MULTITERMINAL SECRECY AND TREE PACKING

With Imre Csiszár, Sirin Nitinawarat, Chunxuan Ye, Alexander Barg and Alex Reznik
Information Theoretic Security

- A complementary approach to computational security for secret key cryptosystems.

- *Unconditional Security*: A quantifiable and provable notion of security, with no assumption of “one-way” functions and no restrictions on the computational power of an adversary.


  - an adversary does not have access to precisely the same observations as the legitimate users;
  - the legitimate plaintext messages and secret keys are, in effect, “nearly statistically independent” of the observations of the adversary.

- New insights: Innate connections with multiterminal data compression and points of contact with combinatorial tree packing algorithms.

- New algorithms: Potential rests on advances in algorithms for multiterminal data compression and a better understanding of connections with combinatorial tree packing of multigraphs.
SECRET KEY GENERATION
Multiterminal Source Model

- The terminals in $\mathcal{M} = \{1, \ldots, m\}$ observe separate but correlated signals, e.g., different noisy versions of a common broadcast signal or measurements of a parameter of the environment.

- The terminals in a given subset $A \subseteq \mathcal{M}$ wish to generate a “secret key” with the cooperation of the remaining terminals, to which end all the terminals can communicate among themselves – possibly interactively in multiple rounds – over a public noiseless channel of unlimited capacity.

- A secret key:
  - random variables (rvs) generated at each terminal in $A$ which agree with probability $\approx 1$; and
  - the rvs are effectively concealed from an eavesdropper with access to the public communication.

- The key generation exploits the correlated nature of the observed signals.

- The secret key thereby generated can be used as a one-time pad for secure encrypted communication among the terminals in $A$. 
Multiterminal Source Model

- The $m$ legitimate terminals in $\mathcal{M} = \{1, \ldots, m\}$ cooperate in secret key generation.
- $X_1, \ldots, X_m$ are finite-valued random variables (rvs) with (known) joint distribution $P_{X_1,\ldots,X_m}$.
- Each terminal $i$, $i = 1, \ldots, m$, observes a signal comprising $n$ independent and identically distributed repetitions (say, in time) of the rv $X_i$, namely the sequence $X_{i}^n = (X_{i1}, \ldots, X_{in})$.
- The signal components observed by the different terminals at successive time instants are i.i.d. according to $P_{X_1,\ldots,X_m}$. 
Multiterminal Source Model

- All the terminals are allowed to communicate over a noiseless channel of unlimited capacity, possibly interactively in several rounds.
- The communication from any terminal is observed by all the other terminals.
- The communication from a terminal is allowed to be any function of its own signal, and of all previous communication.
- Let $F$ denote collectively all the communication.
Terminals 1, \ldots, k govern the inputs of a secure discrete memoryless channel \( W \), with input terminal \( i \) transmitting a signal \( X_i^n = (X_{i1}, \ldots, X_{in}) \) of length \( n \). Terminals \( k + 1, \ldots, m \) observe the corresponding output signals, with output terminal \( i \) observing \( X_i^n \) of length \( n \).

Following each simultaneous transmission of symbols over the channel \( W \), communication over a public noiseless channel of unlimited capacity is allowed between all the terminals, perhaps interactively, and observed by all the terminals. Let \( F \) denote collectively all such public communication.

Randomization at the terminals is permitted, and is modeled by the rvs \( U_i, i = 1, \ldots, m \), which are taken to be mutually independent.
The Objective

**Objective:** To generate a secret key of the largest “size” for a given set \( A \subseteq \{1, \ldots, m\} \) of terminals, i.e., common randomness shared by the terminals in \( A \), which is

- of near uniform distribution;
- concealed from an eavesdropper that observes the public communication \( F \).

All the terminals 1, \ldots, \( m \) cooperate in achieving this goal.

**Assume:** The eavesdropper is passive and cannot wiretap.
Secret Key (SK): A random variable $K$ is a SK for the terminals in $A$, achievable with communication $F$, if

- $Pr\{K = K_i, i \in A\} \approx 1$ ("common randomness")
- $I(K \land F) \approx 0$ ("secrecy")
- $H(K) \approx \log |\text{key space}|$. ("uniformity")

Thus, a SK is effectively concealed from an eavesdropper with access to $F$, and is nearly uniformly distributed.
What is the largest rate $\lim_{n} \frac{1}{n} \log |\text{key space}|$ of such a SK for $A$ which can be achieved with suitable communication: SK capacity $C_S(A)$?

How to construct such a SK?

