Fundamental Limits on Security and Privacy of Information Sources

Vince Poor
(poor@princeton.edu)

Supported in part by the U.S. National Science Foundation under Grant CMMI-1435778.
Overview

Two topics:

- **Secrecy** in wireless data transmission

- **Privacy** of information sources, with applications in smart grid

Common theme:

- **Information theoretic** characterization of fundamental limits
Outline

1. Physical Layer Security in Wireless Networks

2. Privacy-Utility Tradeoffs, with Applications in Smart Grid

3. Summary
Physical Layer Security
in
Wireless Networks

Joint work with Yingbin Liang, Shlomo Shamai, et al.
Wireless Networks: Layers

Application (APP) → Web Browsing, Voice, etc.

Network (NET) → Routing, Flow Control, etc.

Medium Access Control (MAC) → Scheduling, Access Control, etc.

Physical (PHY) → Data Transmission
Motivation: Exploiting the Physical Layer

- Key Techniques for Improving **Capacity & Reliability:**
  - MIMO (Multiple-Antenna Systems)
  - Cooperation & Relaying
  - Cognitive Radio
Motivation: Exploiting the Physical Layer

- **Key Techniques for Improving Capacity & Reliability:**
  - MIMO (Multiple-Antenna Systems)
  - Cooperation & Relaying
  - Cognitive Radio

- **What About Security?**
  - Traditionally a higher-layer issue (e.g., APP)
  - Encryption can be complex and difficult without infrastructure
  - Information theoretic security examines the fundamental ability of the PHY to provide security (confidentiality)
Shannon [1949]: For cipher, perfect secrecy requires a one-time pad.

[I.e., the entropy of the key must be at least the entropy of the source: $H(K) \geq H(M)$]
Information Theoretic Secrecy: Wyner’s Model

“The Wiretap Channel”

- Tradeoff: reliable rate $R$ to Bob vs. the equivocation $H(M|Z)$ at Eve
Information Theoretic Secrecy: Wyner’s Model

“The Wiretap Channel”

- Tradeoff: reliable rate $R$ to Bob vs. the equivocation $H(M|Z)$ at Eve
- Secrecy capacity = maximum $R$ such that $R = H(M|Z)$
Information Theoretic Secrecy: Wyner’s Model

“The Wiretap Channel”

Message $M$ → Alice $X$ → Noisy Channel $Y$ → Bob $\hat{M}$

Eve $Z$ → Noisy Channel

- Tradeoff: **reliable rate** $R$ to Bob vs. the **equivocation** $H(M|Z)$ at Eve
- **Secrecy capacity** = maximum $R$ such that $R = H(M|Z)$
- **Wyner** [1975]: Secrecy capacity $> 0$ iff. $Z$ is **degraded** relative to $Y$
Physical Layer Security in Wireless Networks

- There has been a resurgence of interest in Wyner’s ideas, as encryption is impractical for emerging wireless networking paradigms.
Physical Layer Security in Wireless Networks

• There has been a resurgence of interest in Wyner’s ideas, as encryption is impractical for emerging wireless networking paradigms.

• The physical properties of radio propagation (diffusion & superposition) provide opportunities for this, via

  – fading: provides natural degradedness over time
  – interference: allows active countermeasures to eavesdropping
  – spatial diversity (MIMO, relays): creates “secrecy degrees of freedom”
Physical Layer Security in Wireless Networks

• There has been a resurgence of interest in Wyner’s ideas, as encryption is impractical for emerging wireless networking paradigms.

• The physical properties of radio propagation (diffusion & superposition) provide opportunities for this, via

  - fading: provides natural degradedness over time
  - interference: allows active countermeasures to eavesdropping
  - spatial diversity (MIMO, relays): creates “secrecy degrees of freedom”

• These phenomena lead to rich secrecy capacity regions for the fundamental channel models used to understand wireless networks.
Paradigm: Broadcast Channel with Confidential (BCC) Messages

Models content distribution with multicast and unicast content

- **Csiszár & Körner** [1978]: Discrete Memoryless BCC
- **Liang, Poor & Shamai** [2008]: Gaussian & Fading BCCs
Gaussian BCC: Secrecy Capacity Regions
Fading BCC: Secrecy Capacity Region

\[
\begin{align*}
R_1 &\sim \mu^2 = \nu^2 = 1 \\
P &\sim 5 \text{dB} \\
|h_1|^2 &\sim e^{-x} \\
|h_2|^2 &\sim \frac{1}{\sigma_2^2} e^{-x/\sigma_2}
\end{align*}
\]

Decreasing $\sigma_2$
Fading BCC: Secrecy Capacity Region
Secrecy in Fundamental Channel Models

- **Multiple-Access Channel:**
  - Message $M_1 \rightarrow Alice_1/Eve_2$
    - $X_1 \rightarrow Bob \rightarrow \hat{M}_1, \hat{M}_2$
  - Message $M_2 \rightarrow Alice_2/Eve_1$
    - $X_2 \rightarrow Bob \rightarrow \hat{M}_1, \hat{M}_2$

- **Interference Channel:**
  - Message $M_1 \rightarrow Alice_1$
    - $X_1 \rightarrow Bob_1/Eve_2 \rightarrow \hat{M}_1, ?$
  - Message $M_2 \rightarrow Alice_2$
    - $X_2 \rightarrow Bob_2/Eve_1 \rightarrow \hat{M}_2, ?$

- **Relay Channel:** Relay cooperates to improve security; or relay is untrusted.

