Secret Key Generation and Secure Computing

Himanshu Tyagi

Department of Electrical and Computer Engineering
and Institute of System Research
University of Maryland, College Park, USA.

Joint work with Prakash Narayan and Piyush Gupta

Secure Computing of a Function of Data

Correlated data is collected and stored at separated locations.

Examples include:

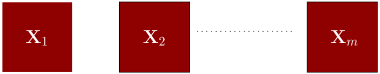
- Data grids and data centers,
- Distributed video coding,
- Sensor networks, etc.

Each location wants to know the value of a function of the data.

- using a communication that keeps the value of the function "secure".

Does there exist a communication protocol to do that?

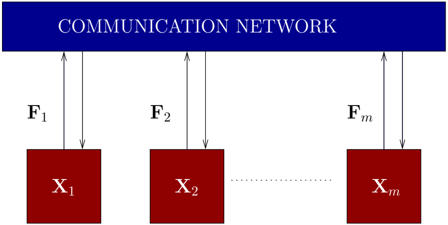
Multiterminal Source Model



Observed data: Correlated rvs $\mathbf{X}_{\mathcal{M}} = (\mathbf{X}_1, ..., \mathbf{X}_m)$.

- Probability distribution of the data is known.

Interactive Communication Protocol



- Terminals communicate over an available network.
- Multiple rounds of interactive communication are allowed.
- Interactive communication: $\mathbf{F} = \mathbf{F}_1, ..., \mathbf{F}_m$.

Assumptions

Assumption on the data

1. *Abundance of data:* Accumulated data grows with time n.
   - $\mathbf{X}_i = X_i^n = (X_{i1}, ..., X_{in})$
   - Data observed at time instance t: $X_{\mathcal{M}t} = (X_{1t}, ..., X_{mt})$.
2. Observations are i.i.d. across time:
   - $X_{\mathcal{M}1}, ..., X_{\mathcal{M}n}$ are i.i.d. rvs.
3. Observations are discrete valued.

Assumptions

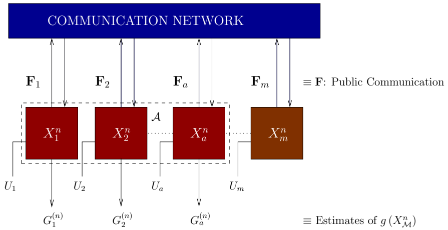
Assumption on the data

1. *Abundance of data:* Accumulated data grows with time n.
   - $\mathbf{X}_i = X_i^n = (X_{i1}, ..., X_{in})$
   - Data observed at time instance t: $X_{\mathcal{M}t} = (X_{1t}, ..., X_{mt})$.
2. Observations are i.i.d. across time:
   - $X_{\mathcal{M}1}, ..., X_{\mathcal{M}n}$ are i.i.d. rvs.
3. Observations are discrete valued.

Assumptions on the protocol

1. Each terminal has access to all the communication.
2. Transmission depends on local data and previous communication.
   - interactive communication over multiple rounds.

Secure Computing of Functions



Secure computability of g by $\mathcal{A}$:

$$\Pr\left(G_i^{(n)} = g\left(X_{\mathcal{M}}^n\right), i \in \mathcal{A}\right) \approx 1 : \quad \text{Recoverability}$$

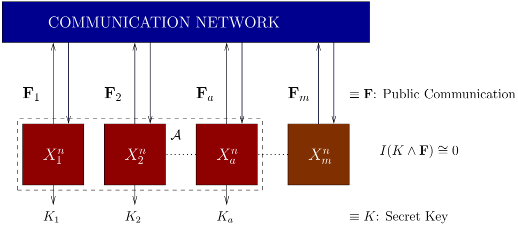
$$I\left(g\left(X_{\mathcal{M}}^n\right) \wedge \mathbf{F}\right) \approx 0 : \quad \text{Secrecy}$$

- Single-letter function: $g\left(X_{\mathcal{M}}^n\right) = \left(g\left(X_{\mathcal{M}1}\right), \ldots, g\left(X_{\mathcal{M}n}\right)\right)$.
- Notation: $G = g\left(X_{\mathcal{M}}\right), \quad G^n = g\left(X_{\mathcal{M}}^n\right)$.