Hereafter, we shall restrict ourselves to the multiterminal source model.
A Toy Example: $\mathcal{M} = \{1, 2, 3\}, \ A = \{1, 2\}$

- $X^n_1, X^n_2$ are $\{0, 1\}$-valued Bernoulli ($\frac{1}{2}$) sequences, and $X^n_1$ is independent of $X^n_2$.

\[ X_{3t} = X_{1t} \oplus X_{2t}, \quad t = 1, \ldots, n. \]

- Scheme (using $n = 1$): Terminal 3 communicates publicly $X_{31} = X_{11} \oplus X_{21}$.

- Terminals 1 and 2 respectively infer $X_{21}$ and $X_{11}$.

- $X_{11}$ is independent of $\mathbf{F} = X_{31} = X_{11} \oplus X_{21}$, and is uniform on $\{0, 1\}$.

- Thus, $X_{11}$ is a perfect SK of rate 1.0 (optimal), so that SK capacity $C_S(A) = 1.0$. 
Some Related Work

- Bennett, Brassard, Crépeau, Maurer 1995.
- Csiszár 1996.
- Maurer - Wolf 1997, 2000, 2003, · · ·
Secret Key Capacity: $M = \{1, 2\} = A$

- SK capacity [Maurer ’93, Ahlswede-Csiszár ’93]:

$$C_S(A) = I(X_1 \land X_2).$$

- An interpretation:

$$C_S(A) = I(X_1 \land X_2)$$
$$= H(X_1, X_2) - [H(X_1|X_2) + H(X_2|X_1)]$$
$$= \text{Entropy rate of “omniscience” –}$$
Smallest aggregate rate of communication, $R_{CO}(A)$,
that enables the terminals in $A$ to become omniscient.
Theorem [I. Csiszár - P. N., '04, '08]:

\[ C_S(A) = H(X_1, \ldots, X_m) - \text{Smallest aggregate rate of overall interterminal communication, } R_{CO}(A), \text{ that enables all the terminals in } A \text{ to become omniscient} \]

\[ = H(X_1, \ldots, X_m) - \max_{\lambda \in \Lambda(A)} \sum_{B \in \mathcal{B}(A)} \lambda_B H(X_B | X_{B^c}) \]

and can be achieved with noninteractive communication.

Remark: \( R_{CO}(A) \) is obtained as the solution to a multiterminal data compression problem of omniscience generation that does not involve any secrecy constraints.

Interpretation: All the terminals cooperate – through public communication – in enabling the terminals in \( A \) to attain omniscience. Then the terminals in \( A \) extract a SK from their omniscience by purging the rate of this communication.
Proposition [I. Csiszár - P. N., ’04]: The smallest aggregate rate of interterminal communication, $R_{CO}(A)$, that enables all the terminals in $A$ to become omniscient, is

$$R_{CO}(A) = \min_{(R_1, \ldots, R_m) \in \mathcal{R}_{SW}(A)} \sum_{i=1}^{m} R_i,$$

where

$$\mathcal{R}_{SW}(A) = \left\{ (R_1, \cdots, R_m) : \sum_{i \in B} R_i \geq H(X_B|X_{B^c}), \forall B \subset \mathcal{M}, B \neq \emptyset, A \not\subseteq B \right\},$$

and can be achieved with noninteractive communication. Furthermore

$$R_{CO}(A) = \max_{\lambda \in \Lambda(A)} \sum_{B \in \mathcal{B}(A)} \lambda_B H(X_B|X_{B^c}).$$
How Can a Secret Key be Constructed?

• **Step 1: Data compression:** The terminals communicate over the public channel using compressed data in order to generate omniscience or some form of “common randomness.” This public communication is observed by the eavesdropper.

• **Step 2: Secret key construction:** The terminals then process this “common randomness” to extract a SK of which the eavesdropper has provably little or no knowledge.
Example: Two Terminals with Symmetrically Correlated Signals

- Terminals 1 and 2 observe, respectively, $n$ i.i.d. repetitions of the correlated rvs $X_1$ and $X_2$, where $X_1, X_2$ are $\{0, 1\}$-valued rvs with

$$P_{X_1 X_2}(x_1, x_2) = \frac{1}{2}(1 - p)\delta_{x_1 x_2} + \frac{1}{2}p (1 - \delta_{x_1 x_2}), \quad p < \frac{1}{2}.$$ 

- $C_S(\{1, 2\}) = I(X_1 \land X_2) = 1 - h_b(p)$ bit/symbol.

