Appunti di
Teoria dell'Informazione e Crittografia

A cura di:
Francesco Refolli
Matricola 865955
Fonte Immagini:
Slide di A. O. Leporati
Appunti di D. Cozzi

Anno Accademico 2023-2024

Part I

Teoria dell'Informazione

Part II

Crittografia

Chapter 1

Introduzione alla crittografia

1.1 Introduzione

La **crittografia** è lo studio di quelle tecniche atte a immagazzinare, processare, trasmettere e in generale proteggere un'informazione su un canale non sicuro. Comprende sia tecniche per nascondere la natura del messaggio che tecniche per proteggerlo da modifiche.

Non va confusa con la *steganografia*, che invece è un insieme di tecniche per nascondere il messaggio in altri media.

La crittografia studia sia le implementazione di crittosistemi a fini di proteggere le comunicazioni, sia le loro vulnerabilità per sfruttarle o per rafforzare i suddetti sistemi.

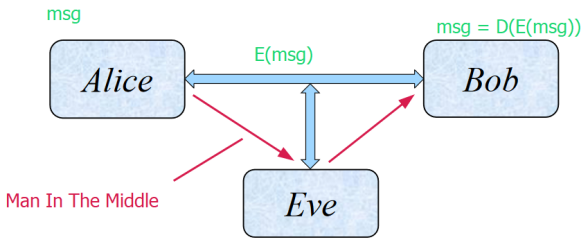

Figure 1.1: Modello di Comunicazione

In Figura 1.1 è riportato una rappresentazione del dominio. Alice vuole inviare un messaggio segreto a Bob e per evitare che Eve lo intercetti lo cifra (è necessario che Bob conosca un algoritmo per decifrare il messaggio). Eve può essere semanticamente sia un malintenzionato che un agente della Polizia Postale, quindi è uno schema molto generale.

Le capacità di Eve possono variare in base alle assunzioni che si fanno: può scrivere e/o leggere, magari nessuno dei due, e tendenzialmente dispone di capacità di computazione arbitraria (generalmente si assume che disponga di una Macchina di Turing Probabilistica).

Kerkhoff's principles Alice e Bob utilizzano due funzioni E, D per cifrare, decifrare ma la segretezza del sistema non dovrebbe risiedere nelle funzioni in se ma in una quantità di informazione chiamata **chiave**. Ovvero visto che cambiare chiave è più semplice che cambiare algoritmi la variabilità della cifratura deve risiedere nella variabilità della chiave.

Definiamo quindi un crittosistema come una tupla $CS = (PT, CT, K, E, D)$.

- PT è lo spazio dei messaggi in chiaro
- CT è lo spazio dei messaggi cifrati
- K è lo spazio delle chiavi
- $E : PT \times K \to CT$ è la funzione di cifratura
- $D : CT \times K \to PT$ è la funzione di decifratura

Alcune considerazioni Lo spazio delle chiavi deve essere abbastanza grande da impedire in pratica un attacco bruteforce. Deve essere **semplice** (tempo polinomiale da una DTM) costruire un messaggio cifrato e **estremamente difficile** (tempo non polinomiale da una DTM) effettuare l'operazione inversa senza conoscere la chiave.

Tipi di Attacchi In generale distinguiamo tra queste categorie in base alle capacità di Eve:

- **ciphertext only**: Eve conosce solo c e vuole computare il messaggio m.
- **known plaintext**: Eve anche alcune coppie (m, c).
- **chosen plaintext**: Eve ha accesso alla cifratura di messaggi.
- **chosen ciphertext**: Eve ha accesso alla decifratura di messaggi.

Inoltre un crittosistema può essere considerato **parzialmente rotto** se possiamo porre alcuni limiti ai possibili valori della chiave riducendo lo spazio su cui fare il bruteforce; **completamente rotto** se possiamo ricavare direttamente m e/o k in modo arbitrario.

Simmetria e Asimmetria Nel contesto della crittografia in essere consideriamo **simmetrico** un crittosistema in cui la chiave di cifratura è uguale a quella di decifratura, **asimmetrico** il contrario. Una definizione più precisa vuole che la chiave di decifratura sia facile da ottenere a partire dalla chiave di cifratura.

