How to Measure Side-Channel Leakage

Aaron Wagner
School of Electrical and Computer Engineering
Cornell University
Collaborators

Ibrahim Issa
Cornell → EPFL → AUB

Sudeep Kamath
Princeton → $$$

Ed Suh
Cornell

Ben Wu
Cornell
Packet-Timing Side Channel
ssh: keystrokes are sent as separate packets.
ssh: keystrokes are sent as separate packets.
Packet timing ↔ keystroke timing ↔ typed letters
Packet-Timing Side Channel

- ssh: keystrokes are sent as separate packets.
- Packet timing ↔ keystroke timing ↔ typed letters
- Packet-sniffing eavesdropper can acquire information about typed characters (e.g. passwords).

[Song, Wagner, and Tian ’01]
Side Channels
Side Channels

- *Side channel*: a mechanism that conveys information inadvertently
Side Channels

- *Side channel*: a mechanism that conveys information inadvertently
- Examples:
Side Channels

- *Side channel*: a mechanism that conveys information inadvertently

- Examples:
  - Packet-timing based:
Side Channels

- **Side channel**: a mechanism that conveys information inadvertently
- Examples:
  - Packet-timing based:
    - typed characters (ssh) [Song et al. ’01]
Side Channels

- *Side channel*: a mechanism that conveys information inadvertently

- Examples:
  - Packet-timing based:
    - typed characters (ssh) [Song et al. ’01]
    - routing [Chaum ’81]
Side Channels

- **Side channel**: a mechanism that conveys information inadvertently

- Examples:
  - Packet-timing based:
    - typed characters (ssh) [Song et al. ’01]
    - routing [Chaum ’81]
    - spoken phrases (VoIP) [Wright et al. ’08]
Side Channels

- **Side channel**: a mechanism that conveys information inadvertently

- Examples:
  - Packet-timing based:
    - typed characters (ssh) [Song et al. ’01]
    - routing [Chaum ’81]
    - spoken phrases (VoIP) [Wright et al. ’08]
  - RSA decryption time [Kocher ’96]
Side Channels

- **Side channel**: a mechanism that conveys information inadvertently

- Examples:
  - Packet-timing based:
    - typed characters (ssh) [Song et al. ’01]
    - routing [Chaum ’81]
    - spoken phrases (VoIP) [Wright et al. ’08]
  - RSA decryption time [Kocher ’96]
  - Cache/memory contention [Ferraiuolo et al. ’16]
Side channels: a mechanism that conveys information inadvertently

Examples:

- Packet-timing based:
  - typed characters (ssh) [Song et al. ’01]
  - routing [Chaum ’81]
  - spoken phrases (VoIP) [Wright et al. ’08]
- RSA decryption time [Kocher ’96]
- Cache/memory contention [Ferraiuolo et al. ’16]
- CPU power consumption [Kocher et al. ’99]
Side Channels

- **Side channel**: a mechanism that conveys information inadvertently

- Examples:
  - Packet-timing based:
    - typed characters (ssh) [Song et al. ’01]
    - routing [Chaum ’81]
    - spoken phrases (VoIP) [Wright et al. ’08]
  - RSA decryption time [Kocher ’96]
  - Cache/memory contention [Ferraiuolo et al. ’16]
  - CPU power consumption [Kocher et al. ’99]
Side Channels

Intel sells off for a second day as massive security exploit shakes the stock

- Newly discovered vulnerabilities could theoretically allow a hacker to steal information stored in the memory of chips themselves.
- Although the exploits affected leading processors in many devices, Intel is bearing most of the fallout.
- Some on Wall Street think that Intel’s loss could mean gains for rivals.

Meltdown (Lipp et al., ’18)
Spectre (Kocher et al., ’18)
Side Channels

- **Side channel**: a mechanism that conveys information inadvertently

Examples:
  - Packet-timing based:
    - typed characters (ssh) [Song et al. ’01]
    - routing [Chaum ’81]
    - spoken phrases (VoIP) [Wright et al. ’08]
  - RSA decryption time [Kocher ’96]
  - Cache/memory contention [Ferraiuolo et al. ’16]
  - CPU power consumption [Kocher et al. ’99]
Side Channels

- **Side channel**: a mechanism that conveys information inadvertently

- **Examples**:
  - Packet-timing based:
    - typed characters (ssh) [Song et al. ’01]
    - routing [Chaum ’81]
    - spoken phrases (VoIP) [Wright et al. ’08]
  - RSA decryption time [Kocher ’96]
  - Cache/memory contention [Ferraiuolo et al. ’16]
  - CPU power consumption [Kocher et al. ’99]

How to measure leakage in this context?
Side Channels

- **Side channel**: a mechanism that conveys information inadvertently

Examples:

- Packet-timing based:
  - typed characters (ssh) [Song et al. ’01]
  - routing [Chaum ’81]
  - spoken phrases (VoIP) [Wright et al. ’08]
- RSA decryption time [Kocher ’96]
- Cache/memory contention [Ferraiuolo et al. ’16]
- CPU power consumption [Kocher et al. ’99]

Given RVs $X$ and $Y$, how much does $Y$ leak about $X$?
Existing Possibilities
Existing Possibilities

- Mutual information (or equivocation) between $X$ and $Y$
Existing Possibilities

- Mutual information (or equivocation) between X and Y
- Eavesdroppers expected distortion in reproducing X
Existing Possibilities

- Mutual information (or equivocation) between $X$ and $Y$
- Eavesdroppers expected distortion in reproducing $X$
- Probability of (approximately) guessing $X$
Existing Possibilities

- Mutual information (or equivocation) between \( X \) and \( Y \)
- Eavesdroppers expected distortion in reproducing \( X \)
- Probability of (approximately) guessing \( X \)
- Expected number of guesses to guess \( X \) correctly
Existing Possibilities

- Mutual information (or equivocation) between $X$ and $Y$
- Eavesdroppers expected distortion in reproducing $X$
- Probability of (approximately) guessing $X$
- Expected number of guesses to guess $X$ correctly
- Maximal correlation between $X$ and $Y$
Existing Possibilities

- Mutual information (or equivocation) between $X$ and $Y$
- Eavesdroppers expected distortion in reproducing $X$
- Probability of (approximately) guessing $X$
- Expected number of guesses to guess $X$ correctly
- Maximal correlation between $X$ and $Y$
- $k$-correlation between $X$ and $Y$
Existing Possibilities

- Mutual information (or equivocation) between $X$ and $Y$
- Eavesdroppers expected distortion in reproducing $X$
- Probability of (approximately) guessing $X$
- Expected number of guesses to guess $X$ correctly
- Maximal correlation between $X$ and $Y$
- $k$-correlation between $X$ and $Y$
- Cryptographic advantage
Existing Possibilities

- Mutual information (or equivocation) between $X$ and $Y$
- Eavesdroppers expected distortion in reproducing $X$
- Probability of (approximately) guessing $X$
- Expected number of guesses to guess $X$ correctly
- Maximal correlation between $X$ and $Y$
- $k$-correlation between $X$ and $Y$
- Cryptographic advantage
- Entropic security
Existing Possibilities

- Mutual information (or equivocation) between $X$ and $Y$
- Eavesdroppers expected distortion in reproducing $X$
- Probability of (approximately) guessing $X$
- Expected number of guesses to guess $X$ correctly
- Maximal correlation between $X$ and $Y$
- $k$-correlation between $X$ and $Y$
- Cryptographic advantage
- Entropic security
- (Local) differential privacy
Existing Possibilities

- Mutual information (or equivocation) between $X$ and $Y$
- Eavesdroppers expected distortion in reproducing $X$
- Probability of (approximately) guessing $X$
- Expected number of guesses to guess $X$ correctly
- Maximal correlation between $X$ and $Y$
- $k$-correlation between $X$ and $Y$
- Cryptographic advantage
- Entropic security
- (Local) differential privacy
- ...

...
Existing Possibilities

- Mutual information (or equivocation) between $X$ and $Y$
- Eavesdroppers expected distortion in reproducing $X$
- Probability of (approximately) guessing $X$
- Expected number of guesses to guess $X$ correctly
- Maximal correlation between $X$ and $Y$
- $k$-correlation between $X$ and $Y$
- Cryptographic advantage
- Entropic security
- (Local) differential privacy
- ...

Wagner and Eckhoff (’15):
Existing Possibilities

- Mutual information (or equivocation) between $X$ and $Y$
- Eavesdroppers expected distortion in reproducing $X$
- Probability of (approximately) guessing $X$
- Expected number of guesses to guess $X$ correctly
- Maximal correlation between $X$ and $Y$
- $k$-correlation between $X$ and $Y$
- Cryptographic advantage
- Entropic security
- (Local) differential privacy
- ...

Wagner and Eckhoff (’15): 81 metrics
The Threat Model
The Threat Model

1. The eavesdropper is interested in a possibly randomized function of $X$ called $U$. 
1. The eavesdropper is interested in a possibly randomized function of $X$ called $U$. 

nominal packet timings
The eavesdropper is interested in a possibly randomized function of $X$ called $U$. 

password
The Threat Model

1. The eavesdropper is interested in a possibly randomized function of $X$ called $U$.

2. The eavesdropper observes $Y$. 
The Threat Model

1. The eavesdropper is interested in a possibly randomized function of $X$ called $U$.

2. The eavesdropper observes $Y$. 

randomized/blurred version of $X$
The Threat Model

1. The eavesdropper is interested in a possibly randomized function of $X$ called $U$.

2. The eavesdropper observes $Y$.

3. The eavesdropper wants to guess, and we want to prevent the eavesdropper from guessing, $U$. 
1. The eavesdropper is interested in a possibly randomized function of $X$ called $U$.

2. The eavesdropper observes $Y$.

3. The eavesdropper wants to guess, and we want to prevent the eavesdropper from guessing, $U$.

brute-force attack
1. The eavesdropper is interested in a possibly randomized function of $X$ called $U$.

2. The eavesdropper observes $Y$.

3. The eavesdropper wants to guess, and we want to prevent the eavesdropper from guessing, $U$.

4. The distribution $P_{U|X}(u|x)$ is unknown to us (but known to the eavesdropper)
The Threat Model

1. The eavesdropper is interested in a possibly randomized function of $X$ called $U$.

2. The eavesdropper observes $Y$.

3. The eavesdropper wants to guess, and we want to prevent the eavesdropper from guessing.

4. The distribution $P_{U|X}(u|x)$ is unknown to us (but known to the eavesdropper).
## Maximal Leakage

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$U$</td>
<td>$X$</td>
<td>$Y$</td>
</tr>
<tr>
<td>[sensitive info]</td>
<td>[nominal process]</td>
<td>[revealed process]</td>
</tr>
</tbody>
</table>
Maximal Leakage

Markov chain:

\[ U \leftrightarrow X \leftrightarrow Y \]

[sensitive info] [nominal process] [revealed process]
Maximal Leakage

Def (Issa-Kamath-Wagner): Given $P_{XY}$, the maximal leakage from $X$ to $Y$ is

$$\sup_{\tilde{u}(\cdot)} \Pr(U = \tilde{u}(Y))$$
Maximal Leakage

**Markov chain:**

\[ U \leftrightarrow X \leftrightarrow Y \]

- [sensitive info]
- [nominal process]
- [revealed process]

**Def (Issa-Kamath-Wagner):** Given \( P_{XY} \), the maximal leakage from \( X \) to \( Y \) is

\[
\sup_{\tilde{u}(\cdot)} \Pr(U = \tilde{u}(Y)) \quad \frac{\sup_{\tilde{u}} \Pr(U = \tilde{u})}{\sup_{\tilde{u}} \Pr(U = \tilde{u})}
\]
Markov chain:  
$\begin{align*}
\text{[sensitive info]} & \leftrightarrow U \leftrightarrow X \leftrightarrow [\text{nominal process}] \\
& \leftrightarrow [\text{revealed process}] \\
& \leftrightarrow Y
\end{align*}$

**Def** (Issa-Kamath-Wagner): Given $P_{XY}$, the **maximal leakage** from $X$ to $Y$ is

$$
\log \frac{\sup \tilde{u}(.) \Pr(U = \tilde{u}(Y))}{\sup \tilde{u} \Pr(U = \tilde{u})}
$$
Maximal Leakage

