Information Theoretic Secrecy

Imre Csiszár

MTA Alfréd Rényi Institute of Mathematics,
Budapest, Hungary

ESIT 2015

Introduction

Basic problems of information theory (IT): information transmission

- reliably over unreliable channels (Shannon 1948)
- securely over insecure channels (Shannon 1949)

Contemporary cryptography relies mostly on computational complexity for security. Shannon mentioned this as an alternative, but his IT approach offers provable security even against adversaries of unlimited computational power.

Variety of security tasks: authentication, commitment, secure computation, watermarking, etc. We focus on generating a secret key, using some kind of common randomness as resource.

We present mathematical techniques admitting to derive fundamental limits called secret key capacities, expected to gain practical relevance soon.

Overview

- Brief history
- One-time pad, security index
- Common randomness, secret key, privacy amplification
- Secret key capacities for two-user source models
- Extractor lemma
- Two-user channel models
- Wiretap channel
- Multiuser models
- Outlook
- Applications

Lecture based on Csiszár and Körner: Information Theory, Second Edition, Cambridge University Press, 2011, Chapter 17.
Broader coverage: Liang, Poor and Shamai, Information Theoretic Secrecy, Now Publishers, 2009.

Brief history

- First IT approach to secrecy: Shannon 1949
- Wiretap channel: Wyner 1975, Csiszár and Körner 1978
- Public communication as a resource for secrecy: Bennet, Brassard and Robert 1988, Maurer 1993, Ahlswede and Csiszár 1993
- Security against active adversaries: Maurer and Wolf 1997
- Secret key generation for multiple users: Csiszár and Narayan 2004, 2008

In this lecture the adversary will be assumed passive: listens to the legal parties' communication, but unable to interfere with it.

One-time pad

Traditionally, to securely transmit a message M one encrypts it using a key K known to the receiver but not to the adversary.

Suppose M and K are independent random variables (RVs) with values in a finite group $\mathcal{K}$. Typically $\mathcal{K} = \{0, 1\}^\ell$. If K is uniformly distributed on $\mathcal{K}$ then encrypting M by $M + K$ guarantees perfect security: the ciphertext $M + K$ is stochastically independent of the message M. If an adversary ignorant of the key learns the cyphertext, she learns nothing about the message.

Disadvantage: A secret key is an expensive resource. To encrypt several messages, a new key is needed for each.

Notation

A (probability) distribution P on a finite set, say $\mathcal{X}$, is a collection of nonnegative numbers (probabilities) $P(x), x \in \mathcal{X}$ with $\sum_{x \in \mathcal{X}} P(x) = 1$. Its entropy is

$$H(P) \triangleq -\sum_{x \in \mathcal{X}} P(x) \log P(x); \quad 0 \leq H(P) \leq \log |\mathcal{X}|.$$

log denotes base 2 logarithm (base e logarithm is ln), entropy is measured in bits.

Measures of difference of distributions P, Q on a set, say $\mathcal{X}$:

variation distance

$$|P - Q| \triangleq \sum_{x \in \mathcal{X}} |P(x) - Q(x)|; \quad 0 \leq |P - Q| \leq 2$$

(some authors use a factor $1/2$), and

I-divergence or relative entropy

$$D(P||Q) \triangleq \sum_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{Q(x)}; \quad D(P||Q) \geq 0.$$

Random variables, their values, resp. sets of possible values (assumed finite sets) are denoted by corresponding upper and lower case, resp. script letters. E.g., a RV X has possible values $x \in \mathcal{X}$. For several RVs we often write XYZ etc. for (X, Y, Z) etc. Entropy, joint and conditional entropies of RVs are defined via their (joint and conditional) distributions.

$$H(X) \triangleq H(P_X), \quad P_X(x) \triangleq \Pr\{X = x\}$$

$$H(X, Y) = H(XY) \triangleq H(P_{XY}), \quad P_{XY}(x, y) \triangleq \Pr\{X = x, Y = y\}$$

$$H(Y|X) \triangleq H(X, Y) - H(X) = \sum_{x \in X} P_X(x) H(Y|X = x)$$

$$H(Y|X = x) \triangleq H(P_{Y|X=x}), \quad P_{Y|X=x}(y) = \Pr\{Y = y|X = x\}$$

Mutual information

$$I(X \wedge Y) \triangleq H(Y) - H(Y|X) = H(X) + H(Y) - H(X, Y)$$

$$= D(P_{XY} || P_X \times P_Y)$$

$$I(X \wedge Y|Z) \triangleq H(Y|Z) - H(Y|XZ) = D(P_{XYZ} || \tilde{P}_{XYZ}),$$

$$\tilde{P}_{XYZ}(x, y, z) \triangleq P_{X|Z=z}(x) P_{Y|Z=z}(y) P_Z(z).$$

Properties of information measures

- Invariance, monotonicity: (conditional) entropies and mutual informations are invariant to replacing either RV by a one-to-one function of it. For any functions f, g

$$H(f(X)|Z) \leq H(X|Z) \leq H(X|g(Z)),$$

$$I(f(X) \wedge Y|Z) \leq I(X \wedge Y|Z),$$

but $I(X \wedge Y|g(Z))$ may be smaller or greater than $I(X \wedge Y|Z)$.

- Chain rules:

$$H(X_1, \dots, X_n|Z) = \sum_{i=1}^{n} H(X_i|X_1, \dots, X_{i-1}Z)$$

$$I(X_1, \dots, X_n \wedge Y|Z) = \sum_{i=1}^{n} I(X_i \wedge Y|X_1, \dots, X_{i-1}Z)$$

- Data processing: If $U -\!\!\circ\!\!- X -\!\!\circ\!\!- Y -\!\!\circ\!\!- Z$ is a Markov chain then

$$I(U \wedge V) \leq I(X \wedge Y).$$

- Fano inequality

$$H(X|Y) \leq \Pr\{X \neq Y\} \log|\mathcal{X}| + h(\Pr\{X \neq Y\}),$$

$$h(t) \triangleq -t \log t - (1-t) \log(1-t)$$

- $|H(P) - H(Q)| \leq \frac{1}{2}|P - Q| \log|\mathcal{X}| + h(\frac{1}{2}|P - Q|)$
- Pinsker inequality $\qquad |P - Q| \leq \sqrt{2 \ln 2 \cdot D(P||Q)}$

Security index

Formal meaning of assumptions on one-time pad key K, where the RV V represents the adversary's knowledge:

$$H(K|V) = H(K) = \log|\mathcal{K}|.$$

Relaxation: $S(K|V)$ below is small.

Definition

The security index of a RV K against another RV V is

$$S(K|V) \triangleq \log|\mathcal{K}| - H(K|V).$$

Theorem

For M independent of (V, K), knowledge of $M + K$ gives the adversary at most $S(K|V)$ bits of information about M.

Proof.

$I(M \wedge V, M + K) = H(M) + H(V, M + K) - H(M, V, K) \leq$
$H(M) + H(V) + \log|\mathcal{K}| - H(M) - H(V, K) = \log|\mathcal{K}| - H(K|V)$. □

Comments

(i) Even if the adversary's prior knowledge gives some information about the message (M is not independent of V), the new information in $M+K$ about M is bounded by $S(K|V)$:

$$I(M \wedge V, M+K) \leq I(M \wedge V) + S(K|V),$$

whenever K is conditionally independent of M given V.

(ii) Equivalent expression of security index:

$$S(K|V) = \sum_{v \in \mathcal{V}} P_V(v) D(P_{K|V=v} || P_0) = D(P_{KV} || P_0 \times P_V),$$

where P_0 is the uniform distribution on $\mathcal{K}$.

Another frequently used security index:

$$S_{\text{var}}(K|V) \triangleq |P_{KV} - P_0 \times P_V| = \sum_{v \in \mathcal{V}} P_V(v) |P_{K|V=v} - P_0|.$$

Properties of information measures imply

$$\frac{1}{2 \ln 2} S_{\text{var}}^2(K|V) \leq S(K|V) \leq \frac{1}{2} S_{\text{var}}(K|V) \log |\mathcal{K}| + h(\tfrac{1}{2} S_{\text{var}}(K|V)).$$

Properties of security index

- For K and K' with values in $\mathcal{K}$, and any V

  $$S(K'|V) \leq S(K|V) + H(K|K').$$

  Follows from the definition and

  $$H(K'|V) = H(KK'|V) - H(K|K'V) \geq H(K|V) - H(K|K')$$

- For uniformly distributed K_0 with $K_0 -\!\!\circ\!\!- K -\!\!\circ\!\!- V$,

  $$S(K_0|V) = I(K_0 \wedge V) \leq I(K \wedge V) \leq S(K|V)$$

- To given K and V, there exists uniformly distributed K_0 with $K_0 -\!\!\circ\!\!- K -\!\!\circ\!\!- V$ and

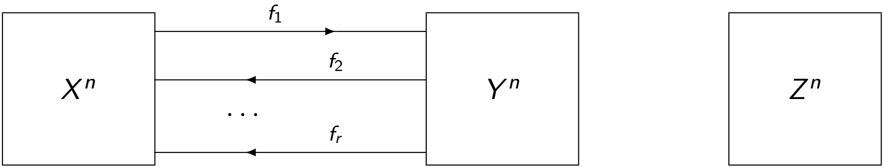
  $$\Pr\{K_0 \neq K\} \leq \sqrt{\frac{\ln 2}{2} S(K|V)}$$

  Indeed, to K one can construct K_0 with $P_{K_0} = P_0$ and $\Pr\{K_0 \neq K\} = \frac{1}{2}|P_K - P_0|$, where the Markov condition is no restriction.
  Then use $D(P_K||P_0) \leq S(K|V)$ and Pinsker inequality.

Two-party source models

Given i.i.d. repetitions of correlated RVs (X, Y, Z), Alice observes $X^n \triangleq (X_1, \ldots, X_n)$, Bob observes $Y^n \triangleq (Y_1, \ldots, Y_n)$ and the eavesdropper Eve observes $Z^n \triangleq (Z_1, \ldots, Z_n)$.

Alice and Bob may be allowed to generate local randomness, and to communicate errorfree . Eve has access to this communication, but she can not interfere with it.



$$f_1 = f_1(X^n),\ f_2 = f_2(Y^n, f_1),\ \ldots,\ f_r = f_r(Y^n, f_1, \ldots, f_{r-1}).$$

"Randomized functions" (also depending on RVs Q_A resp. Q_B generated by Alice resp. Bob, independently of each other and of (X^n, Y^n, Z^n)) may or may not be allowed.

Notation: $||f_i||$ denotes the number of possible values of f_i. The rate of f_i is $\frac{1}{n}\log||f_i||$.

A source model is determined by (X, Y, Z) and by specifying

(i) the permissible communication $\mathbf{F} \triangleq (f_1, \ldots, f_r)$, which may be unrestricted, or one-way ($\mathbf{F} = f_1$), or restricted by rate constraints, etc.

(ii) the allowed randomization.

Special case $Z = const$: Eve has no side information, i.e., no other information than the communication $\mathbf{F}$.

Common randomness (CR)

Definition

A RV K represents ε-CR for two or more parties, achievable under a given model, if K is ε-accessible to the parties, i.e., each can compute a RV equal to K with probability $\geq 1 - \varepsilon$.

