Information Theory and Secure Communications

Øyvind Ytrehus

Simula UiB

Finse, April 25, 2022

Outline

Outline

Cast: Main characters I

- *Alice*  , a talkative sender.

Cast: Main characters I

- *Alice*  , a talkative sender.
- *Bob*  , an eager listener.

What do Alice and Bob want?

Alice wants to send messages to Bob...

- reliably and efficiently

What do Alice and Bob want?

Alice wants to send messages to Bob...

- reliably and efficiently
    - Information theory (What can be achieved)
    - Coding theory (How to achieve it)
    - Communication theory (Physical implementation adapted to channel)

What do Alice and Bob want?

Alice wants to send messages to Bob...

- reliably and efficiently
  - Information theory (What can be achieved)
  - Coding theory (How to achieve it)
  - Communication theory (Physical implementation adapted to channel)
- securely: secretly,privately, authenticated, stealthily

What do Alice and Bob want?

Alice wants to send messages to Bob...

- reliably and efficiently
  - Information theory (What can be achieved)
  - Coding theory (How to achieve it)
  - Communication theory (Physical implementation adapted to channel)
- securely: secretly,privately, authenticated, stealthily
  - Cryptography

What do Alice and Bob want?

Alice wants to send messages to Bob...

- reliably and efficiently
  - Information theory (What can be achieved)
  - Coding theory (How to achieve it)
  - Communication theory (Physical implementation adapted to channel)
- securely: secretly,privately, authenticated, stealthily
  - Cryptography
  - Information Theory
  - Other security techniques
  - Anonymizing networks, Private Information Retrieval

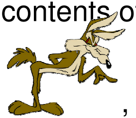
Cast: Main characters II: Adversaries

- *Eve*  , a nosy eavesdropper who wishes to listen *passively* to the contents of the messages from Alice to Bob.

Cast: Main characters II: Adversaries

- *Eve*  , a nosy eavesdropper who wishes to listen *passively* to the contents of the messages from Alice to Bob.
- *Willie*  , a wiley warden who wishes to determine with precision whether at all Alice transmits to Bob. Willie does not care about the content of transmitted messages.

Cast: Main characters II: Adversaries

- *Eve*  , a nosy eavesdropper who wishes to listen *passively* to the contents of the messages from Alice to Bob.
- *Willie*  , a wiley warden who wishes to determine with precision whether at all Alice transmits to Bob. Willie does not care about the content of transmitted messages.
- Fraudsters, impositors, active intruders, repudiators also exist...but beyond the scope

Requirements for secure effective communication (between Alice and Bob)

Bob needs some advantage over the adversaries:

- A secret key (cryptography)
- Other types of knowledge

Requirements for secure effective communication (between Alice and Bob)

Bob needs some advantage over the adversaries:

- A secret key (cryptography)
- Other types of knowledge
- Biometrics

Requirements for secure effective communication (between Alice and Bob)

Bob needs some advantage over the adversaries:

- A secret key (cryptography)
- Other types of knowledge
- Biometrics
- Computation capabilities

Requirements for secure effective communication (between Alice and Bob)

Bob needs some advantage over the adversaries:

- A secret key (cryptography)
- Other types of knowledge
- Biometrics
- Computation capabilities
- Better communication channel: Information theory
  - alternative/addition to cryptography
  - also applied to metadata
  - applied in 5G

Outline

Information theory and noisy channels

- What is Information?

Information theory and noisy channels

- What is Information?
- Discrete stochastic variable (DSV) X, possible outcomes $\mathcal{X}$ distributed as $p(x)$. The *entropy* of X, measured in bits, is

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log_2 p(x)$$

Information theory and noisy channels

- What is Information?
- Discrete stochastic variable (DSV) X, possible outcomes $\mathcal{X}$ distributed as $p(x)$. The *entropy* of X, measured in bits, is

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log_2 p(x) = H(p()). \quad (1)$$

If $p(x)$ is the uniform distribution on $\mathcal{X}$, $H(X) = \log_2(|\mathcal{X}|)$.

Information theory and noisy channels

- What is Information?
- Discrete stochastic variable (DSV) X, possible outcomes $\mathcal{X}$ distributed as $p(x)$. The *entropy* of X, measured in bits, is

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log_2 p(x) = H(p()). \tag{1}$$

  If $p(x)$ is the uniform distribution on $\mathcal{X}$, $H(X) = \log_2(|\mathcal{X}|)$.
- Consider *two* DSVs X and Y, set of joint outcomes $\mathcal{X} \times \mathcal{Y}$ distributed as $p(x, y)$. Then the *equivocation* is

Information theory and noisy channels

- What is Information?
- Discrete stochastic variable (DSV) X, possible outcomes $\mathcal{X}$ distributed as $p(x)$. The *entropy* of X, measured in bits, is

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log_2 p(x) = H(p()). \tag{1}$$

  If $p(x)$ is the uniform distribution on $\mathcal{X}$, $H(X) = \log_2(|\mathcal{X}|)$.
- Consider *two* DSVs X and Y, set of joint outcomes $\mathcal{X} \times \mathcal{Y}$ distributed as $p(x, y)$. Then the *equivocation* is

$$H(X|Y) = \sum_{y \in \mathcal{Y}} p(y) \Big(-\sum_{x \in \mathcal{X}} p(x|y) \log_2 p(x|y) \Big)$$

- The *mutual information* is

$$I(X;Y) = I(Y;X) = H(X) - H(X|Y) = H(Y) - H(Y|X). \tag{2}$$

Some remarks on information theory

- Information and “Information”

Some remarks on information theory

- Information and “Information”
- Information versus entropy

Some remarks on information theory

- Information and “Information”
- Information versus entropy
- Information theory versus probability theory

Some remarks on information theory

- Information and “Information”
- Information versus entropy
- Information theory versus probability theory
- Information versus computation

Some remarks on information theory

- Information and “Information”
- Information versus entropy
- Information theory versus probability theory
- Information versus computation
- The “Bandwagon” - good and bad application areas