When is a given function g securely computable?

Secret Key Generation

- [Maurer 1993, Ahlswede-Csiszár '93, Csiszár-Narayan '04]
  Agreeing on secret bits using public communication.



- Terminals in $\mathcal{A}$ form estimates of the key.
  - Recoverability:

$$\mathtt{Pr}\left(K_1 = K_2 = \ldots = K_a = K\right) \approx 1.$$

  - Security:

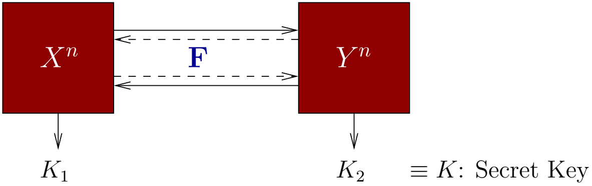
$$I(K \wedge \mathbf{F}) \approx 0.$$

Secret Key Capacity

Rate of the secret key $= \frac{1}{n} H(K)$.

Secret key capacity $C(\mathcal{A})$ = maximum achievable rate of a secret key.

For two terminals



[Maurer '93, Ahlswede-Csiszár '93]

$$C = I(X \wedge Y).$$

Optimum Rate SK for Two Terminals

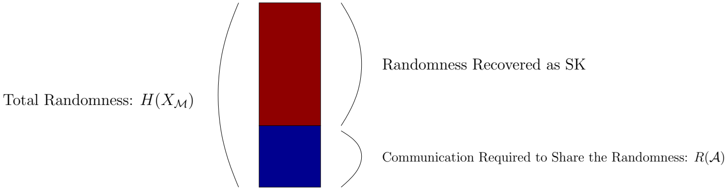
- Maurer-Ahlswede-Csiszár
  - *Common randomness* (CR) generated: X^n or Y^n.
  - Rate of communication required $= \min\{H(X|Y); H(Y|X)\}$.
  - Decomposition:
    $H(X) = H(X \mid Y) + I(X \wedge Y)$,
    $H(Y) = H(Y \mid X) + I(X \wedge Y)$.
- Csiszár-Narayan
  - *Common randomness* generated: X^n, Y^n.
  - Rate of communication required $= H(X|Y) + H(Y|X)$.
  - Decomposition:
    $H(X,Y) = H(X \mid Y) + H(Y \mid X) + I(X \wedge Y)$.
- A generalized decomposition: [Tyagi ISIT '11]

Secret Key Capacity

[Csiszár-Narayan '04]

Omniscience: Having an access to all the randomness.

$R(\mathcal{A}) \equiv$ Communication for omniscience at $\mathcal{A}$.

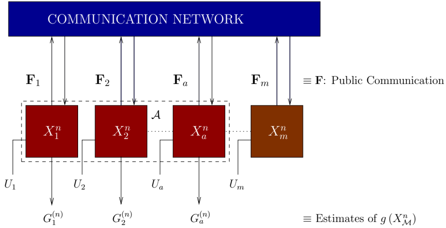


SK capacity:

$$C = H(X_{\mathcal{M}}) - R(\mathcal{A}).$$

$R_{CO}(\mathcal{A})$ can be characterized in a *single-letter-form*.

Secure Computing of Functions



Secure computability of g by $\mathcal{A}$:

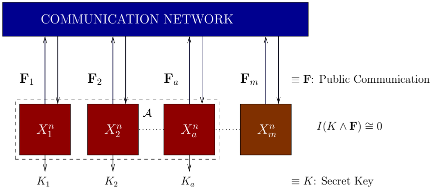
$$\texttt{Pr}\left(G_i^{(n)} = G^n, i \in \mathcal{A}\right) \approx 1 : \quad \text{Recoverability}$$

$$I\left(g\left(X_{\mathcal{M}}^n\right) \wedge \mathbf{F}\right) \approx 0 : \quad \text{Secrecy}$$

When is a given function g securely computable?