Markov chain:

\[ U \leftrightarrow X \leftrightarrow Y \]

[sensitive info] [nominal process] [revealed process]

**Def (Issa-Kamath-Wagner):** Given \( P_{XY} \), the maximal leakage from \( X \) to \( Y \) is

\[
\sup_{\mathcal{U}:U \leftrightarrow X \leftrightarrow Y} \log \left( \frac{\sup_{\tilde{\mathcal{U}}(\cdot)} \Pr(U = \tilde{\mathcal{U}}(Y))}{\sup_{\hat{\mathcal{U}}} \Pr(U = \hat{\mathcal{U}})} \right)
\]
Maximal Leakage

Def (Issa-Kamath-Wagner): Given $P_{XY}$, the maximal leakage from $X$ to $Y$ is

$$\mathcal{L}(X \rightarrow Y) = \sup_{U:U \leftrightarrow X \leftrightarrow Y} \log \frac{\sup \tilde{u}(.) \Pr(U = \tilde{u}(Y))}{\sup \tilde{u} \Pr(U = \tilde{u})}$$
Maximal Leakage

Markov chain:

\[ U \leftrightarrow X \leftrightarrow Y \]

[sensitive info] [nominal process] [revealed process]

**Def (Issa-Kamath-Wagner):** Given \( P_{XY} \), the maximal leakage from \( X \) to \( Y \) is

\[
\mathcal{L}(X \rightarrow Y) = \sup_{U:U \leftrightarrow X \leftrightarrow Y} \log \frac{\sup_{\tilde{u}(\cdot)} \Pr(U = \tilde{u}(Y))}{\sup_{\tilde{u}} \Pr(U = \tilde{u})}
\]

[operationally interpretable]
Maximal Leakage

Markov chain:

\[ \begin{align*}
    U & \leftrightarrow X \leftrightarrow Y \\
    \text{[sensitive info]} & \text{[nominal process]} & \text{[revealed process]} 
\end{align*} \]

**Def (Issa-Kamath-Wagner):** Given \( P_{XY} \), the maximal leakage from \( X \) to \( Y \) is

\[
\mathcal{L}(X \rightarrow Y) = \sup_{U:U \leftrightarrow X \leftrightarrow Y} \log \frac{\sup_{\tilde{u}(\cdot)} \Pr(U = \tilde{u}(Y))}{\sup_{\tilde{u}} \Pr(U = \tilde{u})}
\]

[not evidently computable; Carathéodory?]
Maximal Leakage

**Theorem** (Issa-Kamath-Wagner): For any joint distribution $P_{XY}$ on finite alphabets
**Theorem** (Issa-Kamath-Wagner): For any joint distribution $P_{XY}$ on finite alphabets

$$\mathcal{L}(X \rightarrow Y) = \log \sum_{y \in \mathcal{Y}} \max_{x \in \mathcal{X}} P_{Y|X}(y|x) \quad \text{subject to} \quad P_X(x) > 0$$
**Theorem** (Issa-Kamath-Wagner): For any joint distribution $P_{XY}$ on finite alphabets

$$
\mathcal{L}(X \rightarrow Y) = \log \sum_{y \in \mathcal{Y}} \max_{x \in \mathcal{X} : P_X(x) > 0} P_{Y|X}(y|x)
$$

$$
= I_\infty(X; Y) \quad [\text{Sibson MI of order } \infty]
$$
Theorem (Issa-Kamath-Wagner): For any joint distribution $P_{XY}$ on finite alphabets

$$\mathcal{L}(X \to Y) = \log \sum_{y \in \mathcal{Y}} \max_{x \in \mathcal{X}} P_{Y|X}(y|x)$$

$$= I_{\infty}(X; Y) \quad \text{[Sibson MI of order } \infty\text{]}$$

[depends on $P_X$ only through its support]
The Worst-Case $U$

$$\mathcal{L}(X \to Y) = \sup_{U \leftrightarrow X \leftrightarrow Y} \log \frac{\sup_{\tilde{u}(\cdot)} \Pr(U = \tilde{u}(Y))}{\sup_{\tilde{u}} \Pr(U = \tilde{u})}$$
The Worst-Case $U$

\[
\mathcal{L}(X \to Y) = \sup_{U \leftrightarrow X \leftrightarrow Y} \log \frac{\sup_{\tilde{u}(\cdot)} \Pr(U = \tilde{u}(Y))}{\sup_{\tilde{u}} \Pr(U = \tilde{u})}
\]

\[
P_X
\]
The Worst-Case $U$

$$\mathcal{L}(X \to Y) = \sup_{U \leftrightarrow X \leftrightarrow Y} \log \frac{\sup_{\tilde{u}(.)} \Pr(U = \tilde{u}(Y))}{\sup_{\tilde{u}} \Pr(U = \tilde{u})}$$

$P_X$
The Worst-Case $U$

$$\mathcal{L}(X \to Y) = \sup_{U \leftarrow X \leftarrow Y} \log \frac{\sup \tilde{u}(\cdot) \Pr(U = \tilde{u}(Y))}{\sup \tilde{u} \Pr(U = \tilde{u})}$$
The Worst-Case $U$

$$\mathcal{L}(X \rightarrow Y) = \sup_{U \leftrightarrow X \leftrightarrow Y} \log \frac{\sup \tilde{u}(.) \Pr(U = \tilde{u}(Y))}{\sup \tilde{u} \Pr(U = \tilde{u})}$$

$P_X$ $\rightarrow$ $P_{XU}$
The Worst-Case $U$

$$
\mathcal{L}(X \rightarrow Y) = \sup_{U \leftrightarrow X \leftrightarrow Y} \log \frac{\sup \tilde{u}(\cdot) \Pr(U = \tilde{u}(Y))}{\sup \tilde{u} \Pr(U = \tilde{u})}
$$

$P_X$  $\rightarrow$  $P_{XU}$
The Worst-Case $U$

$$\mathcal{L}(X \rightarrow Y) = \sup_{U \leftrightarrow X \leftrightarrow Y} \log \frac{\sup \tilde{u}(\cdot) \Pr(U = \tilde{u}(Y))}{\sup \tilde{u} \Pr(U = \tilde{u})}$$

$$P_X \xrightarrow{\text{}} \begin{bmatrix} \vdots \end{bmatrix} \rightarrow \begin{bmatrix} \vdots \end{bmatrix} = P_{XU}$$
The Worst-Case $U$

$$\mathcal{L}(X \rightarrow Y) = \sup_{U \leftrightarrow X \leftrightarrow Y} \log \frac{\sup \tilde{u}(\cdot) \Pr(U = \tilde{u}(Y))}{\sup \tilde{u} \Pr(U = \tilde{u})}$$
The Worst-Case $U$

$$\mathcal{L}(X \to Y) = \sup_{U \leftrightarrow X \leftrightarrow Y} \log \frac{\sup \tilde{u}(\cdot) \Pr(U = \tilde{u}(Y))}{\sup \tilde{u} \Pr(U = \tilde{u})}$$
The Worst-Case $U$

$$\mathcal{L}(X \rightarrow Y) = \sup_{U \leftarrow X \leftrightarrow Y} \log \frac{\sup \tilde{u}(\cdot) \Pr(U = \tilde{u}(Y))}{\sup \tilde{u} \Pr(U = \tilde{u})}$$

$P_X$  \hspace{2cm} "shattering" \hspace{2cm} $P_{XU}$
The Worst-Case $U$

$$\mathcal{L}(X \rightarrow Y) = \sup_{U \leftrightarrow X \leftrightarrow Y} \log \frac{\sup_{\tilde{u}(\cdot)} \Pr(U = \tilde{u}(Y))}{\sup_{\tilde{u}} \Pr(U = \tilde{u})}$$

$[U$ is uniform and s.t. $X$ is a deterministic function of $U]$
Upper Bound

$$\sum_{y \in \mathcal{Y}} P_{\mathcal{Y}}(y) \max_{u \in \mathcal{U}} P_{\mathcal{U}|\mathcal{Y}}(u|y)$$

$$= \sum_{y \in \mathcal{Y}} \max_{u \in \mathcal{U}} P_{\mathcal{U}\mathcal{Y}}(u, y)$$

$$= \sum_{y \in \mathcal{Y}} \max_{u \in \mathcal{U}} \sum_{x \in \mathcal{X}} P_{X}(x) P_{\mathcal{U}|X}(u|x) P_{\mathcal{Y}|X}(y|x)$$

$$\leq \sum_{y \in \mathcal{Y}} \max_{u \in \mathcal{U}} \sum_{x \in \mathcal{X}} P_{X}(x) P_{\mathcal{U}|X}(u|x) \max_{x' \in \mathcal{X}} P_{\mathcal{Y}|X}(y|x')$$

$$= \sum_{y \in \mathcal{Y}} \left( \max_{x' \in \mathcal{X}} P_{\mathcal{Y}|X}(y|x') \right) \max_{u \in \mathcal{U}} \sum_{x \in \mathcal{X}} P_{X}(x) P_{\mathcal{U}|X}(u|x)$$

$$= \sum_{y \in \mathcal{Y}} \max_{x \in \mathcal{X}} P_{\mathcal{Y}|X}(y|x) \max_{u \in \mathcal{U}} P_{\mathcal{U}}(u).$$
Maximal Leakage

**Theorem** (Issa-Kamath-Wagner): For any joint distribution $P_{XY}$ on finite alphabets

$$\mathcal{L}(X \rightarrow Y) = \log \sum_{y \in \mathcal{Y}} \max_{x \in \mathcal{X}} P_{Y|X}(y|x) \quad \text{for } P_{X}(x) > 0$$

$$= I_{\infty}(X; Y) \quad [\text{Sibson MI of order } \infty]$$
Properties of Max. Leakage
Corollary: For any joint distribution $P_{XY}$ on finite alphabets
**Corollary**: For any joint distribution $P_{XY}$ on finite alphabets

- Data processing inequality: If $X \leftrightarrow Y \leftrightarrow Z$ then
  $$\mathcal{L}(X \rightarrow Z) \leq \min \{ \mathcal{L}(X \rightarrow Y), \mathcal{L}(Y \rightarrow Z) \}$$
Corollary: For any joint distribution $P_{XY}$ on finite alphabets

- Data processing inequality: If $X \leftrightarrow Y \leftrightarrow Z$ then
  \[ \mathcal{L}(X \rightarrow Z) \leq \min \{ \mathcal{L}(X \rightarrow Y), \mathcal{L}(Y \rightarrow Z) \} \]

- Self-leakage
  \[ \mathcal{L}(X \rightarrow X) = \log |\{x : P_X(x) > 0\}| \]
**Corollary**: For any joint distribution $P_{XY}$ on finite alphabets

- **Data processing inequality**: If $X \leftrightarrow Y \leftrightarrow Z$ then
  \[
  \mathcal{L}(X \rightarrow Z) \leq \min \{ \mathcal{L}(X \rightarrow Y), \mathcal{L}(Y \rightarrow Z) \}
  \]

- **Self-leakage**
  \[
  \mathcal{L}(X \rightarrow X) = \log |\{x : P_X(x) > 0\}|
  \]

- **Cardinality bound**
  \[
  \mathcal{L}(X \rightarrow Y) \leq \min \{ \log |\mathcal{X}|, \log |\mathcal{Y}| \}
  \]
Properties of Max. Leakage
Properties of Max. Leakage

- Independence: $\mathcal{L}(X \rightarrow Y) = 0$ iff $X$ and $Y$ are indep.
Properties of Max. Leakage

- Independence: $\mathcal{L}(X \rightarrow Y) = 0$ iff $X$ and $Y$ are indep.
- Asymmetry: $\mathcal{L}(X \rightarrow Y) \neq \mathcal{L}(Y \rightarrow X)$ in general.
Properties of Max. Leakage

