenő ember-gép, gép-gép kapcsolatok folytán. A számítógépek és az ezekhez hasonló intelligens adat-végberendezések (nyílt rendszerek) közötti adatátvitel a kommunikációs protokollok olyan rendszerét és készleteit igényli, amelyekkel bármely számítógépen futó alkalmazás és bármely számítógép használó igényeit egyaránt ki lehet szolgálni. Ehhez olyan modellt állítottak fel a nemzetközi szabványosító szervezetek (esetünkben az **ITU** és az **ISO**), amely minden kommunikációs részfeladatot is számba véve a legnagyobb rugalmasságot nyújtja egymáshoz képest idegen technikával készült adat-végberendezések (számítógépek, egyszerű terminálok, alkalmazások, stb.) között. Ezt nevezzük "**nyílt rendszerek összekapcsolási**" modelljének (**OSI**, Open Systems Interconnection)[1], amelyet a hálózatok legáltalánosabb – referenciaként alkalmazott - funkcionális hivatkozási modelljének tekintünk.

Az OSI modellje kialakításához több alapelvet fektetett le. Ezek közül a legfontosabbak a következők:

a. A modellt rétegekre tagolva kell megalkotni, ahol az egyes rétegek más-más absztrakciós szinteket képviselnek.

b. Túl sok réteget nem érdemes alkalmazni, mert feleslegesen megnöveli a rendszertechnikai munkát.

c. C. Ott érdemes határt vonni a rétegek közé, ahol a kölcsönhatás a réteghatáron át minimális. Ott célszerű réteghatárt létesíteni, ahol idővel külön előny lehet a réteginterfész szabványossága. A határ megválasztásában előny a korábbi jó tapasztalat.

[1] **ISO 7498-1 | ITU-T X.200**: Nyílt rendszerek összekapcsolása. Hivatkozási modellek. Az alapmodell **[1.11.1]**

d. Külön rétegbe valók az eltérő eljárások és technológiák, azonos rétegbe valók a hasonló feladatok. Külön rétegbe kívánkoznak az adatkezelések eltérő absztrakciós szintjei, szintaktikái, szemantikája.

Az alapelvek szerint szabványosított 7 rétegű OSI modell rétegei az alábbiak:

Fizikai réteg: A fizikai kapcsolat (link) létrehozásának, fenntartásának és megszüntetésének villamos, mechanikai, funkcionális és eljárási jellemzőit írja elő ahhoz, hogy azon át bitfolyamot lehessen transzparens (áttetsző) módon közvetíteni tekintet nélkül a bitfolyam belső szerkezetére. Ez a réteg csak a bitek átvitelével, az átvitel időtartamával és irányával foglalkozik

Adatkapcsolati réteg: Hálózati entitások (egyedek, egységek) közötti adatközlés funkcionális és eljárási eszközeit valamint a hibajelzés és többnyire a hibajavítás alapeszközeit foglalja magában. Gondoskodik adatkapcsolati összeköttetések (data link connection) létesítéséről, fenntartásáról és megszüntetéséről, a bitfolyamok bitjeinek karakterekké és adatkeretekké csoportosításáról, karakter- és keretszinkronizálásról, hibakorlátozásról, közeghozzáférés-vezérlésről valamint adatfolyam-vezérlésről (ennek példái: a HDLC és az Ethernet). Az adatkapcsolati réteg elfedi az átviteli vonal sajátosságait, átviteli hibáit a hálózati réteg elől.

Hálózati réteg: Feladata az adatkeretekből kialakított csomagok eljuttatása a hálózaton keresztül a forrástól a célig. Ehhez ismerni kell a hálózat felépítését és ki kell választani a valamilyen szempontból legkedvezőbb útvonalat. Ezt az eljárást útválasztásnak vagy forgalomirányításnak nevezzük. E réteg teremt függetlenséget az átvitelben alkalmazott technológia és az adatkommunikációs közvetítés (relaying) és forgalomirányítás (routing) között, elfedi a továbbító közeg hiányosságait/kényelmetlenségeit a felsőbb rétegek elől, gondoskodik a kapcsolásról, azaz a hálózati összeköttetések felépítéséről, fenntartásáról és lebontásáról, valamint az adatok használók közti közvetítéséről.

Szállítási réteg:. Feladata, hogy a különböző hálózati összeköttetések között a minőségi eltéréseket eltüntesse. Gondoskodik a nyílt (vég)rendszerek közötti transzparens adatátvitelről, tehermentesíti a felsőbb rétegeket a megbízható és költséghatékony adattovábbítás megszervezésének gondjaitól, az alkalmazói programtól megkívánt szolgáltatásminőségben gondoskodik a végtől végig való ellenőrzésről és információcseréről. Ez a legalsó olyan réteg, amely mindig végrendszerben helyezkedik el.

Viszonyréteg: Feladata az egymással adatokat cserélő végrendszerek közötti kommunikáció ütemezése. A viszonyréteg végzi az alkalmazási folyamatok közötti párbeszéd szervezését és szerkesztését, mechanizmusokat tartalmaz, amelyek révén mindkét irányú és egyidejű, vagy pedig váltakozóirányú működés is megvalósítható, amelyekkel kitűzhetők az adatátvitel egyes részeinek ismétléséhez az alkalmas szinkronizációs (összehangolási) pontok ("kályhák"), valamint amelyekkel az adatcserék strukturálhatók, megszerkeszthetők.

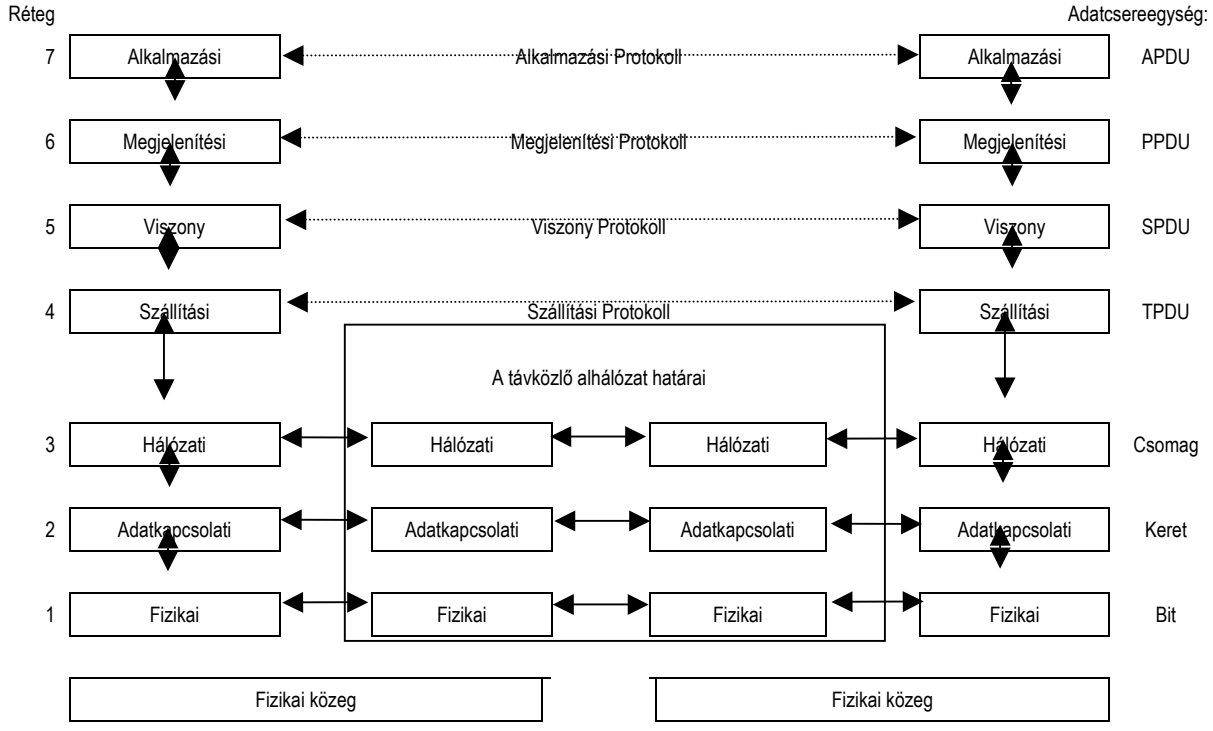
Megjelenítési réteg: Feladata a szállított információ ábrázolásához kapcsolódik, ez a réteg teszi az információ kezelését és megjelenítését egységessé. Függetlenné teszi az alkalmazási folyamatokat az adatábrázolás, azaz szintaktikáik különbségeitől, lehetőséget ad a (transzfer-) szintaktika megválasztására és a szintaktikák közötti átváltásra (konvertálásra) mégpedig azzal, hogy a használó az egyes, választható környezetek egymásba alakításával megválaszthatja megjelenítési környezetét.

Alkalmazási réteg: Az OSI modell legfelső rétegének a feladata, hogy az alkalmazási folyamatok számára hozzáférést nyújtson az OSI környezethez. Közvetlenül foglalkozik az alkalmazások követelményeivel. Az alkalmazási rétegben nyújtott szolgálatelemeket használja valamennyi alkalmazási folyamat. Ezek magukban foglalnak folyamatközi kommunikációt végző könyvtárkezelő rutinokat, alkalmazási (rétegben futó) protokollok összeállítására alkalmas közös eljárásokat, valamint hozzáférési eljárásokat a hálózatban másutt székelő kiszolgálógépek által nyújtott szolgáltatásokhoz.

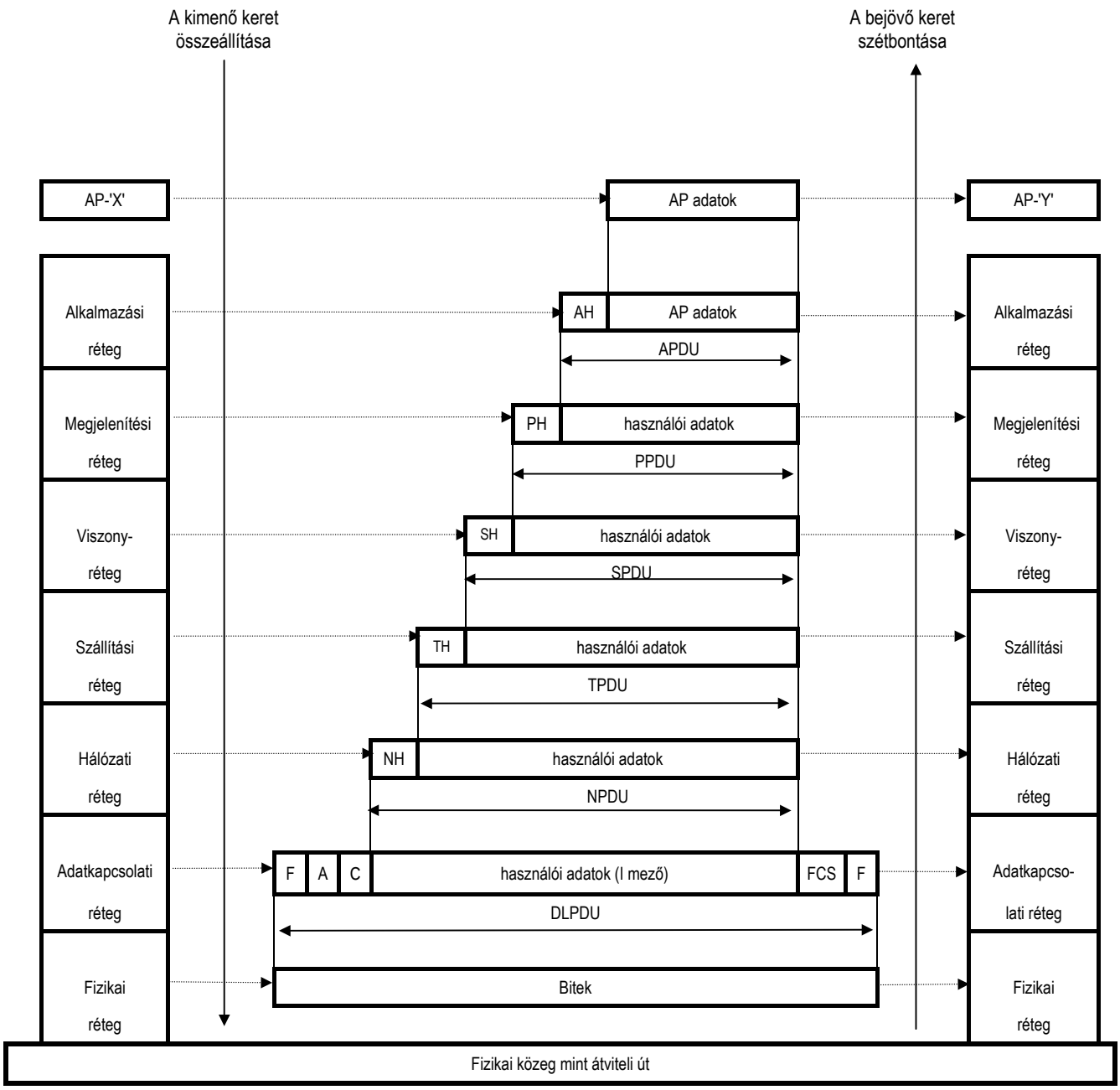
Az OSI rétegek funkcióit (a modellben) rétegentitások hordozzák, más nyílt rendszerek azonos rétegbeli entitásokkal rétegprotokollok szerint protokoll adatelemekkel kommunikálnak (**PDU**-kkal). A kommunikáció az egyes nyílt rendszerek között a fizikai közegben létrejött **adatáramkörön** át folyik, itt haladnak azok az **adatjelek,** amelyek a biteket viszik át az egyik végrendszerről a másikra. Valamennyi réteg részt vesz a kommunikációban, amelynek kezdeményezője mindig a legfelső réteget foglalkoztató alkalmazás. Ez küldi/fogadja az alkalmazási adatelemet (**APDU**), amely az alkalmazási adatokból (**AP adatok**) és az azokat kísérő alkalmazási réteget vezérlő információból (**AH**), röviden fejrészből (header) áll. Az egyes rétegek saját vezérlő információikat a felettük álló rétegtől kapott **PDU**-hoz teszik hozzá: így az **APDU** a megjelenítési rétegben használói adatként jelenik meg,

a hozzáillesztett **PH** vezérlőinformációval együtt alkotja a megjelenítési adatelemet (**PPDU**). Ezt nevezzük egyfajta csomagolástechnikának (packeting) vagy borítékolásnak. Hasonlókép a viszonyrétegben **SPDU** = használói adat (**PPDU) + SH,** a szállítási rétegben **TPDU** = használói adat (**SPDU) + TH,** a hálózati rétegben **NPDU** = használói adat (**TPDU) + NH,** ahogyan azt az 1.11.1 és 1.11.2 ábrák mutatják**.** A 1.11.1. ábrán a vízszintes nyilak a (réteg-)protokollokat, a függőleges nyilak a (réteg-) interfészeket jelképezik. Ezek az **OSI** rendszerben mind szabványosak.