In a two-party source model, a permissible protocol lets Alice or Bob compute those RVs K_A or K_B which are functions of $(X^n, \mathbf{F})$ [or of $(Q_A,\ X^n,\ \mathbf{F})$], respectively of $(Y^n, \mathbf{F})$ [or of $(Q_B,\ Y^n,\ \mathbf{F})$]. A RV K is ε-CR for Alice and Bob if for some such K_A and K_B

$$\Pr\{K \neq K_A\} \leq \varepsilon, \quad \Pr\{K \neq K_B\} \leq \varepsilon.$$

Secret key (SK)

Definition

A RV K with values in a finite set $\mathcal{K}$, which is ε-CR for two or more parties, represents (ε, δ)-SK against an eavesdropper whose knowledge is represented by a RV V (called her view), if

$$S\left(K \mid V\right) \triangleq \log|\mathcal{K}| - H\left(K \mid V\right) \leq \delta.$$

The key length is $\log|\mathcal{K}|$, having in mind $\mathcal{K} = \{0,1\}^{\ell}$.

In a two-party source model, the eavesdropper's view, after communication has been completed, is $V = (\mathbf{F}, Z^n)$. Thus K is an (ε, δ)-SK if K is ε-CR and

$$S(K \mid \mathbf{F}, Z^n) = \log|\mathcal{K}| - H(K \mid \mathbf{F}, Z^n) \leq \delta.$$

SK capacity

Definition

A number R_S is an achievable SK rate for a given model if for each $\eta > 0$ suitable protocols provide $(\varepsilon_n, \delta_n)$-SK K_n with

$$\frac{1}{n} \log |\mathcal{K}_n| \ \geq \ R_S - \eta \, , \quad \varepsilon_n \to 0, \ \delta_n \to 0, \text{ as } n \to \infty.$$

The supremum of achievable SK rates is the SK-capacity C_S.

Remark

In the literature, first a weaker definition of SK capacity was used, requiring $\frac{1}{n} S(K_n | \mathbf{F}, Z^n) \to 0$, rather than $S(K_n | \mathbf{F}, Z^n) \to 0$. Maurer 1994 pointed out that this was too weak for cryptographic purposes. A still stronger definition requires $\varepsilon_n \to 0$, $\delta_n \to 0$ exponentially fast.

For a large class of models, either definition gives the same value of SK capacity (Maurer-Wolf 2000, Csiszár 2011).

SK capacity, continued

An (ε, δ)-SK K need not be errorfree accessible to Alice (Bob). Still, due to $S(K'|V) \leq S(K|V) + H(K|K')$ and Fano's inequality, K_A and K_B errorfree accessible to Alice and Bob are (ε', δ')-SKs with $\varepsilon' = 2\varepsilon$, $\delta' = \delta + \varepsilon \log |\mathcal{K}| + h(\varepsilon)$.

It follows that under either of the "weak" or "exponential" versions of SK capacity, attention may be restricted to $K = K_A$. This need not always hold under our definition, but it does if $\varepsilon_n = o(n)$ can be assumed.

Recall also the existence of a uniformly distributed K_0 with

$$K_0 \ominus K \ominus V, \quad \Pr\{K_0 \neq K\} \leq \sqrt{\frac{\ln 2}{2} S(K|V)}.$$

It follows that attention could be restricted to uniformly distributed SK in the definition of SK capacity, both under our defnition and the stronger, exponential version, but not necessarily under the "weak" version.

From CR to SK

Typical scheme to obtain achievable SK rates:

(i) generate ε-CR for the involved parties, perhaps non-secret

(ii) find a function of this CR that has security index close to 0.

Step (i) is sometimes called information reconciliation
Step (ii) is called privacy amplification.

Example: Find achievable SK rate for two-party source model with only Alice admitted to communicate.
(i) CR generation: By Slepian-Wolf theorem, for any $R > H(X|Y)$ and sufficiently large n there exists $f : \mathcal{X}^n \to \{1, \ldots, 2^{nR}\}$ such that a suitable function of $(Y^n, f(X^n))$ is equal to X^n with probability $\geq 1 - \varepsilon_n$ where $\varepsilon_n \to 0$ exponentially fast as $n \to \infty$.
Then, if Alice sends Bob $\mathbf{F} = f(X^n)$, this one-way communication makes X^n an ε_n-CR for Alice and Bob.

(ii) Privacy amplification. As X^n is ε_n-CR, achieved by (one-way) communication $\mathbf{F} = f(X^n)$, a function $\kappa(X^n)$ of X^n will be $(\varepsilon_n, \delta_n)$-SK if it satisfies

$$S(\kappa(X^n)|f(X^n), Z^n) < \delta_n.$$

By a general result, to any $\beta > 0$ there exists $\xi > 0$ such that if F is any RV with at most 2^{nR} possible values and $|\mathcal{K}| = \lfloor 2^{n(H(X|Z)-R-\beta)} \rfloor$ then a randomly selected mapping $\kappa : \mathcal{X}^n \to \mathcal{K}$ satisfies

$$S(\kappa(X^n)|F, Z^n) < 2^{-n\xi}$$

except with probability going to 0 doubly exponentially fast. As R in the Slepian-Wolf theorem can be arbitrarily close to $H(X|Y)$, it follows that

$$R_S = H(X|Z) - H(Y|X) = I(X \wedge Y) - I(X \wedge Z)$$

is an achievable SK rate.

Private key (PK)

Modified version of source model: Eve reveals Z^n to Alice and Bob. Reasonable if Eve is not the adversary but the adversary can access her observations; or if Eve represents a "trusted center" who helps Alice and Bob to generate SK concealed also from herself. A corresponding SK will be called a private key (PK).

An (ε, δ)-PK is defined as (ε, δ)-SK in the case when Alice and Bob observe Z^n, in addition to X^n resp. Y^n, formally when X, Y in the definition are replaced by XZ, YZ.

Achievable PK rates and PK capacity C_P are defined accordingly. The previous remark on different possible definitions of SK capacity applies to PK capacity, as well.

The mathematical problem of determining PK capacity is less hard than that of determining SK capacity in general.

SK theorems, two-party source models

Theorem (Gács and Körner 1973)

If Alice and Bob are not admitted to communicate, they can not generate CR, let alone SK, at a positive rate, except in trivial cases.

Theorem (Maurer 1993, Ahlswede and Csiszár 1993)

When either unrestricted or one-way communication is allowed, with or without randomization,

$$I(X \wedge Y) - I(X \wedge Z) \leq C_S \leq C_P = I(X \wedge Y|Z).$$

If $X \ominus Y \ominus Z$ form a Markov chain then the equalities hold.

Proof: (i) We have seen that $I(X \wedge Y) - I(X \wedge Z)$ is an achievable SK rate, even with one-way communication. Substituting X and Y by (X, Z) and (Y, Z), this implies that $I(X \wedge Y|Z)$ is an achievable PK rate.

(ii) It remains to show that $C_P \leq I(X \wedge Y|Z)$. Suppose K is an $(\varepsilon_n, \delta_n)$-PK, achieved by communication $\mathbf{F} = (f_1, \ldots, f_r)$, i.e.,

$$S(K|Z^n, \mathbf{F}) = \log |\mathcal{K}| - H(K|Z^n, \mathbf{F}) \leq \delta_n$$

and K equals with probability $\geq 1 - \varepsilon_n$ some function K_A of $(Q_A, X^n, Z^n, \mathbf{F})$ as well as some function K_B of $(Q_B, Y^n, Z^n, \mathbf{F})$.

Here $\varepsilon_n \to 0$, but only $\frac{\delta_n}{n} \to 0$ is assumed, to cover also "weak secrecy".

We may take $K = K_A$, then

$$\begin{aligned} H(K_A|Z^n, F) &\leq H(K_A|K_B) + I(K_A \wedge K_B|Z^n, \mathbf{F}) \\ &\leq H(K_A|K_B) + I(Q_A X^n \wedge Q_B Y^n|Z^n, \mathbf{F}). \end{aligned}$$

A crucial observation is that the conditional mutual information does not decrease by omitting $\mathbf{F}$, hence it is $\leq I(X^n \wedge Y^n|Z^n)$. This can be verified, recalling that $\mathbf{F} = (f_1, \ldots, f_r)$, by checking that successive removal of $f_r, \ldots, f_1$ causes no decrease in either step.

Hence the claim follows, using Fano's inequality.

Special case of the Theorem: When Eve has no side information ($Z = const$), the SK capacity is equal to the mutual information $I(X \wedge Y)$. This provides a new operational meaning of mutual information.
Moreover, the result that when Eve has side information, the PK capacity equals $I(X \wedge Y|Z)$, gives the first explicit operational meaning of conditional mutual information.

For unrestricted communication the secrecy capacity C_S is unknown, in general, so is even the condition for $C_S > 0$.
The last theorem implies (Ahlswede and Csiszár 1993)

$$C_S \leq \inf_{V:(X,Y) \ominus Z \ominus V} I(X \wedge Y|V).$$

An improved bound was given by Renner and Wolf 2003 but as Gohari and Anantharam 2010 showed, it is not always tight, either.

Extractors

Task: generate ℓ "pure random" bits.
Available resource: a drawing from an unknown member of a family $\mathcal{P}$ of distributions on a (finite) set $\mathcal{U}$.

Definition

An (ℓ, δ)-extractor for a family $\mathcal{P}$ of distributions on $\mathcal{U}$ is a mapping $\kappa : \mathcal{U} \to \mathcal{K}$ with $|\mathcal{K}| = 2^\ell$ such that for any RV U with $P_U \in \mathcal{P}$, the distribution of $\kappa(U)$ is δ-close in variation distance to the uniform distribution on $\mathcal{K}$:

$$\sum\nolimits_{k\in\mathcal{K}} \left|\Pr\{\kappa(U) = k\} - 2^{-\ell}\right| \leq \delta \quad \text{if } P_U \in \mathcal{P}.$$

Remark

In the literature, this concept is called deterministic extractor, as opposed to seeded extractors.

Relevance for SK: For any RVs U and V, an extractor κ for the family $\mathcal{P} = \{P_{U|V=v} : \; v \in \mathcal{V}\}$ yields a SK $K = \kappa(U)$.

Information spectrum

An (ℓ, δ)-extractor for a family $\mathcal{P}$ can be expected to exist only if each $P \in \mathcal{P}$ consists of probabilities $P(u) \leq 2^{-\ell}$, perhaps with exceptions of small total probability:

$$\sum_{u:P(u)>2^{-\ell}} P(u) \leq \eta \quad (*)$$

Intuition: An outcome $u \in \mathcal{U}$ of a drawing from a distribution P provides $-\log P(u)$ bits of information. Its expected value is the entropy $H(P)$, and its minimum is

$$H_{\text{min}}(P) \triangleq -\log\left(\max_{u\in\mathcal{U}} P(u)\right) \quad \text{min-entropy.}$$

The distribution function of $-\log P(u)$ is

$$F_P(t) \triangleq P\big(\{u : -\log P(u) < t\}\big) = \sum_{u:P(u)>2^{-t}} P(u).$$

It is called the information spectrum of the distribution P.

Smooth min-entropy

Condition $(*)$ on previous slide is equivalent to

$$l \leq H_{\min,\eta}(P) \triangleq \max\{t : F_P(t) \leq \eta\}.$$

$H_{\min,\eta}(P)$ is sometimes called smooth min-entropy.
It equals the η-quantile of the information spectrum $F_P(t)$.

Conditional information spectrum of RVs U conditioned on V:

$$F_{U|V}(t) \triangleq P_{UV}\big(\{(u,v) : -\log P_{U|V=v}(u) < t\}\big).$$

Its η-quantile is the smooth conditional min-entropy

$$H_{\min,\eta}(U|V) \triangleq \max\{t : F_{U|V}(t) \leq \eta\}.$$

Extractor lemma

Extractor lemma (Ahlswede and Csiszár 1998)

Suppose for some $\eta \geq 0$, $\beta > 0$

$$\ell \leq H_{\min,\eta}(P) - \beta \text{ for each } P \in \mathcal{P}.$$

Then a random mapping $\kappa : \mathcal{U} \to \mathcal{K}$ with $|\mathcal{K}| = 2^{\ell}$ is an $(\ell, \delta + 2\eta)$-extractor for the family $\mathcal{P}$ with probability $\geq 1 - 2^{\ell+1} |\mathcal{P}| \, e^{-\xi \, 2^{\beta}}$, where $\xi = \frac{\delta^2(1-\eta)}{2(1+\delta)}$, $\delta > 0$ arbitrary.