- Independence: \( \mathcal{L}(X \rightarrow Y) = 0 \) iff \( X \) and \( Y \) are independent.
- Asymmetry: \( \mathcal{L}(X \rightarrow Y) \neq \mathcal{L}(Y \rightarrow X) \) in general.
- Additivity: if \( (X_i, Y_i)_{i=1}^n \) are independent over \( i \)

\[
\mathcal{L}(X^n \rightarrow Y^n) = \sum_{i=1}^{n} \mathcal{L}(X_i \rightarrow Y_i)
\]
- Independence: $\mathcal{L}(X \rightarrow Y) = 0$ iff $X$ and $Y$ are independent.
- Asymmetry: $\mathcal{L}(X \rightarrow Y) \neq \mathcal{L}(Y \rightarrow X)$ in general.
- Additivity: if $(X_i, Y_i)_{i=1}^n$ are independent over $i$

$$
\mathcal{L}(X^n \rightarrow Y^n) = \sum_{i=1}^{n} \mathcal{L}(X_i \rightarrow Y_i)
$$

- Convexity: $\exp(\mathcal{L}(X \rightarrow Y))$ is convex in $P_{Y|X}$
Properties of Max. Leakage

- Independence: $\mathcal{L}(X\rightarrow Y) = 0$ iff $X$ and $Y$ are independent.
- Asymmetry: $\mathcal{L}(X\rightarrow Y) \neq \mathcal{L}(Y\rightarrow X)$ in general.
- Additivity: if $(X_i, Y_i)_{i=1}^n$ are independent over $i$
  \[ \mathcal{L}(X^n\rightarrow Y^n) = \sum_{i=1}^n \mathcal{L}(X_i\rightarrow Y_i) \]
- Convexity: $\exp(\mathcal{L}(X\rightarrow Y))$ is convex in $P_{Y|X}$
- Maximal leakage upper bounds mutual info.
  \[ \mathcal{L}(X \rightarrow Y) \geq I(X; Y) = \sum_{x,y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \]
Variations and Extensions

- Multiple guesses
- Approximate guesses
- General gains
- Opportunistic choice of $U$
- Conditional version
- Formula for general measure spaces
- Guessing $X$ itself
**Def (Issa-Kamath-Wagner):** For any positive integer $k$, 

$$L_k(X \rightarrow Y) = \sup_{U \leftrightarrow X \leftrightarrow Y} \log \frac{\sup \tilde{u}_1(\cdot), \ldots, \tilde{u}_k(\cdot) \ P(\cup_i \{U = \tilde{u}_i(Y)\})}{\sup \tilde{u}_1, \ldots, \tilde{u}_k \ P(\cup_i \{U = \tilde{u}_i\})}$$
Extension: Multiple Guesses

**Def** (Issa-Kamath-Wagner): For any positive integer $k$,

$$
\mathcal{L}_k(X \to Y) = \sup_{U \leftrightarrow X \leftrightarrow Y} \log \frac{\sup_{\tilde{u}_1(\cdot), \ldots, \tilde{u}_k(\cdot)} P(U_i \{ U = \tilde{u}_i(Y) \})}{\sup_{\tilde{u}_1, \ldots, \tilde{u}_k} P(U_i \{ U = \tilde{u}_i \})}
$$

**Theorem** (Issa-Kamath-Wagner): If $X$ and $Y$ are discrete then for any positive integer $k$,

$$
\mathcal{L}_k(X \to Y) = \mathcal{L}_1(X \to Y) = \mathcal{L}(X \to Y).
$$
**Definition:** The conditional maximal leakage from $X$ to $Y$ given $Z$ is

$$
\mathcal{L}(X \rightarrow Y|Z) = \sup_{U \leftrightarrow X \leftrightarrow Y|Z} \log \frac{\sup \tilde{u}(.,.) \Pr(U = \tilde{u}(Y, Z))}{\sup \tilde{u}(.) \Pr(U = \tilde{u}(Z))}
$$
**Definition:** The conditional maximal leakage from $X$ to $Y$ given $Z$ is

\[
\mathcal{L}(X \rightarrow Y|Z) = \sup_{U \leftrightarrow X \leftrightarrow Y|Z} \log \frac{\sup_{\tilde{u}(.,.)} \Pr(U = \tilde{u}(Y, Z))}{\sup_{\tilde{u}(\cdot)} \Pr(U = \tilde{u}(Z))}
\]

vs. $U \leftrightarrow X \leftrightarrow (Y, Z)$
**Definition:** The conditional maximal leakage from $X$ to $Y$ given $Z$ is

$$\mathcal{L}(X \rightarrow Y|Z) = \sup_{U \leftrightarrow X \leftrightarrow Y|Z} \log \frac{\sup_{\tilde{u}(\cdot, \cdot)} \Pr(U = \tilde{u}(Y, Z))}{\sup_{\tilde{u}(\cdot)} \Pr(U = \tilde{u}(Z))}$$
**Definition:** The conditional maximal leakage from $X$ to $Y$ given $Z$ is

$$\mathcal{L}(X \rightarrow Y|Z) = \sup_{U \leftrightarrow X \leftrightarrow Y|Z} \log \frac{\sup \tilde{u}(.,.) \Pr(U = \tilde{u}(Y, Z))}{\sup \tilde{u}(.) \Pr(U = \tilde{u}(Z))}$$

**Theorem** (Issa-Wagner):

$$\mathcal{L}(X \rightarrow Y|Z) = \max_z \mathcal{L}(X \rightarrow Y|Z = z)$$
**Corollary**: For any joint distribution $P_{XYZ}$ on finite alphabets

- Data processing inequality: If $X \leftrightarrow Y \leftrightarrow V|Z$ then
  \[ \mathcal{L}(X \rightarrow V|Z) \leq \min\{\mathcal{L}(X \rightarrow Y|Z), \mathcal{L}(Y \rightarrow V|Z)\} \]

- Cond. independence: $\mathcal{L}(X \rightarrow Y|Z) = 0$ iff
  \[ X \leftrightarrow Z \leftrightarrow Y \]

- Mutual information:
  \[ \mathcal{L}(X \rightarrow Y|Z) \geq I(X; Y|Z) \]
Properties of Cond. Max. Leakage
Conditioning reduces max. leakage: if $Z \leftrightarrow X \leftrightarrow Y$ then

$$\mathcal{L}(X \rightarrow Y | Z) \leq \mathcal{L}(X \rightarrow Y)$$
Properties of Cond. Max. Leakage

- Conditioning reduces max. leakage: if $Z \leftrightarrow X \leftrightarrow Y$ then
  \[
  \mathcal{L}(X \rightarrow Y | Z) \leq \mathcal{L}(X \rightarrow Y)
  \]

- Chain rule:
  \[
  \mathcal{L}(X \rightarrow (Y, Z)) \leq \mathcal{L}(X \rightarrow Z) + \mathcal{L}(X \rightarrow Y | Z)
  \]
Properties of Cond. Max. Leakage

- Conditioning reduces max. leakage: if $Z \leftrightarrow X \leftrightarrow Y$ then
  \[ \mathcal{L}(X \rightarrow Y | Z) \leq \mathcal{L}(X \rightarrow Y) \]

- Chain rule:
  \[ \mathcal{L}(X \rightarrow (Y, Z)) \leq \mathcal{L}(X \rightarrow Z) + \mathcal{L}(X \rightarrow Y | Z) \]

- Composition theorem: if $Z \leftrightarrow X \leftrightarrow Y$ then
  \[ \mathcal{L}(X \rightarrow (Y, Z)) \leq \mathcal{L}(X \rightarrow Z) + \mathcal{L}(X \rightarrow Y) \]
Def:

\[ L_I(X \rightarrow Y) = \sup_{P_X} \log \frac{\max_{\hat{X}(\cdot)} P(X = \hat{X}(Y))}{\max_{\hat{X}} P(X = \hat{X})} \]
Guessing $X$

**Def:**

\[
\mathcal{L}_I(X \to Y) = \sup_{P_X} \log \frac{\max_{\hat{x}(\cdot)} P(X = \hat{x}(Y))}{\max_{\hat{x}} P(X = \hat{x})}
\]

**Theorem:**

\[
\mathcal{L}_I(X \to Y) = I_\infty \left[ \equiv \mathcal{L}(X \to Y) \right]
\]
Guessing $X$

**Def:** [Braun et al. '09; Kopf and Smith '10]:

$$\mathcal{L}_I(X \rightarrow Y) = \sup_{P_X} \log \frac{\max_{\hat{x}(\cdot)} P(X = \hat{x}(Y))}{\max_{\hat{x}} P(X = \hat{x})}$$

**Theorem:** [Braun et al. '09; Kopf and Smith '10]:

$$\mathcal{L}_I(X \rightarrow Y) = I_\infty[= \mathcal{L}(X \rightarrow Y)]$$
Guessing $X$

**Def:** [Braun et al. ’09; Kopf and Smith ’10]:

$$\mathcal{L}_I(X \rightarrow Y) = \sup_{P_X} \log \frac{\max_{\hat{x}(\cdot)} P(X = \hat{x}(Y))}{\max_{\hat{x}} P(X = \hat{x})}$$

**Theorem:** [Braun et al. ’09; Kopf and Smith ’10]:

$$\mathcal{L}_I(X \rightarrow Y) = I_\infty [\triangleq \mathcal{L}(X \rightarrow Y)]$$

[**maximal leakage:** not in Wagner and Eckhoff (’15)]
Discrete Examples: BSC

\[ \mathcal{L}(X \rightarrow Y) = \log(2(1 - q)) \]
Discrete Examples: BSC

\[ \mathcal{L}(X \rightarrow Y) = \log(2(1 - q)) \]

\[ [p = \frac{1}{3}] \]
Theorem (Issa-Kamath-Wagner): If $f_X(x)$ and $f_{Y|X}(y|x)$ are continuous then:

$$\mathcal{L}(X \to Y) = \log \int \sup_{x: f_X(x) > 0} f_{Y|X}(y|x) \, dy$$
**Theorem** (Issa-Kamath-Wagner): If \( f_X(x) \) and \( f_{Y|X}(y|x) \) are continuous then:

\[
\mathcal{L}(X \rightarrow Y) = \log \int \sup_{x: f_X(x) > 0} f_{Y|X}(y|x) \, dy
\]

If \( X \) and \( Y \) are jointly Gaussian then

\[
\mathcal{L}(X \rightarrow Y) = \begin{cases} 
0 & \text{if } X, Y \text{ indep.} \\
\infty & \text{otherwise}
\end{cases}
\]
Theorem (Issa-Kamath-Wagner): If $f_X(x)$ and $f_{Y|X}(y|x)$ are continuous then:

$$\mathcal{L}(X \rightarrow Y) = \log \int \sup_{x : f_X(x) > 0} f_{Y|X}(y|x) \, dy$$

If $X$ and $Y$ are jointly Gaussian then

$$\mathcal{L}(X \rightarrow Y) = \begin{cases} 0 & \text{if } X, Y \text{ indep.} \\ \infty & \text{otherwise} \end{cases}$$

[“adding noise” (as opposed to quantizing) leaks]
Theorem (Issa-Kamath-Wagner): If $f_X(x)$ and $f_{Y|X}(y|x)$ are continuous then:

$$\mathcal{L}(X \rightarrow Y) = \log \int_{x : f_X(x) > 0} \sup f_{Y|X}(y|x) \, dy$$

If $X$ and $Y$ are jointly Gaussian then

$$\mathcal{L}(X \rightarrow Y) = \begin{cases} 0 & \text{if } X, Y \text{ indep.} \\ \infty & \text{otherwise} \end{cases}$$
Other Metrics

- Mutual information (or equivocation)
- Expected distortion at eavesdropper
- Probability of (approximately) guessing $X$
- Expected number of guesses to guess $X$ correctly
- Maximal correlation
- $k$-correlation
- Cryptographic advantage
- Entropic security
- (Local) differential privacy
- ...

...
Other Metrics

- Mutual information (or equivocation)
- Expected distortion at eavesdropper
- Probability of (approximately) guessing $X$
- Expected number of guesses to guess $X$ correctly
- Maximal correlation
- $k$-correlation
- Cryptographic advantage
- Entropic security
- (Local) differential privacy
- ...