Some remarks on information theory

- Information and “Information”
- Information versus entropy
- Information theory versus probability theory
- Information versus computation
- The “Bandwagon” - good and bad application areas
    - “Easy to apply”: Digital communications, Experiment design, Compressed sensing

Some remarks on information theory

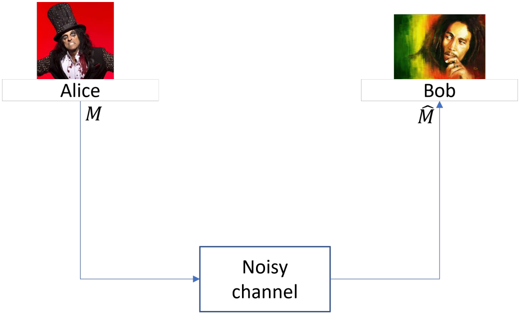
- Information and “Information”
- Information versus entropy
- Information theory versus probability theory
- Information versus computation
- The “Bandwagon” - good and bad application areas
  - “Easy to apply”: Digital communications, Experiment design, Compressed sensing
  - “Hard to apply”: Biology? Medicine? Linguistics? Social Sciences?

Some remarks on information theory

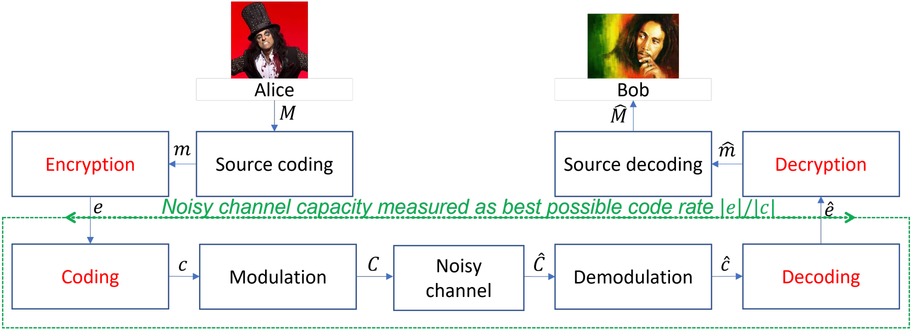
- Information and “Information”
- Information versus entropy
- Information theory versus probability theory
- Information versus computation
- The “Bandwagon” - good and bad application areas
    - “Easy to apply”: Digital communications, Experiment design, Compressed sensing
    - “Hard to apply”: Biology? Medicine? Linguistics? Social Sciences?
- Information versus psychology

Outline

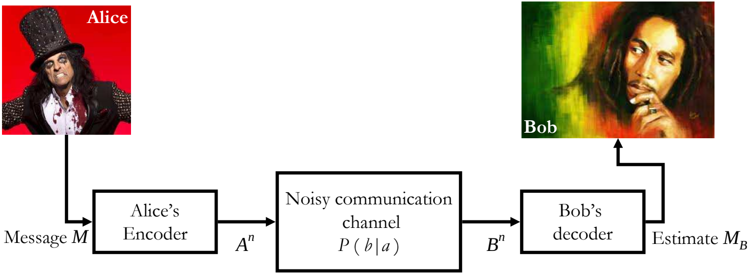
Communication between Alice and Bob



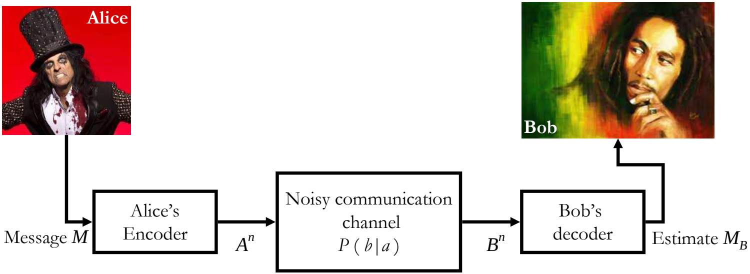
Communication between Alice and Bob



Shannon’s noisy channel



Shannon's noisy channel



$$C_{Shannon} = \max_{p(a)} I(A;B),$$

Shannon's noisy channel: Tools used in proof

- Typical sequences

Shannon's noisy channel: Tools used in proof

- Typical sequences of length n

Shannon's noisy channel: Tools used in proof

- Typical sequences of length n
  $\underline{x}$ typical iff $freq(\underline{x}) \approx p(x)$

Shannon's noisy channel: Tools used in proof

- Typical sequences of length n
  $\underline{x}$ typical iff $freq(\underline{x}) \approx p(x) \Rightarrow p(\underline{x}) \approx 2^{-nH(x)}$

Shannon's noisy channel: Tools used in proof

- Typical sequences of length n
  $\underline{x}$ typical iff $freq(\underline{x}) \approx p(x) \Rightarrow p(\underline{x}) \approx 2^{-nH(x)}$
- Jointly typical sequences

Shannon's noisy channel: Tools used in proof

- Typical sequences of length n
  $\underline{x}$ typical iff $freq(\underline{x}) \approx p(x) \Rightarrow p(\underline{x}) \approx 2^{-nH(x)}$
- Jointly typical sequences
  $\underline{a}$ typical, $\underline{b} \sim P(\underline{b}|\underline{a})$:

Shannon's noisy channel: Tools used in proof

- Typical sequences of length n
  $\underline{x}$ typical iff $freq(\underline{x}) \approx p(x) \Rightarrow p(\underline{x}) \approx 2^{-nH(x)}$
- Jointly typical sequences
  $\underline{a}$ typical, $\underline{b} \sim P(\underline{b}|\underline{a})$: $P((\underline{a}, \underline{b})$ typical$) \approx 1$

Shannon's noisy channel: Tools used in proof

- Typical sequences of length n
  $\underline{x}$ typical iff $freq(\underline{x}) \approx p(x) \Rightarrow p(\underline{x}) \approx 2^{-nH(x)}$
- Jointly typical sequences
  $\underline{a}$ typical, $\underline{b} \sim P(\underline{b}|\underline{a})$: $P((\underline{a}, \underline{b})$ typical$) \approx 1$
  $\underline{a}$ typical, $\underline{b}$ typical :