- Can assume: $X_1^n = X_2^n \oplus V^n$, where $V^n = (V_1, \cdots, V_n)$ is independent of $X_2^n$, and is a Bernoulli ($p$) sequence of rvs.
Step 1: Slepian-Wolf Data Compression

A.D. Wyner, 1974: Scheme for reconstructing $x_1^n$ at terminal 2

- Standard array for $(n, n - m)$ linear channel code with parity check matrix $P$ for a channel with noise $V^n$:

$$
\begin{align*}
&c_1^n \quad c_2^n \quad \ldots \quad c_j^n \quad \ldots \quad c_{2n-m}^n \\
e_2^n &\quad e_2^n+c_2^n \quad \ldots \quad e_j^n+c_j^n \quad \ldots \quad e_{2n-m}^n \\
\vdots &\quad \vdots \\
e_i^n &\quad e_i^n+c_2^n \quad \ldots \quad e_j^n+c_j^n=x_1^n \quad e_i^n+c_{2n-m}^n \\
\vdots &\quad \vdots \\
e_{2m}^n &\quad e_{2m}^n+c_2^n \quad \ldots \quad e_j^n+c_j^n \quad \ldots \quad e_{2n-m}^n
\end{align*}
$$

- Terminal 1 communicates $F = \text{the syndrome } P x_1^n$ to terminal 2.

- Terminal 2 computes the ML estimate $\hat{x}_1^n = \hat{x}_1^n(x_2^n, F)$ as:

$$\hat{x}_1^n = x_2^n \oplus f_P(P x_1^n \oplus P x_2^n),$$

where $f_P(P x_1^n \oplus P x_2^n) = \text{most likely noise sequence } v^n$ with syndrome

$$P v^n = P x_1^n \oplus P x_2^n.$$

- Thus, terminal 2 reconstructs $x_1^n$ with

$$\Pr\{\hat{X}_1^n = X_1^n\} = \cdots = \Pr\{f_P(P V^n) = V^n\} \approx 1.$$
Step 2: Secret Key Construction

C. Ye - P.N., ’05

• SK for terminals 1 and 2
  
  Terminal 1 sets $K_1 = \text{numerical index of } x_1^n \text{ in coset containing } x_1^n$;
  
  Terminal 2 sets $K_2 = \text{numerical index of } \hat{x}_1^n \text{ in coset containing } x_1^n$.

• For a systematic channel code: $K_1$ (resp. $K_2$) = first $(n - m)$ bits of $x_1^n$ (resp. $\hat{x}_1^n$).

• $K_1$ or $K_2$ forms an optimal rate SK, since:
  
  - $\Pr\{K_1 = K_2\} = \Pr\{\hat{X}_1^n = X_1^n\} \approx 1$; \hspace{1cm} \text{(common randomness)}
  
  - $I(K_1 \wedge F) = 0$; \hspace{1cm} \text{(secrecy)}

    as $K_1$ conditioned on $F = PX_1^n \sim \text{uniform } \{1, \cdots, 2^{n-m}\}$;

  - $K_1 \sim \text{uniform } \{1, \cdots, 2^{n-m}\}$; \hspace{1cm} \text{(uniformity)}

  - $\frac{1}{n} H(K_1) = \frac{n-m}{n} \approx 1 - h_b(p)$. \hspace{1cm} \text{(SK capacity)}
TREE PACKING
Pairwise Independent Network (PIN) Model

\[ X_2 = (Y_{21}, Y_{23}, \ldots, Y_{2m}) \]

\[ X_1 = (Y_{12}, \ldots, Y_{1m}) \]

\[ X_m = (Y_{m1}, Y_{m2}, \ldots, Y_{m,m-1}) \]

A special form of a multiterminal source model in which

- \( X_i = (Y_{ij}, j \in \{1, \ldots, m\}\setminus\{i\}), \ i = 1, \ldots, m; \)
- \( Y_{ij} \) is correlated with \( Y_{ji}, \ 1 \leq i \neq j \leq m; \)
- the pairs \( \{(Y_{ij}, Y_{ji})\} \) are mutually independent across \( 1 \leq i < j \leq m. \)
Secret Key Capacity for the PIN Model

Proposition [Nitinawarat et al, ’08]: For a PIN model, the SK capacity for a set of terminals $A \subseteq M = \{1, \ldots, m\}$ is

$$C_S(A) = \min_{\lambda \in \Lambda(A)} \left[ \sum_{1 \leq i < j \leq m} \left( \sum_{B \in B(A): i \in B, j \in B^c} \lambda_B \right) I(Y_{ij} \wedge Y_{ji}) \right].$$

Remark: $C_S(A)$ depends on the underlying joint probability distribution only through a linear combination of $\{I(Y_{ij} \wedge Y_{ji})\}_{i \neq j}$, i.e., the best pairwise SK rates; the corresponding pairwise SKs are mutually independent.