...
\[ I(X; Y) = \sum_{x, y} P_{XY}(x, y) \log \frac{P_{XY}(x, y)}{P_X(x) \cdot P_Y(y)} \]
Mutual Information

\[ I(X; Y) = \sum_{x,y} P_{XY}(x, y) \log \frac{P_{XY}(x, y)}{P_X(x) \cdot P_Y(y)} \]

\[ \iff H(X|Y), \text{ first used by Shannon ('49)} \]
Mutual Information

\[ I(X; Y) = \sum_{x,y} P_{XY}(x, y) \log \frac{P_{XY}(x, y)}{P_X(x) \cdot P_Y(y)} \]

\( \iff H(X|Y) \), first used by Shannon (’49)

solution concept vs. problem formulation
Mutual Information

\[ I(X; Y) = \sum_{x,y} P_{XY}(x, y) \log \frac{P_{XY}(x, y)}{P_X(x) \cdot P_Y(y)} \]

Shannon (’49):

From the point of view of the cryptanalyst, a secrecy system is almost identical with a noisy communication system. The message (transmitted signal) is operated on by a statistical element, the enciphering system, with its statistically chosen key. The result of this operation is the cryptogram (analogous to the perturbed signal) which is available for analysis. The chief differences in the two cases are: first, that the operation of the enciphering transformation is generally of a more complex nature than the perturbing noise in a channel; and, second, the key for a secrecy system is usually chosen from a finite set of possibilities while the noise in a channel is more often continually introduced, in effect chosen from an infinite set.

With these considerations in mind it is natural to use the equivocation as a theoretical secrecy index. It may be noted that there are two significant equivocations, that of the key and that of the message. These will be
Shannon (’49)

Secrecy System

Message $\rightarrow$ Enciphering System $\rightarrow$ Signal available to eavesdropper

key
Shannon (’49)

**Secrecy System**
- Message
- Enciphering System
- Signal available to eavesdropper
- key

**Noisy Communication System**
- Transmitted Signal
- Noisy Channel
- Signal available to decoder
- phy. layer noise
Shannon (’49)

**Secrecy System**
- **Message** → **Enciphering System** → **Signal available to eavesdropper**
  - **key**

**Noisy Communication System**
- **Transmitted Signal** → **Noisy Channel** → **Signal available to decoder**
  - **phy. layer noise**
Shannon (’49)

- “Chief” differences: in secrecy system:
“Chief” differences: in secrecy system:
- Injected randomness is of “more complex nature”
Shannon (’49)

“Chief” differences: in secrecy system:
- Injected randomness is of “more complex nature”
- Injected randomness is discrete
Shannon (’49)

- Other differences: in conventional comm.,
Shannon (’49)

- Other differences: in conventional comm.,
  - Encoder is a willing participant (coding)
Other differences: in conventional comm.,
- Encoder is a willing participant (coding)
- Communication must be reliable
Shannon (’49)

- Other differences: in conventional comm.,
  - Encoder is a willing participant (coding)
  - Communication must be reliable
- Unclear motivation for using MI in secrecy applications
Shannon (’49)

- Other differences: in conventional comm.,
  - Encoder is a willing participant (coding)
  - Communication must be reliable
- Unclear motivation for using MI in secrecy applications
- But isn’t capacity an upper bound?
Folk Theorem: Any reasonable measure of “leakage” from $X$ to $Y$ should be upper bounded by the Shannon capacity of the channel $P_{Y|X}$:

$$\mathcal{L}(X \rightarrow Y) \leq C = \max_{p(x)} I(X; Y).$$
Folk Theorem: Any reasonable measure of “leakage” from $X$ to $Y$ should be upper bounded by the Shannon capacity of the channel $P_{Y|X}$:

$$
\mathcal{L}(X \rightarrow Y) \leq C = \max_{p(x)} I(X;Y).
$$

“Proof:”
Folk Theorem: Any reasonable measure of “leakage” from $X$ to $Y$ should be upper bounded by the Shannon capacity of the channel $P_{Y|X}$:

\[
\mathcal{L}(X \to Y) \leq C = \max_{p(x)} I(X; Y).
\]

“Proof:”

\[
\mathcal{L}(X \to Y) \leq \max_{P_X} \mathcal{L}(X \to Y).
\]
But is Capacity an Upper Bound?

**Folk Theorem:** Any reasonable measure of “leakage” from $X$ to $Y$ should be upper bounded by the Shannon capacity of the channel $P_{Y|X}$:

$$\mathcal{L}(X \to Y) \leq C = \max_{p(x)} I(X; Y).$$

**Proof:**

$$\mathcal{L}(X \to Y) \leq \max_{P_X} \mathcal{L}(X \to Y)$$

$$\leq \lim_{n \to \infty} \max_{P_{X^n}} \frac{1}{n} \mathcal{L}(X^n \to Y^n)$$
But is Capacity an Upper Bound?

**Folk Theorem:** Any reasonable measure of “leakage” from $X$ to $Y$ should be upper bounded by the Shannon capacity of the channel $P_{Y|X}$:

$$\mathcal{L}(X \rightarrow Y) \leq C = \max_{p(x)} I(X; Y).$$

**Proof:**

$$\mathcal{L}(X \rightarrow Y) \leq \max_{p_X} \mathcal{L}(X \rightarrow Y)$$

$$\leq \lim_{n \rightarrow \infty} \max_{P_{X^n}} \frac{1}{n} \mathcal{L}(X^n \rightarrow Y^n)$$

$$\leq C = \max_{p(x)} I(X; Y).$$
But is Capacity an Upper Bound?

**Folk Theorem:** Any reasonable measure of “leakage” from $X$ to $Y$ should be upper bounded by the Shannon capacity of the channel $P_{Y|X}$:

$$\mathcal{L}(X \rightarrow Y) \leq C = \max_{p(x)} I(X; Y).$$

**Proof:**

$$\mathcal{L}(X \rightarrow Y) \leq \max_{P_X} \mathcal{L}(X \rightarrow Y)$$

$$\leq \lim_{n \to \infty} \max_{P_{X^n}} \frac{1}{n} \mathcal{L}(X^n \rightarrow Y^n)$$

$$\leq C = \max_{p(x)} I(X; Y).$$

$C$ is the maximum amortized rate of information transfer over a channel.
But is Capacity an Upper Bound?

**Folk Theorem**: Any reasonable measure of “leakage” from $X$ to $Y$ should be upper bounded by the Shannon capacity of the channel $P_{Y|X}$:

$$L(X \rightarrow Y) \leq C = \max_{p(x)} I(X; Y).$$

“Proof:”

$$L(X \rightarrow Y) \leq \max_P L(X \rightarrow Y)$$

$$\leq \lim_{n \to \infty} \max_{P_{X^n}} \frac{1}{n} L(X^n \rightarrow Y^n)$$

$$\leq C = \max_{p(x)} I(X; Y).$$

[Yet $L(X \rightarrow Y) > C$]

$C$ is the maximum amortized rate of information transfer over a channel.
But is Capacity an Upper Bound?

**Folk Theorem:** Any reasonable measure of “leakage” from $X$ to $Y$ should be upper bounded by the Shannon capacity of the channel $P_{Y|X}$:

$$\mathcal{L}(X \rightarrow Y) \leq C = \max_{p(x)} I(X; Y).$$

**Proof:**

$$\mathcal{L}(X \rightarrow Y) \leq \max_{P_X} \mathcal{L}(X \rightarrow Y) \leq \lim_{n \to \infty} \max_{P_{X^n}} \frac{1}{n} \mathcal{L}(X^n \rightarrow Y^n) \leq C = \max_{p(x)} I(X; Y).$$

$C$ is the maximum amortized rate of **reliable** information transfer over a channel.

[Yet $\mathcal{L}(X \rightarrow Y) > C$]
If $X$ has full support:

$$\mathcal{L}(X \rightarrow Y) = \sup_{U \leftrightarrow X \leftrightarrow Y} \log \frac{\sup_{\tilde{u}(.)} \Pr(U = \tilde{u}(Y))}{\sup_{\tilde{u}} \Pr(U = \tilde{u})}$$
Leakage vs. Capacity

If $X$ has full support:

$$
\mathcal{L}(X \rightarrow Y) = \sup_{U \leftrightarrow X \leftrightarrow Y} \log \frac{\sup_{\tilde{u}(\cdot)} \Pr(U = \tilde{u}(Y))}{\sup_{\tilde{u}} \Pr(U = \tilde{u})}
= \sup_{U \leftrightarrow X \leftrightarrow Y \leftrightarrow \tilde{U}} \log \frac{\Pr(U = \tilde{U})}{\sup_{\tilde{u}} \Pr(U = \tilde{u})}
$$
If $X$ has full support:

$$
\mathcal{L}(X \rightarrow Y) = \sup_{U \leftrightarrow X \leftrightarrow Y} \log \frac{\sup_{\tilde{u}(\cdot)} \Pr(U = \tilde{u}(Y))}{\sup_{\tilde{u}} \Pr(U = \tilde{u})}
$$

$$
= \sup_{U \leftrightarrow X \leftrightarrow Y \leftrightarrow \tilde{U}} \log \frac{\Pr(U = \tilde{U})}{\sup_{\tilde{u}} \Pr(U = \tilde{u})}
$$

$$
= \lim_{n \rightarrow \infty} \sup_{U \leftrightarrow X^n \leftrightarrow Y^n \leftrightarrow \tilde{U}^n} \frac{1}{n} \log \frac{\Pr(U = \tilde{U}^n)}{\sup_{\tilde{u}} \Pr(U = \tilde{u})}
$$
Leakage vs. Capacity

If $X$ has full support:

$$
\mathcal{L}(X \rightarrow Y) = \sup_{U \leftrightarrow X \leftrightarrow Y} \log \frac{\sup_{\tilde{u}(\cdot)} \Pr(U = \tilde{u}(Y))}{\sup_{\tilde{u}} \Pr(U = \tilde{u})}
$$

$$
= \sup_{U \leftrightarrow X \leftrightarrow Y \leftrightarrow \tilde{U}} \log \frac{\Pr(U = \tilde{U})}{\sup_{\tilde{u}} \Pr(U = \tilde{u})}
$$

$$
= \lim_{n \to \infty} \sup_{U \leftrightarrow X^n \leftrightarrow Y^n \leftrightarrow \tilde{U}} \frac{1}{n} \log \frac{\Pr(U = \tilde{U})}{\sup_{\tilde{u}} \Pr(U = \tilde{u})}
$$

$$
= \lim_{n \to \infty} \sup_{P_{X^n}} \sup_{U \leftrightarrow X^n \leftrightarrow Y^n \leftrightarrow \tilde{U}} \frac{1}{n} \log \frac{\Pr(U = \tilde{U})}{\sup_{\tilde{u}} \Pr(U = \tilde{u})}
$$
If $X$ has full support:

$$\mathcal{L}(X \rightarrow Y) = \lim_{n \to \infty} \sup_{p_{X^n}} \sup_{U \leftrightarrow X^n \leftrightarrow Y^n \leftrightarrow \tilde{U}} \frac{1}{n} \log \frac{\Pr(U = \tilde{U})}{\sup_{\tilde{U}} \Pr(U = \tilde{U})}$$
Leakage vs. Capacity

If $X$ has full support:

$$
\mathcal{L}(X \rightarrow Y) = \lim_{n \to \infty} \sup_{p_{X^n}} \sup_{U \leftrightarrow X^n \leftrightarrow Y^n \leftrightarrow \tilde{U}} \frac{1}{n} \log \frac{\Pr(U = \tilde{U})}{\sup_{\tilde{u}} \Pr(U = \tilde{u})}
$$

**Theorem** (Issa-Wagner):

$$
C = \lim_{\epsilon \to 0} \lim_{n \to \infty} \sup_{p_{X^n}} \sup_{U \leftrightarrow X^n \leftrightarrow Y^n \leftrightarrow \tilde{U}} \frac{1}{n} \log \frac{\Pr(U = \tilde{U})}{\sup_{\tilde{u}} \Pr(U = \tilde{u})} \quad \text{subject to} \quad P(U=\tilde{U}) \geq 1-\epsilon
$$
(Local) Differential Privacy

\[ LDP(X \rightarrow Y) := \sup_{x,x',y} \log \frac{P_{Y|x}(y|x)}{P_{Y|x}(y|x')} \]  

[Warner ’65; Evfimievski et al. ’03]
(Local) Differential Privacy

\[
LDP(X \rightarrow Y) := \sup_{x,x',y} \log \frac{P_{Y|X}(y|x)}{P_{Y|X}(y|x')}\]  

[Warner ’65; Evfimievski et al. ’03]

Operational interpretation?
(Local) Differential Privacy

\[ LDP(X \rightarrow Y) := \sup_{x,x',y} \log \frac{P_{Y|X}(y|x)}{P_{Y|X}(y|x')} \]  

[Warner '65; Evfimievski et al. '03]

Operational interpretation?