A Necessary Condition

Secret Key Generation

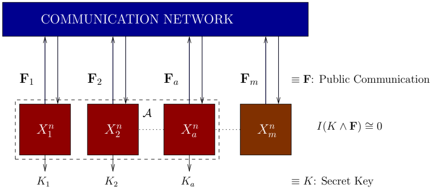


[Csiszár-Narayan ‘04]

$$C(\mathcal{A}) = H\left(X_{\mathcal{M}}\right) - R(\mathcal{A}),$$

A Necessary Condition

Secret Key Generation



[Csiszár-Narayan '04]

$$C(\mathcal{A}) = H\left(X_{\mathcal{M}}\right) - R(\mathcal{A}),$$

If g is securely computable by $\mathcal{A}$,

$$H(G) \leq C(\mathcal{A}).$$

Is $H(G) < C(\mathcal{A})$ sufficient?

All terminals wish to compute: $\mathcal{A} = \mathcal{M}$

[TNG '10]

If $H(G) < C(\mathcal{M}) \Rightarrow$ a protocol for SC of g by $\mathcal{M}$ exists.

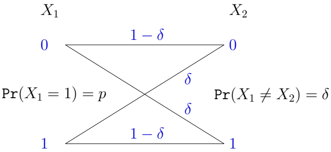
Is $H(G) < C(\mathcal{A})$ sufficient?

All terminals wish to compute: $\mathcal{A} = \mathcal{M}$

[TNG '10]

If $H(G) < C(\mathcal{M}) \Rightarrow$ a protocol for SC of g by $\mathcal{M}$ exists.

An Example for $m = 2$



- $g(x_1, x_2) = x_1 + x_2 \mod 2 \Rightarrow H(G) = h(\delta)$.
- $C(\{1,2\}) = I(X_1 \wedge X_2) = h((1-p)\delta + p(1-\delta)) - h(\delta)$.
- g is securely computable if

$$2h(\delta) < h((1-p)\delta + p(1-\delta)).$$

Example: Secure Computation of Parity

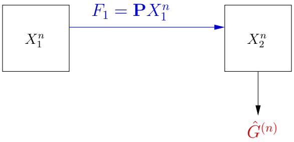
Binary Symmetric Sources: $p = \frac{1}{2}$

- Secure computability condition: $h(\delta) < 1 - h(\delta)$.
- $\mathbf{P}$: parity check matrix of a *linear* SW code for X_1 given X_2.
- $I(G^n \wedge X_1^n) = 0 \Rightarrow I(G^n \wedge F_1) = 0$.
- K: location of X_1^n in the coset of the standard array (for $\mathbf{P}$).
- Rate of $K = 1 - h(\delta)$.
- $I(K \wedge F_1) = 0$.
- Can show: $I(K \wedge F_1, G^n) = 0$.
- $I(G^n \wedge F_2, F_1) = I(G^n \wedge F_2 \mid F_1) \tilde{\leq} I(K \wedge F_1, G^n) = 0$.

Example: Secure Computation of Parity

Binary Symmetric Sources: $p = \frac{1}{2}$

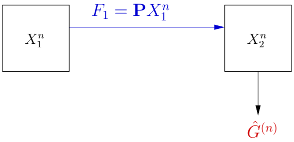
- Secure computability condition: $h(\delta) < 1 - h(\delta)$.
- $\mathbf{P}$: parity check matrix of a *linear* SW code for X_1 given X_2.
- $I(G^n \wedge X_1^n) = 0 \Rightarrow I(G^n \wedge F_1) = 0$.
- K: location of X_1^n in the coset of the standard array (for $\mathbf{P}$).
- Rate of $K = 1 - h(\delta)$.
- $I(K \wedge F_1) = 0$.
- Can show: $I(K \wedge F_1, G^n) = 0$.
- $I(G^n \wedge F_2, F_1) = I(G^n \wedge F_2 \mid F_1) \tilde{\leq} I(K \wedge F_1, G^n) = 0$.