?? Can a SK for the set of terminals $A$ be formed by propagating independent and locally generated pairwise SKs, for instance, by some form of tree packing in an associated multigraph??
Steiner Tree Packing

\[ \mu(A, G) = 3 \]

\[ G(\mathcal{M}, E) = \text{multigraph with vertex set } \mathcal{M} \text{ and edge set } E. \]

**Definition**

- For \( A \subseteq \mathcal{M} \), a *Steiner tree* of \( G \) is a subgraph of \( G \) which is a tree and whose vertex set contains \( A \).

- A *Steiner tree packing* of \( G \) is any collection of edge-disjoint Steiner trees of \( G \). Let \( \mu(A, G) \) denote the maximum size of such a packing.
How to Generate a Secret Key by Steiner Tree Packing?

• Given a PIN model, calculate \( \{I(Y_{ij} \land Y_{ji})\}_{i \neq j}. \)

• With the given PIN model, associate a multigraph \( G^{(n)}(\mathcal{M}, E^{(n)}) \) with vertex set \( \mathcal{M} = \{1, \ldots, m\} \) and edge set \( E^{(n)} = \{e_{ij}^{(n)} = nI(Y_{ij} \land Y_{ji})\}_{i \neq j}. \)

• **Local SK generation**: For every pair of vertices \((i, j) \in G^{(n)}\), the terminals \(i, j\) generate a pairwise SK of size \(nI(Y_{ij} \land Y_{ji})\) bits; these pairwise SKs are mutually independent.

• **SK propagation by Steiner tree packing**:
  
  – **Claim**: Every Steiner tree corresponds to 1 bit of SK for the terminals in \(A\).
  – A Steiner packing of size \(p\) yields \(p\) SK bits shared by the terminals in \(A\).

**Remark**: For \(m\) fixed, this algorithm can be implemented in linear time (in \(n\)).
Secret Key Capacity and Maximal Steiner Tree Packing

**Theorem** [Nitinawarat *et al*, ’08]: For a PIN model, the SK capacity satisfies

\[ C_S(A) \geq \sup_n \frac{1}{n} \mu(A, G^{(n)}). \]

A consequence of independent interest: Given a multigraph \( G = G^{(1)} \),

- the SK capacity of an associated PIN model with

  \[ I(Y_{ij} \land Y_{ji}) = e_{ij}, \quad 1 \leq i < j \leq m \]

  provides a new (information theoretic) upper bound for the maximum rate of Steiner tree packing \( \sup_n \frac{1}{n} \mu(A, G^{(n)}) \);

- this bound is tight when \(|A| = 2\) and \(|A| = m\) but can be loose otherwise.
When $A = M = \{1, \ldots, m\}$, a Steiner tree becomes a spanning tree.

**Theorem** [Nitinawarat et al, ’08]: For a PIN model,

$$C_S(M) = \sup_n \frac{1}{n} \mu(M, G^{(n)}).$$

**Idea of proof:** By a result of Nash-Williams and Tutte,

$$\sup_n \frac{1}{n} \mu(M, G^{(n)}) = \min_{\mathcal{P}: \mathcal{P} \text{ a partition of } M} \frac{1}{|\mathcal{P}| - 1} \left( \text{No. of edges of } G^{(1)} \text{ that cross } \mathcal{P} \right),$$

which coincides with an upper bound for $C_S(M)$ in [I. Csiszár-P.N., ’04].

**Remarks:**

(i) Thus, maximal spanning tree packing attains the SK capacity $C_S(M)$.

(ii) There exists a polynomial-time algorithm (in both $m, n$) for finding a maximal collection of edge-disjoint spanning trees for $G^{(n)}$ [Gabor-Westermann] and forming an optimal rate SK.
VARIANT MODELS FOR SECRET KEY GENERATION
The legitimate user terminals in $A$ wish to generate a secret key $K$ with the cooperation of the remaining legitimate terminals, which is concealed from an eavesdropper with access to the public interterminal communication $F$ and wiretapped side information $Z^n = (Z_1, \ldots, Z_n)$.

The \textit{secrecy condition} is now strengthened to

$$I(K \land F, Z^n) \approx 0.$$
The wiretapped terminal *cooperates* in the secrecy generation by “revealing” its observations to all the legitimate terminals; the resulting key must be concealed from the eavesdropper which knows \((\mathbf{F}, Z^n)\).

