On the Communication Complexity for SK Generation in the Multiterminal Source Model

Manuj Mukherjee Navin Kashyap

Chung Chan Qiaoqiao Zhou Yogesh Sankarasubramaniam

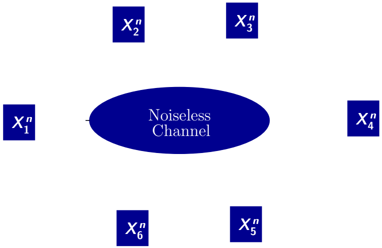
Indian Institute of Science, Bangalore

EECS Divisional Symposium

Outline

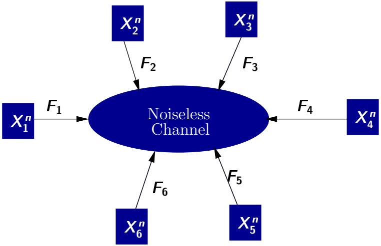
1. The Multiterminal Source Model
2. Lower Bound on R_{SK}
3. R_{SK}-maximal Sources

Preliminaries



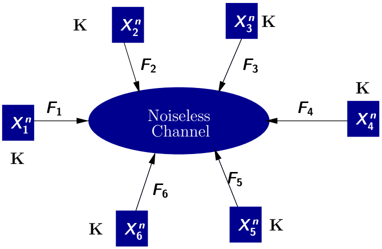
- A set of terminals $\mathcal{M} = \{1, 2, \ldots, m\}$ wants to generate a group secret key.
- Each terminal has a component of a discrete memoryless multiple source, X_i^n, $\forall 1 \leq i \leq m$.

Preliminaries



- Interactive communication is allowed among the terminals.
- $\mathbf{F} = \{F_1, F_2, \cdots, F_r\}$ is the *interactive communication* taking values in $\mathcal{F}$.
  Here F_j sent by some terminal i is a function of X_i^n and all the previous communication.
- Communication rate $= \frac{1}{n} \log |\mathcal{F}|$.
- After the communication the terminals compute a group secret key (SK) $\mathbf{K}^{(n)} = \mathbf{K}^{(n)}(X_{\mathcal{M}}^n)$.

Preliminaries



- Interactive communication is allowed among the terminals.
- $\mathbf{F} = \{F_1, F_2, \cdots, F_r\}$ is the *interactive communication* taking values in $\mathcal{F}$.
  Here F_j sent by some terminal i is a function of X_i^n and all the previous communication.
- Communication rate $= \frac{1}{n}\log|\mathcal{F}|$.
- After the communication the terminals compute a group secret key (SK) $\mathbf{K}^{(n)}{=}\mathbf{K}^{(n)}(X_{\mathcal{M}}^n)$.

Preliminaries(contd.)

- The secret key $\mathbf{K}^{(n)}$ satisfies the following property: for any $\epsilon > 0$ and for all sufficiently large n,
    - $\exists$ some function $g_i^{(n)}(X_i^n, \mathbf{F})$ such that $\Pr(\mathbf{K}^{(n)} \neq g_i^{(n)}(X_i^n, \mathbf{F})) \leq \epsilon, \forall 1 \leq i \leq m$. (*Recoverability*)
    - $I(\mathbf{K}^{(n)}; \mathbf{F}) \leq \epsilon$ (*Strong secrecy*)
    - $\log|\mathcal{K}^{(n)}| - H(\mathbf{K}^{(n)}) \leq \epsilon$, where $\mathcal{K}^{(n)}$ is the range of $\mathbf{K}^{(n)}$. (*Uniformity*)
- If $\frac{1}{n}H(\mathbf{K}^{(n)}) \rightarrow R$ as $n \rightarrow \infty$, then R is an achievable secret key rate.
- Secret key capacity $\mathcal{C}(\mathcal{M}) = \sup R$.

Preliminaries(contd.)