Több változatban is léteznek OSI megvalósítások, különösen sok a változat az alsó 3 réteg megvalósításában. Ezek összehasonlítása és összerendelése néha egészen összetett feladatot jelent, a hálózati rétegek tekintetében mégis legtöbbször az OSI referenciamodell fogalmi rendszerében lehet a legjobban közös nyelvet találni az eltérő rendszerek között.



1.11.1. ábra. A kommunikáció OSI-modellre alapozott architektúrája



Jelölések-rövidítések:

AP	alkalmazási program	('X' az egyik oldalon és 'Y' a másik oldalon)	
AH	alkalmazási fejrész	PH	megjelenítési fejrész
SH	viszony fejrész	TH	szállítási fejrész
NH	hálózati fejrész	FCS	keretellenőrző sorozat
APDU	alkalmazási protokoll adatelem	PPDU	megjelenítési protokoll adatelem
SPDU	viszony protokoll adatelem	TPDU	szállítási protokoll adatelem
NPDU	hálózati protokoll adatelem	DLPDU	adatkapcsolati protokoll adatelem
F	zászló	A	címzés
C	vezérlés	I mező	információmező

Megjegyzés: az absztrakt fizikai rétegben a fizikai protokoll adatelem (PhPDU) a bit, amit a közegben jelek hordoznak az abban mindig jelenlévő adatáramkörön.

1.11.2. ábra. Információegységek felépítése az OSI modellben

Irodalomjegyzék

[1.11.1] MSZ EN ISO/ IEC 7498-1: 1995 Információfeldolgozó rendszerek. Nyílt rendszerek összekapcsolása. Referenciamodell. 1. rész: Alapmodell; p.92. Information processing systems. Open systems interconnection. Basic reference model. Part 1: The basic model; p92. (korábban MSZ 7808-1: 1986, majd MSZ ISO 7498-1: 1995; idt ISO/ IEC 7498-1: 1994; idt EN ISO/ IEC 7498-1: 1995)

[1.11.2] MSZ ISO 7498-2: 1994 Információfeldolgozó rendszerek. Nyílt rendszerek összekapcsolása. Referenciamodell. 2. rész: Biztonsági architektúra; p37. Information processing systems. Open systems interconnection. Basic reference model. Part 2: Security architecture; p 37.

[1.11.3] MSZ ISO 7498-3: 1994 Információfeldolgozó rendszerek. Nyílt rendszerek összekapcsolása. Referenciamodell. 3. rész: Névadás és címzés; p.25. Information processing systems. Open systems interconnection. Basic reference model. Part 3: Naming and addressing; p.25. (idt ISO 7498- 3: 1989)

[1.11.4] MSZ ISO 7498-4: 1994 Információfeldolgozó rendszerek. Nyílt rendszerek összekapcsolása. Referenciamodell. 4. rész: Menedzselési keretrendszer; p.16. Information processing systems. Open systems interconnection. Basic reference model. Part 4: Management framework; p.16. (idt ISO 7498-4: 1989)

1.12. A TÁVKÖZLÉS GAZDASÁGTAN ALAPJAI

Szerző: dr. Lajtha György

Lektor: Konkoly Lászlóné

A távközlés történetének első évszázadában (1880-1980), a legtöbb európai országban csak a telefon és a távíró létezett, mint közcélú távközlési szolgáltatás. Az utolsó évtizedben ez gyökeresen megváltozott: új távszolgáltatások, üzleti és szórakoztató alkalmazások jelentek meg. Ennek jelentős hatása van a távközlési hálózatok tervezésének optimalizálására. Ezen kívül a méretezési, üzemeltetési és fenntartási elveket is befolyásolja. Mindemellett az elmúlt 20 évben volt néhány olyan döntő változás, mely a gazdasági számításokat alapjaiban módosította. Ezek közül megemlítünk néhányat:

- Az emberi erőforrások költsége, összehasonlítva az elektronikus eszközök és alkatrészek költségével, jelentősen megnövekedett. Ugyancsak magasabb lett az épületek és a telkek ára, ha összevetjük azokat egy csatorna vagy átviteli út beruházási költségével.
- A különböző tarifák lényegesen magasabbak lettek mint korábban voltak, ha a távközlési berendezések árával hasonlítjuk össze.
- A számításokat befolyásolja az is, hogy a korábbi állami monopóliumból a verseny egy jellegzetes területe lett.

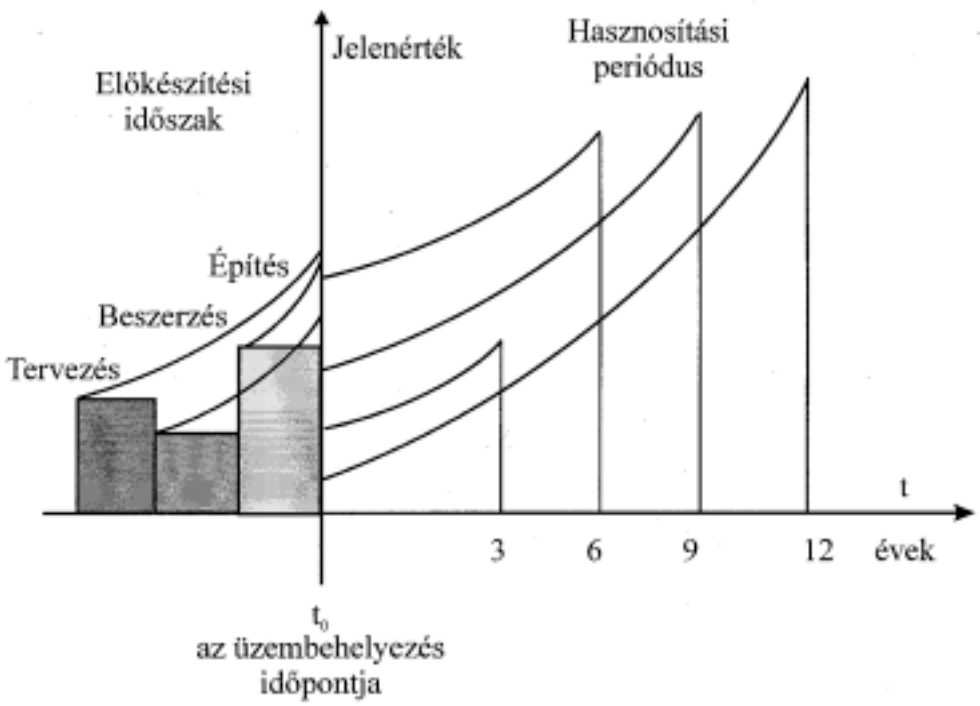
  Mindezen változások indokolják, hogy áttekintsük ezt a témakört.

1.12.1. A hálózattervezés klasszikus gazdasági számítási folyamata

A világszerte alkalmazott eljárás első lépése az aktuális forgalmi mátrix meghatározása. Ezt követően különböző prognózis módszerekkel elkészítjük ugyanezt pl. 5, 10 és akár 20 éves távlatra. Ez képezi az alapját a forgalomirányítási stratégiának. Az ehhez illeszkedő hálózati struktúrát a lehetséges csillag, többszörös csillag, busz, fa, gyűrű és kettősgyűrű elrendezések közül választhatjuk ki. (A struktúrák a 4.1.1. ábrán láthatók.) Ezek közül a választást nemcsak a forgalmi tényezők szabják meg, hanem a terület kulturális, adminisztratív, oktatási és egészségügyi struktúrája is. Ezek meghatározzák a forgalomkoncentráció helyeit, a

szükséges használhatóságot (mely az üzemképes idő és a teljes naptári idő hányadosát jelenti) és a kerülőutak kiépítését. Ha a forgalmi (logikai) struktúrát már meghatároztuk, akkor következhet a megvalósítás tervezése.

A megvalósítás során figyelembe kell venni a meglévő hálózatot, a rendelkezésre álló épületeket, a lehetséges eszközválasztékot, és ezzel kell az új tervet összehangolni. A gazdasági optimum keresése a jelenérték számításon alapul, ami azt jelenti, hogy minden kiadást az időskálán visszavetítünk az üzembe helyezés időpontjára. A vetítést a távközlési kamatláb hatásával végezzük el. Az 1.12.1. ábrán vázlatosan bemutatjuk az idő-költség görbét, amelyen az origó az üzembe helyezés pillanatát jelenti. Hangsúlyoznunk kell, hogy t_0, a koordináta rendszer kezdőpontja az üzembe helyezés pillanata. Ez azt jelenti, hogy a tervezési, fejlesztési, építési és szerelési munkákhoz negatív idő tartozik, aminek következtében a negatív időben kiadott összeg nagyobb értékkel jelentkezik a jelenértékben.



1.12.1. ábra. A jelenérték hatása a kifizetésekre (bevételre)

A beruházás Ki jelenértéke, ha ni kapacitást valósítunk meg a ti időpontban, az alábbiak szerint írható fel:

$$K_i = \frac{C_0 + C_{n_i} n_i}{(1+r)^{t_i}} + \left[S_0 + n_i S_{n_i}\right] \frac{(1+r)^{T-t_i} - 1}{(1+r)^T \cdot r} - \frac{M}{(1+r)^T} \quad (1.12.1.)$$

ahol $\frac{1}{(1+r)^{t_i}}$ { > 1 if $t_i < 0$; = 1 if $t_i = 0$; < 1 if $t_i > 0$

A képletben szereplő M a maradék érték, melynek értéke attól függ, hogy a tervezési periódus (T) végén hogyan lehet értékesíteni vagy újra felhasználni a beszerzett eszközöket. Ki a t=ti időpontban beruházott eszközök jelenértéke, C0 a beruházások kapacitástól független része, Cn a létesített kapacitással arányos beruházási egység költség. S0 és Sn az üzemeltetési költségek hasonló két összetevőjét jelölik. A tervezési időtávlat (T) nevezhető névleges élettartamnak, mely a gyakorlatban az eszközöktől függően pl. 5, 10 vagy akár 20 év. A távközlési kamatláb (r) nagyobb mint a bankkamat, ugyanis kompenzálni kell a befektetés nagyobb kockázatát, a rövidebb elavulási időt és azt, hogy a létesített rendszerek kezdetben nincsenek teljesen kihasználva. Így az rtávközlés 10-25 % is lehet, amikor a bank kamatláb mindössze 4 %. Általában egy beruházást több lépésben valósítunk meg, és így a teljes időtartamra vonatkozó jelenérték az egyes létesítések jelenértékeinek az összege, azaz:

$$K = \sum_i K_i \quad (1.12.2.)$$

1.12.2. Üzletközpontú optimalizálás

A vállalat tulajdonosainak célja, hogy maximális osztalékot érjenek el. Természetesen kérdés, hogy az évenként megjelenő osztalékot D(t) hány éves távlatban összegezve törekszenek maximalizálni. Feltételezve, hogy minden évben más osztalékot tudnak elérni, D(t) az idő függvénye. Ezeknek a meghatározott időre vonatkoztatott jelenértékeknek az összege (folytonos esetben az integrálja) a teljes nyereség, amely folytonos esetben az alábbi formában írható fel:

$$D = \int_0^T D(t) \cdot \frac{1}{(1+\alpha)^t} dt \quad (1.12.3.)$$

Az osztalék minden évben arányos a *(B)* bevétel és a *(K)* költség különbségével. Ennek értékét valamilyen pénzegységben fejezzük ki. A különbségnek csak egy adott részét lehet osztalékként szétosztani, mert tartalékot is