Corollary

For any RVs U, V, a random mapping $\kappa : \mathcal{U} \to \mathcal{K}$ with $\log |\mathcal{K}| = \ell \leq H_{\min,\eta^2}(U|V) - \beta$ gives

$$S(\kappa(U)|V) \leq (\alpha + 2\eta)\ell + h(\alpha + \eta)$$

with probability $\geq 1 - 2^{\ell+1} |\mathcal{V}| \, e^{-\alpha^2 2^{\beta}}$, where $0 \leq \alpha \leq 1/6$ is arbitrary.

Extractor lemma, proof

Random mapping $\kappa : \mathcal{U} \to \mathcal{K}$: the values $\kappa(u)$ are randomly chosen from $\mathcal{K}$, independently for each $u \in \mathcal{U}$, with equal probabilities $\frac{1}{|\mathcal{K}|} = 2^{-\ell}$.
Also called random binning: each $u \in \mathcal{U}$ is randomly assigned to one of $|\mathcal{K}| = 2^{\ell}$ bins. Random binning is a standard technique in IT, the Slepian–Wolf theorem is also proved in this way.

Proof (sketch). For fixed P and $k \in \mathcal{K}$,

$$P(\{u : \kappa(u) = k\}) = \sum_{u \in \mathcal{U}} P(u) \mathbb{1}_{\{\kappa(u)=k\}}$$

is a RV, a weighted sum of i.i.d. RVs $\mathbb{1}_{\{\kappa(u)=k\}}$ equal to 1 or 0 with probabilities $2^{-\ell}$ and $1 - 2^{-\ell}$. Assume first that P satisfies $H_{\min}(P) \geq \ell + \beta$, i.e., $P(u) \leq 2^{-\ell-\beta}$ for each $u \in \mathcal{U}$.

Proof continued

Then Chernoff bounding gives by routine calculation

$$\Pr\left\{\left|P(\{u : \kappa(u) = k\}) - 2^{-\ell}\right| > \frac{\delta}{K}\right\} \leq 2\, e^{-\frac{\delta^2}{1+\delta}\cdot 2^{\beta}}.$$

It follows that if $H_{\min}(P) \geq \ell + \beta$ holds for each $P \in \mathcal{P}$ (i.e., $\eta = 0$ in the Lemma), then

$$\max_{k\in\mathcal{K}} \left|P(\{u : \kappa(u) = k\}) - 2^{-\ell}\right| \leq \frac{\delta}{K}$$

for each $P \in \mathcal{P}$ with probability $\geq 1 - 2^{\ell+1}\,|\mathcal{P}|\,2^{-\frac{\delta^2}{1+\delta}2^{\beta}}$, a stronger result than claimed.

If only $\ell \leq H_{\min,\eta}(P) - \beta$ is assumed for all $P \in \mathcal{P}$, apply this to the distributions P' obtained by conditioning each $P \in \mathcal{P}$ on $\{u : P(u) \leq 2^{-\ell-\beta}\}$. Note that $|P' - P| \leq 2\eta$.

Corollary: apply the Extractor Lemma to $\mathcal{P} = \{P_{U|V=v} : v \in \mathcal{V}'\}$,

$$\mathcal{V}' \triangleq \{v : H_{\min,\eta}(P_{U|V=v}) \geq H_{\min,\eta^2}(U|V)\}.$$

Proof continued

This gives that with probability $\geq 1 - 2^{\ell+1}\,|\mathcal{V}|\,2^{-\xi\,2^{\beta}}$

$$\sum_{k \in \mathcal{K}} \left| \Pr\{\kappa(u) = k | V = v\} - 2^{-\ell} \right| \leq \delta + 2\eta, \quad v \in \mathcal{V}'.$$

Bound the I-divergences $D(P_{\kappa(U)|V=v}||P_0)$, $v \in \mathcal{V}'$ via these variation distances. Noting that $P_V(\mathcal{V}') \geq 1 - \eta$, the Corollary follows by simple algebra.

Supplement If $\mathcal{B} \subset \mathcal{U} \times \mathcal{V}$ is a set with

$$P_{UV}(\mathcal{B}) \geq 1 - (\eta^2 - \alpha^2), \quad P_{UV}(u, v) < \frac{1}{\alpha|\mathcal{B}|} \text{ for } (u, v) \in \mathcal{B}$$

then, denoting $\mathcal{B}_v = \{u : (u, v) \in \mathcal{B}\}$,

$$H_{\min,\eta^2}(U|V) \geq \log \min_{v:\mathcal{B}_v \neq \emptyset} |\mathcal{B}_v| + 3 \log \alpha.$$

This follows by simple algebra, the details are omitted.

Memoryless sources

Lemma

Let $X^n = X_1 \dots X_n$ *and* $Z^n = Z_1 \dots Z_n$ *be i.i.d. repetitions of a pair of RVs* (X, Z)*. To any* $\beta > 0$ *there exists* $\xi > 0$ *such that for n sufficiently large*
(i) some mapping $\kappa : \mathcal{X}^n \to \mathcal{K}_n$ *with* $\frac{1}{n} \log |\mathcal{K}_n| = H(X|Z) - \beta$ *satisfies* $S(\kappa(X^n)|Z^n) < e^{-\xi n}$
(ii) if $F^{(n)}$ *is any RV jointly distributed with* (X^n, Z^n) *that has at most* 2^{nr} *possible values, some* $\kappa : \mathcal{X}^n \to \mathcal{K}_n$ *with* $\frac{1}{n} \log |\mathcal{K}_n| = H(X|Z) - r - \beta$ *satisfies* $S(\kappa(X^n)|Z^n, F^{(n)}) < e^{-\xi n}$*.*
Indeed, a random mapping $\kappa : \mathcal{X}^n \to \mathcal{K}_n$ *is suitable except with probability approaching* 0 *doubly exponentially.*

Proof: By Corollary of the Extractor Lemma.

Intuitive interpretations

Slepian and Wolf 1973: The infimum of rates R for which, when n is large, an encoder $f : \mathcal{X}^n \to \{1, \ldots, 2^{nR}\}$ exists such that X^n can be decoded from $f(X^n)$ and Y^n with small probability of error, equals the conditional entropy $H(X|Y)$.
In addition, for $R > H(X|Y)$ the probability of error can be made exponentially small, even with a universal decoder not depending on P_{XY}.
The SW theorem and (i) of the last Lemma (writing there Y instead of Z) admit complementary interpretations of $H(X|Y)$: it is a measure of the amount of information needed to specify X when Y is known, as well as of that part of the information in X which is independent of Y.

These intuitive concepts coincide only in limiting sense: the coding rate can be arbitrarily close to $H(X|Y)$ from above, random bits independent of Y^n can be extracted from X^n at rate arbitrarily close to $H(X|Y)$ from below.

Interpretation of the result on SK capacity for $Z =$const: The amount of information in X that can be shared with a party knowing Y, with no leak to a third party (who has no side information), is the mutual information $I(X \wedge Y)$.

Attractive intuitive interpretation of the identity $H(X) = H(X|Y) + I(X \wedge Y)$.

Restricted one-way communication

Only Alice communicates, assume randomization is not allowed: $F = F_n = f(X^n)$.
Rate constraint: $\limsup_{n\to\infty} \frac{1}{n} \log ||F_n|| \le R$.
First, address CR without secrecy.
Goal: construct $f(X^n)$, $g(X^n)$ such that if Alice sends Bob $f(X^n)$, this makes $g(X^n)$ ε_n-accessible to him, i.e., some function of $(Y^n, f(X^n))$ equals $g(X^n)$ with probability $\ge 1 - \varepsilon_n$.
Random construction will be used that refines the following rate-distortion theory result.
Given RVs U, X, if $N = 2^{n[I(U\wedge X)+\delta]}$ sequences $\mathbf{u}_i \in \mathcal{U}^n$ are randomly drawn from the distribution P_U, each typical $\mathbf{x} \in \mathcal{X}^n$ is jointly typical with some $\mathbf{u}_i$ with probability $\to 1$ as $n \to \infty$.

Typical sequences

Given RVs U, X, Y, and $\xi > 0$, sequences $\mathbf{u} = u_1 \dots u_n \in \mathcal{U}^n$ or $\mathbf{x} \in \mathcal{X}^n$, $\mathbf{y} \in \mathcal{Y}^n$ or their pairs, triples are called (strongly) typical if the relative frequencies

$$\tfrac{1}{n}|\{i : u_i = u\}|,\ \tfrac{1}{n}|\{i : u_i = u, x_i = x\}|,\ \tfrac{1}{n}|\{i : u_i = u, x_i = x, y_i = y\}|$$

differ from the probabilities $P_U(u)$, $P_{UX}(u, x)$, $P_{UXY}(u, x, y)$ by no more than ξ, and are equal to 0 if the probabilities are.
The typical sets $\mathcal{T}^n_{U,\xi}$, $\mathcal{T}^n_{UX,\xi}$, $\mathcal{T}^n_{UXY,\xi}$ consist of all typical sequences or pairs or triples. The upper index n is often omitted.
Further notation: $\mathcal{T}^n_{UX,\xi}(\mathbf{u}) \triangleq \left\{\mathbf{x} : (\mathbf{u}, \mathbf{x}) \in \mathcal{T}^n_{UX,\xi}\right\}$.
For i.i.d. repetitions of (X, Y), the probabilities

$$\Pr\{X^n \notin \mathcal{T}^n_{X,\xi}\},\ \Pr\{(X^n, Y^n) \notin \mathcal{T}^n_{XY,\xi}\},\ \Pr\{Y^n \notin \mathcal{T}^n_{XY,\xi}(\mathbf{x}) \mid X^n = \mathbf{x}\},$$

and when $U \ominus X \ominus Y$ then also $\Pr\{Y^n \notin \mathcal{T}^n_{UXY,\xi}(\mathbf{u}, \mathbf{x}) \mid X^n = \mathbf{x}\}$ go to 0 exponentially fast, the last two at least when $\mathbf{x} \in \mathcal{T}^n_{X,\zeta}$, resp. $(\mathbf{u}, \mathbf{x}) \in \mathcal{T}^n_{UY,\zeta}$ for some $0 < \zeta < \xi$.

CR construction

Suppose $U \ominus X \ominus Y$, fix $\delta > 0$, let

$$N_1 = 2^{n[I(U \wedge X|Y)+3\delta]}, \quad N_2 = 2^{n[I(U \wedge Y)-2\delta]}$$

Lemma

If $N_1 N_2$ sequences $\mathbf{u}_{ij}$, $1 \leq i \leq N_1$, $1 \leq j \leq N_2$ are randomly drawn from the distribution P_U, for sufficiently small $\sigma > 0$ and $0 < \zeta < \xi < \sigma$ there exist, with probability approaching 1 as $n \to \infty$, mappings $f : \mathcal{X}^n \to \{1, \ldots, N_1\}$, $g : \mathcal{X}^n \to \{1, \ldots, N_2\}$ such that

(i) $(\mathbf{u}_{f(\mathbf{x}),g(\mathbf{x})}, \mathbf{x}) \in \mathcal{T}_{UX,\xi}$ whenever $\mathbf{x} \in \mathcal{T}_{X,\zeta}$

(ii) except for $(\mathbf{x}, \mathbf{y})$ in a set of exponentially small P^n_{XY}-probability, $(\mathbf{u}_{f(\mathbf{x}),g(\mathbf{x})}, \mathbf{x}, \mathbf{y}) \in \mathcal{T}_{UXY,\sigma}$, and only $j = g(\mathbf{x})$ satisfies $(\mathbf{u}_{f(\mathbf{x}),j}, \mathbf{y}) \in \mathcal{T}_{UY,|\mathcal{X}|\sigma}$.

