Capitolo 4

L'ENTROPIA E L'EQUIVOCAZIONE NELLA SEGRETEZZA E NELL'AUTENTICAZIONE

1. Crittografia storica

> ... dans les choses de physique il n'y a pas d'autres demonstration que dans le dechiffrement d'une lettre. Ou ayant fait des suppositions sur quelques legeres conjectures, si l'on trouve qu'elles se verifient en suite, de sorte que suivant ces suppositions de lettres on trouve des paroles ben suivies dans la lettre, on tient d'une certitude tres grande que les suppositions sont vraies, quoy qu'il n'y ait pas autrement de demonstration, et qu'il ne soit pas impossible q'on n'en puisse y avoir d'autres plus veritables (Ch. Huygens, 1673)

1. Lo scarabeo d'oro e il gioco della zara.

In uno dei più famosi racconti di Edgar Allan Poe, *The golden bug* o *Lo scarabeo d'oro*, il protagonista si serve del principio statistico di *massima verosimiglianza* per forzare un enigmatico crittogramma e impadronirsi del favoloso tesoro del pirata Kidd. Il sistema usato da Kidd per occultare le istruzioni che portavano al tesoro è un *cifrario a sostituzione*: ogni lettera del testo in chiaro viene sostituita da un simbolo in cifra secondo una regola fissa; ad esempio le A in chiaro diventano altrettanti 5 nel testo in cifra. Il difetto di questo sistema è che il simbolo in cifra "eredita" la frequenza del suo "progenitore" in chiaro, ciò che porta alla decifrazione illegittima (senza che sia nota la chiave) non appena la lunghezza del testo sia tale da aver permesso alle frequenze di assestarsi sui loro valori tipici; in pratica bastano una trentina di lettere. Nel racconto di Poe c'è perfino una lista delle singole lettere ordinate secondo la loro frequenza decrescente in inglese:

$$EAOIDHNRSTUYCFGLMWBKPQXZ$$

con una presumibile fusione grafica di I e J, e di U e V. Un vero e proprio istogramma delle frequenze tuttavia manca; ad esempio Solomon Kullback, un illustre statistico di questo secolo, dà i dati seguenti in una vecchia monografia destinata alla Signal Intelligence School statunitense:

$$E: \ 12,996\%, \ T: \ 9,190\%, \ N: 7,950\%, \ R: 7,576\%, \ O: 7,528\%$$

e così via. L'idea di base del principio di massima verosimiglianza è presto detta: è verosimile che il simbolo in cifra che come ordine di frequenza occupa la n-esima posizione provenga dal simbolo in chiaro che occupa la stessa posizione nell'istogramma "naturale": ad esempio, se come frequenza il terzo simbolo in cifra è l'1, esso proviene verosimilmente da una N in chiaro, o, in seconda battuta, da una lettera in chiaro le cui frequenza tipica sia vicina a quella della N, come la R o la O. La spiegazione "scientifica" del perchè questo metodo crittanalitico (ossia di scardinamento dei sistemi di cifra) funzioni è la legge dei grandi numeri, che predica appunto l'assestarsi progressivo sulle probabilità delle frequenze osservate sperimentalmente; essa venne enunciata e dimostrata rigorosamente da Jacob Bernoulli nel quarto e ultimo libro dell'*Ars conjectandi*, pubblicata postuma a Basilea nel 1713. Del resto un'enunciazione qualitativa sì, ma molto precisa, del principio di massima verosimiglianza era già stata data da Galileo nel suo *Dialogo sopra i due massimi* sistemi del 1632, in epoca di "preistoria", come si usa dire, del calcolo della probabilità e della statistica: ricordiamo che il calcolo delle probabilità nasce "ufficialmente" nel 1654, anno in cui Blaise Pascal e Pierre Fermat si scambiarono sette famose lettere sul cosiddetto *problema dei punti* nei giochi d'azzardo (come si deve dividere in maniera equa la posta se la partita viene interrotta?); tre anni dopo comparve il primo trattato "moderno" di probabilità, il *De ratiociniis in ludo aleae* di Christiaan Huygens, autore della lettera da cui abbiamo tratto il nostro motto; la data di nascita della statistica è invece il 1662, anno di pubblicazione delle *Natural and political observations made upon the bills of mortality* di John Graunt, dove vengono elaborate estese tabelle di dati demografici. Il *Golden bug* comparve nel 1843, quando la statistica "scientifica" camminava ormai da tempo sulle proprie gambe: più o meno di quegli anni è un'opera-chiave che segna il trionfo della "mentalità statistica" (e della sua invadenza), e cioè la tesi di baccellierato *Sur l'homme et le développement de ses facultés* di Adolphe Quetelet, che fa dell'homme moyen, e non del singolo individuo in carne e ossa, il legittimo protagonista della "fisica sociale" (oggi diremmo della sociologia). Il racconto di Poe, così leggibile e popolare, porta tuttavia a sospettare che il principio di massima verosimiglianza, e dunque la legge dei grandi numeri che gli è soggiacente, abbiano corso anche al di fuori delle sedi ufficiali della statistica paludata: del resto lo stesso Bernoulli, nel capitolo quarto dell'ultima parte dell'*Ars conjectandi*, asserisce che anche gli uomini semplici e incolti hanno una percezione "istintiva" della validità della legge dei grandi numeri.

Queste premesse portano a chiedersi quale sia il rilievo dei metodi crittanalitici nella preistoria della statistica. Non si tratta di quisquiglie, visto che nella storia "ufficiale" della statistica e della probabilità la crittanalisi viene completamente ignorata, il che ci pare uno stravolgimento della realtà dei fatti, stravolgimento tanto più incredibile se si pensa all'attenzione che gli storici della probabilità e della statistica hanno dedicato ad argomenti di ben minor portata, come l'astragalo, un dado d'osso asimmetrico con cui si divertivano gli antichi Greci, oppure, per essere irriverenti, i commenti d'epoca al passo del Purgatorio che accenna al gioco della zara (canto VI, vv.1-3):

Quando si parte il gioco della zara,
colui che perde si riman dolente,
repetendo le volte, e tristo impara ...

2. Le tredici regole di Cicco Simonetta.

La più nota storia della crittografia è senza dubbio quel voluminoso e leggibilissimo capolavoro della "storiografia giornalistica" che sono i *Codebreakers* di David Kahn, pubblicati nel 1967 ("to break a code" significa "forzare un codice"; l'attributo di "giornalistico" non è peggiorativo, ma vuole solo sottolineare la disinibita mescolanza, nei *Codebreakers*, di storia e di pettegolezzo, con divagazioni un po' da ora di matematica, un po' da ora del tè). Il più antico trattato europeo di crittanalisi menzionato da Kahn sono le tredici *Regulae ad extrahendum litteras ziferatas sine exemplo* di Cicco Simonetta, segretario e consigliere degli Sforza a Milano. Esse furono scritte nel 1474 e sono oggi conservate nella Biblioteca nazionale di Parigi. Chi le sfogli viene subito colpito dalla presenza di termini tipici della statistica, e più in particolare tipici del principio di massima verosimiglianza: *estima!*, *verisimile est*, *maxime repetuntur*, *conjectura!*, *raro contingunt*, *frequenter representant*, *verisimilius est*. Siamo tuttavia al livello dello *Scarabeo d'oro* di Poe: le regole hanno carattere meramente qualitativo; per servirci dell'espressione di Bernoulli, di Cicco va più apprezzato l'istinto che la scienza. Il suo trattato ha sì un posto nella preistoria della statistica (e male fanno gli storici della statistica a tacerne), ma è appunto preistoria, è "pre-scientifico". Nonostante la limitatezza del contributo di Cicco, la crittografia europea nasce proprio in quegli anni con la chiara visione (lo sottolineiamo con forza) che i cifrari a sostituzione semplice, come quello usato dall'incolto pirata Kidd, sono gracili, tant'è che le regole statistiche di Cicco Simonetta già bastano a forzarli. A partire dal Cinquecento, il compito di chi costruisce sistemi di cifra è dunque, molto coscientemente, quello di "imbrogliare la statistica". Le strade seguite sono grosso modo due: quella dei cifrari polialfabetici, in cui più regole di sostituzione vengono usate a turno, e quella dei cifrari omofonici, in cui c'è sì un'unica regola di sostituzione, ma a ciascuna lettera in chiaro corrispondono più simboli in cifra fra i quali si può scegliere di volta in volta "a caso": così, alla lettera in chiaro molto frequente E potrebbero corrispondere i tre omofoni % , $*$ e &, da usare a turno nel crittogramma (va da sè che in un sistema omofonico il numero dei simboli in cifra a disposizione del crittografo deve superare quello dei simboli in chiaro). Se vogliamo insistere con le date di nascita convenzionali, e forse eccedere, possiamo ben fissare la nascita della crittografia europea attorno al 1470, quando Leon Battista Alberti, nella veste di funzionario dei servizi segreti vaticani, scrive il *Modus scribendi in ziferas*, in cui illustra appunto il principio della polialfabeticità.