Example: Secure Computation of Parity

Binary Symmetric Sources: $p = \frac{1}{2}$

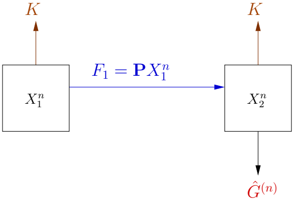
- Secure computability condition: $h(\delta) < 1 - h(\delta)$.
- $\mathbf{P}$: parity check matrix of a *linear* SW code for X_1 given X_2.
- $I(G^n \wedge X_1^n) = 0 \Rightarrow I(G^n \wedge F_1) = 0$.
- K: location of X_1^n in the coset of the standard array (for $\mathbf{P}$).
- Rate of $K = 1 - h(\delta)$.
- $I(K \wedge F_1) = 0$.
- Can show: $I(K \wedge F_1, G^n) = 0$.
- $I(G^n \wedge F_2, F_1) = I(G^n \wedge F_2 \mid F_1) \tilde{\leq} I(K \wedge F_1, G^n) = 0$.



Example: Secure Computation of Parity

Binary Symmetric Sources: $p = \frac{1}{2}$

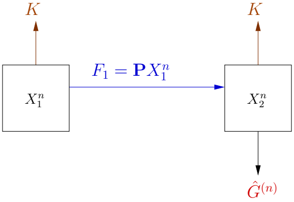
- Secure computability condition: $h(\delta) < 1 - h(\delta)$.
- $\mathbf{P}$: parity check matrix of a *linear* SW code for X_1 given X_2.
- $I(G^n \wedge X_1^n) = 0 \Rightarrow I(G^n \wedge F_1) = 0$.
- K: location of X_1^n in the coset of the standard array (for $\mathbf{P}$).
- Rate of $K = 1 - h(\delta)$.
- $I(K \wedge F_1) = 0$.
- Can show: $I(K \wedge F_1, G^n) = 0$.
- $I(G^n \wedge F_2, F_1) = I(G^n \wedge F_2 \mid F_1) \tilde{\leq} I(K \wedge F_1, G^n) = 0$.



Example: Secure Computation of Parity

Binary Symmetric Sources: $p = \frac{1}{2}$

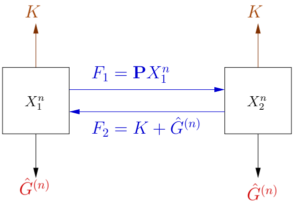
- Secure computability condition: $h(\delta) < 1 - h(\delta)$.
- $\mathbf{P}$: parity check matrix of a *linear* SW code for X_1 given X_2.
- $I(G^n \wedge X_1^n) = 0 \Rightarrow I(G^n \wedge F_1) = 0$.
- K: location of X_1^n in the coset of the standard array (for $\mathbf{P}$).
- Rate of $K = 1 - h(\delta)$.
- $I(K \wedge F_1) = 0$.
- Can show: $I(K \wedge F_1, G^n) = 0$.
- $I(G^n \wedge F_2, F_1) = I(G^n \wedge F_2 \mid F_1) \tilde{\leq} I(K \wedge F_1, G^n) = 0$.



Example: Secure Computation of Parity

Binary Symmetric Sources: $p = \frac{1}{2}$

- Secure computability condition: $h(\delta) < 1 - h(\delta)$.
- $\mathbf{P}$: parity check matrix of a *linear* SW code for X_1 given X_2.
- $I(G^n \wedge X_1^n) = 0 \Rightarrow I(G^n \wedge F_1) = 0$.
- K: location of X_1^n in the coset of the standard array (for $\mathbf{P}$).
- Rate of $K = 1 - h(\delta)$.
- $I(K \wedge F_1) = 0$.
- Can show: $I(K \wedge F_1, G^n) = 0$.
- $I(G^n \wedge F_2, F_1) = I(G^n \wedge F_2 \mid F_1) \overset{\sim}{\leq} I(K \wedge F_1, G^n) = 0$.

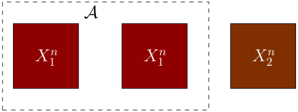


Is $H(G) < C(\mathcal{A})$ sufficient?

All terminals wish to compute: $\mathcal{A} = \mathcal{M}$ [TNG '10]

If $H(G) < C(\mathcal{M}) \Rightarrow$ a protocol for SC of g by $\mathcal{M}$ exists.