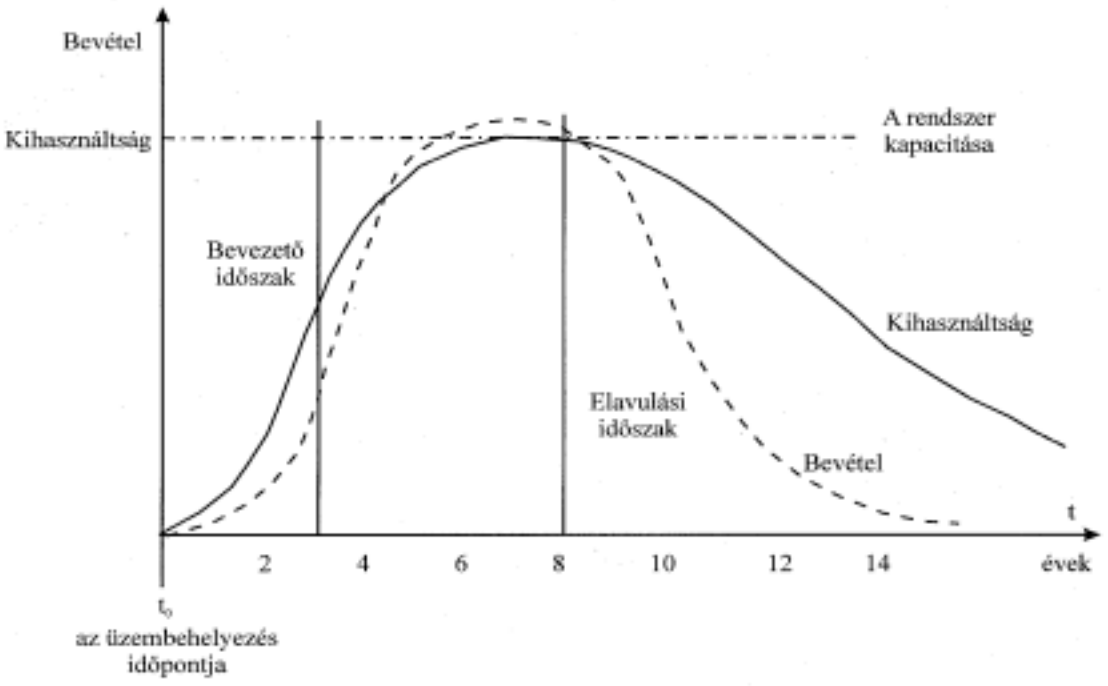
képeznek, fejlesztésre, kutatásra is kell pénzt fordítani, és az elavult eszközök pótlására is kell alapot képezni. Azaz az osztalék:

$$D(t) = \beta[B-K] \qquad (1.12.4.)$$

A tulajdonosok döntésén múlik, hogy a teljes különbség hányadrészét (β) osztják szét. Ezt műszaki, gazdasági és politikai tényezők is befolyásolják a vállalati stratégián keresztül. A teljes bevétel az infrastruktúra kihasználtságától és a tarifáktól függ. Egy új szolgáltatás indításánál a kihasználtság általában alacsony. A bevétel csak 3-5 év múlva éri el a tervezett értéket (1.12.2. ábra). Amikor a forgalom megközelíti a tervezett értéket, akkor sok esetben előnyös lehet a tarifát csökkenteni, mert ezzel az együttes bevétel növekedhet. A tarifa meghatározásában szerepet játszik természetesen a minőség, a használhatóság, a garantált sávszélesség és a szolgáltató rugalmassága.

A hálózat méretezését két tulajdonosi döntés befolyásolja:

- az első, hogy milyen távlatban kívánja a nyereségét maximálni, vagy más megfogalmazásban mennyi legyen az osztalék vagy ß (1.12.4.)
- a másik a tarifa meghatározása. Ez függ a versenytársak stratégiájától (1.12.3. fejezet) és a tulajdonosok üzletpolitikájától. Az olcsóbb tarifa növelheti a jövőben a kihasználtságot amikor már újabb, modernebb rendszerek



1.12.2. ábra. Gazdaságos élettartam jellemző paraméterei

jelennek meg (1.12.2. ábra).

E tényezők figyelembevételével kialakul, hogy mely hálózati részt milyen kapacitásra kívánják tervezni. Az adott mennyiségi és minőségi követelményekhez illeszkedően optimalizáljuk a hálózat struktúráját, határozzuk meg a csomópontok helyét, és a forgalomirányítás módszerét [1.12.1., 3., és 7.].

Ha megvan a keresett optimum, akkor megvalósítjuk a hálózatot, melynek első lépése a fizikai nyomvonalak meghatározása. Ennek során figyelembe vesszük a földrajzi körülményeket, a szükséges használhatóságot, az ennek érdekében megvalósítandó tartalék áramköröket és átviteli kapacitásokat. Végül megtervezendő az üzemeltetés-fenntartási rendszer [1.12.1÷7.].

Az üzletközpontúság az alábbi 10 pont figyelembe vételét jelenti. Ezek befolyásolják a tarifát, a beruházás nagyságát és a megvalósítandó szolgáltatásokat.

- Új technológiák bevezetése: csomagkapcsolás, demokratikus (egyenlő jogú pontokat tartalmazó, igény szerinti irányítású) hálózat
- Különböző jellegű információk: a beszéd már nem meghatározó, mert az adat és képátvitel, valamint a felhasználó technológiájába beépülő távközlés mennyisége és jelentősége növekszik.
- Hálózatszervezés: az új technológiának megfelelően a rugalmasabb, demokratikus hálózat veszi át a hierarchikus struktúra helyét.
- Átviteli utak: az új közegek átviteli kapacitása szinte korlátlan.
- Távszolgálatok: az OSI felső 4 rétegének a szerepe megnövekedett.
- Verseny: számos szolgáltató és hálózat üzemeltető küzd a felhasználókért és a hálózati kapacitásokért.
- Szabályozás: itt a változás nem a konkrét részletekben, hanem a szabályozás koncepciójában mutatkozik.
- Költségtényezők: az eszközök árában a szoftvernek van meghatározó szerepe.
- Rövidebb elavulási idők: a gyors amortizáció miatt nagy kihasználtságra van szükség.
- Nyereséghatárok: egyrészt a verseny megszabja a bevétel felső korlátját, ugyanakkor a gyors megtérülés magasabb árakat követel.

A változásokat és azok hatását tanulmányozva egyértelműnek látszik, hogy a bitek átvitele egyszerű feladat, de már nem elegendően jó üzlet. A távközlési vállalatoknak újabb szolgáltatásokat kell kínálni, kielégítve a felhasználó egyedi igényeit. Az üzleti célok tehát újra definiálandók. Ennek során megnő a használhatóság és a személyre szabott távközlés fontossága.

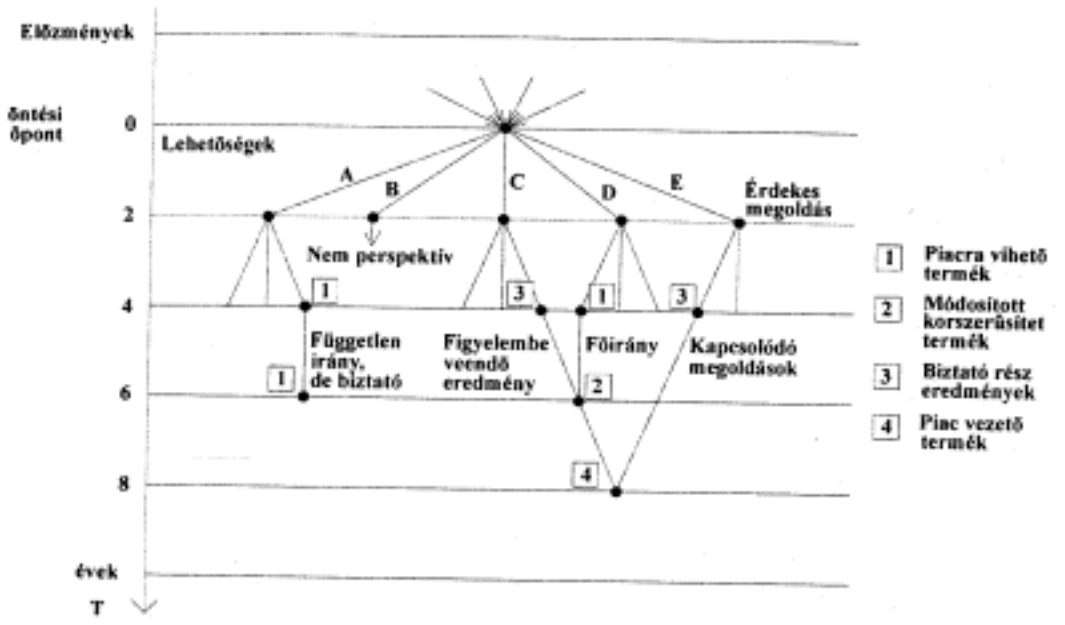
1.12.3. Versenykörnyezet

A gazdasági megfontolásokat nem csak a felhasználók szabják meg, hanem ennél általánosabb korlát a nemzeti jövedelem. Meghatározó ezen kívül a versenytársak stratégiája és az érvényes törvények által szabott korlátok. Egy általános közelítő irányelv volt, hogy a családok jövedelmének maximum 5 %-át tudják távközlésre fordítani. Ez ma is érvényes, de csak a klasszikus távközlési szolgálatokra, mert új lehetőségekért a használók hajlandók többet fizetni.

A felhasználók döntését befolyásolja a versenytársak kínálata és a szolgáltatások minősége, megbízhatósága. Az értékelésnél figyelembe veszik a létesítés gyorsaságát, az árakat és az új szolgálatok használatának érthetőségét. A különböző tényezőket, melyek a versenyre kihatással vannak, és a különböző szereplők kapcsolatát az 1.12.3. ábrán mutatjuk be.

A versenyben két különböző célt lehet kitűzni, az első, hogy növeljük a bevételünket, a második, hogy legyőzzük a versenytársakat. Ez utóbbit azonban nem célszerű olyan áron elérni, hogy saját vállalatunk veszteséges lesz. A szociális, gazdasági és műszaki körülmények vizsgálatával lehet olyan stratégiát kialakítani, ami mind a két cél elérését segíti. Mivel a körülmények változnak és a nyereség hosszú időre vett együttes jelenértéke a lényeges, ezért időről időre újra kell gondolni a stratégiát (1.12.4. ábra). Ebből látszik, hogy olyan stratégiát érdemes választani, melynek megváltoztatása az idők folyamán gyökeres átalakítás nélkül lehetséges. A tervezési periódus végéig előre tekintve kell pillanatnyi döntéseinket meghozni, gondolva arra, hogy a lehetséges változások hatását követni tudjuk.

1.12.3. ábra



1.12.4. ábra Döntési folyamat

A verseny igen előnyös, mert a vállalatok kényszerítve vannak a kreatív gondolkodásra, és arra, hogy többet fordítsanak kutatásra, fejlesztésre. A fejlődés elősegítése érdekében a kormányok mindent elkövetnek, hogy a távközlési piacon ne csak egy szolgáltató legyen. Számos szolgáltató és hálózatüzemeltető kiegyensúlyozott versenye segíti az ország távközlésének és informatikai képességének fejlődését. A verseny során követendő stratégiák a játékelmélet eszközeivel határozhatók meg.

1.12.4. Játékelmélet

Általános célkitűzés, hogy a résztvevők a piaci versenyben a maximális profitot érjék el. A játékelmélet terminológiáját használva a kifizető függvényt kívánják maximálni. Ha mindössze csak két játékos van, akkor azt duopol piaci modellnek nevezzük. Általában azonban többen vannak a piacon. A számítások és a modellezés egyszerűsítése érdekében a kisebb jelentőségű versenytársakat elhanyagolhatjuk, vagy a hasonlókat összevonhatjuk egyetlen versenytárssá.

A versenykörnyezetben a résztvevő vállalatok azt szeretnék elérni, hogy minél több felhasználójuk legyen, és minél nagyobb forgalom átvitelére kapjanak megbízást. Ez egyben azt is jelenti, hogy valamennyi távközlési vállalatnak célja, hogy növeljék a távközlési kedvet, és különböző új szolgáltatásokat tudjanak

ajánlani. A közös célok elérése érdekében szükséges lehet a kooperáció, amellyel a teljes piac távközlési üzleti forgalmának (és ezzel a saját bevételeiknek) növelését akarják elérni.

A távközlés területén különösen hasznosnak mutatkozik a játékelmélet, ugyanis korlátozott számú vállalat lehet jelen csak a piacon, és számos külső tényező befolyásolja a versenyt. A bizonytalanságok figyelembe vételét is lehetővé teszi a játékelmélet. Ez azt jelenti, hogy pl. beruházási döntések előkészítésénél a versenytársak várható viselkedését is figyelembe tudjuk venni. Vannak módszerek melyek figyelembe tudnak venni korlátozott mértékű együttműködést, vagy információcserét a versenytársak között, és ehhez illesztik a javasolt stratégiát.

Hangsúlyoznunk kell, hogy a játékelmélet nem csak egy analitikus számításon alapuló eredményt ad, hanem valamennyi versenytárs optimális stratégiáját is megkaphatjuk. Ez egyben azt jelenti, hogy feltételezi minden versenyzőről az értelmes célratörő játékot. A kölcsönösen feltételezett logikus optimumra törekvés az alapelve a játékelmélet módszerével megadható stratégiáknak.

A kifizető függvény (P) kifejezhető pl. mint a B-K (1.9.3.) jelenértéke. Különböző modellek léteznek a játék kimenetelének vizsgálatára. Az egyes modellek abban különböznek egymástól, hogy mire irányul a stratégia választás, az árra, a hálózat méretére, a reklámra költött összegre, stb.

A játék általános jellemzői

- az egymással kölcsönhatásban lévő játékosok
- a szabályok, amelyek meghatározzák a játékosok számára a megengedett lépéseket
- a stratégiák, melyek különböző döntésekre vezethetnek a verseny során, attól függően, hogy a körülményeket és a célokat (a hálózat mérete, tarifapolitika, reklám és a vállalat hagyományai) hogyan választjuk meg, vagy milyen módon súlyozzuk azokat
- a kifizető függvény, amely lehet a szokásos jelenérték, a készpénzforgalom, a nyereség, az osztalék stb.
- a játék leírásának és megoldásának elvei, melyekre a következőkben részletesebben kitérünk.