Nei crittosistemi che non dipendono da una chiave esplicita di solito si considera la *chiave* come la funzione di cifratura/decifratura e vice versa.

1.2 Cifrari a Sostituzione

Un cifrario a sostituzione prevedere la sostituzione (sorpreso?) di una o più lettere con un altro set di lettere sulla base di una chiave K.

Sono detti **monoalfabetici** i sistemi che cifrano una lettera del messaggio sempre con la stessa lettera del messaggio cifrato.

1.2.1 Cifrario di Cesare

È un cifrario simmetrico a sostituzione monoalfabetica dove presa una chiave $K \in [0, |\Sigma|]$ si shifta a sinistra (o a destra) ogni simbolo del messaggio. È celebre la versione con $K = 3$.

Problemi Come si è intuito il problema principale è lo spazio delle chiavi troppo piccolo che rendere possibile un attacco bruteforse.

Osservazioni Essendo un cifrario lineare monadico, le operazioni sono commutative e simmetriche.

1.2.2 Crittanalisi dei Sistemi Monoalfabetici

In generale questi sistemi sono molto vulnerabili, specie se il messaggio che si deve de/cifrare appartiene al linguaggio naturale (e in generale a qualche linguaggio strutturato). Nel primo caso spesso si utilizzano le frequenze delle lettere che variano di lingua in lingua e che permettono di ricavare facilmente spesso la chiave tramite l'assunzione che nei sistemi monoalfabetici i simboli associati alle lettere con maggior frequenza si ripetono a loro volta con maggior frequenza.

Contrattacco A volte si inseriscono delle lettere a bassa frequenza nel messaggio in modo strutturato per creare rumore. Più efficace è invece il fatto di associare alle lettere più simboli e scegliere randomicamente quale usare a runtime. Esempio: la lettera E ha 12% di frequenza? bene, allora scelgo 12 simboli dall'alfabeto CT e li associo ad E. La scelta del simbolo da usare può anche essere non casuale per interferire con le frequenze dei simboli.

1.2.3 Cifrario di Hill

È un crittosistema polialfabetico che sfrutta l'algebra lineare per cifrare un messaggio. La chiave è una matrice M che viene moltiplicata ad un messaggio disposto in forma matriciale. Chiaramente la chiave di decifrazione è M^{-1} È il primo crittosistema lineare a blocchi che vedremo.

- Example: assume that $m = HELP$, where M is the matrix given above.
  Then:

$$P_1 = \begin{bmatrix} H \\ E \end{bmatrix} = \begin{bmatrix} 7 \\ 4 \end{bmatrix} \qquad P_2 = \begin{bmatrix} L \\ P \end{bmatrix} = \begin{bmatrix} 11 \\ 15 \end{bmatrix}$$

$$C_1 = M \cdot P_1 = \begin{bmatrix} 3 & 3 \\ 2 & 5 \end{bmatrix} \cdot \begin{bmatrix} 7 \\ 4 \end{bmatrix} = \begin{bmatrix} 7 \\ 8 \end{bmatrix} = \begin{bmatrix} H \\ I \end{bmatrix}$$

$$C_2 = M \cdot P_2 = \begin{bmatrix} 3 & 3 \\ 2 & 5 \end{bmatrix} \cdot \begin{bmatrix} 11 \\ 15 \end{bmatrix} = \begin{bmatrix} 0 \\ 19 \end{bmatrix} = \begin{bmatrix} A \\ T \end{bmatrix}$$

Hence, $c = C_1 C_2 = HIAT$

Figure 1.2: Esempio

Problemi Per cominciare creare una chiave non è semplicissimo perchè ci si deve limitare all'uso di matrici invertibili (per ovvie ragioni).

Inoltre essendo un sistema lineare potrei creare una permutazione che invalida il messaggio o lo altera in qualche modo.

1.2.4 Cifrario di Playfair

La chiave è costruita prendendo le lettere dell'alfabeto e disponendole in una matrice 5x5 in modo casuale.