Counterexample for $\mathcal{A} \subsetneq \mathcal{M}$



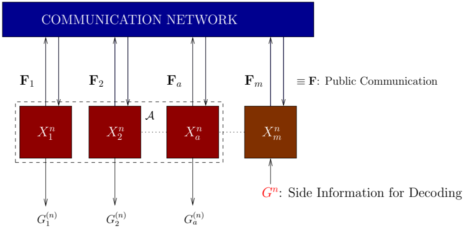
$X_1 \perp\!\!\!\perp X_2$

- $g(x_1, x_1, x_2) = x_2$.
- Let $H(X_2) < H(X_1) = C(\mathcal{A})$.$\rightarrow$ $H(G) < C(\mathcal{A})$ is satisfied.

However, g is clearly **not securely computable**.

A New Necessary Condition

If G^n is securely computable by $\mathcal{A}$:



Provide G^n as side information to terminals in $\mathcal{A}^c$.

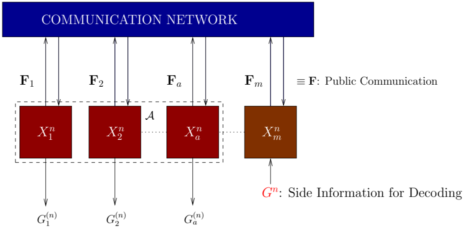
- Available *only for decoding* but *not for communicating*.

G^n forms a secret key for all terminals, termed an aided secret key.

- Let $C_{g,\mathcal{A}}(\mathcal{M})$ be that largest achievable rate of such a key.

A New Necessary Condition

If G^n is securely computable by $\mathcal{A}$:



Provide G^n as side information to terminals in $\mathcal{A}^c$.

- Available *only for decoding* but *not for communicating*.

G^n forms a secret key for all terminals, termed an aided secret key.

- Let $C_{g,\mathcal{A}}(\mathcal{M})$ be that largest achievable rate of such a key.

For a g securely computable by $\mathcal{A}$,

$$H(G) \leq C_{g,\mathcal{A}}(\mathcal{M})$$

Aided Secret Key Capacity

Theorem

The aided secret key capacity is

$$C_{g,\mathcal{A}}(\mathcal{M}) = H(X_{\mathcal{M}}) - R_{g,\mathcal{A}}(\mathcal{M}),$$

where

$R_{g,\mathcal{A}}(\mathcal{M}) =$ *min. sum rate of communication for omniscience at* $\mathcal{M}$
when G^n *is available as side information for decoding to terminals in* $\mathcal{A}^c$*.*

Characterization of Securely Computable Functions

Theorem

If g is securely computable by $\mathcal{A}$: $H(G) \leq C_{g,\mathcal{A}}(\mathcal{M})$.

Conversely, g is securely computable by $\mathcal{A}$ if: $H(G) < C_{g,\mathcal{A}}(\mathcal{M})$.

For securely computable function g:

- *Omniscience can be obtained at $\mathcal{A}$ using $\mathbf{F} \underset{\sim}{\perp\!\!\!\perp} G^n$.*
- *Noninteractive communication suffices.*
- *Randomization is not needed.*

Sketch of the Proof

Consider random binning *of appropriate rate* at each terminal:

- To allow omniscience at $\mathcal{M}$,
  with G^n *given to the terminals in* $\mathcal{A}^c$ *for decoding.*
- To keep bin indices independent of G^n.

Sketch of the Proof

1. $H(G) < C_{g,\mathcal{A}}(\mathcal{M}) = H(X_{\mathcal{M}}) - R_{g,\mathcal{A}}(\mathcal{M})$
   $\Leftrightarrow H\left(X_{\mathcal{M}} \mid G\right) > R_{g,\mathcal{A}}(\mathcal{M})$.

Sketch of the Proof

1. $H(G) < C_{g,\mathcal{A}}(\mathcal{M}) = H(X_{\mathcal{M}}) - R_{g,\mathcal{A}}(\mathcal{M})$
   $\Leftrightarrow H\left(X_{\mathcal{M}} \mid G\right) > R_{g,\mathcal{A}}(\mathcal{M})$.

2. Generate random mappings $F_i = F_i(X_i^n)$ of rate R_i:
   $\sum_i R_i \approx R_{g,\mathcal{A}}(\mathcal{M})$ with $(R_1, ..., R_m)$ s.t.
   - it enables omniscience at $\mathcal{M}$ with side information G^n given to the terminals in $\mathcal{A}^c$ only for decoding.