Al solito ogni storia ha la sua preistoria: nelle *Vite dei Cesari* Svetonio attribuisce l'uso di cifrari a sostituzione semplice a Giulio Cesare e a Cesare Augusto; ma si può risalire più indietro: per esempio allo *sheshach* della Bibbia, anch'esso un cifrario a sostituzione semplice. Nell'antichità, tuttavia, se c'è la *crittografia*, che insegna a cifrare, non ci sembra esser traccia della *crittanalisi*, che insegna a decifrare senza la chiave (crittografia e crittanalisi sono le due facce, opposte ma inscindibili, della *crittologia*). E' interessante chiedersi quale sia stato il contributo arabo, visto che la rinascita scientifica nell'Europa rinascimentale è su quel contributo che si innesta. Che esso sia stato notevole è ovvio già dalla lettura dei *Codebreakers*. Recenti ritrovamenti hanno mostrato che il contributo arabo è stato in realtà eccezionale, tale da mettere in dubbio la legittimità delle nostre date di nascita: il 1662 rimane sì la data di nascita della statistica scientifica europea, ma forse la data di nascita della statistica scientifica tout court va anticipata di un millennio.

3. Da al-Kindi ad al-Kadi: un millenio di crittanalisi statistica.

Abu Yusuf Yaqub ibn Ishaq al-Kindi, filosofo e scienziato di Bagdad del nono secolo, chiamato dai suoi correligionari "il filosofo degli arabi", fu un commentatore di Aristotele; scrisse anche un trattato di ottica, *De aspectibus*, che fu tradotto tre secoli dopo da Gherardo di

Cremona. Il suo *Manoscritto sulla decifrazione dei messaggi cifrati*, dato per perduto da Kahn, è stato ritrovato nell'Archivio Ottomano Sullaimaniyyah di Istanbul, ed è stato pubblicato nell'87 a cura dell'Accademia Araba di Damasco (davvero perduto sembra invece il *Libro sui messaggi in cifra*, del linguista al-Khalil, antecedente al Manoscritto di al-Kindi). Sull'opera di al-Kindi e su altri contributi dei crittologi arabi riferisce con solida competenza tecnica Ibrahim al-Kadi, un ingegnere elettronico che ha lavorato al *Communication Satellite Planning Center* dell'Università di Stanford, e che si occupa di crittografia informatica. Il portavoce scelto è *Cryptologia*, una rivista che presumibilmente non è molto letta dagli storici della probabilità. Nel *Manoscritto* di al-Kindi la legge dei grandi numeri, pur essendo presentata come un fatto meramente empirico, ha un carattere consciamente quantitativo: il trattato comprende una tabella di frequenze delle lettere arabe, valutate su un campione di sette pagine scritte, ossia di 3667 lettere: i risultati numerici reggono sorprendentemente bene se confrontati con quelli di al-Kadi, posteriori di più di un millennio. Che i metodi statistici non funzionino sui testi brevi è ad al-Kindi perfettamente chiaro.

Anche per i crittologi arabi l'imperativo è dunque quello di imbrogliare la statistica: ed essi giungono ai cifrari polialfabetici ben prima dell'Alberti. La tabella di Vigenère, che onora un crittografo francese del Cinquecento e che è ben nota a chi si occupa di cifrazione polialfabetica, compare già nell'opera di ibn ad-Duraihim due secoli prima. Questa stessa opera segna tuttavia un'involuzione rispetto alla maturità statistica di al-Kindi: la frequenza viene sostituita da un'informazione più povera, il rango, ossia il numero d'ordine nell'ordinamento per frequenze decrescenti (come nello *Scarabeo d'oro*). Ciò toglie flessibilità all'applicazione del principio di massima verosimiglianza. Possiamo solo speculare se di questa statistica ormai "degenerata" i crittografi europei del Quattrocento fossero o meno venuti a conoscenza.

Nell'articolo di al-Kadi molte priorità vengono orgogliosamente attribuite agli arabi: ad essi spetterebbe la paternità della crittanalisi, della statistica, della linguistica computazionale e perfino della combinatorica, che già al-Khalil poneva con sicurezza al servizio delle sue indagini linguistiche. Il contributo degli Arabi all'algebra è ben noto e traspare anche nell'uso di temini come appunto *algebra*, o *cifra*, *zero* e *algoritmo*. Ma in questo caso gli Arabi spesso fanno più che altro da tramite di conoscenze preesistenti. Molto più originale appare il loro contributo alla statistica e alla crittanalisi, due discipline indubbiamente congeniali al loro spirito pratico, da matematica applicata, che spesso viene contrapposto al disinteresse e alla "purezza" dei Greci.

4. Un filo spezzato, o un filo consunto?

Rispetto ai modelli arabi, il trattato di Cicco è dunque una copia sbiadita e degenere; la crittanalisi statistica, qualitativa e gracile, non "comunica" in alcun modo con la cultura scientifica dell'Europa del Cinque, Sei e Settecento, e in particolare è del tutto irrilevante nel processo di formazione e coagulo delle idee probabilistiche e statistiche che proprio allora si vanno affinando e consolidando. Questa, almeno, è l'impressione che si ricava dalla lettura dei trattati di storia della probabilità e di storia della statistica; vogliamo però fare qualche osservazione che potrebbe istillare il germe del dubbio. Attualmente, per merito o demerito della pirateria informatica, la crittografia è un argomento alla moda; fino a pochi anni fa, tuttavia, di essa si occupavano solo gli specialisti (essenzialmente i militari), o chi aveva l'hobby dell'enigmistica, come a suo tempo Edgar Allan Poe. La crittologia, insomma, non faceva parte del bagaglio culturale della persona colta. Ben diversa era la situazione fino al Seicento: abbiamo già ricordato Leon Battista Alberti, ma l'elenco dei crittologi illustri contiene altri nomi familiari nella storia della cultura europea; ne citeremo alcuni, un po' a caso: Girolamo Cardano, il cui *Liber de ludo aleae* è uno dei trattati più noti nella preistoria della probabilità, Giambattista Della Porta, Francesco Bacone, Athanasius Kircher, che fra l'altro si occupò della decifrazione dei geroglifici, John Wallis, uno dei precursori del calcolo infinitesimale. La crittografia insomma ha l'aria di essere stata un argomento conosciuto, di cui gli intellettuali-poligrafi discutevano: probabilmente l'interesse era alimentato da un'atmosfera in cui la magia e il mistero erano ancora apprezzati (e difatti nel razionale Settecento la crittografia decade, per poi riprendersi in una nuova veste, quella "ingegneristica" delle macchine cifranti). Si dice che lo sviluppo della statistica sia stato ritardato dalla scarsità di estese tabelle di dati, tabelle che oggi sono diventate fin troppo familiari. Ma un testo in cifra è per l'appunto un'estesa tabella di dati che in maniera del tutto spontanea induce all'analisi statistica: talmente spontanea che la crittanalisi si è rivelata un'ottimo sistema per introdurre la statistica nelle scuole dei primi gradi, a partire dalle elementari. Che la crittanalisi statistica fosse insomma "invisibile" nella cultura del Seicento, il secolo-chiave nella formazione della probabilità e della statistica moderne, non ci sembra così scontato: forse è la storiografia moderna che non è riuscita a vedere.