By the last property, $g(X^n)$ is ε_n-accessible to Bob, with exponentially small ε_n.

Sketch of proof

(i) Standard, as $N_1 N_2 = 2^{n[I(U \wedge X)+\delta]}$
(ii) Joint typicality part: implied by last stated property of typical sequences
Last part: it suffices to show that only with exponentially small probability can such $i \in \{1, \dots, N_1\}$ and $j \neq \ell$ in $\{1, \dots, N_2\}$ exist for which $\mathbf{u}_{ij} \in \mathcal{T}_{UX}(\mathbf{x})$, $\mathbf{u}_{i\ell} \in \mathcal{T}_{UY}(\mathbf{y})$.
Now, for fixed i, j, ℓ

$$\Pr\{\mathbf{u}_{ij} \in \mathcal{T}_{UX}(\mathbf{x}),\ \mathbf{u}_{i\ell} \in \mathcal{T}_{UY}(\mathbf{y})\} \leq 2^{-n[I(U \wedge X)-\tau]}\, 2^{-n[I(U \wedge Y)-\tau']},$$

with $\tau > 0, \tau' > 0$ arbitrary small if the typicality constants are.
Multiplying this bound by $N_1 N_2^2$ gives $2^{-n(\delta-\tau-\tau')}$.

Achievable SK rates

Proposition

For one-way communication with rate constraint R, the maximum of $I(U \wedge Y) - I(U \wedge Z)$ for auxiliary RVs U with

$$U \mathrel{-\!\!\circ\!\!-} X \mathrel{-\!\!\circ\!\!-} YZ, \; I(U \wedge X|Y) \leq R$$

is an achievable SK rate. The maximum is attained for some U with $|\mathcal{U}| \leq |\mathcal{X}| + 1$. Moreover, achievability holds with $\varepsilon_n \to 0$, $\delta_n \to 0$ exponentially fast.

Proof (sketch). Apply privacy amplification to the CR $g(X^n)$ constructed before (with Alice's public message $f(X^n)$ of rate $I(U \wedge X|Y) + 3\delta$). The Extractor Lemma is employed, using its supplement with $g(X^n)$ and $(f(X^n), Z^n)$ in the role of U and V. One can check that the set

Proof continued

$$\mathcal{B} = \{(j,(i,\mathbf{z})) : (\mathbf{u}_{ij},\mathbf{x},\mathbf{z}) \in \mathcal{T}_{UXZ} \text{ for some } \mathbf{x} \in \mathcal{X}^n\}$$

meets the conditions of the supplement, and for $v = (i,\mathbf{z})$

$$|\mathcal{B}_v| \geq |\{j : \mathbf{u}_{ij} \in \mathcal{T}_{UZ}(\mathbf{z})\}| \geq 2^{n[I(U\wedge Y)-I(U\wedge Z)-2\delta-\tau]}.$$

The bound on $|\mathcal{U}|$ is an essential part of the Proposition.
The Caratheodory-Fenchel theorem implies, as standard in IT (details omitted): Any RV U as in the Proposition may be substituted by some $\tilde{U}$ with $\tilde{U} \ominus X \ominus YZ$ and $|\tilde{\mathcal{U}}| \leq |\mathcal{X}| + 1$, without changing $I(U \wedge X|Y)$ and $I(U \wedge Y) - I(U \wedge Z)$.
The bound on $|\mathcal{U}|$ ensures, by continuity, that the maximum in the Proposition is attained, and also that it is computable, at least in principle.

Intuitive considerations

In previous construction, the public message $f(X^n)$ and the CR $g(X^n)$ are nearly independent and uniformly distributed, in the weak sense:

$$\frac{1}{n}H(f(X^n), g(X^n)) \approx \frac{1}{n}\log N_1 N_2 = I(U \wedge X) + \delta$$

(intuitively plausible, easily proved formally).
The full CR achieved is $(f(X^n), g(X^n))$, though only $g(X^n)$ can contribute to SK. If Eve has no side information ($Z =$ const) then $g(X^n)$ itself qualifies as weak SK, and privacy amplification yields strong SK of effectively the same rate $I(U \wedge Y)$. This is best possible when $Z =$ const. Then (see below) the SK capacity is

$$C_S(R) = \max\{I(U \wedge Y) : U \ominus X \ominus Y, I(U \wedge X|Y) \leq R\}.$$

CR capacity

Not hard to show optimality also for constructing CR via one-way communication of rate R (Ahlswede and Csiszár 1998): This model has CR capacity (largest entropy rate of ε_n-CR with $\varepsilon_n \to 0$)

$$C_{CR}(R) = \max\{I(U \wedge X) : U \ominus X \ominus Y,\ I(U \wedge X|Y) \leq R\},$$

for all R when randomization is not allowed, and for $R \leq H(X|Y)$ when it is. When randomization is allowed, the CR capacity equals $I(X \wedge Y) + R$ if $R \geq H(X|Y)$. Intuitive identity:

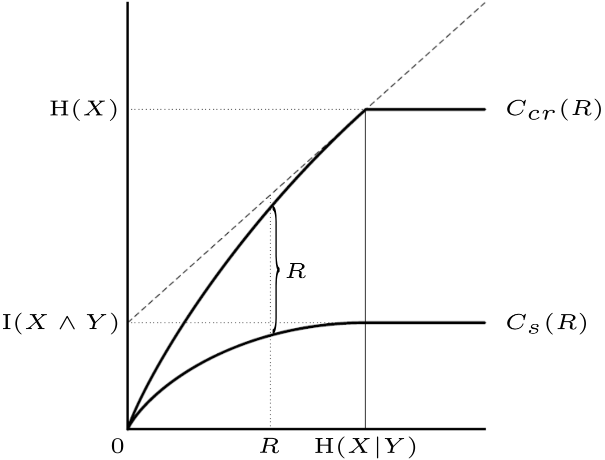
$$C_{CR}(R) = C_S(R) + R \quad \text{if} \quad R \leq H(X|Y).$$

Remark

$$\lim_{R \to 0} C_{CR}(R) = \max\{I(U \wedge X) : U \ominus X \ominus Y, U \ominus Y \ominus X\}.$$

This equals 0 unless there exists a common function of X and Y. Sharpens the Gács-Körner non-existence theorem.

CR and SK capacities for one-way communication, Z=const



SK capacity, Bob silent

The achievable SK rates in the last Proposition may be improved in some cases, though not when $Z = \text{const}$. The general result involves two auxiliary RVs.

Theorem

For the source model with arbitrary X, Y, Z, when only Alice is admitted to communicate, at most with rate R, the SK capacity equals (whether randomization is allowed or not)

$$\max\left[I(U \wedge Y|V) - I(U \wedge Z|V)\right],$$

subject to $UV \circleddash X \circleddash YZ$, $I(UV \wedge X|Y) \le R$, $|\mathcal{U}|, |\mathcal{V}| \le |\mathcal{X}| + 1$.

[Ahlswede and Csiszár 1993 in unrestricted case ($R = +\infty$), Csiszár and Narayan 2000 in general]

Achievability

If the maximum is attained for $V = \text{const}$, the last Proposition suffices for achievability. Otherwise, a more complex two-step random construction is needed (omitted).

Simple achievability proof for the unconstrained case $R = +\infty$ with randomization allowed (Renner and Wolf 2003):

Alice generates i.i.d. U^n, V^n, and reveals V^n. By previous theorem applied to U, YV, ZV in the role of X, Y, Z it follows that

$$I(U \wedge YV) - I(U \wedge ZV) = I(U \wedge Y|V) - I(U \wedge Z|V)$$

is an achievable SK rate.

Technical lemma

Lemma (Csiszár and Körner 1978)

For arbitrary RVs S, T and $Y^n = Y_1 \dots Y_n$, $Z^n = Z_1 \dots Z_n$

$$I(S \wedge Y^n|T) - I(S \wedge Z^n|T)$$

$$= \sum_{j=1}^{n} [I(S \wedge Y_j|Y_1 \dots Y_{j-1} Z_{j+1} \dots Z_n T) - I(S \wedge Z_j|Y_1 \dots Y_{j-1} Z_{j+1} \dots Z_n T)]$$

$$= n[I(S \wedge Y_J|V) - I(S \wedge Z_J|V)]$$

where J denotes a RV independent of the others, uniformly distributed on $\{1, \dots, n\}$, and $V = JY_1 \dots Y_{J-1} Z_{J+1} \dots Z_n T$.

Proof. The j'th term of the sum equals

$$H(S|Y_1 \dots Y_{j-1} Z_{j+1} \dots Z_n T) - H(S|Y_1 \dots Y_j Z_{j+1} \dots Z_n T)$$

$$-H(S|Y_1 \dots Y_{j-1} Z_{j+1} \dots Z_n T) + H(S|Y_1 \dots Y_{j-1} Z_j \dots Z_n T).$$

After cancellations, the sum of the n terms is

$$H(S|Z^n T) - H(S|Y^n T) = I(S \wedge Y^n|T) - I(S \wedge Z^n|T).$$

Converse proof

Suppose $K = K_n$, a function of X^n, Q_A, is an $(\varepsilon_n, \delta_n)$-SK achieved via communication $F = F(X^n, Q_A)$.

$$\delta_n > S(K|Z^n, F) = \log|\mathcal{K}| - H(K|Z^n F)$$

$$\geq \log|\mathcal{K}| - H(K|Z^n F) + H(K|Y^n F) - \varepsilon_n \log|\mathcal{K}| - h(\varepsilon_n)$$

$$= (1-\varepsilon_n)\log|\mathcal{K}| - [I(K \wedge Y^n|F) - I(K \wedge Z^n|F)] - h(\varepsilon_n).$$

Using technical lemma to K, F in the role of S, T:

$$\frac{1}{n}\log|\mathcal{K}| \leq \frac{1}{1-\varepsilon_n}[I(K \wedge Y_J|V) - I(K \wedge Z_J|V)] + \frac{\delta_n + h(\varepsilon_n)}{n(1-\varepsilon_n)},$$

where J is uniformly distributed on $\{1, \ldots, n\}$, independent of Q_A, X^n, Y^n, Z^n, and $V = JY_1 \ldots Y_{J-1} Z_{J+1} \ldots Z_n F$.
This already proves the converse for $R = +\infty$, if the Markov relation $KV \ominus X_J \ominus Y_J Z_J$ is checked. Note that $\delta_n \to 0$ is not needed, $\delta_n/n \to 0$ suffices.
For full converse, $\frac{1}{n}\log||F||$ has to be bounded below.