S	Y	D	W	Z
R	I	P	U	L
H	C	A	X	F
T	N	O	G	E
B	K	M	Q	V

Figure 1.3: Esempio

Quindi la cifratura avviene dividendo il messaggio in blocchi da due lettere(senza blocchi con due lettere uguali) e se le due lettere sono nella stessa riga nella chiave allora esse vengono shiftate verso destra (sempre nella chiave). Se sono nella stessa colonna vengono mosse verso il basso. Altrimenti visto che formano un rettangono di cui sono gli angoli, si considerano gli altri due angoli e sostituiscono con quelli.

La decifratura avviene eseguendo le operazioni senso opposto.

Problemi A parte il fatto che la chiave ha una struttura regolare ed è corta, quindi è piuttosto semplice da indovinare, il problema principale consiste nel fatto che ogni blocco viene cifrato sempre nello stesso modo nel messaggio. Lasciando quindi spazio ad attacchi sulle frequenze su digrammi.

1.2.5 Cifrario di Vigenere

È un cifrario di Cesare con una chiave di lunghezza $|K| > 1$, eventualmente lunga quanto il plaintext (ma può essere ripetuta ciclicamente per poter cifrare tutto).

Problemi È già più robusto ma si possono fare delle assunzioni grazie al fatto che se la chiave è più corta del messaggio allora più lettere saranno cifrate con la stessa chiave. Una volta indovinata la lunghezza della chiave si può sfruttare questa assunzione per fare un attacco basato su frequenze.

Per trovare la lunghezza della chiave si può usare il metodo di Kasisky che consiste nel notare ripetizioni di sequenze di lettere nel ciphertext e usare le distanze relative per identificare i possibili candidati per la lunghezza della chiave.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	T	S	U	V	W	X	Y	Z	A
C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F
H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y

Figure 1.4: Tabella di Appoggio

1.2.6 Macchina Enigma

È una macchina dotata di rotori che applicano sequenzialmente una sostituzione monoalfabetica. Dopo aver cifrato ogni lettera del messaggio i rotori compiono uno scatto di una posizione come un orologio, rendendolo complessivamente polialfabetico. La chiave è la posizione iniziale dei rotori.

La decifrazione avviene ponendo la stessa posizione iniziale e grazie ad un riflettore posto alla fine dell'ultimo rotore.

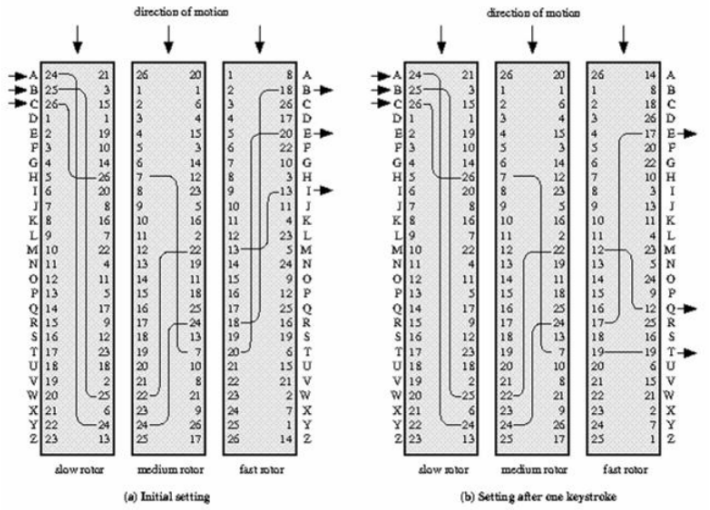

Figure 1.5: Struttura

Problemi Il sistema è molto robusto e la tecnica di Kasisky non può essere applicata essendo il ciphertext un prodotto di sostituzioni polialfabetiche.

Prima dei rotori c'è anche una plugboard che effettua una ulteriore sostituzione, ma l'effetto può essere annullato usando due macchine enigma in serie.

Alla fine il sistema è stato rotto costruendo più copie della macchina enigma in parallelo che effettuano un attacco bruteforce su un sottospazio delle chiavi.