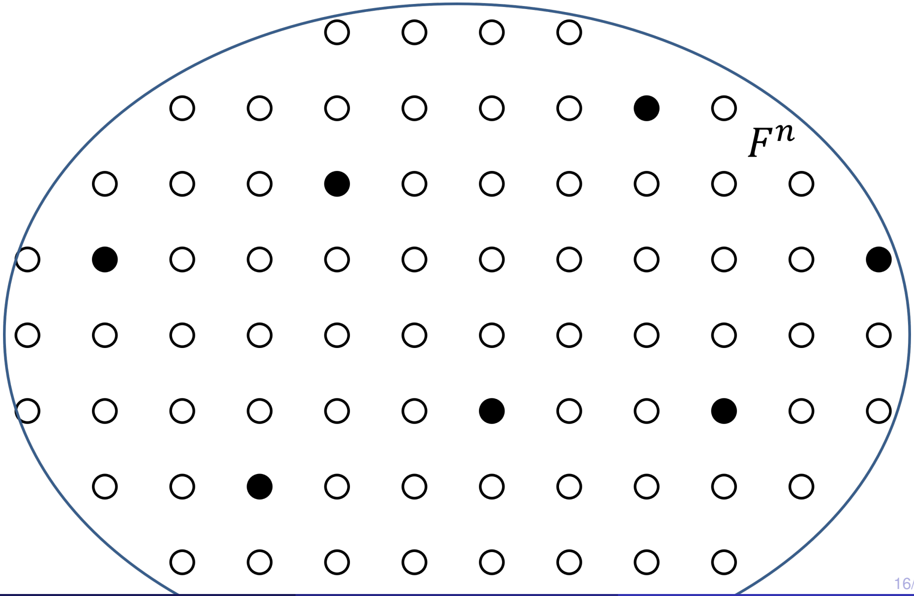
Shannon's noisy channel: Tools used in proof

- Typical sequences of length n
  $\underline{x}$ typical iff $freq(\underline{x}) \approx p(x) \Rightarrow p(\underline{x}) \approx 2^{-nH(x)}$
- Jointly typical sequences
  $\underline{a}$ typical, $\underline{b} \sim P(\underline{b}|\underline{a})$: $P((\underline{a}, \underline{b})$ typical$) \approx 1$
  $\underline{a}$ typical, $\underline{b}$ typical : $P($ random $\underline{a}, \underline{b})$ typical$) \lesssim 2^{-(nI(a;b))})$

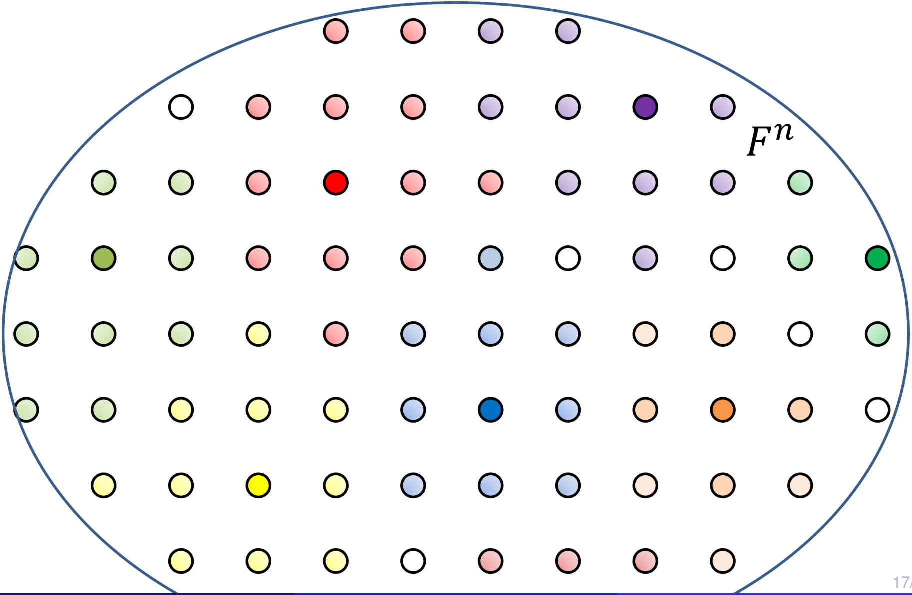
Shannon's noisy channel: Tools used in proof

- Typical sequences of length n
  $\underline{x}$ typical iff $freq(\underline{x}) \approx p(x) \Rightarrow p(\underline{x}) \approx 2^{-nH(x)}$
- Jointly typical sequences
  $\underline{a}$ typical, $\underline{b} \sim P(\underline{b}|\underline{a})$: $P((\underline{a}, \underline{b})$ typical$) \approx 1$
  $\underline{a}$ typical, $\underline{b}$ typical : $P($ random $\underline{a}, \underline{b})$ typical$) \lesssim 2^{-(nI(a;b))})$
- Random coding