Proof continued

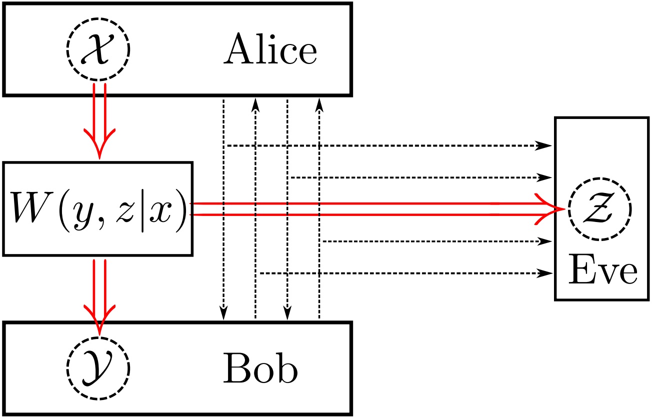
$$\frac{1}{n}\log||F|| \geq \frac{1}{n}H(F|Y^n) \approx \frac{1}{n}H(KF|Y^n)$$

$$\geq \frac{1}{n}[H(KF|Y^n) - H(KF|X^n)] = \frac{1}{n}[I(KF \wedge X^n) - I(KF \wedge Y^n)]$$

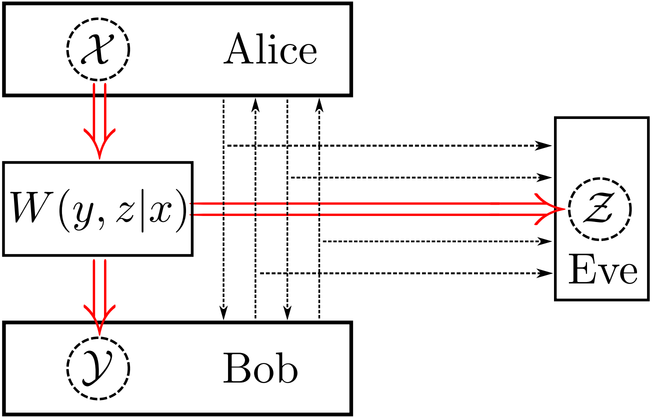
This can be bounded below by $I(KV \wedge X_J|Y_J)$ with V as above, using the technical lemma and calculations omitted here.

To check $KV \ominus X_J \ominus Y_J Z_J$, we have to show that the conditional distribution of $Y_J Z_J$ given the values of $K, J, Y_1, \ldots, Y_{J-1}, Z_{J+1}, \ldots, Z_n, F$ and X_J depends only on the value of X_J, say x. When $J = j$, this conditional distribution equals that of $Y_j Z_j$ given the values of $K, Y_1, \ldots, Y_{j-1}, Z_{j+1}, \ldots, Z_n, F$ and $X_j = x$. Since X^n, Y^n, Z^n are i.i.d. repetitions of (X, Y, Z), this conditional distribution equals $P_{YZ|X=x}$.

Channel model with public communication



Channel model with public communication



Two-party channel models

Main resource: a noisy channel over which Alice may transmit information to Bob, with partial leakage to Eve.

Alice and Bob may be also allowed to communicate, perhaps, interactively, over a noiseless public channel. This public communication is fully accessible to Eve but she can not interfere with it.

The noisy channel is a discrete memoryless channel (DMC) with matrix $W = \{W(y,z|x) : x \in \mathcal{X}, y \in \mathcal{Y}, z \in \mathcal{Z}\}$ of transition probabilities. At each instant when Alice sends input symbol $x \in \mathcal{X}$, Bob and Eve receive outputs $y \in \mathcal{Y}$ and $z \in \mathcal{Z}$ with probability $W(y,z|x)$, independently of all prior communication.

Model defined by W and the permissible public communication.
Essential that Alice is permitted to randomize.
For $1 \leq i \leq n$, Alice sends DMC input X_i, a function of her randomizing RV Q_A and of $F^{i-1} = (F_1, \ldots, F_{i-1})$; Bob and Eve receive Y_i, Z_i.
$F_i = (f_{i1}, \ldots, f_{ir})$, f_{ij} function of Q_A or $Y^{i-1}Q_B$ (for j odd or even) and of the previous communication $F^{i-1}f_{i1} \ldots f_{i(j-1)}$.

SK capacity, channel model

CR, SK: as for source model, but now
$X^n = X_1 \dots X_n$, $Y^n = Y_1 \dots Y_n$, $Z^n = Z_1 \dots Z_n$
denote sequences of input and output RVs of the DMC, not i.i.d. in general.
A RV K is ε-CR achievable by (permissible) communication $\mathbf{F}$ if there exist $K_A = K_A(Q_A, \mathbf{F})$, $K_B = K_B(Y^n, Q_B, \mathbf{F})$ with

$$\Pr\{K \neq K_A\} \leq \varepsilon, \; \Pr\{K \neq K_B\} \leq \varepsilon.$$

(ε, δ)-SK: an ε-CR K with $S(K|Z^n, \mathbf{F}) \leq \delta$.
SK capacity: same definition as for source model.
PK capacity: like for source model, the variant of SK capacity for the case when Eve reveals Z^n (she reveals each Z_i immediately upon receipt).

Source emulation

Simple but useful technique to obtain achievable SK rates for channel models from those for source models.

Suppose the given DMC is used in the simple way of Alice generating and transmitting i.i.d. channel inputs $X_1, \ldots, X_n$. Public communication (if any) is performed only after completing this transmission, including receipt of the corresponding outputs. Thereby a source model is emulated, given by i.i.d. repetitions of RVs X, Y, Z with $P_{XYX}(x, y, z) = P_X(x)W(y, z|x)$, and permissible public communication as in the channel model.

The supremum over P_X of the SK capacities of the emulated source models is a lower bound to the SK capacity of the channel model.

Channel model SK capacity theorem

Analogue of previous source model theorem. also due to Maurer 1993 and Ahlswede and Csiszár 1993.

Theorem

For two-party channel model with unrestricted public communication,

$$\max_{P_X}[I(X \wedge Y) - I(X \wedge Z)] \leq C_s \leq C_P \leq \max_{P_X} I(X \wedge Y|Z),$$

where $P_{XYZ}(x,y,z) = P_X(x)W(y,z|x)$*. The lower bound is achievable without any public communication. The lower bound is tight if* W *is a physically degraded channel, i.e. of form* $W(y,z|x) = W_1(y|x)\tilde{W}(z|y)$*.*

The lower bound holds also replacing $I(X \wedge Z)$ by $I(Y \wedge Z)$, and that bound is tight if the channel W has independent components, i.e. $W(y,z|x) = W_1(y|x)W_2(z|x)$.

Proof

Achievability: immediate via source emulation, except for achievability of the lower bound without public communication, which will be verified later.

Converse: One has to prove $C_P \leq \max_{P_X} I(X \wedge Y|Z)$. Key step is similar to that for source models, to verify the bound

$$I(Q_A \wedge Y^n Q_B | Z^n \mathbf{F}) \leq \sum_{i=1}^{n} I(X_i \wedge Y_i | Z_i)$$

for public communication
$\mathbf{F} = F^n = (F_1, \ldots, F_n), \quad F_i = (f_{i1}, \ldots, F_{ir})$.

This is done by repeating n times the bounding

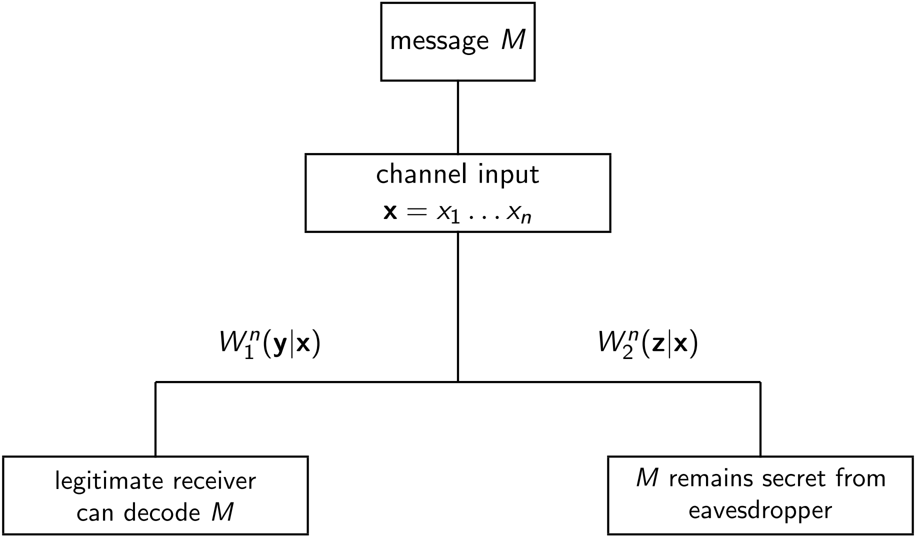
$$\begin{aligned} I(Q_A \wedge Y^n Q_B | Z^n F^n) &\leq I(Q_A \wedge Y^n Q_B | Z^n F^{n-1}) \\ &\leq I(Q_A \wedge Y^{n-1} Q_B | Z^{n-1} F^{n-1}) + I(X_n \wedge Y_n | Z_n). \end{aligned}$$

The first inequality follows by the argument used for source models, the calculation yielding the second one is on the next slide.

Calculation

$$
\begin{gathered}
I(Q_A \wedge Y^n Q_B | Z^n F^{n-1}) \\
= I(Q_A \wedge Y^n Q_B Z^n | F^{n-1}) - I(Q_A \wedge Z^n | F^{n-1}) \\
= I(Q_A \wedge Y^{n-1} Q_B Z^{n-1} | F^{n-1}) + I(Q_A \wedge Y_n Z_n | Y^{n-1} Q_B Z^{n-1} F^{n-1}) \\
- I(Q_A \wedge Z^{n-1} | F^{n-1}) - I(Q_A \wedge Z_n | Z^{n-1} F^{n-1}) \\
= I(Q_A \wedge Y^{n-1} Q_B | Z^{n-1} F^{n-1}) + A \\
A = H(Y_n Z_n | Y^{n-1} Q_B Z^{n-1} F^{n-1}) - H(Y_n Z_n | Q_A Y^{n-1} Q_B Z^{n-1} F^{n-1}) \\
- H(Z_n | Z^{n-1} F^{n-1}) + H(Z_n | Q_A Z^{n-1} F^{n-1}) \\
\leq H(Z_n | Y_n) - H(Y_n Z_n | X_n) + H(Z_n | X_n) = I(X_n \wedge Y_n | Z_n)
\end{gathered}
$$

Wiretap channel, Wyner 1975, Cs-K 1978



message M

channel input

$\mathbf{x} = x_1 \ldots x_n$

$W_1^n(\mathbf{y}|\mathbf{x})$

$W_2^n(\mathbf{z}|\mathbf{x})$

legitimate receiver can decode M

M remains secret from eavesdropper

Wiretap channel

Original formulation unrelated to SK.
Wyner's seminal discovery was that secure transmission over an insecure channel does not necessarily require a key. Nevertheless, the problem turns out equivalent to that of SK capacity of a channel model where public communication is not allowed.
The DMCs $W_1,\ W_2$ of the wiretap model are conveniently regarded as component channels of a two-output DMC $W : \mathcal{X} \to \mathcal{Y} \times \mathcal{Z}$:

$$W_1(y|z) = \sum_{z \in \mathcal{Z}} W(y, z|x), \quad W_2(z|x) = \sum_{y \in \mathcal{Y}} W(y, z|x).$$

This physical channel W does not enter the mathematical problem formulated below, only W_1 and W_2 matter.
Wyner (1975) considered the case when a physically degraded W, i.e., $W(y, z|x) = W_1(y|x)\widetilde{W}(z|y)$ could be chosen.
Then W_2 is called a (stochastically) degraded version of W_1.

Secure transmission problem

Random message M_n uniformly distributed over $\mathcal{M}_n$
Stochastic encoder assigns to M_n length-n channel input $X^n = \varphi(M_n, Q_A)$ where Q_A denotes a RV generated by Alice for randomization.
Bob and Eve receive channel outputs Y^n, Z^n, Bob tries to recover the message via a (deterministic) decoder: $\hat{M}_n = \psi(Y^n)$.
Error probability: $\Pr\{\hat{M}_n \neq M_n\}$.
Information leakage: $I(M_n \wedge Z^n)$.

