

INTRODUZIONE ALLE SORGENTI DI INFORMAZIONE ①

Prof. FRANCESCO A. N. PALMIERI

Dip. di INGEGNERIA
UNIV. DELLA CATANIA

(Corso di
TRASMISSIONE ED ELABORAZIONE
NUMERICA DEI SEGNALE)

Introduzione

AA 2019-20

Un sistema di comunicazione è rivolto allo scambio affidabile di informazione tra due o più enti. Anche se sarà necessario formalizzare meglio cosa si intende per "informazione," figura 1 mostra lo schema base di un sistema di comunicazione nella topologia *punto-punto* in cui l'obiettivo è trasferire il contenuto informativo emesso dalla sorgente ad un utilizzatore.

Una *sorgente* genera il messaggio da trasmettere nella forma di un segnale. Esempi tipici sono: un testo scritto, un segnale vocale, un segnale musicale, una immagine fissa, un flusso numerico proveniente da un calcolatore, un file, una immagine in movimento, ecc. La informazione va opportunamente convertita e convogliata su un *canale* che sarà responsabile del trasporto. Un canale può essere un campo elettromagnetico opportunamente modulato (canale radio), la carta su cui si memorizza un testo scritto, un nastro magnetico, un cavo coassiale, un'onda acustica che si propaga nell'aria, o altro. Il destinatario dell'informazione, detto *utilizzatore* può essere un ascoltatore, un sistema di percezione visiva (un occhio), un calcolatore, un altro sistema di comunicazione, un altro sistema di memorizzazione, eccetera.

Le topologie di un sistema comunicazione possono essere molteplici. Infatti sorgenti e utilizzatori, possono essere distribuiti in moltissimi tipi di configurazioni. La figura 2 mostra alcune topologie di connessioni in cui ogni nodo può essere o una sorgente, o un destinatario, o entrambi. Infatti



Figura 1: Sistema di comunicazione *punto-punto*

2

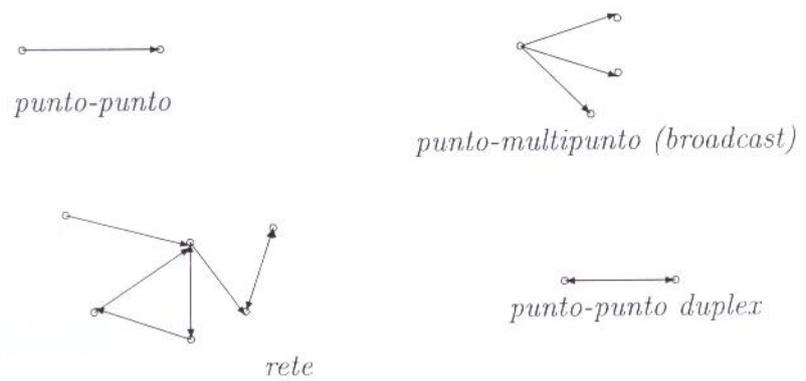


Figura 2: Varie topologie di comunicazione.

le comunicazioni possono avvenire in entrambi le direzioni su ognuno dei rami del grafo. Ad esempio una stessa sorgente può diffondere lo stesso segnale a più utilizzatori (*modalità broadcast*) mediante canali diversi (come nella radiodiffusione televisiva e sonora). Oppure tanti utenti possono essere collegati ad altrettanti utenti mediante delle connessioni bi-direzionali (*internet*). Varie modalità di trasporto nei vari rami del sistema possono coesistere: una rete può essere costituita da canali radio terrestri e satellitari, cavi coassiali, doppi telefonici, eccetera.