**Theorem** (cf. Dwork et al. '06):

\[ LDP(X \rightarrow Y) = \sup_{f,P_X,y} \left| \log \left( \frac{P(f(X) = 1|Y = y)}{P(f(X) = 1)} \right) \right| \]
Theorem (cf. Dwork et al. ’06):

\[ LDP(X \rightarrow Y) = \sup_{f, P_X, y} \left| \log \left( \frac{P(f(X) = 1|Y = y)}{P(f(X) = 1)} \right) \right| \]
(Local) Differential Privacy

**Theorem** (cf. Dwork et al. ’06):

\[
LDP(X \rightarrow Y) = \sup_{f,P_x,y} \left| \log \left( \frac{P(f(X) = 1|Y = y)}{P(f(X) = 1)} \right) \right|
\]

**Theorem** (Issa-Wagner):

\[
LDP(X \rightarrow Y) = \sup_{P_x} \sup_{U \leftrightarrow X \leftrightarrow Y} \frac{\sup_y \sup_{\tilde{u}} \Pr(U = \tilde{u}|Y = y)}{\sup_{\tilde{u}} \Pr(U = \tilde{u})} \log \left( \frac{P(f(X) = 1|Y = y)}{P(f(X) = 1)} \right)
\]
Given $p(x)$ and $c(x,y)$, solve

$$\min_{p(y|x)} \sum_y \max_x p(y|x)$$

subject to

$$\sum_{x,y} p(x)p(y|x)c(x, y) \leq C$$

$$\sum_y p(y|x) = 1 \quad \forall \ x$$

$$p(y|x) \geq 0 \quad \forall \ x, y$$
Optimal Mechanisms

- Given $p(x)$ and $c(x,y)$, solve

  \[
  \min_{p(y|x)} \sum_y \max_x p(y|x) \sum_{x,y} p(x)p(y|x)c(x, y) \leq C
  \]

  subject to

  \[
  \sum_y p(y|x) = 1 \quad \forall x
  \]

  \[
  p(y|x) \geq 0 \quad \forall x, y
  \]

  “exp-leakage”
Formulation as an LP

\[
\begin{align*}
\min & \quad \sum_y q_y p(y|x), q_y \\
\text{subject to} & \quad \sum_{x,y} p(x)p(y|x)c(x, y) \leq C \\
& \quad \sum_y p(y|x) = 1 \quad \forall \ x \\
& \quad p(y|x) \geq 0 \quad \forall \ x, y \\
& \quad p(y|x) \leq q_y \quad \forall \ x, y
\end{align*}
\]
A Structural Assumption

\[ c(x,y): \text{nondecreasing} \]

\[ y \text{ nondecreasing} \]

\[ x \text{ nondecreasing} \]

\[ \infty \]
A Structural Assumption

Examples:

\[ c(x,y) : x \rightarrow y \]

- nondecreasing

- nondecreasing

- nondecreasing

- nondecreasing
A Structural Assumption

Examples:
- Execution time [RSA], power consumption.

\[ c(x, y) : x \]
A Structural Assumption

Examples:
- Execution time [RSA], power consumption
- “Staircase increasing”
Deterministic Mechanisms Are Optimal
Deterministic Mechanisms Are Optimal

**Theorem** (Wu, Wagner, Suh):
If $c(\cdot, \cdot)$ is staircase increasing, then for any $\alpha$ and $P_X$,

$$\sum_y \max_x P_{Y|X}(y|x) + \alpha \cdot \sum_x \sum_y P_X(x)P_{Y|X}(y|x)c(x, y)$$

is minimized by a deterministic (0-1) $P_{Y|X}$.
Theorem (Wu, Wagner, Suh):
If $c(\cdot,\cdot)$ is staircase increasing, then for any $\alpha$ and $P_X$,

$$\sum_y \max_{x} P_{Y|X}(y|x) + \alpha \cdot \sum_x \sum_y P_X(x)P_{Y|X}(y|x)c(x, y)$$

is minimized by a deterministic (0-1) $P_{Y|X}$.

Fails for the cost matrix:

$$\begin{bmatrix}
0 & 1 & 2 \\
2 & 0 & 1 \\
1 & 2 & 0
\end{bmatrix}$$
Corollary (Wu, Wagner, Suh):
The optimal cost/exp-leakage curve is piecewise linear with kink points only at integer exp-leakage values.
Advantages of Deterministic Mechanisms
Advantages of Deterministic Mechanisms

- Do not require randomness (obviously)
Advantages of Deterministic Mechanisms

- Do not require randomness (obviously)
- Easier to describe and store
Advantages of Deterministic Mechanisms

- Do not require randomness (obviously)
- Easier to describe and store
- Immune to averaging attacks
cf. Other Metrics

\[ c(x, y) = \begin{bmatrix}
  1 & 2 & 3 & 4 \\
  \infty & 1 & 2 & 3 \\
  \infty & \infty & 1 & 2 \\
  \infty & \infty & \infty & 1 \\
\end{bmatrix} \quad p(x): \text{uniform} \]

\[
\text{minimize } \{E[c(X,Y)]\} : \text{leakage } \leq 1
\]
cf. Other Metrics

\[ c(x, y) = \begin{bmatrix} 1 & 2 & 3 & 4 \\ \infty & 1 & 2 & 3 \\ \infty & \infty & 1 & 2 \\ \infty & \infty & \infty & 1 \end{bmatrix} \]

\[ p(x) : \text{uniform} \]

Minimize \( \{ E[c(X,Y)] \} : \text{leakage} \leq 1 \)

Maximal Leakage:

\[ p(y|x) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \]
cf. Other Metrics

\[ c(x, y) = \begin{bmatrix} 1 & 2 & 3 & 4 \\ \infty & 1 & 2 & 3 \\ \infty & \infty & 1 & 2 \\ \infty & \infty & \infty & 1 \end{bmatrix} \quad p(x): \text{uniform} \]

\[
\text{minimize } \{ E[c(X,Y)] \} : \text{leakage } \leq 1
\]
cf. Other Metrics

$$c(x, y) = \begin{bmatrix} 1 & 2 & 3 & 4 \\ \infty & 1 & 2 & 3 \\ \infty & \infty & 1 & 2 \\ \infty & \infty & \infty & 1 \end{bmatrix} \quad p(x): \text{uniform}$$

minimize \{E[c(X,Y)]] : \text{leakage} \leq 1\}

Mutual Information:

$$p(y|x) = \begin{bmatrix} 0.52 & 0.27 & 0.14 & 0.07 \\ 0 & 0.56 & 0.29 & 0.15 \\ 0 & 0 & 0.69 & 0.34 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$
**cf. Other Metrics**

\[
c(x, y) = \begin{bmatrix}
1 & 2 & 3 & 4 \\
\infty & 1 & 2 & 3 \\
\infty & \infty & 1 & 2 \\
\infty & \infty & \infty & 1 \\
\end{bmatrix}
\]

\[p(x): \text{uniform}\]

\[
\text{minimize } \{E[c(X,Y)]\} : \text{leakage } \leq 1
\]

**Mutual Information:**

\[
p(y|x) = \begin{bmatrix}
0.52 & 0.27 & 0.14 & 0.07 \\
0 & 0.56 & 0.29 & 0.15 \\
0 & 0 & 0.69 & 0.34 \\
0 & 0 & 0 & 1
\end{bmatrix}
\]

**Local Diff. Privacy:**

\[
p(y|x) = \begin{bmatrix}
0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1
\end{bmatrix}
\]
Maximal Leakage: Other Results

- Shannon cipher system (Issa, Kamath, Wagner ’16)
Maximal Leakage: Other Results

- Shannon cipher system (Issa, Kamath, Wagner ’16)
- Privacy-utility tradeoffs (Liao, Sankar, Calmon, Tan, ’17)
Maximal Leakage: Other Results

- Shannon cipher system (Issa, Kamath, Wagner ’16)
- Privacy-utility tradeoffs (Liao, Sankar, Calmon, Tan, ’17)
- Sibson MI of other orders (Liao, Kosut, Sankar, Calmon, ’18)
Maximal Leakage: Other Results

- Shannon cipher system (Issa, Kamath, Wagner ’16)
- Privacy-utility tradeoffs (Liao, Sankar, Calmon, Tan, ’17)
- Sibson MI of other orders (Liao, Kosut, Sankar, Calmon, ’18)
- Learning ML from trace data (Issa and Wagner, ’18)
Three Takeaways

**Def (Issa-Kamath-Wagner):** Given $P_{XY}$, the *maximal leakage* from $X$ to $Y$ is

$$\mathcal{L}(X \to Y) = \sup_{U:U \leftrightarrow X \leftrightarrow Y} \log \frac{\sup_{\tilde{u}()} \Pr(U = \tilde{u}(Y))}{\sup_{\tilde{u}} \Pr(U = \tilde{u})}$$
Three Takeaways

**Def (Issa-Kamath-Wagner):** Given \( P_{XY} \), the maximal leakage from \( X \) to \( Y \) is

\[
\mathcal{L}(X \rightarrow Y) = \sup_{U:U \leftrightarrow X \leftrightarrow Y} \log \frac{\sup_{\tilde{u}(\cdot)} \Pr(U = \tilde{u}(Y))}{\sup_{\tilde{u}} \Pr(U = \tilde{u})}
\]

Maximal leakage ...
Three Takeaways

**Def** (Issa-Kamath-Wagner): Given $P_{XY}$, the maximal leakage from $X$ to $Y$ is

$$\mathcal{L}(X \rightarrow Y) = \sup_{U:U\leftrightarrow X\leftrightarrow Y} \log \frac{\sup_{\tilde{u}(\cdot)} \Pr(U = \tilde{u}(Y))}{\sup_{\hat{u}} \Pr(U = \hat{u})}$$

Maximal leakage ...

1. ... captures the increase in guessing probability of secrets
Three Takeaways

**Def (Issa-Kamath-Wagner):** Given $P_{XY}$, the maximal leakage from $X$ to $Y$ is

$$
\mathcal{L}(X \to Y) = \sup_{U:U \leftrightarrow X \leftrightarrow Y} \log \frac{\sup_{\tilde{u}(\cdot)} \Pr(U = \tilde{u}(Y))}{\sup_{\tilde{u}} \Pr(U = \tilde{u})}
$$

Maximal leakage ...

1. ... captures the increase in guessing probability of secrets

... is well suited for side channels with keys, passwords.
Three Takeaways

**Def (Issa-Kamath-Wagner):** Given $P_{XY}$, the maximal leakage from $X$ to $Y$ is

$$\mathcal{L}(X \rightarrow Y) = \sup_{U:U \leftrightarrow X \leftrightarrow Y} \log \frac{\sup_{\tilde{u}(\cdot)} \Pr(U = \tilde{u}(Y))}{\sup_{\tilde{u}} \Pr(U = \tilde{u})}$$

Maximal leakage ...

1. ... captures the increase in guessing probability of secrets
   ... is well suited for side channels with keys, passwords.
2. ... is robust to modeling assumptions
Three Takeaways

Def (Issa-Kamath-Wagner): Given $P_{XY}$, the maximal leakage from $X$ to $Y$ is

$$
\mathcal{L}(X \rightarrow Y) = \sup_{U: U \leftrightarrow X \leftrightarrow Y} \log \frac{\sup_{\tilde{u}} \Pr(U = \tilde{u}(Y))}{\sup_{\tilde{u}} \Pr(U = \tilde{u})}
$$

Maximal leakage ...

1. ... captures the increase in guessing probability of secrets
   ... is well suited for side channels with keys, passwords.
2. ... is robust to modeling assumptions
3. ... favors deterministic mechanisms (quantization) over “adding noise” in many contexts.
A Different Question
How many secrecy measures do we need?
How many secrecy measures do we need?
- Probably more than one ...
A Different Question

How many secrecy measures do we need?

- Probably more than one ...

- ML ill-suited for e.g., medical databases
A Different Question

How many secrecy measures do we need?