A játékok leírásának két legismertebb formája a táblázatos és a fa-struktúrával történő megadási mód. A táblázatos formában leírható játékokat normál formájú, a fa-struktúrával leírhatóakat extenzív formájú játékoknak nevezik. A táblázatos formát

általában az egylépéses (statikus) játékoknál alkalmazzák, az extenzív forma a többlépéses (dinamikus) játékok megadására használatos. A játékok megoldása során olyan stratégia együttesek keresése a cél, amelyek esetén valamilyen egyensúly áll fenn. A Nash egyensúly-pont ϑ az, amelynek az eredményét egyoldalúan egyik játékos sem kívánja megváltoztatni, mert a változás hatása számára biztosan kedvezőtlen lenne. A fa struktúrával megadott játékok esetén az egyensúlyi stratégiát pl. úgy is megkaphatjuk, hogy aljátékonként (visszafelé irányban) keressük a lokálisan legjobb stratégiákat. Az egyensúlypont meghatározásához a szokásos optimalizációs apparátuson túlmenően speciális játékmegoldó szoftverek is rendelkezésre állnak (pl. Gambit) [1.12.10.].

A játékok kimenetelének elemzésekor kereshetünk pl. domináns stratégiákat is. Ez egy játékos esetén az a stratégia, amely a többiek bármilyen stratégia választása esetén a legjobb stratégia a szóbanforgó játékos számára.

Minden távközlési modell esetén a gyakorlati felső korlát a fizetőképes kereslet. Ebből következik, hogy a versenyben részt vevő valamennyi játékos számára rendkívül értékes, ha olyan új szolgáltatásokat tud bevezetni, melyek vonzóak a felhasználók számára.

Amikor már minden játékos eldöntötte, hogy milyen stratégiát követ, akkor lényegében eljutottunk az eredményhez. A következő két pontban gyakorlati példákat mutatunk be.

1.12.5. A játékelmélet alkalmazása (1. példa)

Az egyszerűség kedvéért csak 2 játékos legyen (duopol eset). Mind a két vállalat bérelt vonali összeköttetéseken üzemelő ATM hálózatot akar építeni a meglévő SDH áramkörök felhasználásával. Ebben az esetben mind a kettő a Nash egyensúlyi pontot keresi, a másik cég stratégiáját figyelve számítja ki, mekkora hálózatot kell létesítenie ahhoz, hogy maximális legyen a nyeresége. A feladat tehát egy mennyiség meghatározása, amelyet x-szel jelölünk. Ez a hálózat méretét jellemző érték, jelen példában legyen ez az ATM összeköttetések sávszélességeinek összege.

Tételezzük fel, hogy mindkét vállalatnál a beruházási költségek lineáris függvénnyel írhatók le. A kapacitás független részek egyenlők, de a kapacitással

arányos költségek a két vállalatnál különbözőek. Azaz, a méret-gazdaságosság elvének megfelelően, aki nagyobb méretben gondolkodik, annak az egységköltsége kisebb. Feltételezzük, hogy a második cég nagyobb ATM hálózatot akar létesíteni, és ezért a költségarányos összeg kisebb lehet. A példában a működési és fenntartási költségek értékét az egyszerűség kedvéért beolvasztjuk a beruházásba. A kapacitás független költségek azonosak (32).

$$C1(x)=0.08x+32$$

$$C2(x)=0.04x+32$$

Ezzel a két egyenlettel minden felmerülő költséget lefedtünk, az energiafelhasználástól kezdve a szabadalmi díjakig, beleértve a hirdetés, az oktatás és a bérjellegű költségeket is. Tételezzük fel, hogy az inverz keresleti függvény (egységár) a következő alakban adható meg: $p(x1,x2)=0.75-0.00005(x1+x2)$. Az összefüggés mutatja, hogy a mennyiség és az ár fordított arányban áll egymással. Nagyobb mennyiséget csak kisebb áron tudunk értékesíteni. A vállalatok profitját a jelenérték számítás során használatos T időtartamra kívánjuk maximálni.

Az optimalizálási probléma megoldását az alábbi 1.12.2. táblázaton követhetjük, melyből a két vállalat stratégiáját és kifizető függvényét olvashatjuk le. A felső sor mutatja azt a mennyiséget, amelyet a második vállalat létesítene. A függőleges oszlop pedig az egyes vállalat beruházásának mértékét jelzi. A keresztpontokban látjuk a kifizető függvények értékét a különböző megoldásoknál. Minden keresztpontban az első szám az 1. vállalat, a második szám a 2. vállalat adott beruházásához tartozó kifizető függvénye. Láthatjuk, hogy a 3200/4400-as pont jelenti a Nash egyensúlyt, melynél a két kifizető függvény értéke 210 illetve 666. Ennek alapján a kisebbik 3200, a nagyobbik pedig 4400 a fentiek szerinti kapacitású hálózatot épít, és ezzel eljut a közös optimumba.

Ugyanezt bemutathatjuk egy egyszerű grafikus eljárással is. A Nash

	2800	3600	4400	5200	6000
2000	298 706	218 850	138 930	58 946	-22 898
2400	370 650	274 778	178 842	82 842	-14 748
2800	426 594	314 706	202 754	90 738	-22 598
3200	466 538	338 634	210 666	82 634	-46 448
3600	490 482	346 562	202 578	58 530	-86 298
4000	498 426	338 498	178 490	18 426	-142 298

1.12.2. táblázat

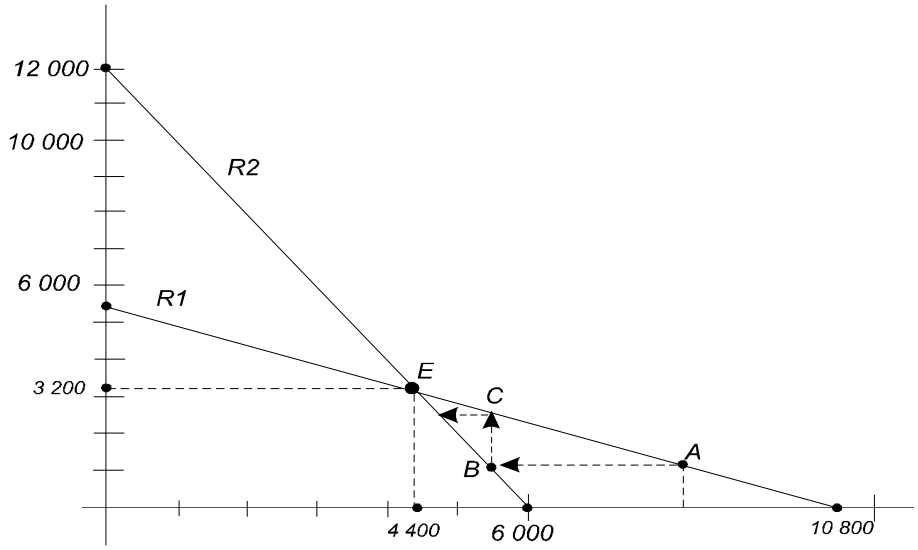
egyensúlyi pontot megtaláljuk, ha felrajzoljuk mindkét vállalatra vonatkozóan a reakció függvényeket, és ahol ez a két egyenes metszi egymást, ott találjuk a Nash egyensúly pontját. (A reakció függvény az egyik cég legjobb stratégiájához tartozó kifizető függvényét mutatja a másik cég stratégiáinak függvényében.) A görbék a kifizető függvények differenciál hányadosai alapján rajzolhatók fel.

Erre egy példa az 1.12.5. ábrán látható, ahol A, B, C mutatja az utat, ahogy eljutunk az E egyensúlyi pontba. Kiindulva az A pontból, a kettes jelű vállalat csökkenti a mennyiséget, hogy közeledjen az egyeneshez. Így eljut a B ponthoz. Ugyanakkor az egyes jelű vállalat elmozdul az A-ból és halad a C felé, és így eljutnak az E egyensúlyi pontba. A példabeli két reakció görbe egyenlete a következő:

$$R1: x1=-0.5x2+5400$$

$$R2: x2=-0.5x1+6000$$

A példával kapcsolatos további részletek [1.12.9.]-ben megtalálhatók.



1.12.5. ábra. A játék megoldása a reakciógörbék segítségével

1.12.6. A játékelmélet alkalmazása (2. példa)

Ebben a példában is két résztvevős esetet vizsgálunk. Feltételezzük, hogy x és y is kínál távközlési szolgáltatásokat. A szabályozó hatóság M értékben rögzítette az egységnyi forgalomhoz, és egységnyi minőséghez rendelhető maximális tarifát.

Mindkét szolgáltató egyöntetűen megállapította, hogy m értéknél alacsonyabb tarifával nem éri meg szolgáltatni. Így mindkét vállalat részére felírhatjuk a tarifa-meghatározás tartományát:

$$m \le bx \le M \text{ és } m \le by \le M$$

A hálózatok kihasználtsága, vagyis a forgalom csökken, ha növekszik a tarifa. Jelölje az A függvény a forgalomnak a két cég tarifájától (bx, by) való függését az alábbiak szerint:

$$A\left(\frac{b_x + b_y}{2}\right) \rangle 0$$

A forgalom megoszlás a két vállalat között függvénye az árnak. Feltételezzük, hogy a forgalom a következő arányban oszlik meg:

$$Px = f(bx - by),\ Py = 1 - f(bx - by),$$

Egyértelmű, hogy ha bx=by akkor a kereslet a két szolgáltató iránt azonos, vagyis mind a kettő 50 %-ban részesül a forgalomból is. Ha eltérőek a tarifák, akkor az alábbiaknak megfelelően határozhatjuk meg a x és y kifizető függvényét:

$$K_x = b_x f\left(b_x - b_y\right) A\left(\frac{b_x + b_y}{2}\right) \qquad 1.12.5\ldots$$

$$K_y = b_y \left[1 - f\left(b_x - b_y\right)\right] A\left(\frac{b_x + b_y}{2}\right) \qquad 1.12.6\ldots$$

Feltételezzük, hogy m=1 és M=4, és a forgalom az alábbiak szerint függ a tarifáktól:

$$A\left(\frac{b_x + b_y}{2}\right) = 20 - \left(b_x + b_y\right)$$

Feltételezzük továbbá, hogy a forgalom eloszlás az alábbi összefüggéssel határozható meg: A következőben mind x, mind y esetére kiszámoljuk a bx=1,2,3,4, és by=1,2,3,4 esetre a kifizető függvények értékeit. Az eredményt az alábbi mátrix mutatja külön-külön a két cégre vonatkozóan. (Valamennyi oszlopban elől x-re és mögötte y-ra vonatkozó kifizető fv értéke olvasható.)

Részletesebben megvizsgálva ezeket az értékeket látjuk, hogy a 9, 16 és 21 az egyensúlyi pontok. Látszólag a 24 is kielégíti ezt a feltételt, azonban ez a pont

	b_y	1	2	3	4
b_x					
1		9 9	$11\frac{1}{3}$ $5\frac{2}{3}$	$13\frac{1}{3}$ 8	15 0
2		$5\frac{2}{3}$ $11\frac{1}{3}$	16 16	20 15	$23\frac{1}{3}$ $9\frac{1}{3}$
3		26 $17\frac{1}{3}$	15 20	21 21	26 $17\frac{1}{3}$
4		0 15	$9\frac{1}{3}$ $23\frac{1}{3}$	$17\frac{1}{3}$ 26	24 24

labilis, mert akár x, akár y el tud mozdulni úgy, hogy ennél nagyobb (26-os értéket) érjen el. Ez viszont együtt jár azzal, hogy a másik versenyző fél lényegesen kisebb értéket tud csak elérni, tehát ebben a pontban nem stabil a rendszer.

Más esetekben az ár mellett a minőség is befolyásolja a két szolgáltató közötti forgalmi arány eloszlását. Ugyancsak hatással van a megoszlásra a cégek közönségkapcsolati munkája.

Számos más eset is vizsgálható, azonban a további részletek helyett csak az irodalomra hivatkozunk [1.12.8., 1.12.11., 1.12.13., 1.12.17], amiből látszik, hogy a különböző szituációk esetén gyakran megtalálható az a játékelméleti módszer, ami a legkedvezőbb stratégia kidolgozását elősegíti. Fontos azonban, hogy a lényeges befolyásoló tényezőket megfelelően súlyozva tudjuk a számításba beépíteni, és valamennyi tényezőt azonos egységre, például pénzre tudjuk transzformálni. Helyes alapadatokból kiindulva, alkalmas játékelméleti modell felállításával a feltételeknek megfelelő gazdasági optimumot ϑ megtalálhatjuk.

1.12.7. Kockázatkezelés

Minden gazdasági döntésnek – vonatkozzon az új beruházásokra, létesítmények megvalósítására, szolgáltatások bevezetésére vagy bármilyen eladási, vételi, kölcsönzési műveletre, pénzügyi tranzakcióra – van valamilyen kockázata. A siker ugyanis számos olyan befolyásoló tényezőtől függ, melynek hatását nem lehet pontosan előrelátni.