Sketch of the Proof

1. $H(G) < C_{g,\mathcal{A}}(\mathcal{M}) = H(X_{\mathcal{M}}) - R_{g,\mathcal{A}}(\mathcal{M})$
   $\Leftrightarrow H\left(X_{\mathcal{M}} \mid G\right) > R_{g,\mathcal{A}}(\mathcal{M})$.

2. Generate random mappings $F_i = F_i(X_i^n)$ of rate R_i:
   $\sum_i R_i \approx R_{g,\mathcal{A}}(\mathcal{M})$ with $(R_1, ..., R_m)$ s.t.
   - it enables omniscience at $\mathcal{M}$ with side information G^n given to the terminals in $\mathcal{A}^c$ only for decoding.

3. Observe: $I(F_{\mathcal{M}} \wedge G^n) \leq \sum_{i}^{m} I\left(F_i \wedge G^n, F_{\mathcal{M}\setminus\{i\}}\right)$.

Sketch of the Proof

1. $H(G) < C_{g,\mathcal{A}}(\mathcal{M}) = H(X_\mathcal{M}) - R_{g,\mathcal{A}}(\mathcal{M})$
   $\Leftrightarrow H\left(X_\mathcal{M} \mid G\right) > R_{g,\mathcal{A}}(\mathcal{M})$.

2. Generate random mappings $F_i = F_i(X_i^n)$ of rate R_i:
   $\sum_i R_i \approx R_{g,\mathcal{A}}(\mathcal{M})$ with $(R_1, ..., R_m)$ s.t.
   - it enables omniscience at $\mathcal{M}$ with side information G^n given to the terminals in $\mathcal{A}^c$ only for decoding.

3. Observe: $I(F_\mathcal{M} \wedge G^n) \leq \sum_i^m I\left(F_i \wedge G^n, F_{\mathcal{M}\setminus\{i\}}\right)$.

4. To prove:
   With high probability $I\left(F_i \wedge G^n, F_{\mathcal{M}\setminus\{i\}}\right) \cong 0$, for each i.

- To prove:
  With high probability $I\left(F_i \wedge G^n, F_{\mathcal{M}\setminus\{i\}}\right) \cong 0$, for each i.

- Shall show:
  For almost all $(\mathbf{y}, \mathbf{z})$:
  $$F_i \mid \{G^n = \mathbf{y}, F_{\mathcal{M}\setminus\{i\}} = \mathbf{z}\} \approx \text{uniform}.$$

- Family of distributions on X_i^n : $\left\{P_{X_i^n \mid \{G^n = \mathbf{y}, F_{\mathcal{M}\setminus\{i\}} = \mathbf{z}\}}\right\}$.

- Seek conditions for random mappings to be uniformly distributed
  - w.r.t. a given family of distributions.

- **Balanced Coloring Lemma:**
  [R. Ahlswede-I. Csiszár, '98], [I. Csiszár-P.N., '04]

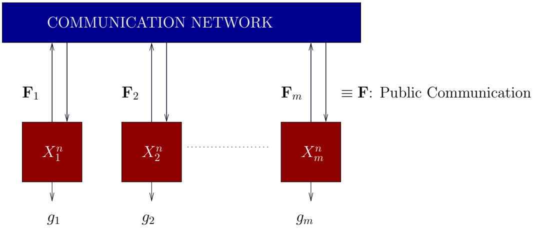
  Given a family of distributions with probabilities uniformly bounded above,

  $$\texttt{Pr}\left(\text{random coloring} \approx \text{uniform, w.r.t. all pmfs in the family}\right) \geq q,$$

  where q depends on the size of the family, the uniform bound and the rate of coloring.

- For the case at hand: a slightly generalized version is applied.
  - $q = q(n)$ grows to 1 super-exponentially in n.

Secure Computability of Multiple Functions



Secrecy Condition: $I\left(\mathbf{F} \wedge G_1^n, ..., G_m^n\right) \approx 0.$

$\boxed{\text{Which functions } g_1, ..., g_m \text{ are securely computable?}}$

Omniscience is not allowed in general

- For $m = 2$: $X_1 \perp\!\!\!\perp X_2 \quad g_i(x_1, x_2) = x_i.$