- The secret key $\mathbf{K}^{(n)}$ satisfies the following property:
  for any $\epsilon > 0$ and for all sufficiently large n,
  - $\exists$ some function $g_i^{(n)}(X_i^n, \mathbf{F})$ such that
    $\Pr(\mathbf{K}^{(n)} \neq g_i^{(n)}(X_i^n, \mathbf{F})) \leq \epsilon, \forall 1 \leq i \leq m$. (*Recoverability*)
  - $I(\mathbf{K}^{(n)}; \mathbf{F}) \leq \epsilon$ (*Strong secrecy*)
  - $\log|\mathcal{K}^{(n)}| - H(\mathbf{K}^{(n)}) \leq \epsilon$, where $\mathcal{K}^{(n)}$ is the range of $\mathbf{K}^{(n)}$. (*Uniformity*)

- If $\frac{1}{n} H(\mathbf{K}^{(n)}) \to R$ as $n \to \infty$, then R is an achievable secret key rate.
- Secret key capacity $\mathcal{C}(\mathcal{M}) = \sup R$.

Evaluating SK Capacity

- $\mathcal{C}(\mathcal{M}) = H(X_{\mathcal{M}}) - \min_{(R_1,R_2,\ldots,R_m)\in\mathcal{R}_{\mathrm{CO}}} \sum_{i=1}^{m} R_i$.
[Csiszár & Narayan, 2004]

  - Here
  $$\mathcal{R}_{\mathrm{CO}} = \Big\{(R_1, R_2, \ldots, R_m) : R_i \geq 0, \forall 1 \leq i \leq m, \sum_{j\in B} R_j \geq H(X_B|X_{B^c}), \forall B \subsetneq \mathcal{M}, B \neq \phi\Big\}$$
  is the achievable communication rate region for all terminals to recover $\mathbf{X}_{\mathcal{M}}^n$.
  - $R_{\mathrm{CO}} = \min_{(R_1,R_2,\ldots,R_m)\in\mathcal{R}_{\mathrm{CO}}} \sum_{i=1}^{m} R_i$ is called the minimum rate of communication for omniscience.

Evaluating SK Capacity

- $\mathcal{C}(\mathcal{M}) = \min_{\mathcal{P}} \Delta(\mathcal{P})$. [Chan & Zheng, 2010]

  - $\Delta(\mathcal{P}) = \frac{1}{\ell-1}\left[H(X_{A_1}) + H(X_{A_2}) + \cdots + H(X_{A_\ell}) - H(X_{\mathcal{M}})\right]$.

  - Here $\mathcal{P} = \{A_1, A_2, \ldots, A_\ell\}$, $\ell \geq 2$, is a partition of $\mathcal{M}$ and $X_A = (X_i : i \in A)$.

  - The quantity $\min_{\mathcal{P}} \Delta(\mathcal{P})$ is called multipartite information.

    Observe for the case of $m = 2$ the quantity $\min_{\mathcal{P}} \Delta(\mathcal{P})$ equals $I(X_1; X_2)$.

  - There exists a unique finest partition $\mathcal{P}^*$ of $\mathcal{M}$ which minimizes $\Delta(\mathcal{P})$.

    We shall refer to $\mathcal{P}^*$ as the fundamental partition.

Communication Complexity

- R_{SK} = Communication complexity,
  is the minimum rate of communication required to achieve SK capacity.
- $R_{\mathrm{SK}} \leq R_{\mathrm{CO}}$. [Csiszár & Narayan, 2004]
- If $R_{\mathrm{SK}} = R_{\mathrm{CO}}$, we call the source R_{SK}-maximal.
  These are thus the worst-case sources in terms of communication rates.
- Can we compute R_{SK}?

Communication Complexity

- R_{SK} = Communication complexity,
  is the minimum rate of communication required to achieve SK capacity.
- $R_{\text{SK}} \leq R_{\text{CO}}$. [Csiszár & Narayan, 2004]
- If $R_{\text{SK}} = R_{\text{CO}}$, we call the source R_{SK}-maximal.
  These are thus the worst-case sources in terms of communication rates.
- Can we compute R_{SK}?

Lower Bound on Communication Complexity

Theorem (Mukherjee & Kashyap, '16)

$R_{SK} \geq CI(X_{\mathcal{M}}) - \mathrm{I}(X_{\mathcal{M}})$.