- Probably more than one ...
  - ML ill-suited for e.g., medical databases
  - DP ill-suited for side channels
A Different Question

How many secrecy measures do we need?

- Probably more than one ...  
  - ML ill-suited for e.g., medical databases  
  - DP ill-suited for side channels  
  - Both ML and DP ill-suited for computationally-bounded eavesdroppers
How many secrecy measures do we need?

- Probably more than one ...
  - ML ill-suited for e.g., medical databases
  - DP ill-suited for side channels
  - Both ML and DP ill-suited for computationally-bounded eavesdroppers
- ... but probably not 80+ either.
Given \( A \subseteq \mathcal{Y} \), the induced deterministic mechanism, \( P_A \), is

\[
p(y|x) = 1 \quad \text{if} \quad y = \text{argmin} \{ c(x, y') : y' \in A \}
\]
Given $A \subseteq \mathcal{Y}$, the *induced deterministic mechanism*, $P_A$, is

$$p(y|x) = 1 \text{ if } y = \text{argmin} \{c(x, y') : y' \in A\}$$
A Greedy Algorithm

- Given $A \subseteq \mathcal{Y}$, the induced deterministic mechanism, $P_A$, is

$$p(y|x) = 1 \text{ if } y = \arg\min\{c(x, y') : y' \in A\}$$

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$
A Greedy Algorithm

- Given $A \subseteq \mathcal{Y}$, the induced deterministic mechanism, $P_A$, is

$$p(y|x) = 1 \quad \text{if} \quad y = \arg\min \{c(x, y') : y' \in A\}$$

$$A = \begin{bmatrix}
1 & 1 & 1 & 1 & 1
\end{bmatrix}$$
A Greedy Algorithm

- Given $A \subseteq \mathcal{Y}$, the induced deterministic mechanism, $P_A$, is

$$p(y|x) = 1 \text{ if } y = \text{argmin} \{c(x, y') : y' \in A\}$$

$$A = \begin{bmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
\end{bmatrix}$$
A Greedy Algorithm
A Greedy Algorithm

- Start with a singleton $A$ that minimizes the cost of $P_A$. 
A Greedy Algorithm

- Start with a singleton $A$ that minimizes the cost of $P_A$.

- Iterate: $A \rightarrow A \cup \{j\}$, where $j \notin A$ is chosen to minimize the cost of $P_{A \cup \{j\}}$. 
Theorem (Wu, Wagner, Suh ’19):
For exp-leakage $k$, let
- $C^*(k)$ denote the optimum cost
- $C_G(k)$ denote the cost obtained by the greedy algorithm

Then $C^*(1) = C_G(1)$, $C^*(2) = C_G(2)$, and

\[
C^*(1) - C_G(k) \geq \left(1 - \left(\frac{k - 2}{k - 1}\right)^{k-1}\right)(C^*(1) - C^*(k)) \\
\geq \left(1 - \frac{1}{e}\right)(C^*(1) - C^*(k)) \\
\geq 0.63(C^*(1) - C^*(k))
\]
Theorem (Wu, Wagner, Suh ’19):

For exp-leakage $k$, let

- $C^*(k)$ denote the optimum cost
- $C_G(k)$ denote the cost obtained by the greedy algorithm

Then $C^*(1) = C_G(1)$, $C^*(2) = C_G(2)$, and

$$C^*(1) - C_G(k) \geq \left(1 - \left(\frac{k-2}{k-1}\right)^{k-1}\right)(C^*(1) - C^*(k))$$

$$\geq \left(1 - \frac{1}{e}\right)(C^*(1) - C^*(k))$$

$$\geq 0.63(C^*(1) - C^*(k))$$

Proof: submodularity of $\text{cost}(P_A)$. 
Theorem (Wu, Wagner, Suh ’19):

For exp-leakage $k$, let

- $C^*(k)$ denote the optimum cost
- $C_G(k)$ denote the cost obtained by the greedy algorithm

Then $C^*(1) = C_G(1)$, $C^*(2) = C_G(2)$, and

$$C^*(1) - C_G(k) \geq \left(1 - \left(\frac{k-2}{k-1}\right)^{k-1}\right)(C^*(1) - C^*(k))$$

$$\geq \left(1 - \frac{1}{e}\right)(C^*(1) - C^*(k))$$

$$\geq 0.63(C^*(1) - C^*(k))$$

Proof: submodularity of $\text{cost}(P_A)$.

Note: leads to a sequence of approximations.
How to Delay Packets?

\[ X(t) \rightarrow ? \rightarrow Y(t) \]

[nominal packet timings] [actual packet timings]
How to Delay Packets?

- Suppose $X(t)$ is a Poisson process with rate $\lambda$
Suppose \( X(t) \) is a Poisson process with rate \( \lambda \)

How to blur the packet timings to minimize leakage?
Try an $M/M/1$ Queue

$X(t)$

$M/M/1$ queue service rate $\mu$

$Y(t)$

[nominal packet timings]

[actual packet timings]
Try an $M/M/1$ Queue

$X(t)$

$M/M/1$ queue

service rate $\mu$

$Y(t)$

[nominal packet timings]

[actual packet timings]

$$\frac{1}{T} \cdot \mathcal{L} \left( \{X(t)\}_{t=0}^{T} \rightarrow \{Y(t)\}_{t=0}^{T} \right) = \mu \text{ nats}$$
Try an $M/M/1$ Queue

$X(t)$

$M/M/1$ queue service rate $\mu$

$Y(t)$

$\left\{ X(t) \right\}_{t=0}^{T} \rightarrow \left\{ Y(t) \right\}_{t=0}^{T}$

$\frac{1}{T} \cdot \mathcal{L} \left( \left\{ X(t) \right\}_{t=0}^{T} \rightarrow \left\{ Y(t) \right\}_{t=0}^{T} \right) = \mu \quad \text{nats}$

[leakage rate is at least $\lambda$]
\[ \frac{1}{T} \cdot \mathcal{L} \left( \{X(t)\}_{t=0}^{T} \rightarrow \{Y(t)\}_{t=0}^{T} \right) \leq \frac{1}{\tau} \log m \]
Accumulate and Dump

\[ \frac{1}{T} \cdot \mathcal{L} \left( \{X(t)\}_{t=0}^{T} \rightarrow \{Y(t)\}_{t=0}^{T} \right) \leq \frac{1}{\tau} \log m \]
Accumulate and Dump

\[
\frac{1}{T} \cdot \mathcal{L} \left( \{X(t)\}_{t=0}^{T} \rightarrow \{Y(t)\}_{t=0}^{T} \right) \leq \frac{1}{\tau} \log m
\]
Accumulate and Dump

\[
\frac{1}{T} \cdot \mathcal{L} \left( \{X(t)\}^T_{t=0} \rightarrow \{Y(t)\}^T_{t=0} \right) \leq \frac{1}{\tau} \log m
\]
Accumulate and Dump

\[
\frac{1}{T} \cdot \mathcal{L} \left( \{X(t)\}_{t=0}^{T} \rightarrow \{Y(t)\}_{t=0}^{T} \right) \leq \frac{1}{\tau} \log m
\]
Accumulate and Dump

\[ \frac{1}{T} \cdot \mathcal{L} \left( \{ X(t) \}_{t=0}^{T} \rightarrow \{ Y(t) \}_{t=0}^{T} \right) \leq \frac{1}{\tau} \log m \]
Accumulate and Dump

\[ \frac{1}{T} \cdot \mathcal{L} \left( \{X(t)\}_{t=0}^T \rightarrow \{Y(t)\}_{t=0}^T \right) \leq \frac{1}{\tau} \log m \]

[quantization leaks less than “adding noise”]
Accumulate and Dump

\[ \frac{1}{T} \cdot \mathcal{L} ( \{X(t)\}_{t=0}^{T} \rightarrow \{Y(t)\}_{t=0}^{T}) \leq \frac{1}{\tau} \log m \]

[quantization leaks less than “adding noise”]

[cf. Kadloor, Kiyavash, and Venkitasubramaniam ’16]
The Shannon Cipher System

\[ X^n \rightarrow f \xrightarrow{M \in \{0, 1\}^{nR}} g \rightarrow \hat{X}^n \]
The Shannon Cipher System

\[ X^n \xrightarrow{f} M \in \{0, 1\}^{nR} \xrightarrow{g} \hat{X}^n \]

Eve
The Shannon Cipher System

$K \in \{0, 1\}^{nr}$

$M \in \{0, 1\}^{nR}$

$X^n \rightarrow f \rightarrow g \rightarrow \hat{X}^n$

Eve
The Shannon Cipher System

\[ K \in \{0, 1\}^m \]

\[ M \in \{0, 1\}^{n_R} \]

\[ f \]

\[ g \]

\[ \hat{X}^n \]

\[ X^n \]

Eve

uniform
The Shannon Cipher System

\[ K \in \{0, 1\}^{nr} \]

\[ M \in \{0, 1\}^{nR} \]

\[ X^n \xrightarrow{f} \rightarrow \hat{X}^n \]

\[ M \xrightarrow{g} \rightarrow \hat{X}^n \]

Eve
Shannon Cipher System

- Shannon ('49): perfect secrecy is possible if the key rate $r$ exceeds the message rate $R$.

\[ K \in \{0, 1\}^{nr} \]
\[ M \in \{0, 1\}^{nR} \]
\[ X^n \rightarrow f \rightarrow M \in \{0, 1\}^{nR} \rightarrow g \rightarrow \hat{X}^n \]

- Eve

\[ \vdash X^n \rightarrow f \rightarrow M \in \{0, 1\}^{nR} \rightarrow g \rightarrow \hat{X}^n \]
The Shannon Cipher System

- Shannon (’49): perfect secrecy is possible if the key rate $r$ exceeds the message rate $R$.
- How to design $f$ and $g$ to minimize leakage when $r < R$?
Leakage and Shannon’s Cipher

\[ K \in \{0, 1\}^{nr} \]
\[ M \in \{0, 1\}^{nR} \]

X^n \quad \text{i.i.d. discrete} \quad \hat{X}^n
Leakage and Shannon’s Cipher

\( K \in \{0, 1\}^{nr} \)

\( M \in \{0, 1\}^{nR} \)

\( X^n \rightarrow f \rightarrow g \rightarrow \hat{X^n} \)

Eve
Leakage and Shannon’s Cipher

\[ K \in \{0, 1\}^{nr} \]

\[ X_1^n \xrightarrow{f} M_1^n \xrightarrow{g} \hat{X}_1^n \]

\[ L_n = \min_{f, g} \frac{1}{n} \cdot \mathcal{L}(X^n \rightarrow M) \]

subject to

\[ f : X_1^n \times \{0, 1\}^{nr} \rightarrow \{0, 1\}^{nR} \]

\[ g : \{0, 1\}^{nR} \times \{0, 1\}^{nr} \rightarrow \hat{X}_1^n \]

\[ \frac{1}{n} \sum_{i=1}^{n} E[d(X_i, \hat{X}_i)] \leq D \]
Leakage and Shannon’s Cipher

\[ K \in \{0, 1\}^{nr} \]

\[ X^n \xrightarrow{f} M \in \{0, 1\}^{nR} \xrightarrow{g} \hat{X}^n \]

\[
L_n = \min_{f, g} \frac{1}{n} \cdot \mathcal{L}(X^n \rightarrow M)
\]

subject to

\[
f : X^n \times \{0, 1\}^{nr} \rightarrow \{0, 1\}^{nR}
\]

\[
g : \{0, 1\}^{nR} \times \{0, 1\}^{nr} \rightarrow \hat{X}^n
\]

\[
\frac{1}{n} \sum_{i=1}^{n} E[d(X_i, \hat{X}_i)] \leq D
\]

\[
L = \lim_{n \to \infty} L_n
\]
Leakage and Shannon’s Cipher

**Theorem** (Issa-Kamath-Wagner): Let $R(D)$ denote the rate-distortion function for the source. If

$$R < R(D),$$

then the problem is infeasible. Otherwise, the min. max. leakage is

$$L = [R(D) - r]^+$$
Theorem (Issa-Kamath-Wagner): Let $R(D)$ denote the rate-distortion function for the source. If

$$R < R(D),$$

then the problem is infeasible. Otherwise, the min. max. leakage is

$$L = [R(D) - r]^+$$

Notes:
Theorem (Issa-Kamath-Wagner): Let $R(D)$ denote the rate-distortion function for the source. If

$$R < R(D),$$

then the problem is infeasible. Otherwise, the min. max. leakage is

$$L = [R(D) - r]^+$$

Notes:
- Using MI instead of leakage gives same result
**Theorem** (Issa-Kamath-Wagner): Let $R(D)$ denote the rate-distortion function for the source. If

$$R < R(D),$$

then the problem is infeasible. Otherwise, the min. max. leakage is

$$L = [R(D) - r]^+$$

**Notes:**
- Using MI instead of leakage gives same result
- Though difference in optimal schemes...
**Theorem** (Issa-Kamath-Wagner): Let $R(D)$ denote the rate-distortion function for the source. If

$$R < R(D),$$

then the problem is infeasible. Otherwise, the min. max. leakage is

$$L = [R(D) - r]^+$$

**Notes:**

- Using MI instead of leakage gives same result
  - Though difference in optimal schemes...
- Large deviations (and a.s.) result
$X^n$
Achievability for Primary User

\[ X^n \rightarrow \text{optimal lossy compressor} \rightarrow nR(D) \text{ bits} \]