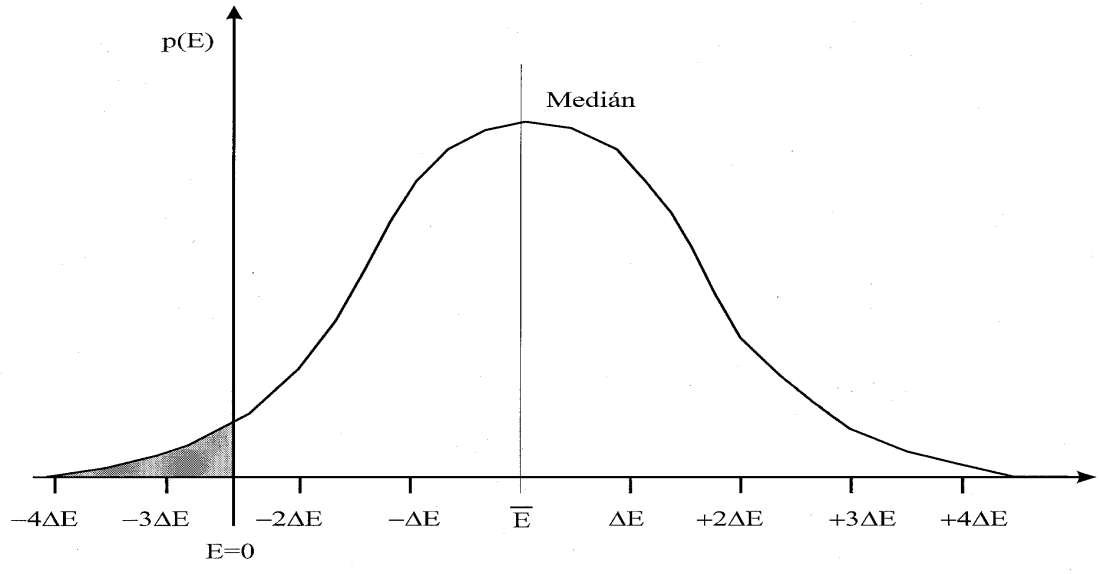
A döntés-előkészítés folyamatában szükség van néhány olyan számítás elvégzésére, amely a legvalószínűbb esetet veszi alapul. A döntés eredményessége egy bizonyos T (3-10 év) időtartam alatt a várható bevételek jelenértékének (B) és az ehhez kapcsolódó kiadások jelenértékének (K) a különbségével (az 1.12.1. alpontban már megfogalmazott módon) az alábbi összefüggés alapján becsülhető:

$$\overline{E}=\overline{B}-\overline{K}$$

A fenti összefüggésben szereplő mennyiségek azonban meglehetősen bizonytalanok. B, K és így E valószínűségi változók. Ezek várható értékét számos olyan tényező is befolyásolja, melyek függetlenek azoktól, akik a döntést előkészítették vagy az elvárt feladatot végrehajtják. Kedvezőbb a helyzet, ha a döntést az E változó sűrűség függvényének ismeretében hozzuk meg. Vagyis keressük a p(E) sűrűségfüggvényt, melynek egy lehetséges alakja az 1.12.6. ábrán látható.

A görbe megkonstruálásához a releváns kockázati tényezőket figyelembe kell venni. Ezek közül az alábbiak jelentik a leggyakoribb kockázatokat:

- a versenytársak stratégiája és sikere
- a szabályozó hatóság korlátozásai (pl. a tarifát illetően)
- a felhasználók viselkedése, tanultsága, kultúrája és a távközlésre fordítható pénzösszege



1.12.6. ábra

- a megfelelő hálózati struktúra és technológia választása, a rendszer használhatósága
- az ország gazdasági helyzetének esetleges változása

A sűrűség függvény ismeretében meghatározható annak valószínűsége, hogy az E értéke negatív. Ez az érték az eloszlásfüggvény “0” helyen felvett értéke, jelölje ezt a P(E=0) kifejezés. Ennek értékét a sűrűség függvény felhasználásával az alábbi integrál adja:

$$P(E=0)=\int_{-\infty}^{0} p(E)dE$$

Ha a P(E=0) valószínűség nagyobb egy előre rögzített ε értéknél, akkor a projekt megvalósítása nem gazdaságos. Ekkor vagy az egész projektet el kell vetni, vagy a kivitelezőknek olyan új tervet kell kidolgozni, melynél a veszteséges kimenetel valószínűsége az előírt határon belül marad.

A kockázat kezelés általában négy fontos lépést tartalmaz:

a) A kockázati tényezőket az adott esetre meg kell határozni. Ehhez lényeges mind a gazdasági, mind a jogi környezet vizsgálata. A piaci versenyt befolyásoló kockázati tényezők az 1.12.3. ábrán láthatók. Minden egyes kockázati tényezőt jellemezni kell a valószínűségi eloszlás (vagy sűrűség) függvényével, figyelembe véve ennek időbeli változását is. Így az adott kockázati tényezőre vonatkozóan egy p(ri, t) sűrűség függvényt kell meghatározni. Ha múltbeli adatokkal rendelkezünk, akkor a matematikai statisztika módszereivel a sűrűség függvény alakja és paraméterei becsülhetők. Más esetben szubjektív szakértői becslések szükségesek. A későbbiekben megvizsgáljuk, hogy a szakértői vélemények alapján becsült sűrűség függvény pontatlansága milyen módszerekkel csökkenthető.

b) Meg kell vizsgálni, hogy a különböző kockázati tényezőknek milyen hatása van a teljes projektre. Azaz, meg kell keresni az összefüggést E változása és az ri különböző lehetséges értékei között. Az alábbi összefüggés a projekt érzékenységét mutatja az adott tényezőre vonatkozóan.

$$S_i=\frac{\partial E}{\partial r_i}$$

Valamennyi kockázati tényezőre megkeressük az Si értékét. Ahol lehet, igyekszünk objektív összefüggések (pl. múltbeli adatok) segítségével meghatározni az érzékenységet, ahol erre nincs lehetőség, ott szakértői vélemények vizsgálatával becsüljük az érzékenységet. Ezt követően sorba rakjuk az Si értékeket, és ahol ez a legnagyobb ott igyekszünk az ri hatását csökkenteni, befolyásolni. Végül előállítjuk p(E)-t. Sokszor a becsült értékeket tartományokra osztjuk és numerikus módszereket

alkalmazunk a sűrűség függvény becslésére, máskor pedig a becslést Monte Carlo szimulációval végezzük el.

c) Miután sikerült az eloszlás függvény veszteséghez tartozó értékét az előírt korlátok alá csökkenteni, megkísérelhetjük a szórás értékét is csökkenteni annak érdekében, hogy a medián, vagyis a legvalószínűbb eset nagy valószínűséggel forduljon elő. A módszer általában az érzékenység további vizsgálata és ennek befolyásolása oly módon, hogy pl. több szállítót, alternatív kivitelezési módszereket, nagyobb biztonságot adó pénzügyi hátteret vagy jobb piaci előkészítést rendelünk a projekthez. Ilyen lehetőség a kockázatok áthárítása, vagy megosztása, úgy, hogy biztosító társaságokat vonunk be. Egy másik lehetőség, hogy a tervezés során további, menetközben elvégezhető döntési lehetőségeket építünk be a folyamatba, melyek az aktuális helyzet függvényében lehetővé teszik újabb megoldások kiválasztását (1.12.4. ábra).

d) Mint a korábbiakban említettük, az ri kockázati tényezők nem determinisztikus értékek. Ezek eloszlásának becslésére szakértői csoportokat szoktak összehívni, esetleg brainstorming-okat rendeznek. Nagyobb csoportok nagyobb biztonságot adnak, és lehetővé teszik pl. a nem konzisztens vélemények kiszűrését. Az előrelátás megbízhatósága növelhető, ha a vélemények összhangjának vizsgálatát is elvégezzük.

A **konzekvencia vizsgálat** során minden kérdést több különböző formában kérünk megválaszolni. Erre vonatkozóan a legismertebb a Guilford módszer [1.12.18], melynél az alábbi négyféle módon kérjük a szakértők válaszát:

Legyen például 5 lehetséges válasz, ill. megvalósítási módja a projektnek. Ezt az öt választ tíz különböző párba állíthatjuk, és minden párra vonatkozóan meg kell mondani, hogy a szakértő melyiket részesítené előnyben.

Ezt követően (kis szünet után) ugyanezen szakértőknek az öt megvalósítási lehetőséget a szakértőnek sorrendbe kell rakniuk úgy, hogy az első legyen az, ami szerinte a legkedvezőbb, az utolsó pedig legyen az, amelyiket elvetni javasol.

A harmadik módszer az, hogy a szakértő mind az öt lehetőségre megadja a siker valószínűségét százalékban, vagyis mindegyikhez hozzárendel egy számot 0 és 100 között.

Végül a következő eljárást kell elvégezniük a szakértőknek. Először el kell vetniük az általuk legrosszabbnak ítélt megoldást, vagyis amelynek a megvalósítását ellenzik. Ezt követően megadják, hogy melyik az az egy vagy két megoldás, amit

egyértelműen támogatnak. Végül a maradék lehetőségekre vonatkozóan nyilatkoznak, hogy azokat egyenértékűnek tartják, vagy azok is sorba állíthatók.

Ha a négyféle módszer során egy adott szakértő esetén az eredmények mindig ugyanazt a preferencia sorrendet adják, akkor ennek a szakembernek a konzekvencia értéke 1. Ha minden esetben különböző megoldásokat választott, akkor a konzekvencia tényezője 0. A következő vizsgálatsorozatból a 0 konzekvenciájú szakértőket ki kell rekeszteni, mert döntéseik nem megalapozottak, inkább csak blöfföltek.

A **konkordancia vizsgálat** során azt kell meghatározni, hogy a különböző szakértők mennyire azonos preferencia sorrendet állapítottak meg. Ha valamennyiüknél ez azonos volt, akkor a csoport konkordanciája 1. Ha minden szakértőnél minden pozícióban más projekt szerepel, akkor ennek a csoportnak 0 minősítést adunk. Mivel nem lehet megállapítani, hogy mely szakértők véleménye volt szakszerűtlen, a csoport egyetlen tagját sem kérjük fel a vizsgálat elvégzésére. Számszerűsítve a konkordancia értéke:

$$Konkordancia = \frac{N_s}{N}$$

ahol N az összes döntési pontok száma, azaz a résztvevő személyek számának és a lehetséges megvalósítások számának szorzata, Ns pedig azon döntési pontok száma, ahol legalább a szakértők 67 %-a azonos véleményen volt. Ez azt jelenti, hogy döntő mértékben azonos preferenciát állapítottak meg a vizsgált kérdésben. Nem remélhető, hogy N=Ns, de ha az egyetértést bizonyító pontok száma 50 %-nál kisebb, akkor ezt a megoldást el kell vetni. Ilyenkor új csapatot kell összehívni, melynek egyetlen tagja sem lehet azonos a korábban felkért szakemberekkel.

A kockázatelemzés és a kockázat csökkentésének folyamata idő és költségigényes. Ezért csak nagy volumenű projektnél érdemes ezt elvégezni, ahol az elemzésre szánt költségek alacsonyabbak, mint a megvalósítás esetleges vesztesége, vagy a várt eredménytől való elmaradás [1.12.15-18.].

Irodalomjegyzék

[1.12.1.] Morgan, T.J.: Telecommunication Economics Mc Donald, London 1958.

[1.12.2.] Rehbein, G = Grunclagen der Ökonomik des Post und Fernmeldewesen. Informationheft der IPF(31) 1960.

[1.12.3.] Lajtha Gy.: Távközlő hálózatok elmélete és tervezése.

(Műszaki Könyvkiadó, Budapest, 1971.)

[1.12.4.] CCITT GAS-3 bizottság: Economic and technical aspects of the choice of transmission systems. (ITU, Genf 1968.)

[1.12.5.] CCITT GAS-5 bizottság: Telecommunication economic studies (ITU, Genf, 1976.)

[1.12.6.] Borsos K. – Lajtha Gy.: Finding the economic optimum in planning telecommunication network. (Budavox Telecomm. Rev. 1971. No. 3. 9-27.)

[1.12.7.] Dr. Sallai Gyula: Távközlő hálózatok tervezésének gazdasági számításai. Budapest, Közdok 1979.

[1.12.8.] Konkoly, R; Fekete, I; Gyürke, A.: Evalvation of uncertainties in Invesment Projects, Third European Workshop on Techno-economics for Multimedia Networks and Services, Aveiro, Portugal, 1999.

[1.12.9.] Konkoly Lászlóné, Gyürke Attila:A játékelmélet alkalmazási lehetőségei a távközlésben, PKI Közlemények, 44.szám, Távközlési kiadó 2000.

[1.12.10.] McKelvey, R. D.: Gambit: An Interactive Extensive Form Game Program, California Institute of Technology, 1997

http://hss.caltech.edu/~gambit/Gambit.html

[1.12.11.] Harsányi, J.: Game with Incomplete Information Played by Bayesian Players-Part II. Management Science, 14, 320…334, 1968.

[1.12.12.] Von Neumann, J. – Morgenstern, O.: Theory of Games and Economic Behaviour, Princeton University Press, 1947.

[1.12.13.] Szép J. – Forgó F.: Bevezetés a játékelméletbe. Közgazdasági és Jogi Könyvkiadó, Budapest, 1972.

[1.12.14.] Leitmann, G.: Cooperative and Non-Cooperative Many-Player Differential Games. CISM Monograph No. 190, Springer Verlag, Vienna, 1974.

[1.12.15.] Dr. K. Géher: Theory of Network Tolerances.

Akadémiai Kiadó, Budapest 1971.

[1.12.16.] Dr. Erdősi Gy.: Innovációs menedzsment

Távközlési Könyvkiadó Budapest 1992.

[1.12.17.] Delbeco A.L. – Van de Ven A.H. – Gust of son D. H.: Group techniques for program planning Scott, Foresman and Company, Glenview, Illionis 1975.

[1.12.18.] Van de Ven A.H.: Group decision-making effectiveness. Kent State University Center for Business and Economic Research Press 1974.

1.13. Hálózat- és szolgáltatásminőség (QoS=Quality of Service)

Szerző: Nándorfi Gyuláné

Lektor: Kesselyák Péter

1.13.1. Minőségügyi trendek (a 21. szd.-ban) [1.13.1]

A minőség filozófiai kategória. A tárgyak lényegi meghatározottsága. A tárgynak, mint egésznek a jellemzője. A minőség már Arisztotelész tíz kategóriájának is egyike. Kantnál pedig a kategóriánál is magasabb szintű fogalom. Tehát a filozófia és a minőségtudomány szoros kapcsolatban vannak.

A minőség értelmezése és jelentősége a piac jellegének változása függvényében alapvetően változott az elmúlt fél évszázadban. A monopóliumok korában a minőség fogalma nagyjából változatlan volt és jelentősége elmaradt a mennyiségi fejlődés és az ár mögött. A liberalizált piacon paradigmaváltás történt. A termékek és szolgáltatások versenypiacán a minőség fogalma alkalmazásfüggő lett és az árral együtt versenyt eldöntő tényező.