- The result is an extension of Tyagi's earlier work for two terminals, i.e., $m = 2$.
- $\mathrm{CI}(X_{\mathcal{M}})$ is the minimum rate of interactive common information.
- Fact: $H(X_{\mathcal{M}}) \geq \mathrm{CI}(X_{\mathcal{M}}) \geq \mathrm{I}(X_{\mathcal{M}})$

  and hence the lower bound is non-negative.
- $\mathrm{CI}(X_{\mathcal{M}})$ is difficult to compute in general.

  Exact computation of $\mathrm{CI}(X_{\mathcal{M}})$ is possible for the special case of the hypergraphical source.

Lower Bound on Communication Complexity

Theorem (Mukherjee & Kashyap, '16)

$R_{SK} \geq CI(X_{\mathcal{M}}) - \mathrm{I}(X_{\mathcal{M}})$.

- The result is an extension of Tyagi's earlier work for two terminals, i.e., $m = 2$.
- $\mathrm{CI}(X_{\mathcal{M}})$ is the minimum rate of interactive common information.
- Fact: $H(X_{\mathcal{M}}) \geq \mathrm{CI}(X_{\mathcal{M}}) \geq \mathrm{I}(X_{\mathcal{M}})$

  and hence the lower bound is non-negative.
- $\mathrm{CI}(X_{\mathcal{M}})$ is difficult to compute in general.

  Exact computation of $\mathrm{CI}(X_{\mathcal{M}})$ is possible for the special case of the hypergraphical source.

Lower Bound on Communication Complexity

Theorem (Mukherjee & Kashyap, '16)

$R_{SK} \geq CI(\mathbf{X}_{\mathcal{M}}) - \mathbf{I}(\mathbf{X}_{\mathcal{M}})$.

- The result is an extension of Tyagi's earlier work for two terminals, i.e., $m = 2$.
- $\mathrm{CI}(\mathbf{X}_{\mathcal{M}})$ is the minimum rate of interactive common information.
- Fact: $H(\mathbf{X}_{\mathcal{M}}) \geq \mathrm{CI}(\mathbf{X}_{\mathcal{M}}) \geq \mathbf{I}(\mathbf{X}_{\mathcal{M}})$

  and hence the lower bound is non-negative.

- $\mathrm{CI}(\mathbf{X}_{\mathcal{M}})$ is difficult to compute in general.

  Exact computation of $\mathrm{CI}(\mathbf{X}_{\mathcal{M}})$ is possible for the special case of the hypergraphical source.

Evaluating $\mathrm{CI}(X_{\mathcal{M}})$: The Hypergraphical Source

- Consider a hypergraph $\mathcal{H} = (\mathcal{V}, \mathcal{E})$.
- $\mathcal{V} = \mathcal{M}$.
- Associate with each hyperedge $e \in \mathcal{E}$ an i.i.d. sequence of n Bernoulli $(1/2)$ random variables ξ_e^n.
- Random variables associated with distinct hyperedges in $\mathcal{E}$ are independent.
- Define a multiterminal source as follows: $X_i^n = (\xi_e^n : e \in \mathcal{E} \text{ such that } i \in e)$.
- The multiterminal source $X_{\mathcal{M}}^n$ is known as the hypergraphical source.

Theorem

For a hypergraphical source we have $\mathrm{CI}(X_{\mathcal{M}}) = |\mathcal{E}_{\mathcal{P}^*}|$, where $\mathcal{E}_{\mathcal{P}^*}$ is the set of hyperedges intersecting with at least two parts of $\mathcal{P}^*$.

Evaluating $\mathrm{CI}(X_{\mathcal{M}})$: The Hypergraphical Source

- Consider a hypergraph $\mathcal{H} = (\mathcal{V}, \mathcal{E})$.
- $\mathcal{V} = \mathcal{M}$.
- Associate with each hyperedge $e \in \mathcal{E}$ an i.i.d. sequence of n Bernoulli $(1/2)$ random variables ξ_e^n.
- Random variables associated with distinct hyperedges in $\mathcal{E}$ are independent.
- Define a multiterminal source as follows: $X_i^n = (\xi_e^n : e \in \mathcal{E}$ such that $i \in e)$.
- The multiterminal source $X_{\mathcal{M}}^n$ is known as the hypergraphical source.