Achievable rate of transmission: number R such that one can achieve for each $\eta > 0$, as $n \to \infty$

$$\frac{1}{n} \log |\mathcal{M}_n| \geq R - \eta, \ \Pr\{\hat{M}_n \neq M_n\} \to 0, \ I(M_n \wedge Z^n) \to 0.$$

Secrecy capacity: supremum of achievable rates.

Better channel?

Intuitively, positive secrecy capacity is expected if Bob's channel W_1 is better than Eve's channel W_2.

Wyner (1975) assumed
(i) W_2 is a degraded version of W_1.
Other concepts of "better channel" (Körner and Marton 1977):
(ii) W_1 more capable than W_2: $I(X \wedge Y) \geq I(X \wedge Z)$ for all P_X
(iii) W_1 less noisy than W_2: $I(V \wedge Y) \geq I(V \wedge Z)$ whenever $V \ominus X \ominus (Y, Z), \quad P_{Y|X=x}(y) = W_1(y|x), \quad P_{Z|X=x}(z) = W_2(z|x)$.
Clearly, $(i) \Rightarrow (iii) \Rightarrow (ii)$.

We will see that wiretap channel secrecy capacity is positive if and only if Eve's channel is not less noisy than Bob's.
Does not mean that Bob's channel has to be better, for two channels need not be comparable (in either sense).

Wiretap channel and secret key

The random message M_n for the wiretap channel can be regarded as SK for the channel model without public communication, defined by any two-output channel DMC that has component channels W_1 and W_2. Its security index is
$S(M_n|Z^n) = I(M_n \wedge Z^n)$.
Conversely, an $(\varepsilon_n, \delta_n)$-SK $K = K_n$ for that channel model qualifies as message M_n for the wiretap channel, providing K is uniformly distributed and equals K_A in the definition of SK. Indeed, then the joint distribution of K_n, X^n can be recovered applying a suitable stochastic encoder to a uniformly distributed RV.

We will determine wiretap channel secrecy capacity, differently from standard, via addressing SK capacity of the channel model without public communication.

Channel model, no public communication

Consider the channel model without public communication, given by a two-output channel $W : \mathcal{X} \to \mathcal{Y} \times \mathcal{Z}$. Let V, X, Y, Z be RVs satisfying $V \ominus X \ominus (Y, Z), \quad P_{YZ|X=x}(y.z) = W(y, z|x)$.

Proposition

For this channel model, $R_S = I(V \wedge Y) - I(V \wedge Z)$ is an achievable SK rate, even with $\varepsilon_n \to 0$ and $\delta_n \to 0$ exponentially fast, taking $K_n = K$ in the definition equal to K_A, i.e., a function of Alice's randomizing RV Q_A.

Proof.

(i) Modify the channel model, allowing Alice (but not Bob) to send a public message $F = F(Q_A)$.

(ii) From the modified channel model, emulate a source model with Bob silent, with generic RVs X, Y, Z, by Alice generating and transmitting over the DMC i.i.d. X^n. □

Proof continued

By previous source model result, all assertions of the Proposition hold for this emulated source model.
Moreover, in the source model both Alice's public message F and the SK $K_n = K$ may be taken as functions of V^n, where $X^n V^n$ represent i.i.d. repetitions of (X, V). Note also that Bob's estimate K_B of the SK K is a function of (Y^n, F).
These assertions carry over to the modified channel model, in which $Q_A = X^n V^n$ may be taken for Alice's randomizing RV.

(iii) Eliminate public communication
For each possible value f of F, condition the joint distribution of all RVs in (ii) on $F = f$. Since $V^n \ominus X^n \ominus Y^n Z^n$. this conditioning preserves the input-output relationship of X^n and Y^n, Z^n.
Hence the joint distribution conditioned on $F = f$ is a valid joint distribution for the channel model, it emerges if Alice generates $X^n V^n$ with distribution $P_{X^n V^n | F=f}$, and transmits this (non-i.i.d.) X^n over the DMC.

Completion of proof

(iv) As F is constant under the distribution conditioned on $F = f$, the latter is, in fact, a valid distribution for the channel model without public communication.
Let $K_n = K$ be an $(\varepsilon_n, \delta_n)$-SK as in (ii). Security index under the conditioned distribution: $S_f(K|Z^n) = \log|\mathcal{K}| - H(K|Z^n, F = f)$.

The proof is completed noting that from

$$\Pr\{K \neq K_B\} = \sum_f \Pr\{F = f\}\Pr\{K \neq K_B | F = f\} \leq \delta_n,$$

$$S(K|Z^n,F) = \log|\mathcal{K}| - H(K|Z^n,F) = \sum_f \Pr\{F = f\} S_f(K|Z^n) \leq \varepsilon_n,$$

it follows that there exists f for which the conditional distribution makes K an $(2\delta_n, 2\varepsilon_n)$-SK for the channel model without public communication.

Wiretap secrecy theorem

Theorem (Wyner 1975, Csiszár and Körner 1978)

The wiretap channel with component channels W_1, W_2 *has secrecy capacity equal to the maximum of* $I(V \wedge Y) - I(V \wedge Z)$ *for RVs*

$$V \mathrel{-\!\!\circ\!\!-} X \mathrel{-\!\!\circ\!\!-} (Y,Z), \quad P_{YZ|X=x}(y.z) = W(y,z|x), \quad |\mathcal{V}| \leq |\mathcal{X}|.$$

The same result holds for weak secrecy, as well as when requiring exponentially fast convergence to 0 *of the error probability* $\Pr\{\hat{M}_n \neq M_n\}$ *and information leakage* $I(M_n \wedge Z^n)$*.*

Corollary

The secrecy capacity is positive unless W_2 *is less noisy than* W_1*.*

Remark

If W_1 is more capable than W_2, the secrecy capacity equals the maximum of $I(X \wedge Y) - I(X \wedge Z)$.

Wiretap channel proof

For technical purposes, introduce a “physical channel” that has components $W_1,,W_2$. Its actual choice does not matter, may be $W(y,z|x) = W_1(y|x)W_2(z|x)$.

Achievability follows from the Proposition about achievable SK rates for the channel model not admitting public communication. Recall the discussion preceding that Proposition. There the SK K has been assumed uniformly distributed, but as seen before, that assumption does not restrict generality.

Constraint $|\mathcal{V}| \leq |\mathcal{X}|$: Essential part of the Theorem, as in other IT results involving auxiliartry RVs.
Standard argument via Caratheodory-Fenchel theorem (not detailed) shows that any RV V with $V \ominus X \ominus YZ$ may be substituted by another one meeting also the range constraint, with no change of $I(V \wedge Y) - I(V \wedge Z)$.

Wiretap converse proof

Suppose message $M = M_n$ is (stochastically) encoded to channel input X^n, yields outputs $Y^n,\ Z^n$, Bob decodes $\hat{M}$, where $\Pr\{\hat{M} \neq M\} \leq \varepsilon_n$, $I(M \wedge Z^n) \leq \delta_n$. Then

$$\log|\mathcal{M}| = H(M) = I(M \wedge Y^n) + H(M|Y^n)$$
$$\leq I(M \wedge Y^n) - I(M \wedge Z^n) + \delta_n + \varepsilon_n \log|\mathcal{M}| + h(\varepsilon_n).$$

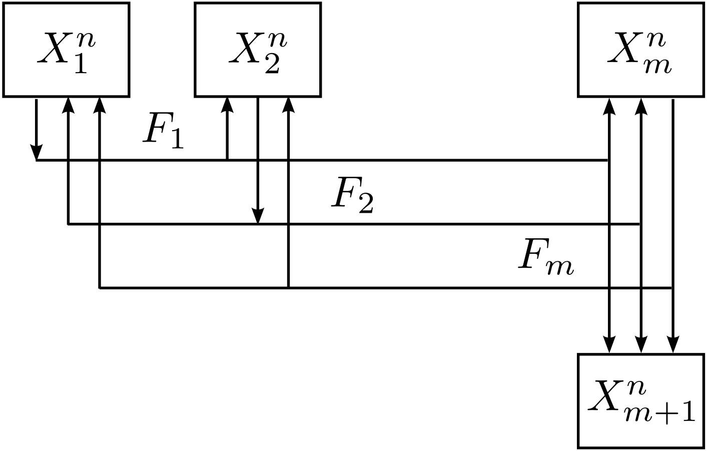
Using technical lemma with $S = M,\ T = \emptyset$, this gives

$$\frac{1}{n}\log|\mathcal{M}| \leq \frac{1}{1-\varepsilon_n}[I(M \wedge Y_J|V) - I(M \wedge Z_J|V)] + \frac{h(\varepsilon_n)+\delta_n}{n(1-\varepsilon_n)}$$

with J as before and $V = JY_1 \dots Y_{J-1}Z_{J+1} \dots Z_n$. One checks as before that $MV \ominus X_J \ominus Y_J Z_J$ with $P_{Y_J Z_J|X_J=x}(y,z) = W(y,z|x)$.

Finally, the maximum of $I(U \wedge Y|V) - I(U \wedge Z|V)$ subject to $UV \ominus X \ominus YZ$ is attained for $V = const$.

Multiterminal source model



Multi-terminal source models

Given i.i.d. repetitions of correlated RVs $(X_1, \dots, X_{m+1})$, the i'th party $(1 \leq i \leq m)$ observes X_i^n, n independent repetitions of X_i.
The eavesdropper Eve observes X_{m+1}^n.
Parties $1, \dots, m$ are allowed to communicate, perhaps interactively, all communication is received errorfree by all parties, as well as by the eavesdropper.

$$\begin{aligned} f_1 &= f_1(X_1^n),\ f_2 = f_1(X_2^n, f_1), \dots \\ f_i &= f_i(X_j^n, f_1, \dots, f_{i-1}) \quad \text{where} \quad j \equiv i\,(mod\,m) \end{aligned}$$

Randomized functions (i.e., $f_i = f_i(X_i^n, Q_i)$ where Q_i is RV generated by party i for randomization) may or may not be allowed.
Total communication: $\mathbf{F} = (f_1, \dots, f_r)$.
Non-interactive communication: $\mathbf{F} = (f_1, \dots, f_m)$, each f_i depends only on X_i^n (possibly randomized).

Goal: Generate (ε, δ)-SK for a set of parties $A \subset \{1, \dots, m\}$, that is, a RV K that represents ε-CR for the parties $i \in A$ and has security index

$$S(K|\mathbf{F}, X^n_{m+1}) = \log|\mathcal{K}| - H(K|\mathbf{F}, X^n_{m+1}) \leq \delta.$$

The parties $j \in \{1, \dots, m\} \setminus A$, if any, serve as "helpers" in generating CR.

"Trusted center" scenario: The "eavesdropper" is a trusted center or a compromised but cooperative party who helps in generating CR, by revealing X^n_{m+1}. Then, as in two-party models, we speak of PK.
Achievable SK (PK) rates and SK (PK) capacity are defined as before. These capacities will be denoted by $C_S(A)$ and $C_P(A)$.

We concentrate on determining $C_P(A)$.