A nemzetközi fejlődésnek a minőségügyet is érintő várható fő területei

- az új vállalati modellek,
- az új technikák és
- a piac irányítása a vevőorientált minőség által.

Az új technikák elsősorban a hálózatminőségre hatnak, a vevőorientált minőség pedig főleg a szolgáltatásminőséggel van kapcsolatban.

Fontos, a minőségügy fejlődését befolyásoló tényező, még a gazdasági szempontok integrálása a pénzügyi- és elszámolási rendszerekbe.

1.13.2. Hálózat- és szolgáltatásminőség fogalma

A minőség egyik legfrissebb és legáltalánosabb definíciója az ISO-tól származik (ISO 9000:2000 DIS) és 2000-ben került megfogalmazásra: "a minőség

valamely termék, rendszer vagy folyamat valamennyi saját jellemzőjének az együttes képessége, hogy kielégítse a vevők és más érdekelt felek követelményeit". A *termék* lehet hagyományos gyártmány vagy szolgáltatás. Az összevont fogalom ellenére a hagyományos termék és a szolgáltatások között a különbség alapvető. A *szolgáltatások* megfoghatatlanok, nem tárgyiasultak, keletkezésük pillanatában elfogyasztásra kerülnek, nem tárolhatók. A termékek és a szolgáltatások közötti különbségek minőségük megfogalmazásában is jelentkeznek. A fizikai termékek minőségét elsősorban a megbízhatóság (hibamentesség, karbantarthatóság, karbantartás-ellátás), használhatóság és teljesítőképesség jellemzi. A *hálózatnak, mint terméknek* a minőségét a hálózatot alkotó összeköttetések minőségének az átlaga, vagy a leggyengébb láncszem minősége határozza meg. "*Az összeköttetések minősége* pedig annak a mértéke, ahogy egy összeköttetés reprodukálja a meghatározott feltételek mellett felajánlott jelet" (ITU Terms and Definitions Database "SANCHO"). *A szolgáltatások minősége* (ITU-T E.800-as ajánlás) "a szolgáltatás képességeinek együttes hatása, amely meghatározza a felhasználó szolgáltatással való megelégedettségének a fokát". Közelebbről a távközlési szolgáltatás-minőséget a szolgáltatás támogatási-, a szolgáltatás működtethetőségi- és a kiszolgálási képességek, valamint a szolgáltatás biztonsága jellemzik.

A QoS definiálható a felhasználó vagy a szolgáltató szemszögéből. Az *ügyfél/felhasználó szempontjából* a QoS-t azok a kritériumok határozzák_meg, amelyek a szolgáltatás használata során számára lényegesek. *A szolgáltató szempontjából* a QoS azokkal a jellemzőkkel adható meg, amelyek hozzájárulnak a felhasználók követelményeit tükröző, végpontok közötti szolgáltatási képességekhez. A szolgáltatásminőség megkülönböztetendő a hálózati képességektől (network performance=NP). A QoS-t a felhasználó tapasztalja, a hálózat képességét pedig a hálózat elemeinek, vagy az egész hálózatnak a műszaki képessége határozza meg. De a hálózati képességek – szolgáltatási képességeken keresztül - a QoS részét képezik, hatnak rá.

Az 1980-as években az általános törekvés az volt, hogy minél általánosabb definíciót fogalmazzanak meg mind a minőségre, mind a szolgáltatás minőségre. Az utóbbi években előtérbe került az a megközelítés, hogy a QoS definíció alkalmazásfüggő. Erre példa az Eurescom meghatározása: "a szolgáltatásminőség annak a mértéke, hogy mennyire felel meg a szolgáltató által a felhasználónak

nyújtott szolgáltatás az ügyfél és a szolgáltató között létrejött megállapodásnak". Zavaró ebben a definícióban, hogy a minőséget a megfelelőség fogalmához közelíti, amivel nem lehet egyetérteni. Ugyanakkor ez a definíció jól rímel az ISO 9000:2000 végleges *minőség* definíciójára, amely szerint "a minőség annak mértéke, hogy mennyire teljesíti a saját jellemzők egy csoportja a követelményeket". Az Eurescom definicióban megjelent az SLA (Service Level Agreement) vagyis a szolgáltatók egymás közötti és az ügyfél és a szolgáltató közötti szerződés fogalma. Ennek keretében vállal a szolgáltató garanciát egy bizonyos szintű QoS-ért (kivételt képeznek a nem-garantált QoS-ű szolgáltatások, ld. 5. fejezet)

1.13.3. Szolgáltatás-minőségi modellek [1.13.2]

A modell alkalmas arra, hogy a vizsgált rendszer vagy folyamat belső összefüggéseit, jellemző sajátosságait elemezzék. A szolgáltatásminőség modelljeit több szempont szerint lehet csoportosítani, például eredetük szerint. Az 1.13.1 táblázat felsorol néhány fontos szolgáltatás-minőségi modellt, a teljesség igénye nélkül.

Példaként bemutatjuk az I.350 ajánlásban szereplő modellt, melynek lényege, hogy definiálja a QoS/NP paraméterek (az ajánlás a mérték és paraméter szavakat

Eredet	Modell	jellemzői/mire vonatkozik
ITU-T	E.800-as, I.350-es, I.380-as, X.140-es, G.109-es,	fogalmak, definíciók, QoS-NP, QoS/NP digitális hálózatokban IPalapú adatkomm.s szolg. nyilvános adat hálózat, beszédátviteli minőség
ETSI	ETR003, ETR138, EG 201 769-1 Tiphon (Telecommuni-cations and Internet Protocol Harmonization Over Networks) ETS 300.416 TBR 21, 38	I.350 előképe, QoS mutatók távbeszélőre, ISDN-re, QoS mutatók távbeszélőre, definíció+mérés, Távközlés és Internet technológia összekapcsolása
ISO/IEC	12326 (ITU-T X.641)-es, 15802-3-as	Információ technológia
Eurescom	P616-os, P806-os	Minőség osztályok QoS/NPsok szolgáltató esetén
IETF	RFC 1633-as, RFC 2475-ös	QoS integrált szolg. esetén, QoS differenciált szolg. esetén

1.13.1 táblázat

szinonimaként használja) 3 x 3-as mátrixát (1.13.2 táblázat)

A mátrix által meghatározott kilenc alap- vagy generik paraméter, amely elsődleges paraméterek (ld. 1.124) és amelyek adott szolgáltatás esetén konkretizálhatók.

Másik példaként megadjuk a Tiphon projekt szolgáltatás minőséggel foglalkozó területének struktúráját, amely tükrözi a modell lényegét is:

- a QoS általános szempontjai,
- a minőségi osztályok meghatározása,
- szolgáltatásminőség ellenőrzés,
- szolgáltatás-minőségi mérések módszerei,
- teszt vizsgálatok,
- tervezési útmutató

Minőségi kritérium ⇒ Kommunikációs funkció ⇓	sebesség	Pontosság	szolgáltatás képtelenség
Hozzáférés			
Információ átvitel			
Felszabadítás			

1.13.2 táblázat

1.13.4. Szolgáltatás-minőségi mutatók [1.13.3]

A szolgáltatásminőség, különböző definícióiból is következően, nem számszerűsíthető mennyiség, de vannak a szolgáltatás minőséget jellemző tényezők, amelyeket a szolgáltatásminőség meghatározóinak-, szolgáltatás-minőségi mutatóknak- vagy szolgáltatás-minőségi mértékeknek nevezünk. Ezek már mérhetők objektíven, mérőeszközzel, vagy szubjektív úton, vélemény-kutatással. Ennek megfelelően beszélünk objektív vagy szubjektív szolgáltatás-minőségi mutatókról. Az objektív méréseket általában maga a szolgáltató végzi. Ezért az objektív szolgáltatás-minőségi mutatókat belső mutatóknak is nevezik. A belső mutatók tehát szolgáltató-orientáltak. A szubjektív mérések alanyai az ügyfelek, akik a szolgáltatást kívülről ítélik meg. Igy a szubjektív mutatók külső mutatók és ügyfél-orientáltak.

Az ETSI, az Európai Unió 98/10/EC direktívájára támaszkodva, a minőségi mutatók képzésénél és kiválasztásánál figyelembeveendő szem-pontokra az alábbiakat javasolja:

- a szolgáltatásminőségi mutatók legyenek könnyen érthetők az ügy-felek számára,
- a mutatók az előfizetői végződésnél legyenek mérhetők,
- a mérések (ahol ennek értelme van) lehetőleg az élőforgalomra és ne vizsgálóhívásokra támaszkodjanak,
- független szervezetek által is legyenek mérhetőek,
- a méréseknél megfelelő pontosság legyen biztosítható,
- a választott mutatók között legyenek statisztikus mértékek (átlagérték, valamint legtöbbször a statisztikai eloszlás 95%-os szintértékéhez (kvantiliséhez) tartozó érték), és legyenek olyan mértékek is, amelyek az egyes előfizetőknek nyújtott szolgáltatásminőség megállapítására alkalmasak.

A szolgáltatás-minőségi mutatók minőségi követelményeket és tevékenységeket rendelnek össze.

Az ETNO (ld. 1.128) a távközlő szolgáltatások esetén az alábbi távközlő tevékenységeket különbözteti meg: szolgáltatásnyújtás, szolgáltatás-támogatás, javítás, maga a hívásfolyamat vagy kommunikáció, számlázás, panaszkezelés. Az ETNO által felsorolt minőségi kritériumok pedig: a sebesség, a pontosság, a használhatóság (elérhetőség), a megbízhatóság, (ezek objektív minőségi kritériumok), a biztonság, az egyszerűség, a rugalmasság, és az elégedettség. Ez utóbbiak szubjektív minőségi kritériumok. Más szerzőknél további meghatározók is szerepelhetnek.

A hivatkozott Európai Uniós direktíva a vezetékes távbeszélő szolgáltatások számára az alábbi szolgáltatás-minőségi mutató készletet ajánlja:

- az első vonal bekapcsolásának ideje,
- *előfizetői vonalak hiba aránya,*
- *hiba javítási idő,*
- *sikertelen hívások aránya *,*
- *hívás felépítési idő*,*
- *a kezelői szolgálat válaszideje,*
- *a tudakozó szolgálat válaszideje,*
- *a működő érmés- és kártyás nyilvános állomások aránya,*
- *a hibás számlákra vonatkozó panaszok mennyisége.*

A minőségi mutatók definícióit és mérésének módszerét az ETSI EG 201 769-1 (2000. április) útmutató szerint kell figyelembe venni.

A csillaggal megjelölt mutatók mérésétől bizonyos esetekben el lehet tekinteni. Ezek a mutatók szoros kapcsolatban vannak a Magyar-országon a vezetékes távbeszélő szolgáltatásra jelenleg érvényes koncessziós követelményekben szereplő minőségi mutatókkal (1.13.3 táblázat). A táblázat a koncessziós szerződésben szereplő aktuális célértékeket is tartalmazza.

Szolgáltatás-minőségi Mutató	2001 év [%]	2002 év [%]
Nagyforgalmú időszakban kezdeményezett, primer körzeten belüli vizsgálóhívások sikertelen-ségi mutatója (a44)	1,40	1,40
Nagyforgalmú időszakban kezdeményezett bel-földi távolsági vizsgálóhívások sikertelenségi mutatója (a44)	2,80	2,70
Nagyforgalmú időszakban kezdeményezett nemzetközi vizsgálóhívások sikertelenségi mutatója (a44)	2,80	2,70
Közönségszolgálati munkahelyekre kezdemé-nyezett vizsgáló hívásokra meghatározott időn belül történő jelentkezések összesített mutatója (a21)	97,00 T < 20 s	97,00 T < 20 s
Meghatározott időn belül tárcsázási hanghoz nem jutó vizsgálóhívások mutatója (a44)	0,80 T > 3 s	0,80 T > 3 s
Nyilvános távbeszélő-állomások üzemkészsége (a23)	92,00	92,00
Távbeszélő-állomások egy hónap időtartam alatt bekövetkezett meghibásodásai, darab/állomás/év (a34)	0,020	0,020
Távbeszélő előfizetői állomások meghibásodásai-nak éves átlagos időtartama (sec.) (a31)	3,10	3,10
24 órán belül elhárított hibáknak a bejelentett tényleges hibákhoz viszonyított aránya (a34)	0,90	0,90
Távbeszélő-számla elleni felszólalások aránya (a52)	0,016	0,016

1.13.3 táblázat

Külön említésre méltó a QuEST Fórum (lásd később) összehasonlítható teljesítmény mutató rendszere, amely ma már távközlési cégek gyakorlati QoS adatainak közös adatbanki rendszerben való statisztikai feldolgozását teszi elérhetővé a Fórum tagjai számára.