5. Battere la statistica si può, ma è una vittoria di Pirro.

La "battaglia contro la statistica" segna il percorso della ricerca crittologica dai trattatisti arabi al giorno d'oggi: ne segnaleremo alcune tappe salienti. La fig.3 mostra un cifrario usato da parte cattolica nella guerra fra Maria Stuarda ed Elisabetta d'Inghilterra: niente di particolarmente raffinato, si tenta di ingannare la statistica introducendo a casacccio nel crittogramma delle zeppe o *nulle*, ossia simboli di disturbo privi di corrispondente (che poi le "lettere" usate per formare il crittogramma siano di fantasia è suggestivo dal punto di vista estetico, ma pressochè irrilevante da quello crittografico). Nella fig.4 compare un vero sistema omofonico, stavolta di parte inglese: a differenza del precedente esso non venne forzato. Le ruote delle figure 5 e 6 servono invece per attuare, o "implementare", sistemi polialfabetici. Scelta una parola-chiave da tenere rigorosamente segreta, come FIORDALISO, essa va scritta sotto il messaggio in chiaro, ripetetendola quante volte serve: per formare il crittogramma, il messaggio in chiaro e la chiave vengono "sommati" lettera dopo lettera. Per sommare una lettera in chiaro e una lettera di chiave, diciamo la D e la F, si pone la F del cerchio interno (cerchio della chiave) sotto la A di quello esterno (cerchio del messaggio in chiaro), e si sostituisce la D in chiaro con la lettera che le corrisponde sul cerchio interno: in questo caso la lettera di crittogramma è la I (stiamo usando i cerchi di Tritemio, che confondono la U e la V; il cerchio interno, quello numerato, è un contatore che serve a non perdere la sincronizzazione; contiamo che il lettore apprezzi la magnificenza barocca della macchina rivale, quella del Della Porta, che da vero secentista non sa rinunciare ai simboli fantastici).

A U U E L E N A I L D U C A S T A S E R A
F I O R D A L I SO F I O R D A L I S O F
F F L X O E A I DB I F Q R X T L DZ H F

Un elemento importante di un sistema polialfabetico è il suo periodo, ossia la lunghezza della parola chiave, che qui è dieci: ogni dieci lettere si ricomincia daccapo. I cifrari polialfabetici resistettero fino alla fine dell'ottocento, quando l'ufficiale prussiano Friedrich von Kasiski propose un metodo per la determinazione del periodo che permette di ricondurne la decrittazione a quella dei cifrari monoalfabetici; il metodo di Kasiski, tuttavia, funziona solo quando il messaggio in chiaro ha una lunghezza molto maggiore di quella del periodo (non funzionerebbe sul nostro messaggio: una chiave lunga dieci diventa debole per messaggi di qualche centinaio di lettere). Nel nostro secolo, macchine elettromeccaniche di cifra, come l'*Enigma* usato dai nazisti nella seconda guerra mondiale, consentirono di attuare cifrari polialfabetici con periodo enorme. Tuttavia l'Enigma venne forzato dagli Alleati: per agevolare il loro lavoro crittanalitico gli inglesi costruirono poderose macchine di calcolo, i *Colossi*, che sono fra i primissimi calcolatori della storia, preceduti solo da quelli realizzati da Konrad Zuse, che per nostra fortuna i nazisti non seppero usare a loro vantaggio. La crittanalisi sta dunque anche alle radici dell'informatica.

I cifrari polialfabetici sono alla base del cifrario perfetto: in questo caso la chiave è lunga quanto il messaggio e va scelta in maniera totalmente casuale, ad esempio scrivendo le venti e passa lettere dell'alfabeto sul bordo di una ruota della fortuna (di una roulette) che va fatta ruotare tante volte quanto serve (in pratica, mittente e ricevente devono generare una lunghissima tabella di lettere a caso, più lunga della lunghezza globale dei messaggi che prevedono di scambiarsi, e girare il mondo ciascuno armato di una copia della tabella: le lettere di chiave usate vanno via via depennate e non possono venir riutilizzate). Il cifrario perfetto è impermeabile a qualunque metodo crittanalitico, sia esso statistico, sia esso di qualunque altro tipo: la statistica e più in generale la crittanalisi si possono dunque sconfiggere! Si badi: la nostra affermazione non è una valutazione ottimistica, bensì un rigoroso teorema di matematica, che vogliamo esplicitamente enunciare: Se si usa il cifrario perfetto, il crittogramma di per sè, in assenza della chiave, non fornisce nessuna informazione sul messaggio in chiaro. Non stupirà che in tempi di guerra fredda era proprio a un cifrario perfetto che si affidava la protezione della "linea calda" fra Washington e Mosca. Stante l'eccessiva lunghezza della chiave, il cifrario perfetto è estremamente farraginoso, e non può venir usato nella vita di tutti i giorni. Purtroppo esiste un altro teorema di matematica, stavolta di sapore negativo, che recita: Se si vogliono ottenere le prestazioni del cifrario perfetto, la complessità della chiave dev'essere perlomeno pari alla complessità del messaggio. E invece una proprietà irrinunciabile per un sistema di cifra realistico è proprio la brevità e la maneggevolezza della chiave. Si è pensato di sostituire le lettere casuali con lettere (o con cifre) pseudocasuali, che potrebbero venir generate in maniera automatica col calcolatore: tuttavia i normali algoritmi di generazione, dei quali si era così orgogliosi, e che in maniera pulita e deterministica portano a sfilze di simboli apparentemente prive di qualsiasi struttura o regolarità statistica, sono miseramente caduti di fronte alle inappagabili esigenze crittografiche. Come si possa sostituire alla casualità effettiva la casualità simulata è uno scottante tema di ricerca della crittologia odierna.

Oggidì la crittografia ha smesso l'uniforme: basta pensare alla necessità di proteggere da una parte le cartelle cliniche memorizzate nelle banche di dati, dall'altra le transazioni elettroniche di denaro, per capire quanto ampia sia la rivoluzione sociologica che ha investito la crittografia contemporanea. Le nuove esigenze hanno posto la ricerca crittologica, un tempo così marginale, al centro del dibattito scientifico, là dove si rimeditano concetti-cardine della nostra cultura come la casualità e la complessità. Essa interseca e feconda i campi più imprevedibili, come la teoria dei numeri, che passa, o passava, per essere la più "pura" delle discipline matematiche: la sicurezza del sistema RSA (dai nomi dei suoi inventori, Rivest, Shamir e Adleman) è fondata sulla complessità intollerabile degli algoritmi che consentono di decomporre un numero intero "elevato" (di duecento cifre decimali, per esempio) nei suoi fattori primi, algoritmi che funzionano solo in linea di principio, perchè in pratica i tempi di calcolo sono da era geologica. Oppure la fisica quantistica: un sistema di cifra che attualmente comincia a entrare in uso è basato sul principio di indeterminazione di Heisenberg. A livello subatomico l'osservazione modifica la realtà: nel nostro caso l'osservatore è il crittanalista, il quale, modificando la realtà, segnala con ciò stesso la sua indesiderata intrusione. La crittologia ha ritrovato il suo antico splendore ed è ridiventata cultura.

2. Multicodifica deterministica e probabilistica (omofonica, frazionata)

> La multicodifica è tipica sia dei codici di segretezza sia di quelli di autenticazione: sono disponibili vari codici di trascrizione, e una *chiave* decide qual codice viene usato.

$\mathcal{X}$ e $\mathcal{Y}$ sono, rispettivamente, l'alfabeto primario e quello secondario; $|\mathcal{X}| \leq |\mathcal{Y}|$. Nel contesto crittografico le lettere primarie vengono chiamate *messaggi (in chiaro)* o *stati della sorgente*, le lettere secondarie *parole di codice*, o *messaggi in cifra*, o *crittogrammi*.