01001011010001000011
Achievability for Primary User

\[ X^n \xrightarrow{\text{optimal lossy compressor}} nR(D) \text{ bits} + 11010111000 \text{ (key)} \]
Achievability for Primary User

$X^n \rightarrow \text{optimal lossy compressor} \rightarrow nR(D) \text{ bits}$

$01001011010001000011 + 11010111000 \text{ (key)}$

$10011100010001000011$
Achievability for Primary User

\[ X^n \rightarrow \text{optimal lossy compressor} \rightarrow nR(D) \text{ bits} \]

\[ 01001011010001000011 + 11010111000 \text{ (key)} \]

\[ \overline{10011100010001000111} \]

\[ M \]
Achievability for Primary User

\[ X^n \rightarrow \text{optimal lossy compressor} \rightarrow nR(D) \text{ bits} \]

\[ \begin{array}{c}
0100101101000100011 \\
+ 110101111000 \text{ (key)} \\
\hline
1001110001000100011
\end{array} \]

\[ M \]

\[ 1001110001000100011 \]
Optimal lossy compressor

\[ X^n \rightarrow \text{optimal lossy compressor} \rightarrow nR(D) \text{ bits} \]

\[ \begin{array}{c}
01001011010001000011 \\
+ 11010111000 \text{ (key)} \\
\hline
10011100010001000100011
\end{array} \]

\[ M \]

\[ 10011100010001000100011 \\
+ 11010111000 \text{ (key)} \]
Achievability for Primary User

\[ X^n \rightarrow \text{optimal lossy compressor} \rightarrow nR(D) \text{ bits} \]

\[
01001011010001000011 + 110101111000 \text{ (key)} \\
\hline
10011100010001000011
\]

\[ M \]

\[
10011100010001000011 + 110101111000 \text{ (key)} \\
\hline
01001011010001000011
\]
Achievability for Primary User

\[ X^n \rightarrow \text{optimal lossy compressor} \rightarrow \]

\[ nR(D) \text{ bits} \]

\[ 01001011010001000011 + 11010111000 \text{ (key)} \]

\[ \underline{10011100010001000011} \]

\[ M \]

\[ 10011100010001000011 + 11010111000 \text{ (key)} \]

\[ \underline{01001011010001000011} \]
Achievability for Primary User

\[ X^n \to \text{optimal lossy compressor} \to \text{(nR(D) bits)} \]

\[ \hat{X}^n \leftarrow \text{optimal lossy decompressor} \]

\[ 0100101101000100011 + 11010111000 \text{ (key)} \]

\[ 10011100010001000011 \]

\[ M \]

\[ 10011100010001000011 + 11010111000 \text{ (key)} \]

\[ 0100101101000100011 \]
Achievability for Eavesdropper

\[ X^n \rightarrow f \rightarrow g \]

\[ K \in \{0, 1\}^{nr} \]

\[ M \in \{0, 1\}^{nR} \]

Eve
1. Consider worst-case $U$

$U \leftrightarrow X^n \xrightarrow{M \in \{0, 1\}^n} g \xrightarrow{Eve}$
Achievability for $Eve$:

1. Initialize with $U \leftrightarrow X^n$
2. Guess key randomly $K \in \{0, 1\}$
3. Generate $M \in \{0, 1\}^{nR}$
4. Process through function $f$
5. Process through function $g$
6. $Eve$ intercepts $M$
Achievability for Eavesdropper

\[ U \leftrightarrow X^n \]

\[ f \]

\[ K \in \{0, 1\}^{nr} \]

\[ M \in \{0, 1\}^{nR} \]

\[ g \]

\[ \tilde{X}^n \]

3. Emulate \( g \) to create \( \tilde{X}^n \)
4. Pick $X^n$ uniformly at random from within distortion ball around $\tilde{X}^n$. 
Achievability for Eavesdropper

5. Generate $U$ from $X^n$. 

$K \in \{0, 1\}^{nr}$

$M \in \{0, 1\}^{nR}$

$\tilde{X}^n$
Suppose $r = 0$ and $R$ is large
Suppose $r = 0$ and $R$ is large.
Quantization vs. Adding Noise

\[ X^n \xrightarrow{f} \hat{X}^n \xrightarrow{\text{Eve}} \]
Quantization vs. Adding Noise

Then

\[ L_n = \min_{\hat{X}^n} \frac{1}{n} \cdot \mathcal{L}(X^n \rightarrow \hat{X}^n) \]

subject to

\[ \frac{1}{n} \sum_{i=1}^{n} E[d(X_i, \hat{X}_i)] \leq D \]
Quantization vs. Adding Noise

Then

\[ L_n = \min_{\hat{X}^n} \frac{1}{n} \cdot \mathcal{L}(X^n \rightarrow \hat{X}^n) \]

subject to

\[ L = \lim_{n \to \infty} L_n \quad \frac{1}{n} \sum_{i=1}^{n} E[d(X_i, \hat{X}_i)] \leq D \]
Quantization vs. Adding Noise

Then \( L_n = \min_{\hat{X}^n} \frac{1}{n} \cdot \mathcal{L}(X^n \rightarrow \hat{X}^n) \)
subject to \( \frac{1}{n} \sum_{i=1}^{n} E[d(X_i, \hat{X}_i)] \leq D \)
Quantization vs. Adding Noise

Optimal scheme:
Quantization vs. Adding Noise

Optimal scheme:

- Compress $X^n$ optimally to rate $R(D)$, then decompress.
Quantization vs. Adding Noise

Optimal scheme:

- Compress $X^n$ optimally to rate $R(D)$, then decompress.
- Leaks $R(D)$ bits per symbol
Optimal scheme:

- Compress $X^n$ optimally to rate $R(D)$, then decompress.
- Leaks $R(D)$ bits per symbol
- Deterministic but noncausal
Quantization vs. Adding Noise

Optimal scheme:

- Compress $X^n$ optimally to rate $R(D)$, then decompress.
- Leaks $R(D)$ bits per symbol
- Deterministic but noncausal

Memoryless scheme:

\[
\begin{align*}
X_1 &\rightarrow \text{Channel}_1 \rightarrow \hat{X}_1 \\
X_2 &\rightarrow \text{Channel}_2 \rightarrow \hat{X}_2 \\
X_3 &\rightarrow \text{Channel}_3 \rightarrow \hat{X}_3 \\
&\vdots
\end{align*}
\]

\[
X_n \rightarrow \text{Channel}_n \rightarrow \hat{X}_n
\]
Quantization vs. Adding Noise

Optimal scheme:
- Compress $X^n$ optimally to rate $R(D)$, then decompress.
- Leaks $R(D)$ bits per symbol
- Deterministic but noncausal

Memoryless scheme:
\[
\begin{align*}
X_1 &\rightarrow \text{Channel}_1 \rightarrow \hat{X}_1 \\
X_2 &\rightarrow \text{Channel}_2 \rightarrow \hat{X}_2 \\
X_3 &\rightarrow \text{Channel}_3 \rightarrow \hat{X}_3 \\
&\vdots \\
X_n &\rightarrow \text{Channel}_n \rightarrow \hat{X}_n
\end{align*}
\]
Memoryless scheme is causal but suboptimal.
Optimal scheme:
- Compress $X^n$ optimally to rate $R(D)$, then decompress.
- Leaks $R(D)$ bits per symbol
- Deterministic but noncausal

Memoryless scheme:
- $X_1 \rightarrow \text{Channel}_1 \rightarrow \hat{X}_1$
- $X_2 \rightarrow \text{Channel}_2 \rightarrow \hat{X}_2$
- $X_3 \rightarrow \text{Channel}_3 \rightarrow \hat{X}_3$
- $\vdots$
- $X_n \rightarrow \text{Channel}_n \rightarrow \hat{X}_n$

Memoryless scheme is causal but suboptimal.

[quantization is preferable to “adding noise”]
Quantization vs. Adding Noise

Optimal scheme:
- Compress $X^n$ optimally to rate $R(D)$, then decompress.
-Leaks $R(D)$ bits per symbol
- Deterministic but noncausal

Memoryless scheme:

Memoryless scheme is causal but suboptimal.

[quantization is preferable to “adding noise”]
[cf. mutual info.]
Def (Issa-Kamath-Wagner): For any metric space \( \mathcal{U} \),

\[
\mathcal{L}_\mathcal{U}(X \rightarrow Y) = \sup_{U: U \leftrightarrow X \leftrightarrow Y} \log \frac{\sup_{\hat{u}(\cdot)} \Pr(U \in B(\hat{u}(Y)))}{\sup_{\hat{u}} \Pr(U \in B(\hat{u}))}.
\]
**Def (Issa-Kamath-Wagner):** For any metric space \( U \),

\[
\mathcal{L}_U(X \rightarrow Y) = \sup_{U: U \leftrightarrow X \leftrightarrow Y} \log \frac{\sup_{\hat{u}(\cdot)} \Pr(U \in B(\hat{u}(Y)))}{\sup_{\hat{u}} \Pr(U \in B(\hat{u}))}
\]

where \( \exists u: \Pr(U \in B(u)) > 0 \).

**Theorem (Issa-Kamath-Wagner):** For any metric space \( U \),

\[
\mathcal{L}_U(X \rightarrow Y) \leq \mathcal{L}(X \rightarrow Y)
\]

with equality if \( U \) has countably many points no two of which are contained in the same unit ball.
**Def** (Issa-Kamath-Wagner):

\[
\mathcal{L}_G(X \rightarrow Y) = \sup_{U: U \leftrightarrow X \leftrightarrow Y} \log \frac{\sup_{\hat{\mu}(\cdot)} E[g(U, \hat{\mu}(Y))]}{\sup_{\hat{\mu}} E[g(U, \hat{\mu})]}
\]

\[
g(\cdot, \cdot): \mathcal{U} \times \hat{\mathcal{U}} \rightarrow [0, \infty):
\]

\[
\sup_{\hat{\mu}} E[g(U, \hat{\mu})] > 0
\]
**Extension: General Gains**

**Def** (Issa-Kamath-Wagner): 
\[ \mathcal{L}_G(X \rightarrow Y) = \sup_{U:U \leftrightarrow X \leftrightarrow Y} \sup_{g(\cdot, \cdot):U \times \hat{U} \rightarrow [0, \infty)} \left( \log \frac{\sup_{\hat{U}}} {\sup_{\hat{U}}} \mathbb{E}[g(U, \hat{U})] \right) \]

**Theorem** (Issa-Kamath-Wagner): If \( X \) and \( Y \) are discrete, then
\[ \mathcal{L}_G(X \rightarrow Y) = \mathcal{L}(X \rightarrow Y). \]
**Definition:** The opportunistic maximal leakage is

\[
\mathcal{L}_O(X \rightarrow Y) = \log E_Y \left[ \sup_{U \leftrightarrow X \leftrightarrow Y} \frac{\sup_{\tilde{u}} P_{U \mid Y}(\tilde{u} \mid y)}{\sup_{\tilde{u}} P(\tilde{u})} \right]
\]
**Definition:** The opportunistic maximal leakage is

\[
\mathcal{L}_O(X \rightarrow Y) = \log E_Y \left[ \sup_{U \leftrightarrow X \leftrightarrow Y} \frac{\sup_{\tilde{u}} P_{U|Y}(\tilde{u}|y)}{\sup_{\tilde{u}} P(\tilde{u})} \right]
\]

**Theorem** (Issa-Wagner): For any joint distribution \( P_{XY} \) on finite alphabets

\[
\mathcal{L}_O(X \rightarrow Y) = \mathcal{L}(X \rightarrow Y)
\]
**Corollary (IKW):** If $X$ and $Y$ are jointly continuous then

$$\mathcal{L}(X \to Y) = \log \int \sup_{x: f_X(x) > 0} f_{Y|X}(y|x) \, dy$$
Corollary (IKW): If $X$ and $Y$ are jointly continuous then

$$\mathcal{L}(X \rightarrow Y) = \log \int \sup_{x: f_X(x) > 0} f_{Y|X}(y|x) \, dy$$

Corollary (IKW): If $X$ and $Y$ are jointly Gaussian then

$$\mathcal{L}(X \rightarrow Y) = \begin{cases} 0 & \text{if } X, Y \text{ indep.} \\ \infty & \text{otherwise} \end{cases}$$
Extension: General Alphabet
Theorem (IKW ’17): Let \((\mathcal{X} \times \mathcal{Y}, \sigma_{X \times Y}, P_{XY})\) be a prob. space with associated prob. spaces \((\mathcal{X}, \sigma_X, P_X)\) and \((\mathcal{Y}, \sigma_Y, P_Y)\).
Theorem (IKW ’17): Let \((\mathcal{X} \times \mathcal{Y}, \sigma_{X \times Y}, P_{XY})\) be a prob. space with associated prob. spaces \((\mathcal{X}, \sigma_X, P_X)\) and \((\mathcal{Y}, \sigma_Y, P_Y)\).