Un ulteriore limite alle chiavi è il fatto che $c \neq m$ by design.

Chapter 2

AES e DES

2.1 DES

DES è il primo crittosistema standard. Utilizza una chiave da 56 bit, interleaved con un bit di non parità in terminazione ad ogni byte, ma noi consideriamo solo i 56 bits.

Generazione delle Chiavi Alla chiave originale viene applicata una permutazione e si divide il blocco di bit in due blocchi da 28 bit ciascuno C_0, D_0.

A partire da questi si ricavano ad ogni iterazione C_n, D_n da C_{n-1}, D_{n-1} tramite uno shift a sinistra seguendo una tabella che indica 1 o 2 shift in base all'iterazione.

La chiave dell'i-esima iterazione K_i è ottenuta tramite una permutazione che seleziona 48 bit a partire dalla concatenazione $C_i D_i$

Cifratura Il messaggio è diviso in blocchi da 64 bits ciascuno, e ogni blocco viene permutato e splittato in due blocchi da 32 bits L_0, R_0. Ad ogni iterazione si calcolano $L_n = R_{n-1}$ e $R_n - L_{n-1} \oplus f(R_{n-1}, K_n)$. Alla fine si applica l'inverso della permutazione iniziale.

La decifrazione segue i passi al contrario.

Funzione f Il blocco da 32 bits in ingresso è espanso a 48 bit secondo una matrice di permutazione e viene xorato con la chiave. Il risultato viene diviso in blocchi da 6 bit e ogni blocco viene trasformato in un blocco di 4 bit tramite delle S-boxes. Il valore da sostituire viene triangolato tramite x, y che sono rispettivamente *primo e ultimo bit*, e i 4 bit interni.

Pregi e Difetti È molto veloce (implementato con hardware dedicato per via delle operazioni sui singoli bit) e opera con permutazioni per introdurre non linearità. La versione corrente resiste alla crittoanalisi lineare e differenziale.

Ma ha una sicurezza di soli 56 bits e negli anni 90 già era stata costruita una macchina per romperlo. Non è più considerato sicuro.

Poi ci sono altri problemi, come chiavi omomorfiche o complementari che permettono di ricostruire il messaggio originale o lo mantengono intatto. (vedi 3DES).

3DES Per aumentare la sicurezza del DES si è pensato di concatenarlo 3 volte cifrando/decifrando (alternatamente) 3 volte il messaggio con 1, 2 o 3 chiavi diverse in base allo schema. Ma la sicurezza in termini di bit non supera i 112.

2.2 AES

AES è la seconda generazione di crittosistemi standard pensato per essere più robusto e versatile (utilizza chiavi da 128 fino a 256 bits) e per essere più efficiente, specie quando implementato come software (opera su byte invece che su bits).

Noi analizziamo l'AES-128.

State L'algoritmo mantiene uno state di 128 bits (16 bytes) che inizialmente contiene il messaggio originale e alla fine contiene il messaggio cifrato. Questo stato è rappresentato come matrice e si fanno operazioni su questa rappresentazione come rotazioni e shifting.

Operazioni

- State = m
- AddRoundKey
  - SubBytes
  - ShiftRows
  - MixColumns
  - AddRoundKey
- SubBytes
- ShiftRows
- MixColumns

AddRoundKey Computa lo xor tra lo state e la chiave derivata.

SubBytes Applica una trasformazione non lineare invertibile, che corrisponde all'inverso nel campo polinomiale $GF(2^8)$ sul polinomio irriducibile $x^8 + x^4 + x^3 + x + 1$.

Questo è equivalente ad utilizzare ancora una volta delle S-boxes che precalcolano il risultato.

ShiftRows È una rotazione verso sinistra delle righe di 0, 1, 2 o 3 posizioni in base all'indice della riga.

MixColumns Moltiplica lo state per una matrice invertibile nel campo $GF(2^8)$.