Error correcting code

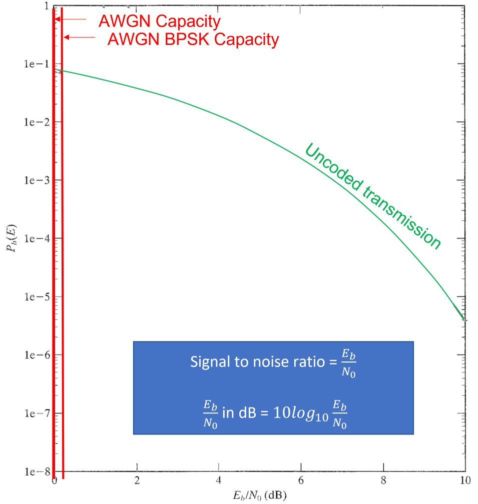


Error correcting code partitioning the space F^n



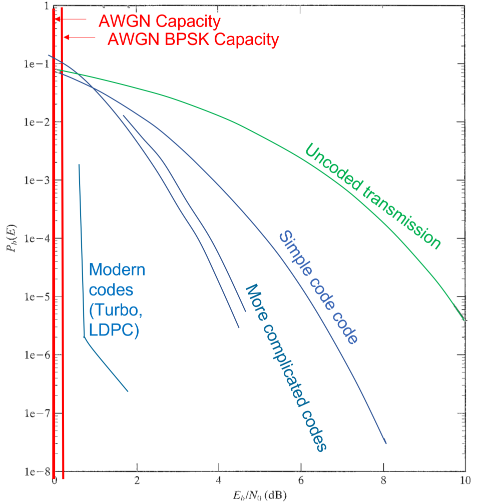
The Additive White Gaussian Noise (AWGN) channel

How?



The Additive White Gaussian Noise (AWGN) channel

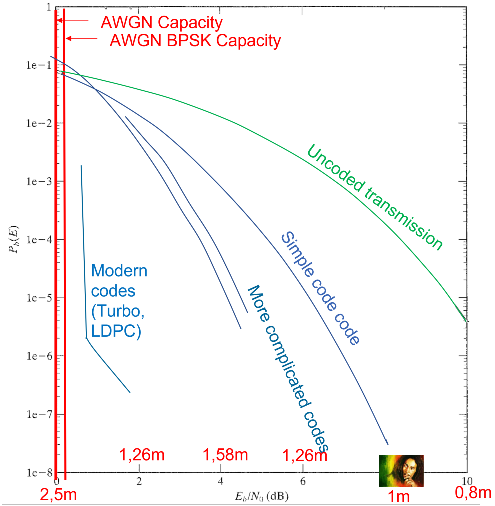
How?



The Additive White Gaussian Noise (AWGN) channel

How?

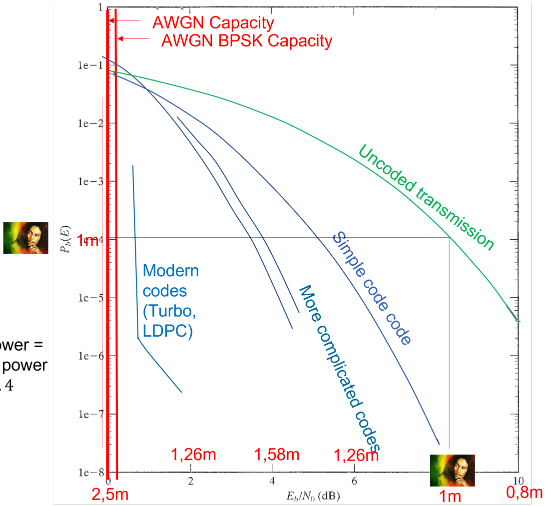
Received power = Transmitted power $\times d^{\alpha}, \alpha = 2..4$



The Additive White Gaussian Noise (AWGN) channel

How?

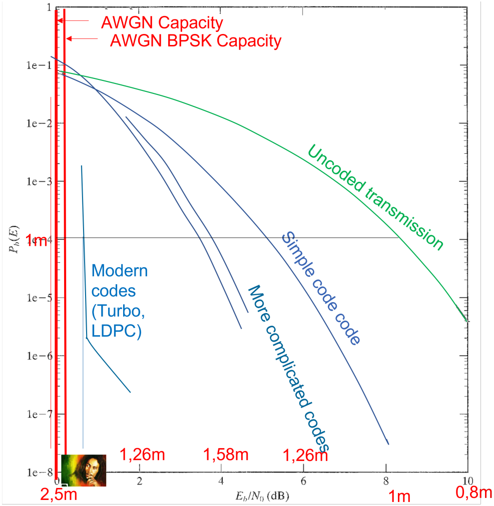
Received power = Transmitted power $\times d^{\alpha}, \alpha = 2..4$



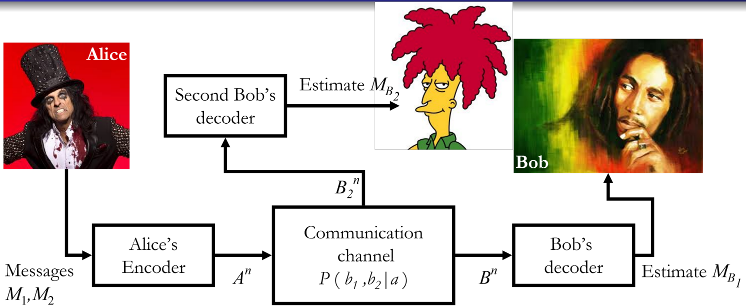
The Additive White Gaussian Noise (AWGN) channel

How?

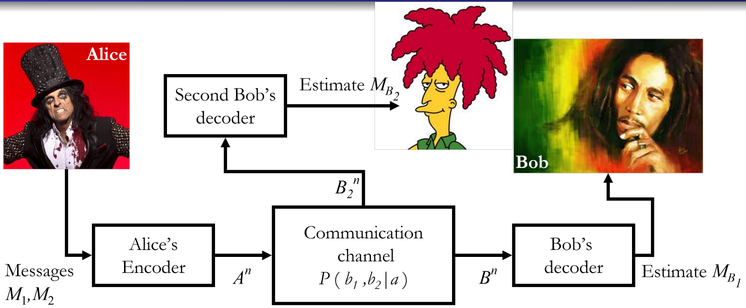
Received power = Transmitted power $\times d^{\alpha}, \alpha = 2..4$