Qualunque sia la topologia del sistema di comunicazione, lo schema di riferimento per studiare le modalità di codifica e trasporto dell'informazione, resta quello della comunicazione punto-punto unidirezionale di figura 1. Il *trasmettitore* deve opportunamente convertire l'informazione in un formato adatto ad essere convogliato sul canale e la natura del sistema dipende dalla tipologia del segnale e del meccanismo di trasporto fisico di cui si dispone. Per esempio: la trasmissione di una sequenza numerica proveniente da un computer su un canale elettromagnetico richiede una appropriata modulazione; la trasmissione di un segnale vocale su un collegamento satellitare necessita di appropriate operazioni di conversione, di codifica e di modulazione; la trasmissione di una immagine in movimento richiede una complessa operazione di adattamento del segnale bidimensionale tempo-variante alle caratteristiche del canale.

Un sistema di comunicazione è responsabile anche di compattare l'informazione proveniente dalla sorgente in una maniera che minimizzi le risorse necessarie

alla trasmissione (*compressione*). E' piuttosto comune infatti che la sorgente esibisca un segnale che è in un formato *ridondante* rispetto alla effettiva informazione in esso contenuta. L'informazione va anche opportunamente protetta rispetto alle degradazioni che saranno introdotte dal meccanismo di trasporto (rumore, perdite, ecc.). Si pensi ad esempio ad un collegamento via cavo, o a un ponte radio, o a un collegamento in fibra ottica, o a una memorizzazione e lettura da un supporto magnetico o ottico e alle relative degradazioni.

Va inoltre menzionato che le modalità di comunicazione non sono limitate ai sistemi che operano in tempo reale. Si pensi alle registrazioni su memorie quali nastri magnetici e dischi ottici in cui il sistema deposita l'informazione per poi prelevarla in un tempo successivo. Gli strumenti di analisi e di sintesi sono gli stessi utilizzati per i sistemi che operano in tempo reale.

Nella pratica ingegneristica delle telecomunicazioni la variabilità delle sorgenti e le degradazioni introdotte sui canali sono raramente quantificabili in maniera certa. E' necessario pertanto, nella maggioranza dei casi, un approccio probabilistico-statistico che modelli le caratteristiche medie del sistema. Ecco perchè le principali metodologie di studio e di manipolazione di sorgenti e canali di comunicazioni sono di natura probabilistica. In effetti è proprio l'imprevedibilità dei fenomeni connessi alle sorgenti e ai canali che porta al concetto di "informazione" come sarà spiegato in seguito.

0.1 Sorgenti di informazione

Una sorgente emette informazione mediante un *segnale* che è una funzione (anche multidimensionale) del tempo

$$Seg : T \rightarrow \mathcal{A} \quad (1)$$

Per parlare di segnali di informazione è opportuno introdurre una classificazione basata sulla struttura del dominio di definizione dei tempi T e del codominio \mathcal{A} .

Rispetto alla struttura dell'insieme temporale T adotteremo la seguente classificazione:

- 1. Segnale tempo-continuo:** T è un sottoinsieme continuo dell'asse reale. Una ulteriore classificazione può essere introdotta rispetto alla natura di tale sottoinsieme:

- *illimitato bilatero*: il dominio di definizione dei tempi è tutto l'asse reale, ovvero $\mathcal{T} =] - \infty, +\infty[$;
- *illimitato monolatero*: il dominio di definizione dei tempi è un semiasse reale. Per esempio, $\mathcal{T} = [t_0, +\infty[$; oppure $\mathcal{T} =] - \infty, t_0]$;
- *limitato*: il dominio di definizione dei tempi è un intervallo di lunghezza finita, ovvero $\mathcal{T} = [t_0, t_1]$.

2. Segnale tempo-discreto: \mathcal{T} è un insieme discreto numerabile. Anche qui una ulteriore classificazione può essere introdotta rispetto alla natura di tale insieme:

- *illimitato bilatero*: l'insieme discreto di definizione è associato a tutta la sequenza dei numeri interi, ovvero $\mathcal{T} = \{\dots, t_{-2}, t_{-1}, t_0, t_1, t_2, \dots\}$;
- *illimitato monolatero*: l'insieme discreto di definizione è associato ad una semisequenza di interi. Per esempio $\mathcal{T} = \{t_0, t_1, t_2, \dots\}$; oppure $\mathcal{T} = \{\dots, t_{-2}, t_{-1}, t_0\}$;
- *limitato*: l'insieme discreto di definizione è essere associato ad una sequenza finita di interi, ovvero $\mathcal{T} = \{t_0, t_1, \dots, t_{n-1}\}$.