- If \(P_{XY} \ll P_X \times P_Y\) and \(\sigma_X\) is generated by a countable set then

\[
\mathcal{L}(X \to Y) = \log \int_{\mathcal{Y}} \text{ess sup}_{x} \left\{ \frac{dP_{XY}}{dP_X \times dP_Y}(x, y) \right\} dP_Y
\]
Theorem (IKW ’17): Let \((\mathcal{X} \times \mathcal{Y}, \sigma_{X \times Y}, P_{XY})\) be a prob. space with associated prob. spaces \((\mathcal{X}, \sigma_{X}, P_{X})\) and \((\mathcal{Y}, \sigma_{Y}, P_{Y})\).

- If \(P_{XY} \ll P_{X} \times P_{Y}\) and \(\sigma_{X}\) is generated by a countable set then

\[
\mathcal{L}(X \rightarrow Y) = \log \int_{\mathcal{Y}} \text{ess sup}_{x} \left\{ \frac{dP_{XY}}{dP_{X} \times dP_{Y}}(x, y) \right\} dP_{Y}
\]
Theorem (IKW ’17): Let \((\mathcal{X} \times \mathcal{Y}, \sigma_{\mathcal{X} \times \mathcal{Y}}, P_{XY})\) be a prob. space with associated prob. spaces \((\mathcal{X}, \sigma_X, P_X)\) and \((\mathcal{Y}, \sigma_Y, P_Y)\).

- If \(P_{XY} \ll P_X \times P_Y\) and \(\sigma_X\) is generated by a countable set then

\[
\mathcal{L}(X \rightarrow Y) = \log \int_{\mathcal{Y}} \text{ess sup}_x \left\{ \frac{dP_{XY}}{dP_X \times dP_Y}(x, y) \right\} dP_Y
\]
Theorem (IKW '17): Let \((\mathcal{X} \times \mathcal{Y}, \sigma_{X \times Y}, P_{XY})\) be a prob. space with associated prob. spaces \((\mathcal{X}, \sigma_X, P_X)\) and \((\mathcal{Y}, \sigma_Y, P_Y)\).

- If \(P_{XY} \ll P_X \times P_Y\) and \(\sigma_X\) is generated by a countable set then

\[
\mathcal{L}(X \rightarrow Y) = \log \int_{\mathcal{Y}} \text{ess sup}_x \left\{ \frac{dP_{XY}}{dP_X \times dP_Y}(x, y) \right\} dP_Y
\]
**Theorem (IKW '17):** Let \((\mathcal{X} \times \mathcal{Y}, \sigma_{X \times Y}, P_{XY})\) be a prob. space with associated prob. spaces \((\mathcal{X}, \sigma_X, P_X)\) and \((\mathcal{Y}, \sigma_Y, P_Y)\).

- If \(P_{XY} \ll P_X \times P_Y\) and \(\sigma_X\) is generated by a countable set then

\[
\mathcal{L}(X \to Y) = \log \int_{\mathcal{Y}} \text{ess sup}_x \left\{ \frac{dP_{XY}}{dP_X \times dP_Y}(x, y) \right\} dP_Y
\]
Theorem (IKW '17): Let \((\mathcal{X} \times \mathcal{Y}, \sigma_{X \times Y}, P_{XY})\) be a prob. space with associated prob. spaces \((\mathcal{X}, \sigma_X, P_X)\) and \((\mathcal{Y}, \sigma_Y, P_Y)\).

- If \(P_{XY} \ll P_X \times P_Y\) and \(\sigma_X\) is generated by a countable set then

\[
\mathcal{L}(X \rightarrow Y) = \log \int_{\mathcal{Y}} \text{ess sup}_x \left\{ \frac{dP_{XY}}{dP_X \times dP_Y} (x, y) \right\} dP_Y
\]

- If then

\[
\mathcal{L}(X \rightarrow Y) = \log \int_{\mathcal{Y}} \text{ess sup}_x \left\{ \frac{dP_{XY}}{dP_X \times dP_Y} (x, y) \right\} dP_Y
\]
Theorem (IKW '17): Let \((\mathcal{X} \times \mathcal{Y}, \sigma_{\mathcal{X} \times \mathcal{Y}}, P_{\mathcal{X} \mathcal{Y}})\) be a prob. space with associated prob. spaces \((\mathcal{X}, \sigma_X, P_X)\) and \((\mathcal{Y}, \sigma_Y, P_Y)\).

- If \(P_{\mathcal{X} \mathcal{Y}} \ll P_X \times P_Y\) and \(\sigma_X\) is generated by a countable set then

\[
\mathcal{L}(X \to Y) = \log \int_{\mathcal{Y}} \text{ess sup}_x \left\{ \frac{dP_{\mathcal{X} \mathcal{Y}}}{dP_X \times dP_Y}(x, y) \right\} dP_Y
\]

- If \(P_{\mathcal{X} \mathcal{Y}} \ll P_X \times P_Y\) then

\[
\mathcal{L}(X \to Y) = \infty
\]
Extension: General Alphabet
Theorem (IKW ’17): Let \((\mathcal{X} \times \mathcal{Y}, \sigma_{XY}, P_{XY})\) be a prob. space with associated prob. spaces \((\mathcal{X}, \sigma_X, P_X)\) and \((\mathcal{Y}, \sigma_Y, P_Y)\).
Theorem (IKW ’17): Let \((\mathcal{X} \times \mathcal{Y}, \sigma_{X \times Y}, P_{XY})\) be a prob. space with associated prob. spaces \((\mathcal{X}, \sigma_X, P_X)\) and \((\mathcal{Y}, \sigma_Y, P_Y)\).

- If \(P_{XY} \ll P_X \times P_Y\) and \(\sigma_X\) is generated by a countable set then

\[
\mathcal{L}(X \rightarrow Y) = \log \int_{\mathcal{Y}} \text{ess sup}_x \left\{ \frac{dP_{XY}}{dP_X \times dP_Y}(x, y) \right\} dP_Y
\]
Theorem (IKW '17): Let \((\mathcal{X} \times \mathcal{Y}, \sigma_{\mathcal{X} \times \mathcal{Y}}, P_{\mathcal{X} \mathcal{Y}})\) be a prob. space with associated prob. spaces \((\mathcal{X}, \sigma_\mathcal{X}, P_\mathcal{X})\) and \((\mathcal{Y}, \sigma_\mathcal{Y}, P_\mathcal{Y})\).

- If \(P_{\mathcal{X} \mathcal{Y}} \ll P_\mathcal{X} \times P_\mathcal{Y}\) and \(\sigma_\mathcal{X}\) is generated by a countable set then

\[
\mathcal{L}(X \to Y) = \log \int_{\mathcal{Y}} \text{ess sup}_x \left\{ \frac{dP_{\mathcal{X} \mathcal{Y}}}{dP_\mathcal{X} \times dP_\mathcal{Y}}(x, y) \right\} dP_\mathcal{Y}
\]
Theorem (IKW ’17): Let \((\mathcal{X} \times \mathcal{Y}, \sigma_{X \times Y}, P_{XY})\) be a prob. space with associated prob. spaces \((\mathcal{X}, \sigma_X, P_X)\) and \((\mathcal{Y}, \sigma_Y, P_Y)\).

- If \(P_{XY} \ll P_X \times P_Y\) and \(\sigma_X\) is generated by a countable set then

\[
\mathcal{L}(X \to Y) = \log \int_{\mathcal{Y}} \operatorname{ess sup}_x \left\{ \frac{dP_{XY}}{dP_X \times dP_Y}(x, y) \right\} dP_Y
\]
Theorem (IKW ’17): Let \((\mathcal{X} \times \mathcal{Y}, \sigma_{X \times Y}, P_{XY})\) be a prob. space with associated prob. spaces \((\mathcal{X}, \sigma_X, P_X)\) and \((\mathcal{Y}, \sigma_Y, P_Y)\).

- If \(P_{XY} \ll P_X \times P_Y\) and \(\sigma_X\) is generated by a countable set then

\[
\mathcal{L}(X \to Y) = \log \int_{\mathcal{Y}} \text{ess sup}_x \left\{ \frac{dP_{XY}}{dP_X \times dP_Y}(x, y) \right\} dP_Y
\]
Theorem (IKW ’17): Let \((\mathcal{X} \times \mathcal{Y}, \sigma_{X\times Y}, P_{XY})\) be a prob. space with associated prob. spaces \((\mathcal{X}, \sigma_X, P_X)\) and \((\mathcal{Y}, \sigma_Y, P_Y)\).

- If \(P_{XY} \ll P_X \times P_Y\) and \(\sigma_X\) is generated by a countable set then

\[
\mathcal{L}(X \rightarrow Y) = \log \int_{\mathcal{Y}} \text{ess sup}_x \left\{ \frac{dP_{XY}}{dP_X \times dP_Y}(x, y) \right\} dP_Y
\]
Theorem (IKW ’17): Let \((\mathcal{X} \times \mathcal{Y}, \sigma_{\mathcal{X} \times \mathcal{Y}}, P_{XY})\) be a prob. space with associated prob. spaces \((\mathcal{X}, \sigma_X, P_X)\) and \((\mathcal{Y}, \sigma_Y, P_Y)\).

- If \(P_{XY} \ll P_X \times P_Y\) and \(\sigma_X\) is generated by a countable set then

\[
\mathcal{L}(X \to Y) = \log \int_{\mathcal{Y}} \text{ess sup}_x \left\{ \frac{dP_{XY}}{dP_X \times dP_Y}(x, y) \right\} dP_Y
\]

- If \(\ldots\) then \(\ldots\).
Theorem (IKW ’17): Let \((\mathcal{X} \times \mathcal{Y}, \sigma_{\mathcal{X} \times \mathcal{Y}}, P_{XY})\) be a prob. space with associated prob. spaces \((\mathcal{X}, \sigma_{\mathcal{X}}, P_X)\) and \((\mathcal{Y}, \sigma_{\mathcal{Y}}, P_Y)\).

- If \(P_{XY} \ll P_X \times P_Y\) and \(\sigma_{\mathcal{X}}\) is generated by a countable set then

\[
\mathcal{L}(X \rightarrow Y) = \log \int \text{ess sup}_{x} \left\{ \frac{dP_{XY}}{dP_{X} \times dP_{Y}}(x, y) \right\} dP_{Y}
\]

- If \(P_{XY} \ll P_X \times P_Y\) then

\[
\mathcal{L}(X \rightarrow Y) = \infty
\]