The Broadcast Channel



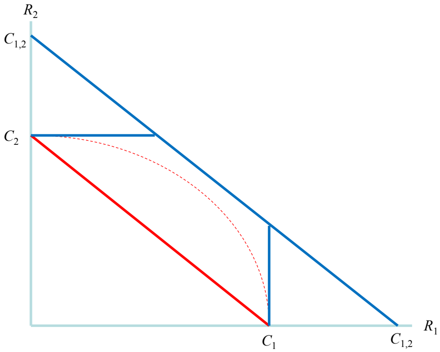
The Broadcast Channel



$$R_1 \leq C_1 = \max_{p_a} I(A; B_1), \; R_2 \leq C_2 = \max_{p_a} I(A; B_2),$$
$$R_1 + R_2 \leq C_{1,2} = \max_{p_a} I(A; B_1, B_2),$$

The Broadcast Channel

$$R_1 \leq C_1 = \max_{p_a} I(A;B_1), R_2 \leq C_2 = \max_{p_a} I(A;B_2),$$
$$R_1 + R_2 \leq C_{1,2} = \max_{p_a} I(A;B_1,B_2),$$



The broadcast channel: Sketch of proof, and example

- Jointly typical sequences

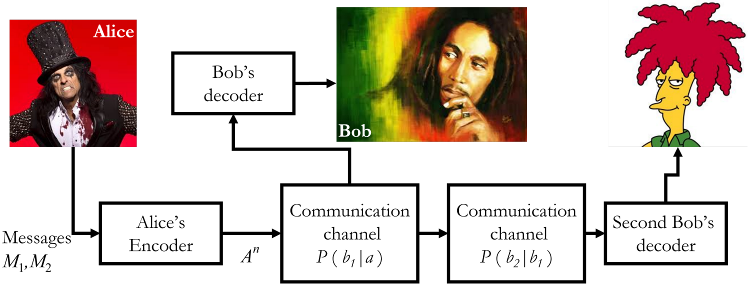
The broadcast channel: Sketch of proof, and example

- Jointly typical sequences
- Superposition coding

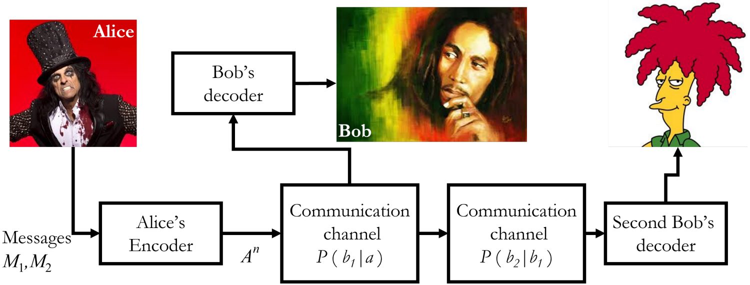
The broadcast channel: Sketch of proof, and example

- Jointly typical sequences
- Superposition coding

The Broadcast Channel, Degraded



The Broadcast Channel, Degraded



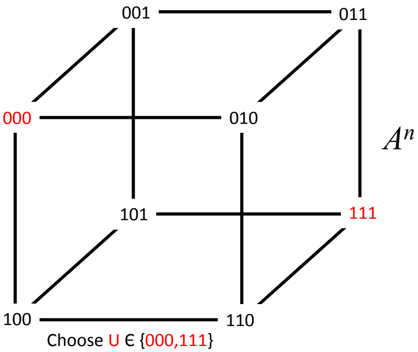
$$R_1 \leq I(A; B_1 | U),$$

$$R_2 \leq I(U; B_2),$$

for some pmf $p(u, a)$ and conditions on U.

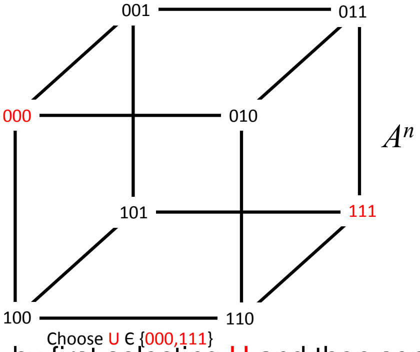
What does this mean? What is U?

Assume $n = 3$ and that Bob_1 has an error free channel, while Bob_2 typically will see at most one bit error for each 3 sent.



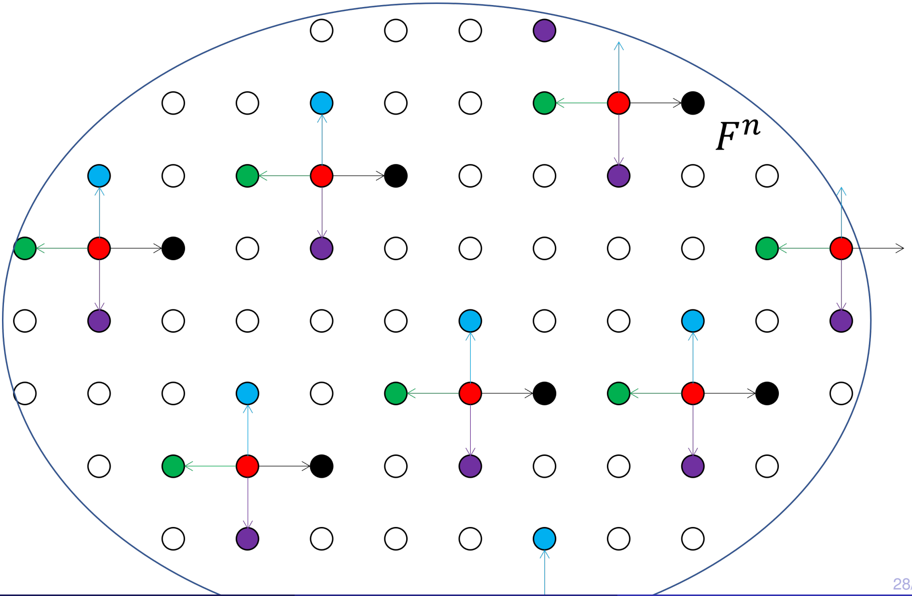
What does this mean? What is U?

Assume $n = 3$ and that Bob_1 has an error free channel, while Bob_2 typically will see at most one bit error for each 3 sent.