??? Largest rate of a *private key* for \(A\): Known.
Largest rate of a secret key for $A$: Unknown in general but for special cases and bounds.
IN CLOSING
A Few Questions

• Information theoretic secrecy generation in a network is intertwined with multiterminal data compression and channel coding for certain network models.
  - What are the explicit connections for general network models?
  - What are the corresponding best rates of secret keys?
  - New algorithms for secret key construction?

• Multiuser secrecy generation for the PIN model has connections to the combinatorial problem of tree packing in multigraphs.
  - Tree packing algorithms for global secret key generation?
  - Information theoretic tools for tackling combinatorial tree packing problems?
Idea of Proof of SK Capacity Theorem

**Achievability**

If \( L \) represents "common randomness" for all the terminals in \( A \), achievable with communication \( F \) for some (signal) observation length \( n \), then \( \frac{1}{n}H(L|F) \) is an achievable SK rate for the terminals in \( A \).

- The terminals communicate publicly using compressed data in order to generate common randomness for the terminals in \( A \) equalling
  \[
  L \equiv \text{omniscience} = (X_1^n, \ldots, X_m^n), \text{ with } F = F_{CO} = F_{CO}(X_1^n, \ldots, X_m^n).
  \]
- The terminals in \( A \) then process this \( L \) to extract a SK of rate
  \[
  \frac{1}{n}H(L|F) \equiv \frac{1}{n}H(X_1^n, \ldots, X_m^n|F_{CO}) = H(X_1, \ldots, X_m) - \frac{1}{n}H(F_{CO})
  \]
  and of which the eavesdropper has provably little or no knowledge.

**Converse**

Tricky, since interactive communication is not excluded a priori.

**Decomposition interpretation:**

\[
\text{Omniscience} = (X_1^n, \ldots, X_m^n) \equiv (\text{Optimum secret key for } A, F_{CO}).
\]
Private Key Capacity

**Theorem** [I. Csiszár - P. N., ’04, ’08]:

\[ C_P(A|Z) = H(X_1, \ldots, X_m, Z) - H(Z) - \text{Smallest aggregate rate of public communication which enables the terminals in } A \text{ to become omniscient when all terminals additionally know } Z^n \]

\[ = H(X_1, \ldots, X_m|Z) - \max_{\lambda \in \Lambda(A|Z)} \sum_{B \in \mathcal{B}(A|Z)} \lambda_B H(X_B|X_B^c, Z) \]

and can be achieved with noninteractive communication.

**Remarks:**

- Clearly, WSK capacity \( C_W(A|Z) \leq \text{PK capacity } C_P(A|Z) \) with equality in special cases.

- Better upper bounds on WSK capacity are available due to Renner-Wolf (’03) and Gohari-Anantharam (’07, ’08).
Private Key Generation

**Achievability:**

- The terminals in $A$ generate common randomness $L$ such that

$\left( L, Z^n \right) \cong (\text{omniscience}, Z^n) = (X_1^n, \ldots, X_m^n, Z^n),$

using public interterminal communication $F_{CO} = F_{CO}(X_1^n, \ldots, X_m^n, Z^n)$ that is independent of $Z^n$.

- The terminals in $A$ then extract secrecy of rate

$\frac{1}{n} H(L, Z^n | F_{CO}, Z^n) \cong \cdots \cong H(X_1, \ldots, X_m | Z) - \frac{1}{n} H(F_{CO}).$

**Decomposition interpretation:**

$\left( X_1^n, \ldots, X_m^n, Z^n \right) \cong (\text{Optimum private key for A}, F_{CO}, Z^n).$
Open Problem: The General Wiretapper Model with \( M = \{1, 2\} = A \)

![Diagram of wiretapper model]

Gohari-Anantharam, ’07, ’08

- Terminals 1, 2 generate common randomness \( L \) using public interterminal communication \( F = F(X^n_1, X^n_2) \), such that

\[
(L, Z^n) \equiv (\text{omniscience}, Z^n) = (X^n_1, X^n_2, Z^n).
\]

Note that that \( F \) is not a function of \( Z^n \).

- The “nonsingle-letter” characterization of WSK capacity is

\[
C_W(A|Z) = \lim_{n} \max_{L, F} \frac{1}{n} H(L|F, Z^n) = \cdots = H(X_1, X_2|Z) - \lim_{n} \min_{F} \frac{1}{n} H(F|Z^n).
\]

**Question:** If \( L \) is the Slepian-Wolf codeword for the joint source \( (X^n_1, X^n_2) \) with “decoder side information” \( Z^n \), what is \( \lim_n \min_{F} \frac{1}{n} H(F|Z^n) \) where \( F \) is the interterminal communication needed to form \( L \) by distributed processing?