Rispetto alla struttura del codominio \mathcal{A} adotteremo la seguente classificazione:

- a. Segnale a valori continui:** \mathcal{A} è un sottoinsieme continuo di \mathcal{R}^n , dove n è la dimensionalità del segnale.
- b. Segnale a valori discreti:** \mathcal{A} è un insieme numerabile associabile ad un sottoinsieme di \mathcal{Z}^n , dove n è la dimensionalità del segnale ed \mathcal{Z} è l'insieme dei numeri interi. L'insieme \mathcal{A} è spesso denominato *alfabeto sorgente*.

Nei sistemi di trasmissione numerica i segnali più comuni hanno una struttura tempo-discreto con istanti uniformemente spaziatati: $\dots, t_0 - 2T, t_0 - T, t_0, t_0 + T, t_0 + 2T, \dots$. In tal caso, se il segnale è a valori continui, la spaziatura temporale T è detta *intervallo di campionamento*. Il suo reciproco, $f_c = 1/T$ è detto *frequenza di campionamento* e si misura in *campioni/sec* o in *Hz*. Se il segnale è a valori discreti, la spaziatura temporale T è detta *intervallo di simbolo*, e il suo reciproco $R_s = \frac{1}{T}$ è detto *frequenza di simbolo* o *baud-rate* (da Emile Baudot) e si misura in *simboli/sec* o in *bauds*. Una sorgente tempo-discreto, a valori discreti e alfabeto finito è più sinteticamente detta *sorgente discreta*.

Esaminiamo alcuni esempi tipici.

Esempio 0.1 Si consideri una sequenza di caratteri prelevati dall'alfabeto italiano di 21 caratteri, come mostrato in figura 3 (abbiamo escluso le maiuscole e i segni particolari). Ogni lettera è emessa dalla sorgente in istanti ugualmente spaziatati di $T = 10\text{ms}$. Trattasi di un segnale tempo-discreto a valori discreti con alfabeto finito \mathcal{A} di dimensionalità 21. La frequenza di simbolo è $R_s = 100$ simboli/sec.



Figura 3: Una sequenza di caratteri dell'alfabeto italiano

Esempio 0.2 Si consideri un segnale $s(t)$ che rappresenta l'andamento istantaneo della pressione dell'aria nei pressi delle labbra di un parlatore. Tale segnale proporzionale a quello prelevato all'uscita di un microfono posto in tale posizione è detto *segnale vocale*. Un segmento appartenente alla vocale "e" è mostrato in figura 4. Tale segnale è tempo-continuo a valori continui.

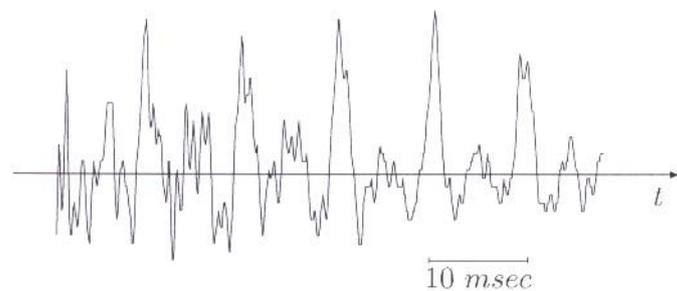


Figura 4: Un segmento di segnale vocale

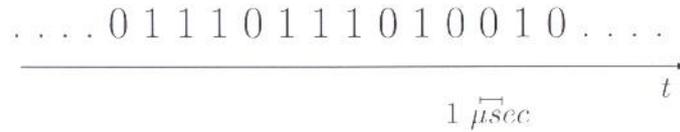


Figura 5: Una sequenza binaria

Esempio 0.3 Si consideri una sorgente che emette una sequenza di bit “0” e “1” come in figura 5. Si supponga che ogni bit sia emesso a intervalli regolari di 10^{-6} sec. Il segnale è tempo-discreto a valori discreti con alfabeto finito $\mathcal{A} = \{0, 1\}$. La frequenza di simbolo è $R_s = 10^6$ simboli/sec = 10^6 bit/sec.