Definizione. Un *multicodice deterministico* è una famiglia di applicazioni iniettive (di codici di trascrizione) $f_z : \mathcal{X} \to \mathcal{Y}$; l'indice z è il valore osservato di una variabile aleatoria $Z \in \mathcal{Z}$, chiamata *chiave* o *regola di codifica*; $|\mathcal{Z}| < +\infty$.

In altri termini è l'estrazione della chiave Z che determina quale codice f_z venga in effetti usato. Si noti che, fissata z, rimane fissata la funzione di *decodifica*, o *decifrazione*, $g_z : \mathcal{Y} \to \mathcal{X} \cup \{?\}$, che (essenzialmente) inverte la funzione di *codifica*, o *cifrazione* f_z. Si pone $g_z(y) = ?$ se $y \neq f_z(x)\ \forall x$; ? è un simbolo esterno a $\mathcal{X}$, il cui significato è: sotto la chiave z il crittogramma y non si presenta mai. Se $\mathcal{X} = \mathcal{Y}$, o comunque se $|\mathcal{X}| = |\mathcal{Y}|$, il multicodice si dice *endomorfo* (senza *espansione dei dati*).

> Beninteso anche i messaggi in chiaro hanno un loro comportamento statistico, poichè sono generati da una sorgente d'informazione. In accordo con la pratica corrente supporremo sempre che la variabile aleatoria chiave sia *indipendente* dal processo stocastico che genera i messaggi (insomma

la chiave viene estratta senza sapere cosa combini la sorgente). Nella trattazione che segue escludiamo i modelli *sequenziali*, in cui la chiave evolve anch'essa nel tempo. Nel seguito, per evitare precisazioni banali escluderemo messaggi in chiaro inutili (di probabilità zero), chiavi inutili (di probabilità zero) e crittogrammi inutili (che non codificano nulla).

La nozione di multicodice deterministico può venir generalizzata a quella di *multicodice probabilistico* (o anche multicodice *omofonico*, o multicodice *frazionato*): fissata la chiave z, ogni messaggio in chiaro x ha più parole di codice y che lo possono codificare; la scelta è determinata da una legge probabilistica detta *strategia di frazionamento*. Per consentire univocamente la decodifica, sono proibite le *collisioni*: una volta fissata la chiave, la stessa parola di codice non può codificare due messaggi in chiaro diversi. Nella definizione che segue $f_z(x)$ indica l'insieme degli *omofoni* di x sotto z, ossia l'insieme dei crittogrammi y che, scelta la chiave z, hanno probabilità positiva di codificare il messaggio in chiaro x.

Definizione. Un *multicodice probabilistico* indiciato dalla chiave aleatoria Z è una famiglia di variabili aleatorie (di codici probabilistici) $F_z(x)$ a valori in $\mathcal{Y}$ tali che per ogni chiave z e per ogni coppia di messaggi distinti x e x' si abbia $f_z(x) \cap f_z(x') = \emptyset$.

Rimangono ben definite le funzioni di decodifica g_z, che sono comunque *deterministiche*. Pensando alla sorgente dei messaggi, il messaggio x è anch'esso l'uscita di una variabile aleatoria X (del messaggio in chiaro aleatorio X, che, lo rammentiamo, è indipendente dalla chiave aleatoria Z): rimane allora ben definita la *terna aleatoria* XYZ, dove Y è il crittogramma aleatorio. La strategia di frazionamento è determinata da probabilità condizionate del tipo $\text{Prob}\{Y = y|X = x, Z = z\}$; si ha $\text{Prob}\{Y = y|X = x, Z = z\} > 0$ sse $y \in f_z(x)$. Una politica di solito seguita è quella di estrarre le chiavi del tutto a caso, per cui, per così dire, "una regola di codifica vale l'altra":

Definizione. Un multicodice si dice *canonico* se la chiave aleatoria Z è uniforme: $\text{Prob}\{Z = z\} = \frac{1}{|\mathcal{Z}|}$.

Esercizio. Un multicodice è deterministico sse $| \cup_x f_z(x)| = |\mathcal{X}|$ per ogni chiave z. Un multicodice endomorfo è sempre deterministico.

3. Il modello shannoniano della crittografia: segretezza incondizionata e segretezza computazionale

L'articolo con cui Claude Shannon fondò la crittografia "scientifica", *Communication theory of secrecy systems*, precedette (e influenzò) quello con cui Shannon fondò nel 1948 la teoria dell'informazione senza segretezza, *A mathematical theory of communication*. Tuttavia, per la comprensibile severità dei vincoli di segreto militare in tempo di guerra, l'articolo sulla crittografia, riveduto e corretto, venne pubblicato per secondo, nel 1949.

Nel modello teorico-informazionale di Shannon un *sistema di cifra* o *cifrario* viene modellato mediante un multicodice, o, se si vuole, mediante una terna aleatoria XYZ, dove $X \in \mathcal{X}$ è il *messaggio in chiaro* aleatorio, $Y \in \mathcal{Y}$ è il *crittogramma* (il messaggio in cifra) aleatorio e $Z \in \mathcal{Z}$ è la *chiave* aleatoria. L'incertezza della spia, che auspicabilmente è "grande", viene valutata mediante tre *equivocazioni* (entropie condizionate):

$H(X|Y)$, *equivocazione sul messaggio*
$H(Z|Y)$, *equivocazione sulla chiave*
$H(Z|XY)$, *equivocazione in chiaro sulla chiave*, o *ambiguità della chiave*, utile nel caso di attacchi "con testo in chiaro".

Sviluppando in due modi $H(XZ|Y)$ e ricordando che $X = g_Z(Y)$, per cui $H(X|ZY) = 0$, si dimostra una prima significativa identità che lega le tre equivocazioni:

$$H(Z|Y) = H(X|Y) + H(Z|XY)$$

Poichè $H(Z|XY) \geq 0$, l'equivocazione sulla chiave non può essere inferiore a quella sul messaggio, una disuguaglianza tanto immediata quanto profonda:

$$H(Z|Y) \geq H(X|Y)$$

Veniamo alla seconda identità fondamentale. E' sempre $H(X|YZ) = 0$ (la cifrazione è deterministica), per cui $H(YZ) = H(XYZ)$, per cui ancora, messaggio e chiave essendo variabili aleatorie indipendenti:

$$H(YZ) = H(X) + H(Z) + H(Y|XZ)$$

L'entropia condizionata $H(Y|XZ)$ è nulla sse il multicodice è deterministico; si noti che essa è un'"equivocazione spuria", nel senso che non contribuisce in alcun modo alla sicurezza del cifrario.

Le due definizioni che seguono, se soddisfatte, fanno la felicità del crittografo:

Definizione. Se $H(X|Y) = H(X)$ il cifrario si dice *perfetto.*

In un cifrario perfetto il crittogramma di per sè, in assenza della chiave, "non dice nulla sul messaggio". L'obiettivo sembra troppo ambizioso: tuttavia, come vedremo nella sezione 6, esso è ottenibile, sia pure a un prezzo molto alto.

Definizione. Se $H(Z|Y) = H(Z)$ il cifrario si dice *ideale.*

Sciogliendo $H(YZ)$ al primo membro della seconda identità fondamentale si trova:

Corollari. Se $H(X) + H(Y|XZ) \geq H(Y)$ il cifrario è ideale. In particolare, se $H(X) \geq H(Y)$ il cifrario è ideale.

Di solito il messaggio e il crittogramma "crescono nel tempo", e la loro lunghezza n aumenta via via che la sorgente emmette messaggi; ciò nonostante, come nel caso del cifrario del pirata Kidd, il sistema può rimanere *a chiave fissa*, potenzialmente per l'eternità. Se il messaggio in chiaro X_n è lungo n (la sorgente è stata attiva n istanti di tempo), altrettanto lungo è il crittogramma Y_n. Rimangono allora definite le equivocazioni

$$H(X_n|Y_n),\ H(Z|Y_n),\ H(Z|X_n, Y_n)$$

il cui andamento in funzione del parametro temporale n è essenziale per valutare la validità del sistema di cifra adottato. Per un'ovvia proprietà dell'entropia condizionata, che decresce in senso debole quando si tolgono i condizionamenti, in un sistema a chiave fissa le due equivocazioni sulla chiave al crescere di n non possono che *decrescere*. Purtroppo quando l'equivocazione sulla chiave è diventata "trascurabile", a maggior ragione lo è quella sul messaggio (per la disuguaglianza fondamentale).