A toy example

Let $X_1, \ldots, X_m$ be binary RVs either $m-1$ of them i.i.d. $(1/2, 1/2)$, and $X_1 + \cdots + X_m = 0 \pmod 2$.
Let $X_{m+1} = \text{const}$ and $A = \{1, \ldots, m\}$.
Claim: The SK (and PK) capacity equals $1/(m-1)$, achievable with perfect SK.
Protocol: Let $n = m-1$, and let the parties $1 \leq i \leq m-1 = n$ communicate their observations $X_i^n = (X_{i1}, \ldots, X_{in})$ excluding X_{ii}, while party m communicates $(X_{m1} + X_{m2}, \ldots, X_{m1} + X_{mn})$.
This communication $\mathbf{F}$ makes the whole $(X_1^n, \ldots, X_m^n)$ a perfect CR for all parties $1 \leq i \leq m$. As, for example, X_{11} is independent of the communication, $K = X_{11}$ will be a (perfect) SK. This shows that $1/(m-1)$ is an achievable SK rate. The non-achievability of larger SK rates appears non-trivial but follows from subsequent general results.

Case $m = 5, n = 4$

$m = 5, n = 4$

$$\begin{array}{ll} X_1^4 = (X_{11}, X_{12}, X_{13}, X_{14}) & f_1 = (X_{12}, X_{13}, X_{14}) \\ X_2^4 = (X_{21}, X_{22}, X_{23}, X_{24}) & f_2 = (X_{21}, X_{23}, X_{24}) \\ X_3^4 = (X_{31}, X_{32}, X_{33}, X_{34}) & f_3 = (X_{31}, X_{32}, X_{34}) \\ X_4^4 = (X_{41}, X_{42}, X_{43}, X_{44}) & f_4 = (X_{41}, X_{42}, X_{43}) \\ X_5^4 = (X_{51}, X_{52}, X_{53}, X_{54}) & f_5 = (X_{51} + X_{52}, X_{51} + X_{53}, X_{51} + X_{54}) \end{array}$$

Note: f_5 also gives the sums $X_{52} + X_{53}$, $X_{52} + X_{54}$, $X_{53} + X_{54}$ using that $X_{5i} = X_{1i} + X_{2i} + X_{3i} + X_{4i}$, all parties can recover each X_{ij} from those they see and from the other parties' public communication. $K = X_{11}$ is perfect SK.

PK-capacity and omniscience

Steps of determining PK capacity of multiterminal source models:

- Generate CR, in this case equal to $(X_1^n, \dots, X_m^n)$ (omniscience)
- Privacy amplification to obtain PK
- Proof of optimality of the achieved PK rate

Definition

The omniscience rate $R_{OS}(A)$ is the smallest number R for which, assuming Eve reveals X_{m+1}^n, the whole $X_1^n, \dots, X_m^n$ can be made ε_n-CR for the parties $i \in A$, with $\varepsilon_n \to 0$, via communication $\mathbf{F} = (f_1, \dots, f_r)$ of total rate $\frac{1}{n}\sum_{i=1}^r \log ||f_i|| \to R$.

First, non-interactive communication without randomization, $\mathbf{F} = (f_1(X_1^n), \dots, f_m(X_m^n))$ is addressed that makes $X_1^n, \dots, X_m^n$ an ε-CR for the parties $i \in A$, i.e., some function of $(\mathbf{F}, X_i^n, X_{m+1}^n)$ equals $(X_1^n, \dots, X_m^n)$ with probability $\geq 1 - \varepsilon$, for each $i \in A$.

A classical source coding result

Lemma (Wyner, Wolf and Willems 2002)

For an m-tuple $(R_1, \ldots, R_m)$*, there exists for each* $\varepsilon > 0$, $\delta > 0$ *and sufficiently large n, communication with rates*

$$\frac{1}{n} \log ||f_i|| \leq R_i + \delta,\ 1 \leq i \leq m,$$

if and only if for all sets $B \subset \{1, \ldots, m\}$ *not containing A*

$$\sum_{i \in B} R_i \geq H(X_B | X_{B^c}).$$

Notation: $X_B \triangleq \{X_i, i \in B\}$, $B^c \triangleq \{1, \ldots, m+1\} \setminus B$.

Corollary

The omniscience rate $R_{OS}(A)$ *is bounded above by the minimum of* $R_1 + \cdots + R_m$ *subject to the inequalities in the Lemma.*

Moreover, under these conditions, the error probability can be made exponentially small.

PK-capacity of multiterminal model

Recall the general result that for n independent repetitions of RVs (X, Z), any $\beta > 0$, and RV F with at most 2^{nr} possible values, for sufficiently large n there exist mappings $\kappa : \mathcal{X}^n \to \mathcal{K}$ with

$$\frac{1}{n} \log |\mathcal{K}| = H(X|Z) - r - \beta, \quad S(\kappa(X^n)|F, Z^n) < 2^{-n\xi}.$$

This, applied to $(X_1, \ldots, X_m)$ and X_{m+1} in the roles of X and Z, implies the $C_P(A) \geq \ldots$ part of the next Theorem.

Theorem (Csiszár-Narayan 2004)

$C_P(A) = H(X_1, \ldots, X_m|X_{m+1}) - R_{OS}(A)$,

where $R_{OS}(A)$ is equal to the upper bound in the last Corollary. This PK-capacity can be achieved via non-interactive communication without randomization, and with PK equal to a function of X_i^n, for either $i \in A$.

Remark: This theorem holds, with the same proof, under either version of the definition of PK capacity.

PK capacity, completion of proof

The still missing part of the theorem follows from the next Lemma, applied to the case when the ε-CR K is actually (ε, δ)-PK.

Lemma (Csiszár and Narayan 2004)

If K represents ε-CR for the parties $i \in A \subset \{1, \ldots, m\}$, achievable with (perhaps interactive) communication $\mathbf{F} = (f_1, \ldots, f_r)$, then

$$\frac{1}{n} H(K|\mathbf{F}, X_{m+1}^n) = H(X_1, \ldots, X_m | X_{m+1}) - \sum_{i=1}^{m} R_i',$$

with $(R_i', \ldots, R_m')$ such that the Slepian-Wolf inequalities in the previous Lemma hold for $R_i = R_i' + \frac{\varepsilon \log |\mathcal{K}| + 1}{n}$.

Proof of Lemma: We claim that a suitable choice for R_i' is

$$\frac{1}{n} \sum_{l \equiv i (mod\ m)} H(f_l | f_1 \ldots f_{l-1} X_{m+1}^n) + \frac{1}{n} H(X_i^n | \mathbf{F}, K, X_{[i+1, m+1]}^n).$$

Proof of Lemma continued

For any $B \subset \{1, \ldots, m\}$

$$\begin{aligned} nH(X_B|X_{B^c}) &= H(X_B^n|X_{B^c}^n) = H(\mathbf{F}, K, X_B^n|X_{B^c}^n) \\ &= \sum_{\nu=1}^{r} H(f_\nu|f_1 \ldots f_{\nu-1}\, X_{B^c}^n) + H(K|\mathbf{F}, X_{B^c}^n) \\ &\quad + \sum_{j \in B} H(X_j^n|\mathbf{F}, K, X_{B\cap[i+1,m]}^n X_{B^c}^n). \end{aligned}$$

Taking $B = \{1, \ldots, m\}$, this gives the claimed identity. For any $B = \{1, \ldots, m\}$,

$$H(f_\nu|f_1 \ldots f_{\nu-1}\, X_{B^c}^n) = 0 \ \text{ if } \ \nu \equiv i \in B^c.$$

If B does not contain A then additionally

$$H(K|\mathbf{F}, X_{B^c}^n) \leq \varepsilon \log|\mathcal{K}| + 1.$$

hence in this case

$$nH(X_B|X_{B^c}) \leq n \sum_{i \in B} R_i' + \varepsilon \log|\mathcal{K}| + 1.$$

Toy example: optimality

$X_1, \dots, X_m$ binary, $X_{m+1} = \text{const}$, $A = \{1, \dots, m\}$.
$\Pr\{X_1 = x_1, \dots, X_m = x_m\} = 2^{-(m-1)}$ if $x_1 + \cdots + x_m = 0$, otherwise 0.
For $B \subset \{1, \dots, m\}$, $H(X_B | X_{B^c}) = |B| - 1$.
Slepian-Wolf condition for $B = \{1, \dots, m\} \setminus \{j\}$:

$$\sum_{i \neq j} R_i \geq m - 2$$

Summation for $j \in \{1, \dots, m\}$ implies

$$R_{OS} \geq \frac{m(m-2)}{m-1} = m - 1 - \frac{1}{m-1}.$$

This proves that

$$C_S = H(X_1, \dots, X_m) - R_{OS} \leq \frac{1}{m-1}.$$

Three sources

Let $m = 3$, and $X_4 =$ const (Eve has no side information). The omniscience rates $R_{OS}(\{1,2,3\})$ and $R_{OS}(\{1,2\})$ equal the maximum of $R_1 + R_2 + R_3$ subject to

$$R_1 \geq H(X_1|X_2,X_3), \quad R_2 \geq H(X_2|X_1,X_3), \quad R_3 \geq H(X_3|X_1,X_2),$$

$$R_1{+}R_2 \geq H(X_1,X_2|X_3),\ R_1{+}R_3 \geq H(X_1,X_3|X_2),\ R_2{+}R_3 \geq H(X_2,X_3|X_1)$$

or same constraints omitting $R_1 + R_2 \geq H(X_1X_2|X_3)$. Then $C_S(\{1,2,3\}) = H(X_1X_2X_3) - R_{OS}(\{1,2,3\})$ equals, by simple algebra, the smallest one of $I(X_1 \wedge X_2X_3)$, $I(X_2 \wedge X_1X_3)$, $I(X_3 \wedge X_1X_2)$ and $\frac{1}{2}[H(X_1) + H(X_2) + H(X_3) - H(X_1X_2X_3)]$.

Similarly, $C_S(\{1,2\}) = \min[I(X_1 \wedge X_2X_3), I(X_2 \wedge X_1X_3)]$. Recall: $C_S(\{1,2\})$ is the largest SK rate achievable for parties 1 and 2, with help of party 3 from whom the key need not be secret.

Intuitive interpretation

If parties 2 and 3 were merged, the largest SK rate achievable for party 1 and "party 23" would be $I(X_1 \wedge X_2X_3)$; similarly for $I(X_2 \wedge X_1X_3)$ and $I(X_3 \wedge X_1X_2)$.
This makes intuitive the formula of $C_S(\{1,2,3\})$, and that the three mutual informations are upper bounds to $C_S(\{1,2,3\})$. The emergence of multiinformation (divided by 2) appears, however, less intuitive.
Interpretation of the identity

$$C_S(\{1,2\}) = I(X_1 \wedge X_2|X_3) + \min[I(X_1 \wedge X_3), I(X_2 \wedge X_3)] :$$

First term is achievable with additional secrecy from party 3 (two-user PK capacity). Second term achievable for the three parties with 1 and 2 silent.