Esempio 0.4 Si consideri una sequenza di valori provenienti da un sistema di misurazione della temperatura ambiente. Ogni valore è prelevato a intervalli di 1 ora. Un segmento del segnale è mostrato in figura 6. Il segnale è tempo-discreto a valori continui. La frequenza di campionamento è $f_c = 1/3600$ campioni/sec.

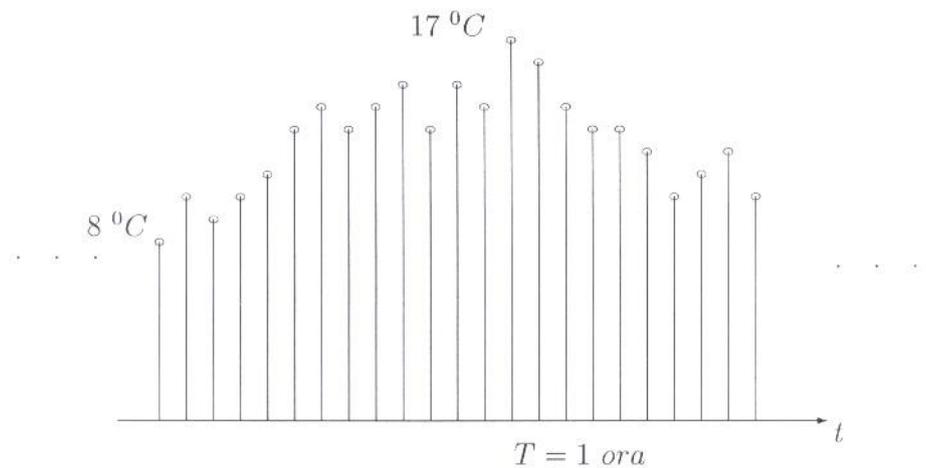


Figura 6: Una sequenza di valori di temperatura misurati

Esempio 0.5 Si consideri una sequenza di immagini di luminanza (bianco e nero) come mostrato in figura 7. Ogni immagine è una matrice di $M \times N$ pixel generata dalla sorgente ogni $\frac{1}{25}$ sec. Ad ogni pixel è associato solo uno di $2^8 = 256$ valori di intensità luminosa. Il segnale è tempo-discreto, ha dimensione $M \times N$ ed ha un alfabeto finito (anche se molto grande). Si noti che anche se ogni immagine è rappresentata da un matrice, quindi da una struttura bidimensionale, è in effetti associabile ad un unico vettore di dimensione $M \times N$. Infatti, il numero di immagini rappresentabili è pari a tutte le possibili configurazioni ordinate di $M \times N$ simboli ottali, ovvero $256^{M \cdot N}$. Assumendo che ogni pixel venga associato a un simbolo, abbiamo una frequenza di simbolo $R_s = M \cdot N \cdot 25$ simboli/sec. Per dei valori tipici $M = 400$ e $N = 500$, abbiamo $R_s = 5$ Msimboli/sec. Se invece il simbolo è il bit, poiché ogni pixel necessita di 8 bit per essere rappresentato, abbiamo $B_r = 40$ Mbit/sec.