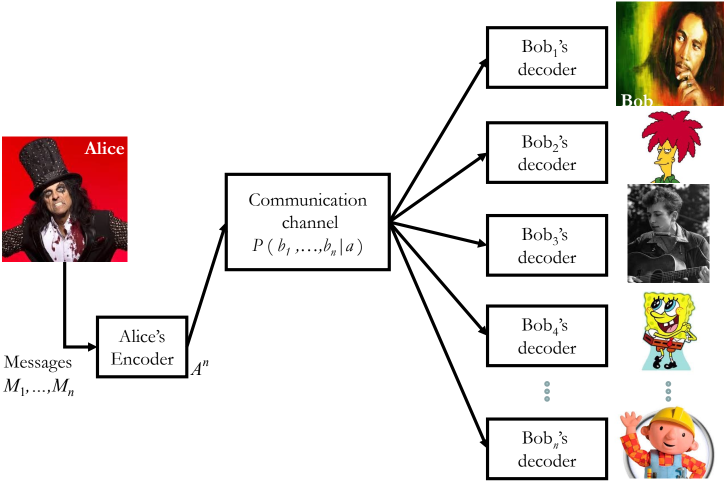


Then send three bits by first selecting U and then sending A as U plus one of the bit patterns $\{000, 001, 010, 100\}$.

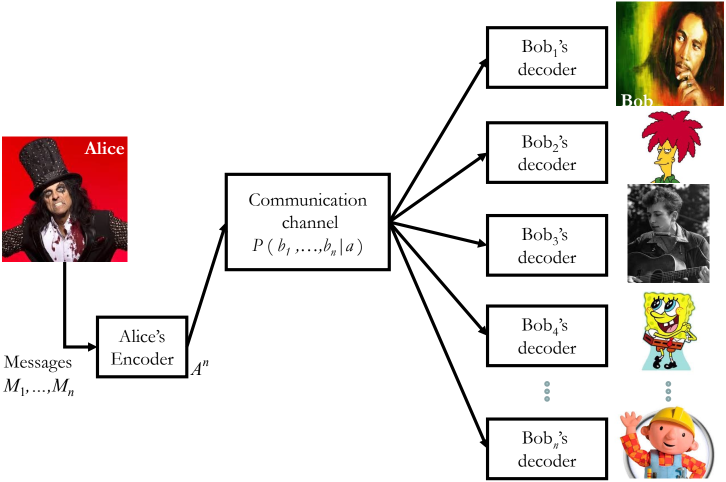
Error correcting code and coset



The Broadcast Channel, Generalized

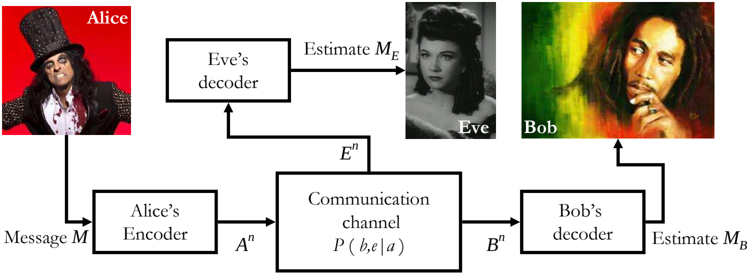


The Broadcast Channel, Generalized

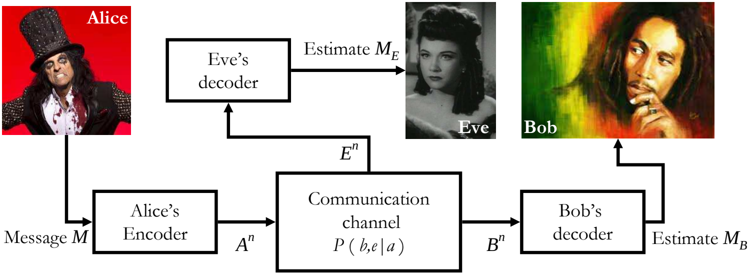


Outline

The wiretap channel (Type I)

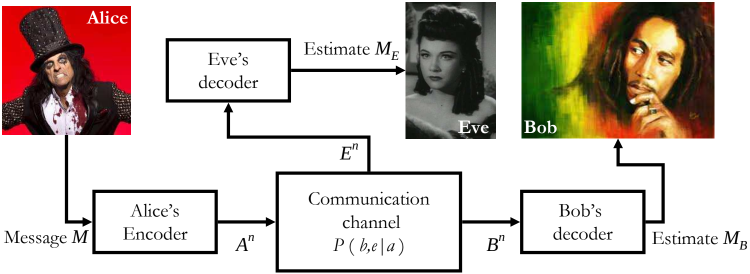


The wiretap channel (Type I)



$$C_{DM-WTC} = \max\{0, \max_{p(u,m)} (I(U;B) - I(U;E))\},$$

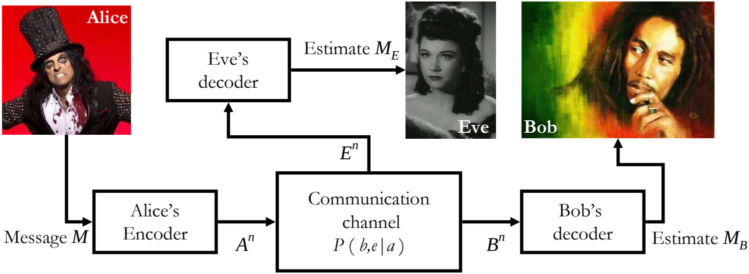
The wiretap channel (Type I)



$$C_{DM-WTC} = \max\{0, \max_{p(u,m)} (I(U;B) - I(U;E))\},$$

$$C_{DM-WTC-degr} = \max_{p(m)} (I(M;B) - I(M;E)),$$

The wiretap channel (Type I)