Összehasonlítható teljesítménymutatók: Az Egyesült Királyság szolgáltatóinak gyakorlatában már múltja van az "összehasonlítható teljesítmény- mutatók"-nak. A szolgáltatók az Ipari Fórumon megegyeztek arról, hogy bizonyos szolgáltatásminőségi mutatóik (1.13.4 táblázat) általuk teljesített értékét önként az Egyesült Királyság Szabályozójának rendelkezésére bocsátják és ezeket publikálják. Ennek a folyamatnak elsősorban az a célja, hogy a fogyasztók megismerjék a lehetséges szolgáltatásválaszték minőségét és az adatokat mérlegelve döntsenek a szolgáltató választásakor. Látható, hogy közvetve a szolgáltatásminőségi mutatóknak fogyasztóvédelmi szerepük is van. Másrészt egy ilyen mutatórendszer

távközlési tevékenység	objektív minőségi mutató	szubjektív minőségi mutató
szolgáltatásnyújtás	a megrendelések ígért határidőre való teljesítése, a megrendelések időbeli teljesítésének eloszlása	⇐ az ezzel való elégedettség
szolgáltatás-helyreállítás	a bejelentett hibák ígért határidőre való elháritása, a hibaelhárítás időbeli teljesítésének eloszlása	⇐ az ezzel való elégedettség
a szolgáltatás műszaki minősége	a bejelentett hibák száma, összeköttetés-megszakadások által érintett ügyfelek száma	a szolgáltatás ⇐ megbízhatóságával való elégedettség
számlázás	1000 számlára jutó számlapanaszok száma	⇐ az ezzel való elégedettség
panaszkezelés	20 munkanapon belül megoldott panaszok aránya	⇐ az ezzel való elégedettség

1.13.4. táblázat

lehetőséget ad a szolgáltatók számára is a szolgáltatások összehasonlítására (benchmarking), ami szolgáltatásaik továbbfejlesztése szempontjából fontos információ számukra is.

Az ilyen típusú mutatórendszer elemeit összehasonlítható mutatóknak (CPI = Comparable Performance Indicators) nevezzük.

Megjegyezzük, hogy a QoS mutatóknak léteznek további csoportosítási szempontjai is. Például vannak:

- elsődleges (közvetlenül mérhető) paraméterek, és
- másodlagos (elsődlegesből származtatott) paraméterek. Valamint léteznek alap paraméterek (ld.1.123)

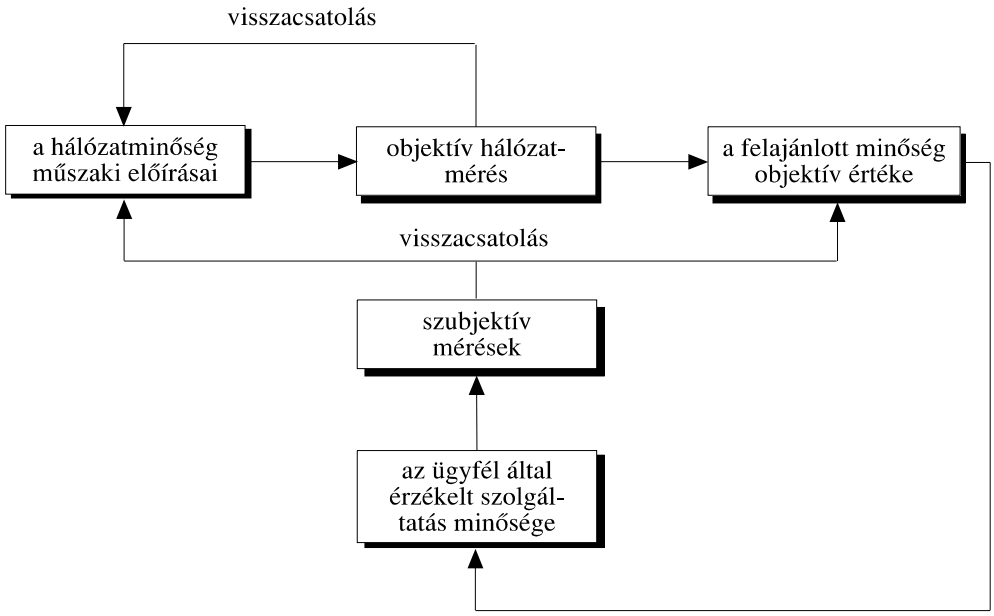
1.13.5. A szolgáltatásminőség mérése [1.13.4]

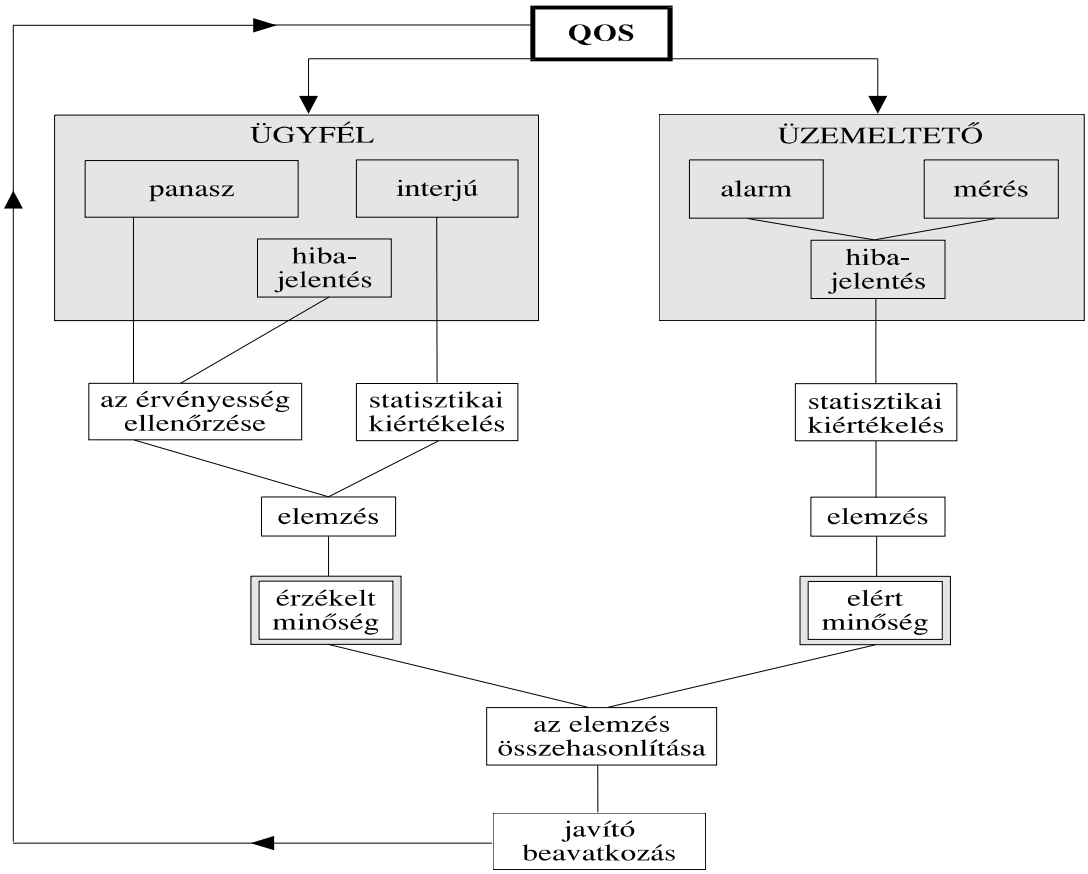
A szolgáltatásminőség javíthatóságának feltétele a mérhetőség. Az 1.13.1 ábra mutatja a minőségi mértékek (szubjektív és objektív) kettősségén alapuló minőségellenőrzés alapelvét.

Természetes elvárás, hogy ha egy szolgáltatás jó, akkor a szubjektív és objektív mérések egyaránt jó eredményt nyújtsanak. Ha viszont az ügy-feleknek nyújtott szolgáltatás rossz, akkor mind a szubjektív, mind az objektív mérések rossz eredménnyel zárulnak. De a tapasztalatok jelentős eltérést mutatnak a kétfajta mérés között. Ha egyszerre akarjuk kihasználni mind a szubjektív mind az objektív mérések előnyeit, akkora legjobb, ha kombináljuk a két módszert, amint azt az 1.13.2. ábra mutatja.

Más csoportosítás szerint a szolgáltatásminőség mérésének alapvetően két módszere van: a próbahívások alkalmazásával történő mérések és az élőforgalom alapján végzett mérések.

A szolgáltatásminőség mérhető a *szolgáltató aspektusából*. Ezek a mérések a hálózati képességek mérésére irányulnak. Másrészt vannak *felhasználócentrikus* szolgáltatásminőség mérések, amelyek a végfelhasználók közötti mérések. Számos automatikus mérő rendszert fejlesztettek ki a hálózati képességek és szolgáltatás-minőségi mutatók mérésére. Ezek a rendszerek a távközlő hálózat elemeihez kapcsolt adatgyűjtőkből, az adat gyűjtést vezérlő hardverből és szoftverből, valamint az adatokat feldolgozó központi egységből állnak.





1.13.2 ábra

1.13.6. A szolgáltatásminőség közgazdasági vonatkozásai [1.13.5],[1.13.6],[1.13.7]

- A minőség és költség kapcsolata:

A fogyasztó, ha szolgáltatást választ, vannak minőség követelményei. A szolgáltatás választással együtt az ügyfél szolgáltatót is választ. Ekkor figyelembe veszi múltbeli tapasztalatait, azokat az írott információkat, amelyek a szolgáltatásról rendelkezésre állnak, a szolgáltató hírnevét, imázsát a piacon és nem utolsó sorban az árat is. A tapasztalatok és a tanulmányok azt mutatják, hogy szoros összefüggés van a szolgáltatás minősége és ára között. A gyakorlatban ez azt jelenti, hogy az ügyfél értékalapon választ, ahol az értéket a minőség és ár együtt jellemzi. A fogyasztónak el kell döntenie, hogy mennyit tud áldozni a jó minőség érdekében. A klasszikus minőségelméletben egy emelkedő görbe írja le a költség és a minőség kapcsolatát. A kérdés új megközelítésében a szolgáltatás költségét leíró formulába

változóként beépítik a szolgáltatásminőséget, egy Q tényező bevezetésével. A Q faktor a gyakorlatban szolgáltatás-minőségi mutatók célértékeit jelenti. Ezt a módszert az USA-ban már a gyakorlatban alkalmazzák.

- A minőség (a jó és rossz minőség) költségei:

Valamely vállalat vagy vállalkozás üzleti eredményeit meghatározó adottságok között lényeges szerepe van a minőség költségeinek is. Az EFQM (European Foundation of Quality Management) modellje jól mutatja a minőségköltségekkel való kapcsolatot.

A minőségköltségeket két csoportra bontjuk: a követelményeknek való *megfelelés* és a *nemmegfelelés* költségeire. Ezen költség-összetevők összege folyamatköltség. A két költség-összetevő a minőség függvényében ellentétes irányban változik, ezért összegüknek, a folyamatköltségnek bizonyos minőségnél minimuma van, ez az optimális minőség). Ez a felosztás bármely folyamat vagy szervezet bármely szintjén értelmezhető.

A klasszikus irodalomban az optimális minőségköltséget általánosan egy abszolút minimummal rendelkező, parabolaszerű görbe írja le. Az új kutatások szerint az optimális minőségköltség görbéje egy leszálló ágú hiperbolaszerű görbe.

A minőségköltségek egy másik felfogása (BS 6143 modell) az alábbi minőségköltség összetevőket különbözteti meg: megelőzési-, értékelési-, (belső- és külső) hibaköltségek. A tapasztalat szerint ezek aránya:

5 : 30 : 65.

Az Európai Unio egyik projektjében kidolgoztak egy gyakorlati alkalmazásra megfelelő minőségköltség csökkentési módszert.

- A gazdasági élet, a piac területén való elfogadhatóság (acceptability):

Ez a fogalom vagy mérték fontos a verseny piacon és annak jellemzésére szolgál, hogy a felhasználó hogyan viszonyul a technológiához, az új alkalmazásokhoz. Az elfogadhatóság szoros kapcsolatban van a szolgáltatás-minőséggel. Definició szerint az elfogadhatóság egy adott szolgáltatást alkalmazó felhasználói csoport egyedeinek száma viszonyítva az alkalmazások szempontjából lehetséges teljes populáció méretéhez.

Az elfogadhatóságnak vannak szolgáltatás-specifikus-, felhasználó-specifikus-, vállalatspecifikus- és környezetspecifikus meghatározói. A vállalatspecifikus meghatározók függetlenek a szolgáltatástól, például ilyenek maga a szervezet, a személyzet képzettsége, a verseny intenzitása, stb. Környezetspecifikus meghatározók: a szolgáltatás kompatibilitása, a fejlesztés gyorsulása, az infrastruktúra, a hirdetések, stb. Felhasználó-specifikus tényezők: a szubjektív előnyök, a motiváció, az elvárások, a kockázati tényezők, a technológiai előrejelzések, stb. Végül szolgáltatás-specifikus a tervezés, az alkalmazhatóság, a funkcionalitás, a díjazás, az alkalmazásbavétel szabványosítása és a szolgáltatásminőség, stb.

1.13.7. A szolgáltatásminőség szabályozása [1.13.5], [1.13.8]

A szolgáltatásminőség szabályozásával kapcsolatos kulcskérdések az alábbiak:

- a szabályozási környezet jellege (monopol-vagy versenyhelyzet),
- a szolgáltató által publikálandó szolgáltatás-minőségi mutatók meghatározása,
- a közzétett adatok auditálása,
- a nem teljesítő szolgáltatóval szemben alkalmazott politika (jutalmazás és büntetés szükséges-e),
- új szolgáltatások esetén a minőségi paraméterek bevezetésének ütemezése,
- a szolgáltató által nyújtott minőség fenntartásának módja, lépései.

A monopóliumok korszakában, különösen a nyolcvanas években a szolgáltatásminőség szabályozásának két alap típusa létezett:

- a szabályozó orientált típus, amelynek jellemzői:
  - a szolgáltatóminőségi mutatók mért értékeit a szolgáltatók és/vagy szabályozók, beszámolók formájában nyilvánosságra hozzák,
  - a szabályozók célértékeket tűznek ki e mutatókra,
  - a célértékek nem-teljesülése esetén a szolgáltató büntetést fizet,
- a felhasználó orientált szabályozás, amelynek jellemzői:
  - a szolgáltató nyilvánosságra hozza szolgáltatásminőségi mutatóinak teljesített értékeit,
  - a szolgáltató az előfizetői szerződésben vállalja bizonyos értékű szolgáltatásminőségi mutató biztosítását,

- ha a szolgáltató nem teljesíti vállalását, akkor az ügyfélkompenzációt kap.