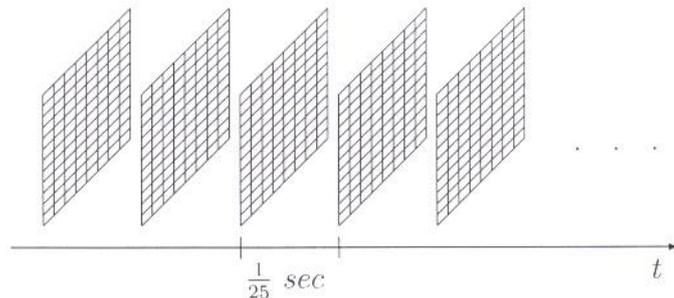


Figura 7: Una sequenza di immagini

Esempio 0.6 Si consideri il segnale di figura 8 costituito da campioni di un segnale vocale. I campioni sono prelevato con campionamento uniforme ogni $T = 1/8000$ sec. I campioni sono anche *quantizzati*, ovvero i loro valori sono arrotondati a soli 256 livelli. Si tratta di un segnale tempo-discreto ad alfabeto discreto di cardinalità 256. La frequenza di simbolo è pari alla frequenza di campionamento ed è $\frac{1}{T} = f_c = 8000$ simboli/sec.

La caratterizzazione di un segnale costituisce un argomento centrale nella teoria e nella pratica ingegneristica delle telecomunicazioni. E' infatti la quantificazione delle caratteristiche dei segnali che ci consente di utilizzarli o analizzarli rispetto al loro contenuto informativo. Ai fini della

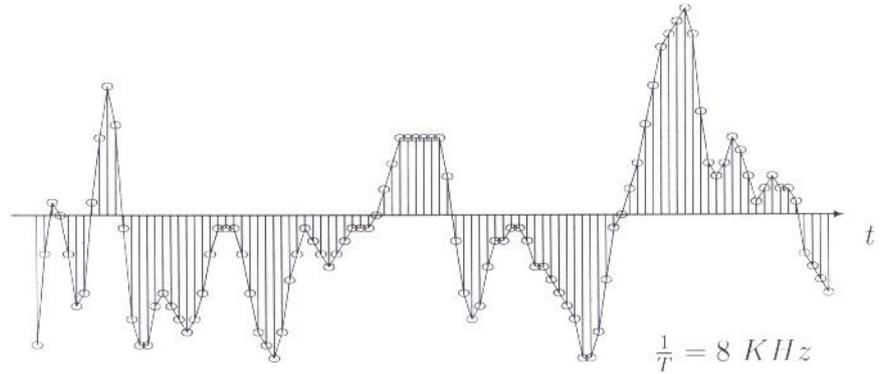


Figura 8: Una sequenza di campioni di un segnale vocale

modellistica è pertanto opportuno adottare una ulteriore classificazione. Diremo che un segnale è *deterministico* o *determinato*, se esso costituisce una fissata funzione del tempo, magari dipendente da un insieme di parametri, ma che non presenta alcuna caratteristica di incertezza. Per esempio sono segnali deterministici: un segnale sinusoidale, una sequenza periodica con fase e periodo fissati, un segnale risultante da una sola misura, eccetera. Diremo viceversa che un segnale è *aleatorio*, o *casuale*, o *random*, o *stocastico* se esso non corrisponde ad una sola funzione del tempo, ma ad una collezione (anche infinita) di funzioni associate agli elementi di uno spazio campione su cui è definita una misura di probabilità. In altre parole, quando si osserva un segnale aleatorio, bisogna pensare che ciò che vediamo è solo una delle possibili *realizzazioni* (o *funzioni membro*) del segnale ed è proprio il meccanismo aleatorio a sceglierne una secondo la misura di probabilità. La teoria dei segnali aleatori, detti anche *processi aleatori*, è ovviamente al di là degli scopi di queste note, ma è utile puntualizzare come essa sia di importanza centrale per la caratterizzazione e il controllo del contenuto informativo trasportato dai segnali. Infatti, come vedremo nel seguito, l'informazione contenuta in un segnale è proprio collegata alla sua struttura probabilistica.