4. Cifrari elementari

> Supporre che il comportamento statistico della sorgente dei messaggi in chiaro sia noto e controllabile è in crittografia un'ipotesi tanto ottimistica quanto pericolosa; per ragioni didattiche esamineremo tuttavia il comportamento dei due cifrari più elementari che si possano concepire, quello a sostituzione e quello a trasposizione, supponendo che la sorgente dei messaggi sia bernoulliana. I cifrari che consideriamo sono *completi*, nel senso che *tutte* le chiavi disponibili vengono effettivamente usate, e *canonici*.

Nel cifrario a sostituzione le chiavi sono $K!$, tante quante sono le permutazioni dell'alfabeto in chiaro; stante la canonicità $H(Z) = \log K!$. Nel cifrario a trasposizione bisogna fissare un *periodo* $\Pi \geq 1$; le chiavi sono tante quante sono le permutazioni (chiamate *trasposizioni* in questo contesto) dei blocchi di lunghezza Π in cui viene spezzato il messaggio (per semplificare la trattazione, senza perdere nulla dell'essenziale supporremo che la lunghezza n del messaggio sia proprio un multiplo di Π). Per canonicità $H(Z) = \log \Pi!$. La sorgente in chiaro è bernoulliana, e dunque $H(X_n) = nH(P)$.

Cifrari a sostituzione. Fra le componenti di P potrebbero esserci dei "pareggi": dichiariamo *equivalenti* due chiavi che differiscono solo in quanto rimescolano lettere equiprobabili: $z_i \sim z_j$ sse $z_i(a) = z_j(b) \Rightarrow P(a) = P(b)\ \forall a, b \in \mathcal{X}$. Rimane allora definita la *chiave aleatoria ingrossata* $\hat{Z}$; ogni chiave ingrossata $\hat{z}$ è costituita da una classe di chiavi equivalenti z. Più in dettaglio, se in P le probabilità distinte sono $\kappa \leq K$, e compaiono rispettivamente $k_1,\ k_2,\ \ldots,\ k_\kappa$ volte $(k_1 + k_2 + \ldots + k_\kappa = K)$, ciascuna classe ingrossata $\hat{z}$ è costituita da esattamente $\chi = \chi(P) \doteq k_1!\ k_2!\ \ldots\ k_\kappa!$ chiavi "semplici". Il parametro $\chi = \chi(P)$ si rivelerà essenziale; esso può venir calcolato a partire da P.

Esercizio $1 \leq \chi(P) \leq K!$; $\chi(P) = 1$ sse tutte le K componenti di P sono distinte, $\chi(P) = K!$ sse P è uniforme.

Esercizio. $H(Z|\hat{Z}) = \log \chi(P)$.

Esercizio. L'entropia condizionata $H(\hat{Z}|Y_n)$ è infinitesima con n (usa la legge dei grandi numeri, come nello *Scarabeo d'oro*, e la disuguaglianza di Fano).

Esercizio. L'equivocazione $H(Z|X_n, Y_n)$ è infinitesima con n (l'unico "incidente" possibile è la presenza in $\underline{x}$, e dunque in $\underline{y}$, di meno di K lettere distinte, ciò che è via via meno probabile per la legge dei grandi numeri; usa di nuovo la disuguaglianza di Fano).

Esercizio. $H(Z|\hat{Z}, Y_n) = H(Z|\hat{Z})$ (calcola $\mathrm{Prob}\{Z = z|\hat{Z} = \hat{z}, Y_n = \underline{y}\} = \frac{\mathrm{Prob}\{Z=z, Y_n=\underline{y}\}}{\mathrm{Prob}\{\hat{Z}=\hat{z}, Y_n=\underline{y}\}}$ e verifica che il risultato è indipendenta da y; in effetti si trova 0 oppure $\frac{1}{\chi(P)}$ a seconda che sia $z \notin \hat{z}$ oppure $z \in \hat{z}$).

$\hat{Z}$ essendo una funzione deterministica di Z, si ha $H(Z|Y_n) = H(Z, \hat{Z}|Y_n) = H(\hat{Z}|Y_n) + H(Z|\hat{Z}, Y_n)$; l'uguaglianza fondamentale che lega le tre equivocazioni dà allora:

$$H(X_n|Y_n) = \log \chi(P) + \epsilon_0(n)$$

$$H(Z|Y_n) = \log \chi(P) + \epsilon_1(n)$$

$$H(Z|X_n, Y_n) = \epsilon_2(n)$$

dove s è posto $H(\hat{Z}|Y_n) \doteq \epsilon_1(n) \to 0$, $H(Z|X_n, Y_n) \doteq \epsilon_2(n) \to 0$, $\epsilon_0(n) \doteq \epsilon_1(n) - \epsilon_2(n) \to 0$ (si ha $\epsilon_0(n),\ \epsilon_2(n) > 0$; $\square$!).

Alla lunga la sicurezza del sistema è legata solo alla numerosità dei pareggi; il sistema è comunque fragile contro gli attacchi con testo in chiaro. Se la probabilità P è uniforme, $H(Z) = \log K!$, e il cifrario è ideale.

Cifrari a trasposizione. L'esercizio che seque mostra che, limitatamente alle sorgenti bernoulliane, i cifrari a trasposizione sono *ideali*:

Esercizio. $H(Z|Y_n) = H(Z) = \log \Pi!$. Cenni: dimostra dapprima che si ha $\mathrm{Prob}\{Y_n = \underline{y}\} = \mathrm{Prob}\{X_n = \underline{y}\}$; usa l'uguaglianza $\mathrm{Prob}\{X_n = g_z(\underline{y})\} = \mathrm{Prob}\{X_n = \underline{y}\}$, ricordando che $\underline{y} = f_z(\underline{x})$ traspone $\underline{x} = g_z(\underline{y})$, e che la sorgente è bernoulliana. Ne viene $H(Y_n) = H(X_n)$; usa il corollario della seconda identità fondamentale.

Nel seguito avremo bisogno dell'*indice di coincidenza* $C(P) \doteq \sum_i p_i^2$.

Esercizio $\frac{1}{K} \le C(P) \le 1$; $C(P) = \frac{1}{K}$ sse P è uniforme, $C(P) = 1$ sse P è deterministico. Cenni: a sinistra calcola $0 \le \sum_i (p_i - \frac{1}{K})^2 = C(P) - \frac{1}{K}$. A destra eleva al quadrato $1 = \sum_i p_i$ e trascura i doppi prodotti; non ci sono doppi prodotti da trascurare solo nel caso deterministico.

L'indice di coincidenza è stato chiamato *energia informazionale* da O. Onicescu; la coincidenza formale con la formula dell'energia in fisica, accompagnata al parallelismo fra le due entropie, informazionale e termodinamica, è allettante, ma si è rivelata, almeno fino a ora, piuttosto inconcludente. Il complemento a 1 dell'indice di coincidenza si chiama anche *indice di Gini-Simpson*, ed è un rivale dell'entropia di Shannon come indice adeguato per valutare la *diversità*, o *varietà*, di un popolazione (ecologica, sociologica). Si noti che, a differenza dell'indice di Gini-Simpson, meno il *logaritmo* dell'indice di coincidenza ha gli stessi estremi dell'entropia di Shannon, con le stesse condizioni per il raggiungimento dei loro valori.