$$C_{DM-WTC} = \max\{0, \max_{p(u,m)} (I(U;B) - I(U;E))\},$$

$$C_{DM-WTC-degr} = \max_{p(m)} (I(M;B) - I(M;E)),$$

$$C_{DM-WTC-K} = \max_{p(m)} \min(I(M;B) - I(M;E) + R_K, I(M;B)).$$

The wiretap channel (Type I), Simple case

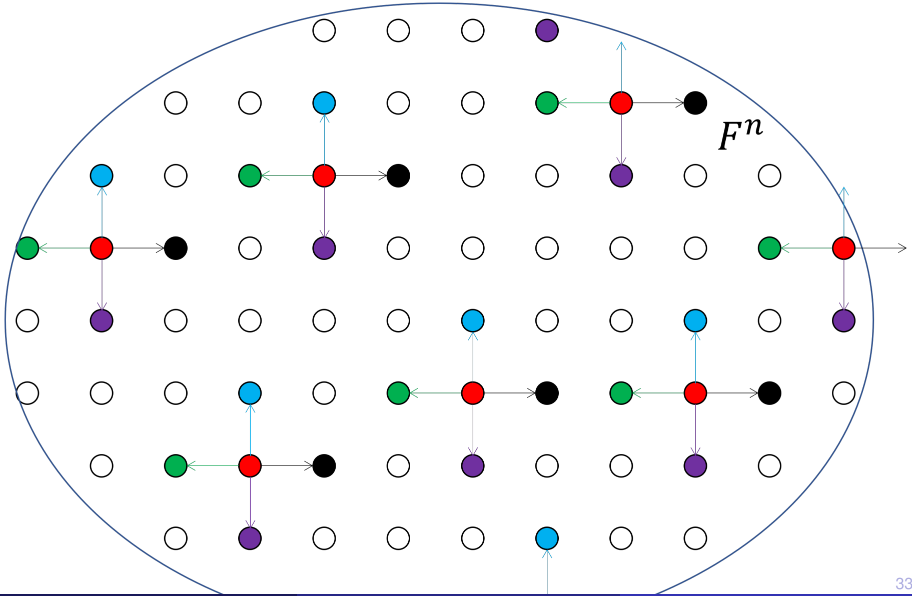
Simplest example: Degraded Bob's channel is noiseless, Eve's channel has noise. How to encode?

The wiretap channel (Type I), Simple case

Simplest example: Degraded Bob's channel is noiseless, Eve's channel has noise. How to encode?

- select $[n, k]$ error correcting code according to $(n-k)/n < C_{DM-WTC}$
- Represent message by a *coset*
- Alice sends U = *random* codeword + corresponding coset leader

Error correcting code and coset



Degrees of Secrecy

Encoded message M is n-bit vector, Eve sees noisy n-bit E^n, n large

- Perfect secrecy: $I(M; E^n) = 0$

Degrees of Secrecy

Encoded message M is n-bit vector, Eve sees noisy n-bit E^n, n large

- Perfect secrecy: $I(M; E^n) = 0$
- Strong secrecy: $I(M; E^n) \leq \epsilon$ for some small $\epsilon > 0$

Degrees of Secrecy

Encoded message M is n-bit vector, Eve sees noisy n-bit E^n, n large

- Perfect secrecy: $I(M; E^n) = 0$
- Strong secrecy: $I(M; E^n) \leq \epsilon$ for some small $\epsilon > 0$
- Weak secrecy: $I(M; E^n)/n \leq \epsilon$ for some small $\epsilon > 0$

Degrees of Secrecy

Encoded message M is n-bit vector, Eve sees noisy n-bit E^n, n large

- Perfect secrecy: $I(M; E^n) = 0$
- Strong secrecy: $I(M; E^n) \leq \epsilon$ for some small $\epsilon > 0$
- Weak secrecy: $I(M; E^n)/n \leq \epsilon$ for some small $\epsilon > 0$
- Semantic secrecy: $\max_{p_m} I(M; E^n) \leq \epsilon$ for some small $\epsilon > 0$ *regardless* of distr. p_m

Degrees of Secrecy

Encoded message M is n-bit vector, Eve sees noisy n-bit E^n, n large

- Perfect secrecy: $I(M; E^n) = 0$
- Strong secrecy: $I(M; E^n) \leq \epsilon$ for some small $\epsilon > 0$
- Weak secrecy: $I(M; E^n)/n \leq \epsilon$ for some small $\epsilon > 0$
- Semantic secrecy: $\max_{p_m} I(M; E^n) \leq \epsilon$ for some small $\epsilon > 0$ *regardless* of distr. p_m
- Perfect $\Rightarrow$ Semantic $\Rightarrow$ Strong $\Rightarrow$ Weak

Degrees of Secrecy

Encoded message M is n-bit vector, Eve sees noisy n-bit E^n, n large

- Perfect secrecy: $I(M; E^n) = 0$
- Strong secrecy: $I(M; E^n) \leq \epsilon$ for some small $\epsilon > 0$
- Weak secrecy: $I(M; E^n)/n \leq \epsilon$ for some small $\epsilon > 0$
- Semantic secrecy: $\max_{p_m} I(M; E^n) \leq \epsilon$ for some small $\epsilon > 0$ *regardless* of distr. p_m
- Perfect $\Rightarrow$ Semantic $\Rightarrow$ Strong $\Rightarrow$ Weak

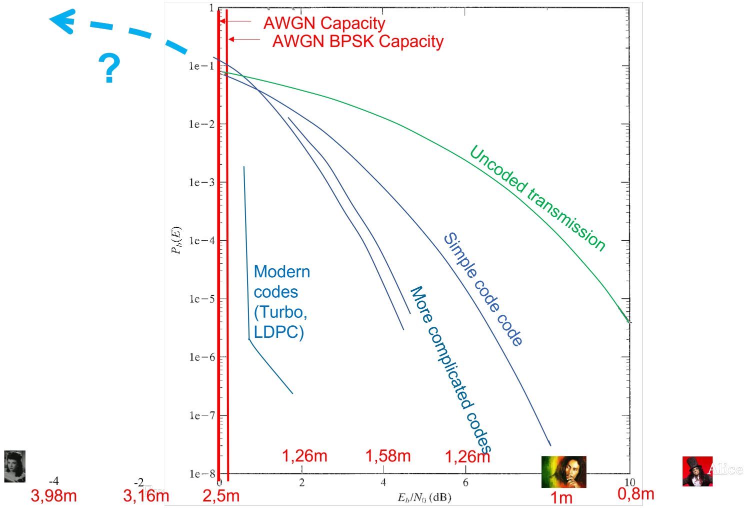
Note: $I(* : *)$ can be replaced by Renyi information.
Note: Independent of Eve's computational resources.
Note: Category of secrecy may depend on code and coding scheme