Una larga parte della modellistica sui segnali è dedicata allo studio della loro struttura di dipendenza temporale. Con ciò si intende la quantificazione delle relazioni esistenti tra i valori del segnale a diversi istanti di tempo. In un segnale deterministico tale dipendenza può essere caratterizzata da equazioni differenziali o alle differenze, mentre in un segnale aleatorio si

parla di vere e proprie dipendenze probabilistiche. La teoria dei processi aleatori tempo-continuo sarà discussa altrove ed è sicuramente al di là degli scopi di queste note. Per quanto riguarda i processi tempo-discreto invece ci interesseremo solo di segnali a valori discreti e che mostrano una struttura temporale di dipendenza piuttosto semplice. In particolare, parleremo della caratterizzazione e della codifica di sorgenti discrete (tempo-discrete e a valori discreti) per *sorgenti senza memoria*, ovvero senza alcuna struttura di dipendenza tra valori del segnale in istanti successivi.

0.2 Canali di informazione

I moderni meccanismi di trasporto dell'informazione sono spesso piuttosto complessi in quanto costituiti da sottosistemi che possono anche essere molto eterogenei. Nello studio ingegneristico dei sistemi di telecomunicazioni è necessario pertanto effettuare delle schematizzazioni che ci consentano di manipolare le grandezze in gioco in maniera semplice. La schema a blocchi più generale di un canale di informazione è rappresentato in figura 9. Si tratta di un sistema che ha all'ingresso un segnale $Seg1$ e all'uscita un segnale $Seg2$.

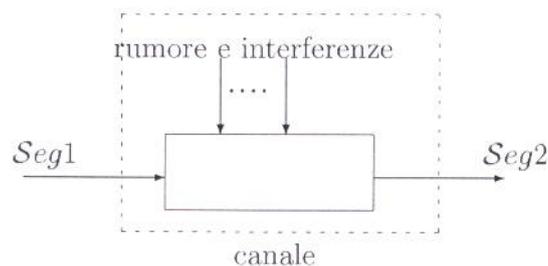


Figura 9: Lo schema generale di un canale di trasmissione

La tipologia dei segnali di ingresso e uscita del sistema caratterizzerà il tipo di sistema con il quale abbiamo a che fare. Inoltre la presenza di rumore e interferenze impone una modellistica del canale che tenga conto della parziale imprevedibilità di tali disturbi. La situazione più tipica è chiaramente quella in cui all'ingresso del canale c'è un segnale che si vuole riprodurre, secondo un preassegnato criterio di fedeltà, dall'altra parte. La modellistica generale ci consente comunque di pensare al nostro canale come

uno dei blocchi costitutivi di una catena di trasmissione anche più complessa, che magari vede al suo ingresso un segnale già codificato e alla sua uscita un segnale che sarà ulteriormente elaborato. Pertanto la modellistica del canale è da intendersi come la modellistica di un sistema deterministico o aleatorio che ha al suo ingresso e alla sua uscita segnali anche di tipo eterogeneo in tutte le combinazioni continuo e discreto possibili fornite dalla classificazione dei segnali. Inoltre il sistema può anche essere *tempo-variante*, ovvero seguire una evoluzione temporale anch'essa deterministica o aleatoria. Il sistema può anche modificare la struttura di dipendenza temporale del segnale di ingresso. Infatti è possibile pensare a dei sistemi discreti che hanno all'ingresso una sorgente senza memoria e all'uscita un segnale che mostra campioni o simboli dipendenti.

In queste note introduttive, per evitare di confondere il lettore con eccessivi dettagli, limiteremo la nostra attenzione alle sorgenti e ai canali *tempo-discreto e senza memoria*. Tali sistemi rappresentano lo scenario più tipico, e che comunque riflette il principale schema di riferimento per molti dei moderni sistemi di comunicazione, che stanno evolvendo verso una modalità quasi totalmente *numerica*.

Anche se i dettagli di un sistema di telecomunicazione sono tantissimi e richiederebbero una preparazione a largo raggio che va dall'informatica alla teoria dei sistemi, alla elettronica alla conoscenza dei problemi di elettromagnetismo, in queste note introduttive tenteremo di dare agli studenti una visione generale basata su delle schematizzazioni equivalenti. Ciò ci consente di trattare il problema della codifica e del trasporto della informazione già dai primi anni indipendentemente dai dettagli dei meccanismi fisici su cui esso si basa.