- **MIMO Channel:** Allows simultaneous secure transmission without rate penalty.
A Rich Area

- Coding Theory
  - code design
- Cryptography
  - key generation & management
- Networking
  - cross-layer design
- Game Theory
  - adversarial model

Information Theoretic Security
(feedback, side info, etc.)
Privacy-Utility Tradeoffs

with

Applications in Smart Grid

Joint work with Lalitha Sankar, et al.
Motivation: Privacy & Utility of Data

- There are many **electronic information sources** of information about us.
  - Google, Facebook, smart metering, biometric systems, etc.
Motivation: Privacy & Utility of Data

- There are many *electronic information sources* of information about us.
  - Google, Facebook, smart metering, biometric systems, etc.

- The *utility* of these sources depends on their accessibility.
Motivation: Privacy & Utility of Data

• There are many electronic information sources of information about us.
  – Google, Facebook, smart metering, biometric systems, etc.

• The utility of these sources depends on their accessibility.

• But, they can also leak private information.
There are many electronic information sources of information about us.  
- Google, Facebook, smart metering, biometric systems, etc.

The utility of these sources depends on their accessibility.

But, they can also leak private information.

How can we characterize this fundamental tradeoff?
Privacy vs. Secrecy

- Privacy is **not** secrecy:
Privacy vs. Secrecy

- Privacy is *not* secrecy:

- Denial of access (secrecy) makes a data source *useless*. 
A database is a table – rows: individual entries (total of $n$); columns: attributes for each individual (total of $K$).

### Attributes

<table>
<thead>
<tr>
<th>Gender</th>
<th>Visit Date</th>
<th>Diagnosis</th>
<th>...</th>
<th>Medication</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Numeric and non-numeric data

**Query**

**Response**

**User**
• Database with $n$ rows is a sequence of $n$ independent observations of a random vector $\mathbf{X} = (X_1, X_2, \ldots, X_K)$ with a given probability distribution.

• Attributes divided into public (revealed) and private (hidden) variables, possibly not disjoint:

$$k^{th} \text{ entry: } \mathbf{X}_k = (X_{r,k}, X_{h,k})$$
Privacy-Utility Tradeoff

• How can we characterize the tradeoff between utility and privacy in such a setting?
Privacy-Utility Tradeoff

• How can we characterize the tradeoff between utility and privacy in such a setting?
  
  – Measure utility by distortion of the public variables as revealed to a user of the database; and
Privacy-Utility Tradeoff

• How can we characterize the tradeoff between utility and privacy in such a setting?

  – Measure utility by distortion of the public variables as revealed to a user of the database; and

  – Measure privacy by equivocation on the private variables in information revealed to a user.
Privacy-Utility Tradeoff

• How can we characterize the tradeoff between utility and privacy in such a setting?
  
  – Measure utility by distortion of the public variables as revealed to a user of the database; and
  
  – Measure privacy by equivocation on the private variables in information revealed to a user.

• The distortion-equivocation region describes the tradeoff.
Distortion-Equivocation Model

- Encoder maps the original database to a “sanitized” database (SDB):

\[
\text{Encoder} : X^n \rightarrow \mathcal{W} = \{SDB_1, SDB_2, \ldots, SDB_M\}
\]
Distortion-Equivocation Model

• Encoder maps the original database to a “sanitized” database (SDB):

\[
\text{Encoder} : X^n \rightarrow \mathcal{W} = \{SDB_1, SDB_2, \ldots, SDB_M\}
\]
Distortion-Equivocation Model

- Encoder maps the original database to a “sanitized” database (SDB):

\[
\text{Encoder} : X^n \rightarrow \mathcal{W} = \{SDB_1, SDB_2, \ldots, SDB_M\}
\]

Distortion

Distortion between \(\{X_{r,k}\}\) and \(\{\hat{X}_{r,k}\}\) \(\leq D\)
Distortion-Equivocation Model

- Encoder maps the original database to a “sanitized” database (SDB):

  \[
  \text{Encoder} : X^n \rightarrow \mathcal{W} = \{SDB_1, SDB_2, \ldots, SDB_M\}
  \]

**Distortion**

Distortion between \(\{X_{r,k}\}\) and \(\{\hat{X}_{r,k}\}\) ≤ \(D\)