Veniamo all'ambiguità della chiave, $H(Z|X_n, Y_n)$. L'unico "incidente" ($\square$!) che possa pregiudicare la ricostruzione della chiave Z a partire dal messaggio X_n e dal crittogramma Y_n è il fatto che in ciascuno dei $\nu = \frac{n}{\Pi}$ blocchi di X_n ci siano (almeno) due posizioni fisse $i < j$ in cui la lettera in chiaro si ripete; ossia: $x_i = x_j$, $x_{\Pi+i} = x_{\Pi+j}$, $x_{2\Pi+i} = x_{2\Pi+j}$, $x_{3\Pi+i} = x_{3\Pi+j}$, ..., $x_{n-\Pi+i} = x_{n-\Pi+j}$; $1 \leq i < j \leq \Pi$. Indichiamo con $\mathcal{E}_{i,j}$ l'evento appena descritto e poniamo $\mathcal{E} \doteq \cup_{1 \leq i < j \leq \Pi} \mathcal{E}_{i,j}$.

Esercizio. Le probabilità $P^n(\mathcal{E}_{i,j})$ e $P^n(\mathcal{E})$ tendono esponenzialmente a 0 con n. Cenni: calcola la prima probabilità e usa la subadditività della probabilità ($P^n(\mathcal{E}_{i,j}) = C(P)^\nu$, $P^n(\mathcal{E}) \leq \binom{\Pi}{2} C(P)^\nu$).

In base alla disuguaglianza di Fano, tanto basta per concludere che l'ambiguità della chiave è un infinitesimo $\epsilon(n)$ che tende a 0 con n, o meglio con ν; l'uguaglianza fondamentale che lega le tre equivocazioni dà:

$$H(X_n|Y_n) = \log \Pi! - \epsilon(n)$$

$$H(Z|Y_n) = \log \Pi!$$

$$H(Z|X_n, Y_n) = \epsilon(n)$$

Ciò mostra che i cifrari a trasposizioni sono robusti se il periodo è elevato, purchè beninteso la sorgente sia bernoulliana.

Le sorgenti reali sono meno gentili, e l'esperienza mostra che i cifrari a trasposizione danno forti delusioni. I principi "storici" della sostituzione e della trasposizione non sono tuttavia obsoleti. Sostituzioni e trasposizioni sono usate *in serie* nel celebre cifrario DES (= *Data Encryption Standard*), un vero e proprio *standard* della crittografia commerciale contemporanea.

5. Cifrari perfetti e (pseudo)casualità

Si veda il libro sulla Crittografia dell'autore.

6. Cifrari ideali e ridondanza

In questa sezione piuttosto teorica "scomparirà" il parametro *tempo*; in realtà può ben essere che il messaggio nominalmente unidimensionale sia in effetti un testo generato da una sorgente a comportamento statistico capriccioso, e che dunque la complessità degli schemi proposti sia molto maggiore di quanto non sembri a prima vista.

Esempio deprimente. Qualunque cifrario *degenere*, ossia a una sola chiave, è ideale: $H(Z|Y) = H(Z) = O$.

Di solito nella definizione di idealità si aggiunge qualcosa per evitare l'intrusione di banalità come questa. Preferiamo conservare la definizione "pulita" e "simmetrica" rispetto alla definizione di perfezione; il lettore è tuttavia avvertito che un codice ideale è interessante solo se l'entropia della chiave è "elevata", ossia, nella vantaggiosa ipotesi che il codice sia canonico, solo se il numero delle chiavi è "elevato". E' per questo che nei due esempi che seguono dovremo "ottimizzare" la chiave, ossia massimizzarne la numerosità.

Dal corollario alla seconda identità fondamentale segue subito:

Teorema. Qualunque cifrario endomorfo che cifri un messaggio aleatorio X totalmente casuale è ideale.

Dimostrazione. Si ha $H(X) = \log K = \log|\mathcal{Y}| \geq H(Y)$.

Ottimizzazione della chiave. Se usiamo in maniera canonica tutte le chiavi (distinte) disponibili si ottiene $H(Z) = \log K!$; in questo caso il cifrario è anche perfetto.

Il corollario suggerisce il seguente progetto: codifica la sorgente primaria con un codice di sorgente che elimini "quasi tutta" la ridondanza. La sorgente codificata (secondaria) è "quasi" totalmente casuale, e dunque *qualunque* cifrario deterministico endomorfo è "quasi" ideale. A un risultato simile si perviene anche, non con la compressione dei dati, bensì con la loro *espansione omofonica*.

Esercizio-complemento. Supponiamo che le lettere primarie (i messaggi in chiaro di $\mathcal{X}$) abbiano probabilità *razionali*: $\text{Prob}(a_i) = p_i = \frac{n_i}{n}$. A ciascuna lettera in chiaro a_i assegniamo n_i omofoni, ognuno con la stessa probabilità (condizionata) $\frac{1}{n_i}$, per cui $H(Y|X = a_i, Z = z) = \log n_i$. Si ha $H(X) + H(Y|X,Z) = \log|\mathcal{Y}| \geq H(Y)$, e dunque il cifrario è ideale ($\square$! $H(X) + H(Y|X, Z = z) = H(X) + \sum_i \frac{n_i}{n} \log n_i = \log n = \log|\mathcal{Y}|$). Il numero massimo di chiavi distinte è $\binom{n}{n_1\, n_2 \ldots n_K}$ ($\square$! l'ordine in cui vengono specificati gli omofoni è irrilevante); l'ottimizzazione della chiave porta a usare un cifrario canonico in cui $H(Z) = \log\binom{n}{n_1\, n_2 \ldots n_K}$.

Le divagazioni sulla ridondanza e sull'omofonia hanno scarso impatto sul piano pratico: i parametri "tempo" e "complessità" sono stati colpevolmente sottovalutati; i risultati dipendono dal comportamento statistico dei messaggi in chiaro in maniera determinante (non sono affatto "universali").

7. Un'altra divagazione sulle disuguaglianze informazionali

La disuguaglianza fondamentale sulla divergenza (la sua non-negatività), può esser rafforzata nel caso in cui i due vettori di probabilità coinvolti, P e Q, abbiano degli zeri. Siano $\mathcal{X}_P$ e $\mathcal{X}_Q$ i *supporti* di P e di Q, ossia i sottinsiemi di $\mathcal{X}$ su cui le loro componenti sono strettamente positive. Posto $\alpha \doteq \sum_{a_i \in \mathcal{X}_P} q_i$, ipotizziamo per il momento $\alpha > 0$ (ossia $\mathcal{X}_P \cap \mathcal{X}_Q \neq \emptyset$). P può ben esser visto come un vettore di probabilità su $\mathcal{X}_P$; basta ora “gonfiare” le componenti di Q su $\mathcal{X}_P$ (passare da q_i a $\frac{q_i}{\alpha}$ se $a_i \in \mathcal{X}_P$) e applicare la disuguaglianza fondamentale per trovare:

$$D(P\|Q) \geq -\log \alpha$$

con uguaglianza sse $q_i \propto p_i$ per $a_i \in \mathcal{X}_P$ (□ ! la divergenza su $\mathcal{X}$ è di fatto una somma su $\mathcal{X}_P$). La nuova disuguaglianza funziona anche per $\alpha = 0$, nel qual caso assume la forma $+\infty \geq +\infty$.

Sia ora data una coppia di variabili aleatorie A e B. La mutua informazione $I(A \wedge B)$, come ben sappiamo, è una divergenza, e più precisamente è la divergenza $D(P_{AB}\|P_A \times P_B)$. Posto $\rho(a,b) = 1$ sse $\text{Prob}\{A = a, B = b\} > 0$, altrimenti $\rho(a,b) = 0$ (la funzione ρ indica il supporto di P_{AB}), la disuguaglianza diventa:

$$\sum_{a,b:\rho(a,b)=1} \text{Prob}\{A = a\}\, \text{Prob}\{B = b\} \geq 2^{-I(A \wedge B)}$$

o anche, poichè $\sum_{a:\rho(a,b)=1} \text{Prob}\{A = a\} = \text{Prob}\{\rho(A,b) = 1\}$:

$$\sum_{b \in \mathcal{B}} \text{Prob}\{B = b\}\, \text{Prob}\{\rho(A,b) = 1\} \geq 2^{-I(A \wedge B)}$$

con uguaglianza sse A e B sono indipendenti.