Chiavi Derivate Si necessita di 11 chiavi. Si inizializza un array di 10 word a 4 bytes con alcune costanti. Quindi si applicano rotazioni a sinistra all'interno della word e sostituzioni dei byte nella word con SubBytes.

Sicurezza AES adotta la strategia *wide trail* che rende le crittoanalisi lineari e differenziali difficili da applicare. L'unica parte non lineare sono le S-box.

Chapter 3

Crittosistemi Simmetrici

3.1 Modi di Operazione

3.1.1 Electronic CodeBook

È lo schema più semplice e meno sicuro: ogni blocco viene cifrato sempre allo stesso modo con la stessa chiave.

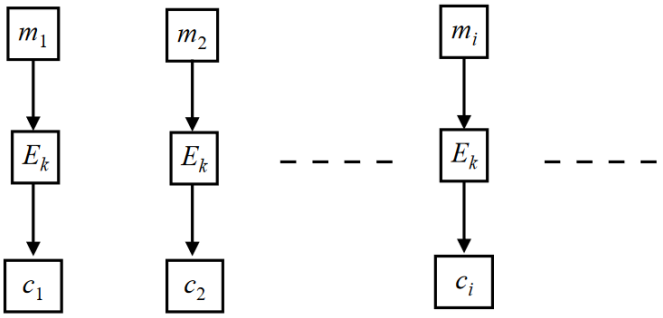

Figure 3.1: Schema

3.1.2 Cipher Block Chaining

Il blocco cifrato precedente è xorato con il blocco in chiaro successivo prima della cifratura.

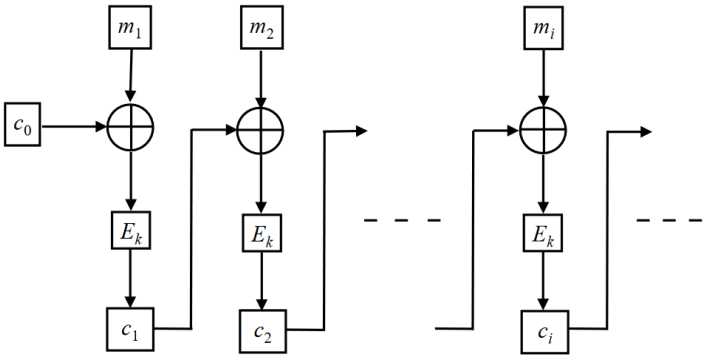

Figure 3.2: Schema

3.1.3 Output FeedBack

Si tiene un generatore di segreti che a partire da uno stato comune genera le chiavi con cui cifrare i blocchi. Nel caso in considerazione la chiave è generata tramite la cifratura ricorsiva delle chiavi dello stream, le quali vengono usate in xor con il blocco in chiaro per produrre il blocco cifrato.

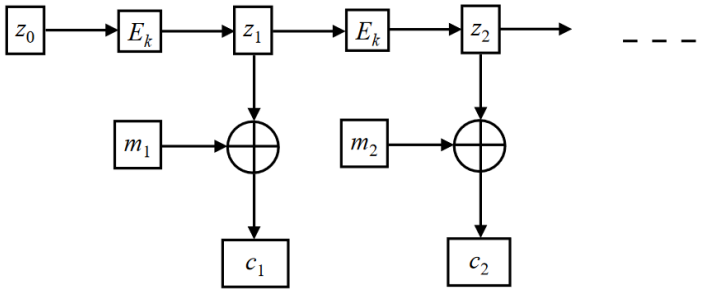

Figure 3.3: Schema

3.1.4 Cipher FeedBack

L'idea è simile alle precedenti, ma in questo caso le chiavi sono generate cifrando i blocchi cifrati precedenti e producendo i nuovi con uno xor come prima.