**Equivocation**

Entropy of \(\{X_{h,k}\}\) given \(W \geq E\)
Distortion-Equivocation Model

- Encoder maps the original database to a “sanitized” database (SDB):

Encoder: \( X^n \rightarrow \mathcal{W} = \{SDB_1, SDB_2, \ldots, SDB_M\} \)

Distortion

Distortion between \( \{X_{r,k}\} \) and \( \{\hat{X}_{r,k}\} \) ≤ \( D \)

Equivocation

Entropy of \( \{X_{h,k}\} \) given \( W \geq E \)

Add a rate constraint

\[ \frac{\log M}{n} \leq R \]
Utility-Privacy/RDE Regions

(a): Rate-Distortion-Equivocation Region

(b): Utility-Privacy Tradeoff Region
Application: Smart Meter Privacy

• Smart meter data is useful for price-aware usage, load balancing

• But, it leaks information about in-home activity
P-U tradeoff leads to a spectral \textit{`reverse water-filling'} solution
Source Coding Solution

P-U tradeoff leads to a spectral ‘reverse water-filling’ solution

Can also use energy storage to aid privacy – results in a control-theoretic solution [Tan-Gunduz-Poor, 2013] [Yang-Chen-Zhang-Poor, 2015]
Competitive Privacy: Motivating Example

- N.A. Grid: interconnected regional transmission organizations (RTOs) which
  - need to share state measurements for **reliability of state estimation** (utility)
  - wish to withhold information for **economic competitiveness** (privacy)
Competitive Privacy: Motivating Example

- N.A. Grid: interconnected regional transmission organizations (RTOs) which
  - need to share state measurements for reliability of state estimation (utility)
  - wish to withhold information for economic competitiveness (privacy)

- Leads to a problem of competitive privacy
Competitive Privacy Model

- Noisy measurements at RTO $k$:

$$Y_k = \sum_{m=1}^{M} H_{k,m} X_m + Z_k, \ k = 1,2,\ldots,M$$

$m^{th}$ system state
Competitive Privacy Model

• Noisy measurements at RTO $k$:

$$Y_k = \sum_{m=1}^{M} H_{k,m} X_m + Z_k, \; k = 1,2,\ldots,M$$

• Utility for RTO $k$: mean-square error for its own state $X_k$
Competitive Privacy Model

• Noisy measurements at RTO $k$:

\[ Y_k = \sum_{m=1}^{M} H_{k,m} X_m + Z_k, \quad k = 1, 2, \ldots, M \]

m\textsuperscript{th} system state

• Utility for RTO $k$: mean-square error for its own state $X_k$

• Privacy for RTO $k$: leakage of information about $X_k$ to other RTOs
Competitive Privacy Model

• Noisy measurements at RTO $k$:

\[
Y_k = \sum_{m=1}^{M} H_{k,m} X_m + Z_k, \quad k = 1, 2, \ldots, M
\]

• Utility for RTO $k$: mean-square error for its own state $X_k$

• Privacy for RTO $k$: leakage of information about $X_k$ to other RTOs

• Wyner-Ziv coding maximizes privacy for a desired utility at each agent.
Competitive Privacy Model

- Noisy measurements at RTO $k$:

\[ Y_k = \sum_{m=1}^{M} H_{k,m} X_m + Z_k, \ k = 1,2,\ldots,M \]

- Utility for RTO $k$: mean-square error for its own state $X_k$

- Privacy for RTO $k$: leakage of information about $X_k$ to other RTOs

- Wyner-Ziv coding maximizes privacy for a desired utility at each agent.

- But, how much to share? We can study this issue via game theory.
Other Potential Applications

Biometric Systems: Tradeoff between security & privacy
Other Potential Applications

Biometric Systems:
Tradeoff between security & privacy

E-Commerce:
Tradeoff between profit & privacy
Summary

• Information theory can help understand the fundamental ability of the radio channel to provide confidentiality of wireless data.
Summary

- **Information theory** can help understand the fundamental ability of the **radio channel** to provide **confidentiality** of wireless data.

- A **fundamental tradeoff** between **privacy** and **utility** of data sources can also be viewed in an **information theoretic** setting.
Summary

• **Information theory** can help understand the fundamental ability of the *radio channel* to provide *confidentiality* of wireless data.

• A *fundamental tradeoff* between *privacy* and *utility* of data sources can also be viewed in an *information theoretic* setting.

• Examples from *smart grid*: *smart metering* and *competitive privacy* give rise to tradeoffs between *fidelity* and *information leakage*.
Summary

• Information theory can help understand the fundamental ability of the radio channel to provide confidentiality of wireless data.

• A fundamental tradeoff between privacy and utility of data sources can also be viewed in an information theoretic setting.

• Examples from smart grid: smart metering and competitive privacy give rise to tradeoffs between fidelity and information leakage.

• These are theoretical constructs, but they point to potential practical solutions.
Thank You!