Il primo membro è il valor medio di $\text{Prob}\{\rho(A,b) = 1\}$ calcolato con i “pesi” $\text{Prob}\{B = b\}$; poichè il valor massimo eccede il valor medio, un ovvio corollario della disuguaglianza è:

$$\max_{b \in \mathcal{B}} \text{Prob}\{\rho(A,b) = 1\} \geq 2^{-I(A \wedge B)}$$

con uguaglianza sse A e B sono indipendenti e se inoltre la probabilità $\text{Prob}\{\rho(A,b) = 1\}$ è costante in b (e se inoltre il valor medio coincide con il valore massimo).

8. Frodi all’autenticità dei messaggi: impersonazioni e sostituzioni

Il multicodice non verrà più usato per ragioni di segretezza, bensì per garantire l’autenticità dei messaggi. L’avversario è ora un *contraffattore* che inserisce nel sistema messaggi codificati nella speranza la chiave corrente si lasci frodare, li *autentichi* e li inoltri lungo le maglie del sistema. Il problema è drammatico nel caso del trasferimento automatico di denaro, tanto per dare un esempio. In accordo con il pessimismo del principio di Kerckhoffs, supporremo di avere a che fare con un nemico intelligente: ciò dà ragione dei massimi che compaiono nelle definizioni delle probabilità di frode.

Definiamo la *matrice di autenticazione* χ, ad entrate binarie $0-1$, le cui righe sono intestate alle chiavi $z \in \mathcal{Z}$ e le cui colonne sono intestate alle parole di codice $y \in \mathcal{Y}$:

$$\chi(z,y) = 1 \;\Leftrightarrow\; \text{Prob}\{Z = z, Y = y\} \neq 0$$

Diremo che la chiave z *autentica* la parola di codice y.

Frode di impersonazione. L'avversario vuole inserire una parola di codice y che venga autenticata dalla chiave corrente Z. La probabilità di frode è:

$$P_I = \max_y \mathrm{Prob}\{\chi(Z, y) = 1\}$$

Frode mirata. L'avversario vuole inserire una parola di codice y che venga autenticata dalla chiave corrente Z e che venga decodificata a un messaggio in chiaro x appartenente a un sottinsieme $\mathcal{M} \subseteq \mathcal{X}$ di messaggi per lui vantaggiosi. La probabilità di frode è:

$$P_{\mathcal{M}} = \max_y \mathrm{Prob}\{\chi(Z, y) = 1, g_Z(y) \in \mathcal{M}\}$$

Frode di sostituzione. L'avversario, dopo aver intercettato la parola di codice legale c (sicuramente autenticata!), vuole inserire una parola di codice y che venga autenticata dalla chiave corrente, ma che *non* venga decodificata allo stesso modo di c. La probabilità di frode è:

$$P_S(c) = \max_y \mathrm{Prob}\{\chi(Z, y) = 1, g_Z(y) \neq g_Z(c)|Y = c\}$$

Se la codifica è deterministica si ha ($\square$!):

$$P_S(c) = \max_{y \neq c} \mathrm{Prob}\{\chi(Z, y) = 1|Y = c\}$$

L'analogia strutturale fra le tre definizioni va esplicitata, per evitare di ripetere tre volte lo stesso lavoro. Introdurremo a tal scopo due nuove matrici binarie *coperte* dalla χ (che hanno gli stessi zeri della χ più eventualmente altri). Nel caso della frode mirata, giriamo a 0 gli 1 di χ in posizione (z, y) con $g_z(y) \notin \mathcal{M}$; sia $\chi_{\mathcal{M}}$ la matrice che si ottiene. Nel caso della frode di sostituzione, giriamo a 0 gli 1 di χ in posizione (z, y) con $g_z(y) = g_z(c)$; sia χ_c la matrice che si ottiene (nel caso deterministico per ottenere χ_c basta girare a 0 gli 1 della colonna c). In altre parole conserviamo gli 1 solo là dove la frode effettivamente riesce. Ciò fatto, possiamo riscrivere:

$$P_{\mathcal{M}} = \max_y \mathrm{Prob}\{\chi_{\mathcal{M}}(Z, y) = 1\}$$

$$P_S(c) = \max_y \mathrm{Prob}\{\chi_c(Z, y) = 1|Y = c\}$$

Grazie all'analogia strutturale, possiamo avvalerci di uno "schema astratto" che sussume molte frodi di autenticazione concrete, non solo le tre da noi elencate.

Consideriamo una variabile aleatoria S sull'insieme $\mathcal{Z}$ delle chiavi, che chiameremo il *selettore* aleatorio, e una matrice binaria Φ a righe intestate in $\mathcal{Z}$ e a colonne intestate in $\mathcal{Y}$.

Definizione. Uno *schema di frode* è una coppia costituita da un *selettore aleatorio* $S \in \mathcal{Z}$ e da una *matrice di frode* $\Phi: \mathcal{Z} \times \mathcal{Y} \to \{0,1\}$. La *probabilità di frode* è:

$$P = \max_y \text{Prob}\{\Phi(S,y) = 1\}$$

L'euristica dietro la definizione dovrebbe ormai esser chiara: il selettore è la chiave del codice di autenticazione nella misura in cui essa è nota al contraffattore che vuole perpetrare la frode (*chiave a posteriori*), gli uno nella matrice di frode indicano quali crittogrammi contraffatti se inseriti illegalmente nel sistema riescono a frodarlo. Sono irrilevanti (e depennabili) gli eventuali selettori di probabilità zero e gli eventuali crittogrammi a colonna di soli zero.

Una matrice stocastica W si dice *simile* alla matrice di frode Φ se ha esattamente gli stessi 0 e gli stessi non-0 di Φ, ossia se può venir ottenuta da Φ trasformando gli 1 in probabilità positive e conservando gli 0. In altre parole le matrici stocastiche simili a Φ coprono e sono coperte da Φ; scriveremo $W \simeq \Phi$. Ogni matrice W simile a Φ consente di estendere la variabile aleatoria S a una *coppia* aleatoria S, F, la cui seconda componente a valori in $\mathcal{Y}$ verrà chiamata la *contraffazione* aleatoria; $\text{Prob}\{F = y | S = z\} = W(z,y)$. Il corollario che chiude la sezione precedente dà subito:

$$P \geq 2^{-I(S \wedge F)} \equiv 2^{-I(P_S, W)} \quad \forall\, W \simeq \Phi$$

Ciò è vero per *qualunque* matrice stocastica W simile a Φ, ma è più interessante per quelle che danno un membro di destra "grande"; la limitazione migliore è $P \geq 2^{-\inf I(S \wedge F)}$, dove l'estremo inferiore viene appunto calcolato sull'insieme delle matrici W simili a Φ (tale insieme è topologicamente aperto, ciò che spiega l'inf al posto del max; $\square$!).

Esercizio. La chiusura topologica dell'insieme $\{W : W \simeq \Phi\}$ è costituta dall'insieme di tutte le matrici stocastiche W che hanno *almeno* tutti gli zeri di Φ ($\square$!), ossia a quelle che sono *coperte* da Φ; in simboli $\{W : W \preceq \Phi\}$ (in una posizione in cui c'è il vincolo della positività possiamo avvicinarci allo zero "quanto ci pare").

Esercizio: $\inf_{W \simeq \Phi} I(S \wedge F) = \min_{W \preceq \Phi} I(S \wedge F)$ (è un esercizio di routine di analisi o di topologia; rammenta che la mutua informazione è una funzione continua di W).

Ne viene un risultato *normativo* il quale asserisce che la probabilità di frode e comunque non inferiore a una certa funzione di P_S e di Φ. Posto:

$$R(S,\Phi) = \min_{W \preceq \Phi} I(S \wedge F) = \min_{W \preceq \Phi} I(P_Z, W)$$

si ha:

Teorema inverso sulle frodi di autenticazione:

$$P \geq 2^{-R(S,\Phi)}$$

Esercizio. Tenendo presente la chiusa della sezione 8, enuncia un criterio di uguaglianza per il teorema inverso.