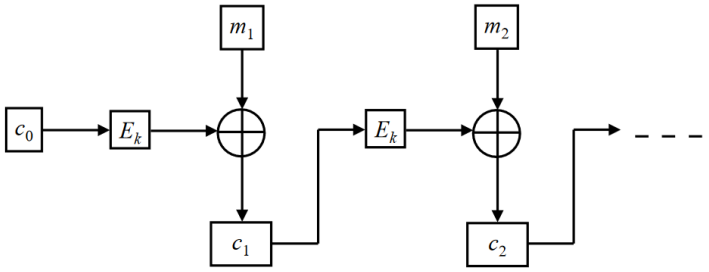

Figure 3.4: Schema

Una variante di questo schema prevede di usare una rotazione a sinistra di S bits e poi di selezionarli con una permutazione dopo la cifratura di questo stream.

3.1.5 Counter Mode

Può essere visto come una variante di OFB visto che in questo caso il segreto cifrato per ottenere le chiavi è ottenuto a sua volta tramite un contatore che parte da uno stato comune.

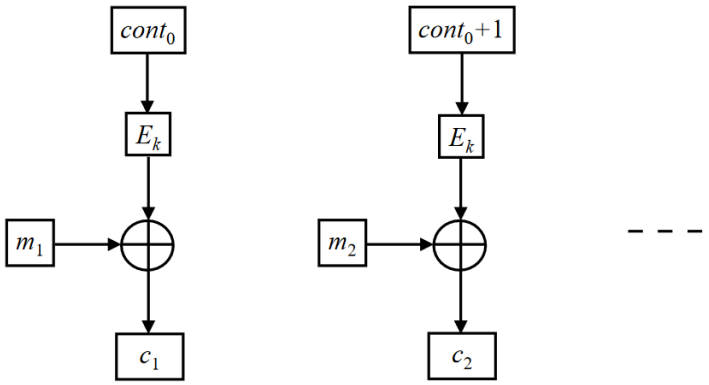

Figure 3.5: Schema

3.1.6 Alcune Considerazioni

Per via della loro natura non sequenziale o para sequenziale, gli schemi ECB, OFB, CTR possono essere implementati con cifrature sui blocchi in parallelo. Sempre per questa ragione se un blocco viene corrotto durante la comunicazione e quelli seguenti sono intatti non ho problemi a leggerli.

3.2 Confusione e Diffusione

Confusione Le modifiche interne al messaggio, allo stato o alla chiave manipolati devono essere imprevedibili. Questo è realizzato tramite delle sostituzioni complesse.

Diffusione Per rendere difficile fare analisi statistica di un crittosistema si distribuisce il peso dei simboli del messaggio in modo che ognuno di essi influisca in qualche molto su molti simboli dell'output. Questo è spesso ottenuto tramite delle permutazioni (possibilmente non lineari, ex: AES, DES).

3.3 Rete di Sostituzioni e Permutazioni

Shannon ha dato un modello di sistema ideale per adottare Conf. e Diff. Di base abbiamo una sequenza di iterazioni in cui si cifra il buffer con la chiave dell'iterazione, poi si applicano delle sostituzioni su sotto porzioni, quindi si permuta il risultato totale. La prima e l'ultima operazione di una SPN è costituita dalla cifratura con la chiave per rendere difficile fin da subito la crittoanalisi.

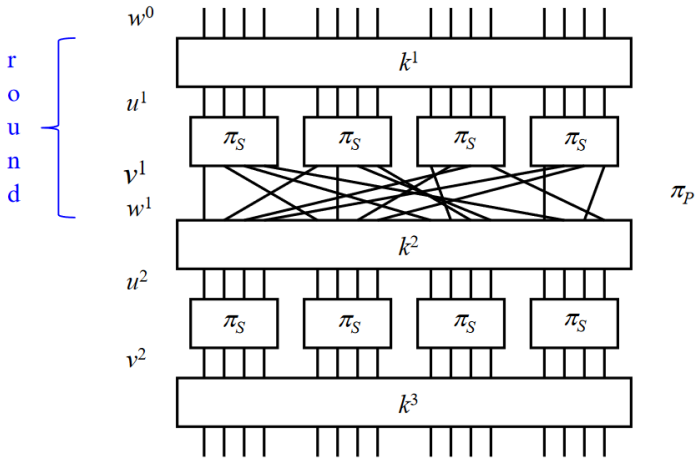

Figure 3.6: Schema

3.4 Rete di Feistel