Markov chain

Let $X_1 \ominus X_2 \ominus \cdots \ominus X_m$ be Markov chain, $X_{m+1} = \text{const}$.
Claim: $C_S(\{1, \ldots, m\}) = \min_{1 \leq i \leq m} I(X_i \wedge X_{i+1})$.
(i) $C_S(\{1, \ldots, n\}) \leq I(X_1 \ldots X_i \wedge X_{i+1} \ldots X_m) = I(X_i \wedge X_{i+1})$
(ii) For t minimizing $I(X_i \wedge I_{i+1})$ take

$$R_i = \begin{cases} H(X_i | X_{i+1}) & \text{if} \quad i \leq t \\ H(X_i | X_{i-1}) & \text{if} \quad i > t. \end{cases}$$

It can be verified that this $(R_1, \ldots, R_m)$ satisfies the constraints

$$\sum_{i \in B} R_i \geq H(X_B | X_{B^c}), \; B \subset \{1, \ldots, m\} \text{ proper}$$

Due to

$$H(X_1,\ldots,X_m)=\sum_{i=1}^{t}H(X_i|X_{i+1})+I(X_t\wedge X_{t+1})+\sum_{i=t+1}^{n}H(X_i|X_{i-1})$$

it follows that

$$C_S(\{1,\ldots,m\})=H(X_1,\ldots,X_m)-R_{OS}(\{1,\ldots,m\})\geq I(X_t\wedge X_{t+1})$$

Similarly, if $X_1 \mathrel{-\!\!\circ\!\!-} X_2 \mathrel{-\!\!\circ\!\!-} \cdots \mathrel{-\!\!\circ\!\!-} X_m \mathrel{-\!\!\circ\!\!-} X_{m+1}$ is Markov,

$$C_P(\{1,\ldots,m\})=\min_{1\leq i\leq m}I(X_i\wedge X_{i-1}|X_{m+1}).$$

Alternate formula for omniscience rate

Notation:

$\mathcal{B}(A)$: family of all sets $B \subset \{1, \ldots, m\}$ that do not contain A
$\Lambda(A)$: set of all vectors $\lambda = \{\lambda_B, B \in \mathcal{B}(A)\}$ with non-negative components satisfying

$$\sum_{B \in \mathcal{B}(A), i \in B} \lambda_B = 1, \quad i = 1, \ldots, m$$

(such vector λ is called a fractional partition of $\{1, \ldots, m\}$.)
In other words, denoting by $\mathbf{A}$ the matrix whose rows are the incidence vectors of the sets $B \in \mathcal{B}(A)$,

$$\Lambda(A) = \{\lambda : \lambda \mathbf{A} = (1, \ldots, 1), \lambda \geq 0\}$$

Proposition (Csiszár and Narayan 2004)

$$R_{OS}(A) = \max_{\lambda \in \Lambda(A)} \sum_{B \in \mathcal{B}(A)} \lambda_B H(X_B | X_{B^c}).$$

Proof of alternate formula

Proof.
By Duality Theorem of Linear Programing, for a row vector $\mathbf{c}$, matrix $\mathbf{A}$, and a column vector $\mathbf{b}$, the minimum of $\mathbf{cx}$ subject to $\mathbf{Ax} \geq \mathbf{b}$, if finite, equals the maximum of $\mathbf{yb}$ subject to $\mathbf{yA} = \mathbf{c}$, $\mathbf{y} \geq 0$.
Take $\mathbf{c} = (1, \ldots, 1)$, $\mathbf{A}$ the matrix on previous side, and $\mathbf{b} = \{H(X_B|X_{B^c}), B \in \mathcal{B}(A)\}$. Then the minimum in the Duality Theorem equals $R_{OS}(A)$, the minimum of $R_1 + \cdots + R_m$ subject to the conditions $\sum_{i \in B} R_i \geq H(X_B|X_{B^c})$, $B \in \mathcal{B}(A)$. The maximum in the Duality Theorem is that in the Proposition. □

Example. For a partition $(D_1, \ldots, D_k)$ of $\{1, \ldots, m\}$, let

$$\lambda_B \triangleq \begin{cases} \frac{1}{k-1} & \text{if } B = \{1, \ldots, m\} \setminus D_i \text{ for some } 1 \leq i \leq k, \\ 0 & \text{otherwise} \end{cases}$$

SK capacity and multi-information

If each D_i intersects A, the last equation defines a vector $\lambda \in \Lambda(A)$. Supposing $X_{m+1} = \text{const}$,

$$\sum_{B \in \mathcal{B}(A)} \lambda_B H(X_B | X_{B^c}) = \sum_{i=1}^{k} \frac{1}{k-1} [H(X_1, \dots, X_m) - H(X_i)]$$

$$= H(X_1, \dots, X_m) - \frac{1}{k-1} I(X_{D_1} \wedge \dots \wedge X_{D_k}),$$

where I denote multi-information: for RVs $Y_1, \dots, Y_k$

$$I(Y_1 \wedge Y_2 \wedge \dots \wedge Y_k) \triangleq \sum_{i=1}^{k} H(Y_i) - H(Y_1, \dots, Y_k).$$

It follows that for each $(D_1, \dots, D_k)$ as above

$$C_{SK}(A) \geq \frac{1}{k-1} I(X_{D_1} \wedge \dots \wedge X_{D_k})$$

(Csiszár and Narayan 2004). If $A = \{1, \dots, m\}$, this bound is tight for some $(D_1, \dots, D_k)$, but not necessarily otherwise (Chan 2008).

Multi-terminal channel models

Given a DMC with one input and m output terminals

$$W = \{W(x_2, \dots, x_{m+1}|x_1) : \; x_i \in \mathcal{X}_i, \; i = 1, \dots, m+1\}.$$

Party 1 controls the input, parties $2, \dots, m$ and the eavesdropper $m+1$ observe the corresponding outputs. Parties $1, \dots, m$ may also communicate publicly, errorfree accessible to all parties including the eavesdropper.
Assume unrestricted public communication is allowed.

Goal: generate SK or PK for a set of parties $A \subset \{1, \dots, m\}$, while those in $\{1, \dots, m\} \setminus A$ (if any) help by taking part in the communication but are not required to learn the key, neither to remain ignorant of it. As before, the term PK refers to an "eavesdropper" (rather, a compromised but cooperative party) who reveals her channel outputs immediately upon receipt.
Same as SK when eavesdropper lacks side information ($|\mathcal{X}_{m+1}| = 1$), otherwise we focus on PK capacities $C_P(A)$.

PK capacity theorem

Recall: $\Lambda(A)$ denotes the set of all vectors (fractional partitions)

$$\lambda = \{\lambda_B : \ B \in \mathcal{B}(A) \text{ with } \lambda_B \geq 0, \ \sum_{B: i \in B} \lambda_B = 1, \ i = 1, \ldots, m\}$$

where $\mathcal{B}(A)$ is the family of subsets of $\{1. \ldots, m\}$ not containing A.

For $\lambda \in \Lambda(A)$ and distribution P on $\mathcal{X}_1$ denote

$$G(P, W, \lambda) \triangleq H(X_1 \ldots X_m | X_{m+1}) - \sum_{B \in \mathcal{B}(A)} \lambda_B H(X_B | X_{B^c})$$

where $P_{X_1 \ldots X_{m+1}}(x_1 \ldots, x_{m+!}) = P(x_1) W(x_2, \ldots, x_{m+1} | x_1)$.

Theorem (Csiszár and Narayan 2008)

$$C_P(A) = \max_P \min_{\lambda \in \Lambda(A)} G(P, W, \lambda) = \min_{\lambda \in \Lambda(A)} \max_P G(P, W, \lambda).$$

This PK capacity is achievable with parties $2, \ldots, m$ each sending at most one public message, and party 1 not sending or listening to any public message.

Proof

Achievability of maxmin: Immediate via source emulation from multiterminal source PK theorem.

Equality of maxmin and minmax: by minimax theorem, since $G(P, W, \lambda)$ is concave in P and affine in λ. Concavity check: $H(X_B|X_{B^c})$ is affine in P when B does not contain 1, and

$$H(X_1 \dots X_m|X_{m+1}) - \sum_{B:1\in B} \lambda_B H(X_B|X_{B^c})$$

$$= \sum_{B:1\in B} \lambda_B [H(X_1 \dots X_m|X_{m+1}) - H(X_B|X_{B^c})]$$

is concave, for the terms in bracket simplify to $H(X_{B^c}|X_{m+1})$.

Converse proof is based on the minmax expression, and a technical lemma on the next slide. The details are cumbersome and omitted.

Final assertions: from multiterminal source PK theorem. Public message of party 1 is dispensed with by a conditioning argument exactly as in the two users case.

Lemma and Remark

Lemma (Csiszár and Narayan 2008)

Let $X_1, \ldots, X_m, Y$ and K be any RVs such that for each $i \in A$ some function of (X_i, Y) is equal to K with probability $\geq 1 - \varepsilon$. Then for each $\lambda \in \Lambda(A)$

$$H(K|Y) \geq H(X_1, \ldots, X_m|Y) - \sum_{B \in \mathcal{B}(A)} \lambda_B H(X_B|X_{B^c} Y) + \eta$$

where $\eta = (m+2)(\varepsilon \log |\mathcal{K}| + h(\varepsilon))$.

Remark The PK capacity theorem also gives a fundamental limit of secure transmission to several receivers. Suppose party 1 sends a message M over a DMC to $m-1$ receivers who, in order to decode M, may communicate publicly but not leaking information about M to an eavesdropper (suppose she does not have side information). Then the secure transmission capacity (largest achievable rate $\frac{1}{n} \log |\mathcal{M}|$) is equal to $C_P(\{1, \ldots, m\})$, in the special case $|\mathcal{X}_{m+1}| = 1$.

Eve controls DMC inputs

As PK is associated with scenarios in which Eve is regarded a cooperative but compromised party rather than the adversary, it is not unreasonable to consider a model where she controls the DMC inputs, and parties $1, \dots, m$ observe the outputs.

In this modified model, the DMC is given by transition probabilities $W(x_1, \dots, x_m | x_{m+1})$. The PK capacity is still given by the previous maxmin or minmax, the only difference is that the joint distribution of $X_1, \dots .X_{m+1}$ is now given by $W(x_1, \dots, x_m | x_{m+1}) P(x_{m+1})$.

In this model, the PK capacity can be attained by Eve transmitting a suitable deterministic sequence over the DMC.

Example Binary alphabets, $m = 3$, outputs X_1, X_2 by coin-tossing, $X_3 = X_1$ or X_2 according as Eve sent 0 or 1. For blocklength $n = 2$, perfect PK of 1 bit is achievable if Eve sends 01 and party 3 reveals the mod 2 sum of his bits. This achieves $C_P(\{1, 2, 3\}) = 1/2$.

Biometric secrecy system

Two stages (i) enrollment (ii) authentication
(i) Biometric features of the person are measured and suitably processed, yielding biometric sequence $X^n = X_1 \dots X_n$.
A key K is generated, as well as helper data F. Both are stored in the system, K securely, F publicly.
(ii) In authentication stage, another biometric sequence $Y^n = Y_1 \dots Y_n$ is taken, differs from X^n due to measurement noise and elapsed time. The person is authenticated if his key K can be reconstructed from Y^n and F.
Goals: possibly large key, secret from public helper data F, possibly small information in F about the personal data X^n.

Information theoretic approach

IT approach: Ignatenko and Willems 2009, Lai, Ho and Poor 2011. Different model versions, we focus on the "unconditional" versions of the "generated" and "chosen secret" models. Actually, source models of SK with one-way communication, without or with randomization. The role of Alice's public communication is played by the helper data F. Rate constraint on F: for current problem not relevant per se but enters via limiting the information F provides about X^n.

Key length and privacy leakage

"Generated secret" model: no randomization, K and F are functions of X^n
In general, a randomizing Q_A may also be used. "Chosen secret" refers to K independent of X^n, actually identifying K with Q_A.

In "generated secret" case, the information in F about X^n (called privacy leakage) is $I(F \wedge X^n) = H(F)$. Constraining this privacy leakage is tantamount to constraining communication rate in the SK model. With $C_S(R)$ denoting the corresponding SK capacity (determined previously), it follows that the optimal ratio of key length to privacy leakage is $\sup \frac{1}{R} C_S(R)$.
Note: this is attained in the limit $R \to 0$.

Generalizations

In "chosen secret" model, in general $I(F \wedge X^n) < H(F)$, and to find the best ratio of key length to privacy leakage is less straightforward. Still, the result remains true.

So far, secrecy of the key K of the biometric system has been required from adversaries knowing the public helper data. More generally, adversaries might have also side information such as biometric data of relatives of the person in question. The above results easily extend, employing SK capacity for the case when Eve has side information.