Funzioni come le R generalizzano l'entropia in un ramo della teoria dei codici di sorgente che si chiama *teoria della distorsione* o *rate-distortion theory*, dove ci si accontenta di decodificare le parole di codici a messaggi in chiaro che siano *simili*, e non necessariamente uguali a quello codificato. I legami fra la teoria della distorsione e la teoria dell'autenticazione sembrano però ridursi al lato formale, ciò che lascia perplessi, visto che le coincidenze formali di solito hanno spiegazioni sostanziali che talora sono illuminanti.

La distribuzione di S è quella che è, e dunque, essendo $I(S \wedge F) \leq H(S)$:

Corollario. $P \geq 2^{-H(S)}$

Corollario. $P \geq \frac{1}{|\mathcal{Z}_S|}$

La numerosità della chiave è dunque essenziale in un buon codice di autenticazione. Se la matrice Φ è *equilibrata*, ossia ha lo stesso numero ν di 1 in ogni riga z di probabilità positiva, fra le matrici $W \simeq \Phi$ c'è quella in cui $\text{Prob}\{F = y|S = z\} = \frac{1}{\nu}$ in ogni posizione non nulla. In questo caso $H(F|S) = \log \nu$ e $I(S \wedge F) = H(F) - \log \nu \leq \log \frac{|\mathcal{Y}^*|}{\nu}$, dove $\mathcal{Y}^* \subseteq \mathcal{Y}$ indica l'insieme delle parole di codice che sono contraffazioni "efficaci", ossia contraffazioni che frodano almeno un selettore di probabilità positiva. Ne viene:

Corollario combinatorio. Se la matrice di frode è equilibrata e ogni riga ha peso ν:

$$P \geq \frac{\nu}{|\mathcal{Y}^*|}$$

Una caratteristica piacevole del corollario è che esso prescinde dalle probabilità del selettore e si accontenta della sua numerosità.

Esercizio. Riformula il teorema inverso nel caso delle frodi di impersonazione, mirata e di sostituzione. Si ha $\Phi = \chi,\ \chi_{\mathcal{M}},\ \chi_c$ nei tre casi; il selettore S e la chiave Z coincidono per l'impersonazione e per la frode mirata, nella sostituzione di c il selettore ha invece la distribuzione $\text{Prob}\{S = z\} = \text{Prob}\{Z = z|Y = c\}$.

Esercizio (disuguaglianza di Simmons per l'impersonazione):

$$P_I \geq 2^{-I(Y \wedge Z)}$$

(basta controllare che la matrice stocastica W delle probablità $\text{Prob}\{Y = y|Z = z\}$ sia simila a χ: ciò è ovviamente vero, visto che la matrice di autenticazione del codice XYZ è costruita a partire proprio da quelle probabilità!)

Nella letteratura si è spesso considerata la probabilità di sostituzione "assoluta" $P_S = \sum_{c \in \mathcal{Y}} \text{Prob}\{Y = c\} P_c$, calcolata come *media* rispetto a tutte le parole di codice intercettabili da parte dell'avversario: una deplorevole deviazione dal principio di Kerckhoffs, che avrebbe richiesto la probabilità di frode *massima*.

Esercizio $P_S \geq 2^{-H(Z|Y)}$ (il primo corollario al teorema dà $P_c \geq 2^{-H(Z|Y=c)}$; applica la disuguaglianza di Jensen alla funzione esponenziale, che è convessa).

Esercizio (codifica deterministica). Se il multicodice è deterministico $\nu = |\mathcal{X}|$ in χ, $\nu = |\mathcal{M}|$ in $\chi_{\mathcal{M}}$, $\nu = |\mathcal{X}| - 1$ in χ_c. La limitazione combinatoria implica allora che in un buon codice di autenticazione il numero delle parole di codice, e dunque il costo del codice, dev'essere "elevato" in rapporto al numero dei messaggi in chiaro. Infatti, tenuto anche conto che nel caso della sistituzione "sparisce" una colonna (nel caso della frode mirata ci siamo dovuti accontentare della $|\mathcal{Y}^*| \leq |\mathcal{Y}|$):

$$P_I \geq \frac{|\mathcal{X}|}{|\mathcal{Y}|}, \quad P_{\mathcal{M}} \geq \frac{|\mathcal{M}|}{|\mathcal{Y}|}, \quad P_c \geq \frac{|\mathcal{X}| - 1}{|\mathcal{Y}| - 1}$$

La limitazione combinatoria per i codici canonici e deterministici.

Se il selettore aleatorio S è uniforme sul suo supporto $\mathcal{Z}_S \subseteq \mathcal{Z}$ e se la matrice di frode Φ ha lo stesso numero ν di 1 per ogni riga di probabilità positiva, alla limitazione combinatoria si può giungere direttamente con calcoli più semplici, che fra l'altro fanno subito capire quale sia il criterio di uguaglianza; (selettori uniformi e matrici di frode equilibrate sono tipici delle codifiche canoniche e deterministiche, che in pratica sono le più usuali). Sotto $\mathcal{Y}^* \subseteq \mathcal{Y}$ indica di nuovo l'insieme delle contraffazioni "efficaci". Si ha:

$$|\mathcal{Y}^*| P \equiv |\mathcal{Y}^*| \max_{y \in \mathcal{Y}^*} \text{Prob}\{\Phi(S, y) = 1\} \geq$$

$$\geq \sum_{y \in \mathcal{Y}^*} \text{Prob}\{\Phi(S, y) = 1\} = \sum_{y \in \mathcal{Y}^*} \frac{|z : \Phi(z, y) = 1|}{|\mathcal{Z}_S|} = \frac{1}{|\mathcal{Z}_S|} \sum_{z \in \mathcal{Z}_S} |y : \Phi(z, y) = 1| = \nu$$

per cui:

$$P \geq \frac{\nu}{|\mathcal{Y}^*|}$$

con uguaglianza sse $\text{Prob}\{\Phi(S, y) = 1\}$ è costante per $y \in \mathcal{Y}^*$.

Il criterio di uguaglianza spinge a cercare matrici di autenticazione equilibrate "in verticale" oltre che in "orizzontale"; strutture combinatorie di questo tipo sono note da tempo nella geometria combinatoria, e hanno la loro origine in applicazioni statistiche (pianificazione degli esperimenti, o *design theory*).

Esercizio-complemento: la limitazione "storica" di Gilbert, MacWilliams e Sloane.

Idealmente un buon codice di autenticazione dovrebbe essere resistente rispetto a tutti i tipi di frode, e non solo, diciamo, all'impersonazione o alla sostituzione. Nella letteratura hanno ricevuto attenzione i codici resistenti *sia* all'impersonazione *sia* alla sostituzione: pessimisticamente si suppone che l'avversario, dei due attacchi, scelga quello rispetto a cui il codice è più fragile. La probabilità di sostituzione cui ci riferiremo è quella *media*.

Di seguito $P_{I,S}$ la probabilità più alta fra P_I e P_S: $P_{I,S} = \max[P_I, P_S]$; già sappiamo che è (cfr. il secondo corollario): $\min[P_I, P_S] \geq \frac{1}{|\mathcal{Z}|}$.

Esercizio. $P_{I,S} \geq 2^{-\frac{1}{2}H(Z)}$. Il massimo è $\geq$ della media aritmetica: $P_{I,S} \geq \frac{1}{2}P_I + \frac{1}{2}P_S$; se minoriamo P_I con $2^{-I(Z\wedge Y)}$ e P_S con $2^{-H(Z|Y)}$ (cfr. sopra), la disuguaglianza di Jensen applicata alla funzione esponenziale dà appunto: $P_{I,S} \geq 2^{-\frac{1}{2}\{I(Z\wedge Y+H(Z|Y)\}} \equiv 2^{-\frac{1}{2}H(Z)}$

Come ovvio corollario ritroviamo una disuguaglianza "combinatoria" del 1974:

$$P_{I,S} \equiv \max[P_I, P_S] \geq \frac{1}{\sqrt{|\mathcal{Z}|}}$$