Theorem

For a hypergraphical source we have $\mathrm{CI}(X_{\mathcal{M}}) = |\mathcal{E}_{\mathcal{P}^*}|$, where $\mathcal{E}_{\mathcal{P}^*}$ is the set of hyperedges intersecting with at least two parts of $\mathcal{P}^*$.

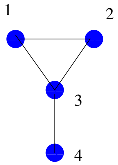
Evaluating $\mathrm{CI}(X_{\mathcal{M}})$: The Hypergraphical Source

- Consider a hypergraph $\mathcal{H} = (\mathcal{V}, \mathcal{E})$.
- $\mathcal{V} = \mathcal{M}$.
- Associate with each hyperedge $e \in \mathcal{E}$ an i.i.d. sequence of n Bernoulli $(1/2)$ random variables ξ_e^n.
- Random variables associated with distinct hyperedges in $\mathcal{E}$ are independent.
- Define a multiterminal source as follows: $X_i^n = (\xi_e^n : e \in \mathcal{E} \text{ such that } i \in e)$.
- The multiterminal source $X_{\mathcal{M}}^n$ is known as the hypergraphical source.

Theorem

For a hypergraphical source we have $\mathrm{CI}(X_{\mathcal{M}}) = |\mathcal{E}_{\mathcal{P}^*}|$, where $\mathcal{E}_{\mathcal{P}^*}$ is the set of hyperedges intersecting with at least two parts of $\mathcal{P}^*$.

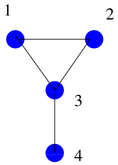
The Lower Bound is Loose



- Consider the following hypergraphical model.
  - $m = 4$ and $\mathcal{E} = \{\{1,2\},\{1,3\},\{2,3\},\{3,4\}\}$.
  - The random variables $(\xi_e)_{e\in\mathcal{E}}$ are Bernoulli (1/2) random variables.
  - $\mathcal{P}^* = \{\{1,2,3\},\{4\}\}$ and $\mathrm{I}(X_{\mathcal{M}}) = 1$.

- Therefore, $\mathrm{CI}(X_{\mathcal{M}}) = 1$ and hence, $\mathrm{CI}(X_{\mathcal{M}}) - \mathrm{I}(X_{\mathcal{M}}) = 0$.
- However, $R_{\mathrm{SK}} > 0$ as (X_1, X_2) is independent of X_4.

The Lower Bound is Loose



- Consider the following hypergraphical model.
  - $m = 4$ and $\mathcal{E} = \{\{1,2\},\{1,3\},\{2,3\},\{3,4\}\}$.
  - The random variables $(\xi_e)_{e \in \mathcal{E}}$ are Bernoulli (1/2) random variables.
  - $\mathcal{P}^* = \{\{1,2,3\},\{4\}\}$ and $\mathsf{I}(X_{\mathcal{M}}) = 1$.
- Therefore, $\mathsf{CI}(X_{\mathcal{M}}) = 1$ and hence, $\mathsf{CI}(X_{\mathcal{M}}) - \mathsf{I}(X_{\mathcal{M}}) = 0$.
- However, $R_{\text{SK}} > 0$ as (X_1, X_2) is independent of X_4.

When is a Source R_{SK}-maximal?

Theorem

A multiterminal source $X_{\mathcal{M}}$ with fundamental partition $\mathcal{P}^*$ is R_{SK}-maximal if for all $A \in \mathcal{P}^*$ we have $H(X_A|X_{A^c}) = 0$.

Theorem

A hypergraphical source $\mathcal{H} = (\mathcal{M}, \mathcal{E})$ is R_{SK}-maximal iff $\mathcal{E} = \mathcal{E}_{\mathcal{P}^*}$.

- **Example:** Hypergraphical source defined on the complete t-uniform hypergraph $K_{m,t}$:

  $\mathcal{V} = \mathcal{M}$.
  $\mathcal{E}$ is the set of all t-subsets of $\mathcal{M}$.


Thank you