A verseny körülményei között a szabályozást új alapokra kell helyezni. Egy lehetséges megközelítés az Európai Union belül:

- Európai Uniós joganyagokra (pl. 98/10/EC direktíva), valamint az ETSI szabványokra kell támaszkodni.
- referencia modellt kell alapul venni, amely szerint
  - a szabályozás érdekeltjei az ügyfelek, a szolgáltatók és a szabályozók, ezek kapcsolatát
  - # az ügyfél hurok (az ügyfél és a szabályozó között),
  - # a piaci hurok (az ügyfél és a szolgáltató között) és
  - # a szabályozói hurok (a szolgáltató és a szabályozó között)
  - jellemzik, valamint
  - # ezen rendszeren belül, a szabályozás szempontjából kulcs-kérdé-
  - sek: az információ, a kötelezettségek és a segédeszközök.

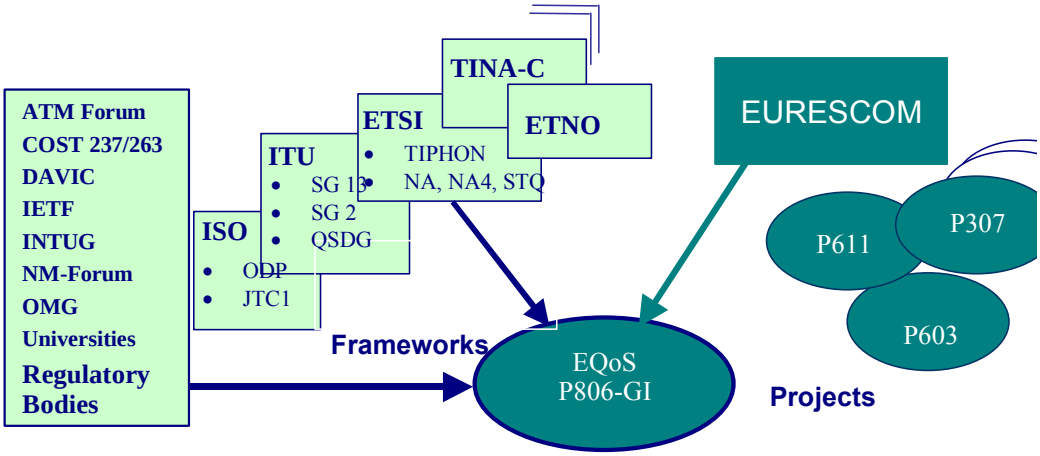
Egy lehetséges, javasolt megoldást tartalmaz a 1.13.5 táblázat.

QoS szabályozás	információ	Kötelezettség	segéd-eszközök
Piaci hurok	-proaktív információkáltalában a képességekről -és a hibás működésre vonatkozó megoldások	egyedi kötelezettségek a szerződések-ben	-kompenzáció,ha QoS a célértéknél rosszabb ,-világos meg-fogalmazások a szerződésben, -lehetőség a párbeszédre
Ügyfél hurok	a szabályozó által nyilvános ságra hozottinformációk		másodfokú párbeszéd
Szabályozói hurok	képesség mérések: kérésre v. kötelezően (pl. egyetemes szolgáltatás)	-kollektív célok kitűzése, törvé-nyi, rendeleti megfogalmazás,-gyenge pont monitorozás	a szabályozó birtokában legyen számos eszköz a szolgáltatóval szemben(hatékony szankcionálás is)

1.13.5 táblázat

1.13.8. Nemzetközi szervezetek [1.13.9]

A szolgáltatásminőség általános és speciális kérdéseivel számos nemzetközi és európai szabványosító szervezet (pl. ISO, IEC, ITU, ETSI), kutatási központ (pl. Eurescom) valamint a távközlésben érdekelt felek: gyártók, szolgáltatók, ügyfelek szervezetei (pl. IETF, INTUG, ETNO, QuEst Fórum,stb.) foglalkoznak. Röviden



1.13.3 ábra

bemutatjuk ezeket a szervezeteket, elsősorban a minőséggel összefüggő céljaikra, tevékenységükre fókuszálva.

Az *ISO* (International Standards Organisation)-t 1947-ben alapította huszonöt nemzeti szabványosítási szervezet. Az ISO célul tűzte ki nemzetközi szabványok kiadását, elősegítendő az árúk és szolgáltatások világméretű cseréjét. Munkáját szoros együttműködésben végzi az IEC (International Electrotechnical Comission)-vel és az ITU-T(International Telecommunications Union-Telecommunication Standardisation Sector)-vel. Számos szabványt azonos szöveggel és saját számozási rendszerüknek megfelelő számozással ad ki. A minőségügy területén a minőségirányítási rendszerekre vonatkozó szabványai a legjelentősebbek (EN ISO 9000:2000).

Az IEC (International Electrotechnical Commission) TC 56 (Dependability) műszaki bizottsága az elektrotechnikai és elektronikai gyártmányok, rendszerek megbízhatósági kérdéseivel foglalkozik és e témakörben bocsát ki a távközlésben is jól használható szabványokat, elsősorban a vizsgálat tervezés, a hibák osztályozása, a statisztikai adatfeldolgozás és a karbantartás szervezés területén, ami a QoS megítéléséhez nagy segítséget nyújt.

Az *ITU-T* kormányok, nemzeti távközlési adminisztrációk, elismert magántársaságok, gyártók, szolgáltatók, szabályozók, tudományos és ipari szervezetek, fogyasztók érdekvédelmi csoportjainak, stb. összefogója. Munkáját Tanulmányi Bizottságokban-, azon belül munkacsoportokban végzi. Egy-egy kérdés

kidolgozásával rapportőr csoportok foglalkoznak. Jelenleg öt Tanulmányi Bizottság (TB2, TB7, TB12, TB13, TB16) munkája kapcsolódik a szolgáltatásminőséghez.

Az öt TB-n kívül még a QSDG (Quality of Service Development Group) speciális munkacsoportnak is ez az érdeklődési területe. Ez a munkacsoport 1982-ben alakult a TB2-n belül. Elsődleges célja a nemzetközi távközlési szolgáltatások minőségének a javítása. Ennek érdekében közel száz ország, több-száz szakértőjét fogja össze, akik rendszeresen kicserélik elméleti és gyakorlati eredményeiket. A szolgáltatásminőség összes aspektusával foglalkoznak.

Az *ETSI* hatásköre Európára terjed ki, de szabványdokumentumainak elfogadottsága Európán kívül is növekszik. Tevékenységi területe a távközlés szabványosítása, beleértve a rádiótávközlést és a műsor-szórást is, valamint az információ technológiát. 1988-ban alakult. Műszaki szabványokat hoz létre és hagy jóvá. Tevékenységének egy részét projekt munkában végzi. A szolgáltatás-minőséggel a SPAN (Services and Protocol for Advanced Networks) műszaki bizottság, az STQ (Speech Processing, Transmission and Quality aspects) bizottság, valamint a Tiphon projekt foglalkozik.

Az *Eurescom* húsz európai hálózat-üzemeltető kutatás-fejlesztési tevékenységét fogja össze, team munka keretében.

Az *IETF* (Internet Engineering Task Force) tervezők, üzemeltetők, kutatók, stb. nemzetközi közössége, amely az Internettel kapcsolatos műszaki tevékenységeit munkacsoportokban végzi. Ezek tématerületek szerint alakulnak. A szolgáltatásminőség is egy ilyen tématerület.

Az *INTUG* (International Telecommunications User's Group) a távközlési felhasználók érdekeit képviseli, többek között a szolgáltatásminőség területén is.

Az 1991-ben alakult *ETNO* (European Telecommunications Network Operators) az európai nyilvános hálózatüzemeltetők fóruma. Az üzemvitel és szabályozás különválása során jött létre. Munkacsoportjainak egyike a QoS WG, amely ETNO mátrixos reprezentációval egy jól használható szolgáltatás-minőségi modellt alkotott. Jelentős még az ETNO abból a szempontból is, hogy az európai távközlési szervezetek közül elsőnek kezdeményezték a szubjektív szolgáltatásminőség mutatók egységesítését.

1997-ben a világ vezető távközlési cégei egy új szervezetet hoztak létre, a *QuEST* (Quality Excellence for Suppliers of Telecommunication) *Fóru*-mot, hogy

- meghatározzák az ISO 9000 szerinti minősítés speciális távközlési követelményeit és mérési módszereit TL 9000 (Telecommunication Leadership) jelzettel,
- auditorokat képezzenek ki a távközlési gyártó, üzemeltető és szolgáltató cégek minőségügyi rendszereinek TL 9000 szerinti minősítéséhez, valamint
- korlátozott számban minősítő (regisztrátor) intézményeket akkreditáljanak.

A TL 9000 többszintű rendszer. Az alap az ISO 9000:2000. Erre épülnek az *általános* távközlési követelmények. A következő szint a hardver, a szoftver és a szolgáltatások *speciális* minőségi követelményeit tartalmazza. A következő szintre kerülnek a minőségi követelmények ellenőrzéséhez szükséges mértékek és mérési módszerek. A TL 9000 kézikönyv első kötete tartalmazza a minőségi követelményeket, a második kötet a minőségi mértékeket.

A QuEST Fórumnak 2001-ben már több, mint 160 cég és intézmény a tagja világszerte, akik kiváló minőségük demonstrálása céljából háromhavonta rendszeresen megküldik a TL 9000 szerint egységes módon értelmezhető és így összehasonlításra alkalmas QoS adataikat a Fórum központi adatbankjába, ahol ezek az adatok statisztikai feldolgozásra kerülnek és a tagok számára elektronikus úton hozzáférhetők, saját minőségfejlesztési programjaik továbbfejlesztéséhez és benchmarking célokra. A QuEST Fórum adatvérkeringése az első a világon, amely tapasztalati úton nyert és globálisan közös nevezőre hozott QoS adatokat „forgalmaz” a távközlés területén.

1.13.9. Szolgáltatásminőség-/minőséggel foglalkozó szabványok, projektek [1.13.10]

Szabványok:

- ISO szabványok:
  - MSZ EN ISO 9000:2000 Alapok és szótár
  - MSZ EN ISO 9001:2000 Követelmények
  - MSZ EN ISO 9004:2000 Útmutató a működés fejlesztéshez
- ITU-T ajánlások:

- E ajánlások

QoS ellenőrzése (E.420 - E.489)

QoS fogalmak és definíciók (E.800 - E.810)

QoS modellek (E.810 - E.845)

QoS tervezés (E.845 - E.899)

- G ajánlások

általában (G.100 -G.109)

távbeszélő átviteli minőség (G.110 - G.119)

Q és A célértékek (G.820 - G.830)

- I és P ajánlások

minőségi célértékek (I.350 - I.359)

általános hálózati követelmények (I.370 - I.399)

távbeszélő átviteli minőség (P.10 - P.80)

minőség értékelés (P.80 - P.800)

multimédia szolgáltatások hang-

minősége (P.900 - …)

- M és Q ajánlások

méréstechnika (Kiegészítés az M sorozathoz)

távbeszélő átviteli minőség („ „ „)

tervezési célértékek Q.543

- X és Y ajánlások

adatátvitel,

IP alapú szolgáltatások

- ETSI ajánlások:
  - EG 202 086 Hagyományos távbeszélő szolgáltatás minőségének célértékei összekapcsolt hálózatokban,
  - EG 201 377-1 Beszédátviteli minőség, összehasonlító mérések,

- EG 201 769, 769-1 A 98/106/EC ONP direktíva alá tartozó QOS beszédátviteli követelmények
- TR 101 329 Tiphon; A QoS általános aspektusai
- TR 101 329 - 1 Végpontok közötti QoS Tiphon rendszerekben,

  A QoS általános aspektusa
- TR 101 329 - 2 A QoS osztályok definiciója
- TR 101 329 - 5 QoS mérési metodológia
- TR 101 329 - 6 Hálózat és végberendezés jellemzők aktuális mérései
- TR 101 329 - 7 Tervezési irányelv

Projektek:

- Eurescom projektek:
  - P 307 Hibamentesség tervezés
  - P 514 Jelzés hálózatok megbizhatósági tervezése
  - P 603 QoS mérések. Módszer választás
  - P 806 - GI QoS/NP a sokszolgáltatós környezetben
  - P 905 - PF Interneten át haladó audiovizuális jelek minőségének értékelése
  - P 906 - GI : QoS methodológia
  - P1008- PF Együttműködési interfész a vég - vég IP QoS biztosítására
  - P1003- PF IP QoS keretei összekapcsolás esetén
- ITU-T projektek:
  - GII -N.8,
  - IP - 8
- ETSI projektek:
  - Tiphon.

Irodalomjegyzék:

[1.13.1] dr. A.V. Feigenbaum: Quality trends in the new Millennium.

44-th European Quality Congress, 2000.jun.

[1.13.2] Eurescom P906-GI projekt: QUASIMODO - Quality of service methodolgies 1999, www.eurescom.de

[1.13.3] dr. Buzás Ottó és szerző társai: Távközlési kultúra PressCon kiadó 2001

[1.13.4] dr. Veress Gábor és szerző társai : Minőségügyi jegyzet sorozat a Veszprémi Egyetem posztgraduális oktatása számára, 1998.

[1.13.5] A.P. Oadan at all: Quality of service in telecommunications IEE Telecommunications series 39, 1997.

[1.13.6] BS 6143 angol szabvány

[1.13.7] Telecommunications Regulation Handbook World Bank, Washington 2000.

[1.13.8] Sagatel study for the EC. 2000

[1.13.9] www.iso.ch, www.iec.ch, www.itu.int, www.etsi.org, www.eurescom.de, www.ietf. org About us

[1.13.10] www.eurescom.de, www.etno.be