Now, the practice...The Type I AWGN channel

- The different roles of Bob and Eve
  - Bob wants a simple and efficient decoding to get the best decoding solution: Bit error rate reasonable metric
  - Eve willing to spend more efforts, maybe try out different option: Mutual information

Now, the practice...The Type I AWGN channel

How?



Now, the practice...The Type I AWGN channel

- The different roles of Bob and Eve
  - Bob wants a simple and efficient decoding to get the best decoding solution: Bit error rate reasonable metric
  - Eve willing to spend more efforts, maybe try out different option: Mutual information
- How to compute the mutual information (or, equivalently the equivocation)?

Now, the practice...The Type I AWGN channel

How to compute the mutual information (or, equivalently the equivocation)?

- Syndrome coding
  - Can be computed with complexity 2^{n-k}

Now, the practice...The Type I AWGN channel

How to compute the mutual information (or, equivalently the equivocation)?

- Syndrome coding
  - Can be computed with complexity 2^{n-k}
- Direct communication
  - Bounds exist (not very tight)

Now, the practice...The Type I AWGN channel

How to compute the mutual information (or, equivalently the equivocation)?

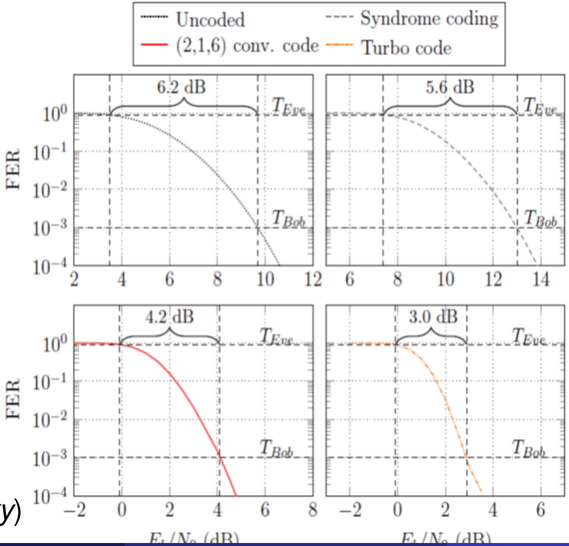
- Syndrome coding
  - Can be computed with complexity 2^{n-k}
- Direct communication
  - Bounds exist (not very tight)
  - Exact computation: Trellis computation Complexit $\leq \min\{2^k, 2^{n-k}\}$
  - *Joakim Algrøy, Angela Isabel Barbero and Øyvind Ytrehus, "Determining the Equivocation in Coded Transmission Over a Noisy Channel", accepted for IEEE International Symposium on Information Theory, June 26-July 1, 2022*

Now, the practice...The Type I AWGN channel

- The different roles of Bob and Eve
  - Bob wants a simple and efficient decoding to get the best decoding solution: Bit error rate reasonable metric
  - Eve willing to spend more efforts, maybe try out different option: Mutual information
- How to compute the mutual information (or, equivalently the equivocation)?

Now, the practice...The Type I AWGN channel

Comparison between syndrome coding and regular communication coding:



(*Figure: Joakim Algrøy*)

Now, the practice...The Type I AWGN channel

- The different roles of Bob and Eve
  - Bob wants a simple and efficient decoding to get the best decoding solution: Bit error rate reasonable metric
  - Eve willing to spend more efforts, maybe try out different option: Mutual information
- How to compute the mutual information (or, equivalently the equivocation)?
- Bob's channel is not noiseless

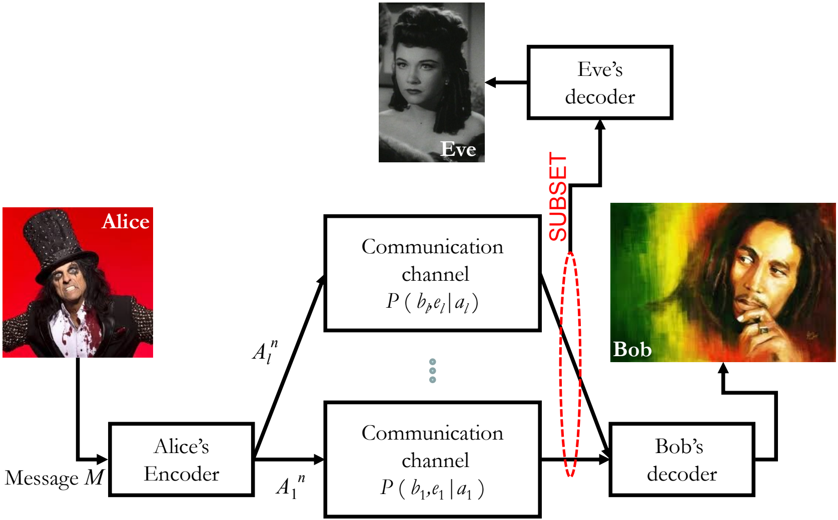
Now, the practice...The Type I AWGN channel

- The different roles of Bob and Eve
  - Bob wants a simple and efficient decoding to get the best decoding solution: Bit error rate reasonable metric
  - Eve willing to spend more efforts, maybe try out different option: Mutual information
- How to compute the mutual information (or, equivalently the equivocation)?
- Bob's channel is not noiseless
- Alice's message is not infinitely long

Now, the practice...The Type I AWGN channel

- The different roles of Bob and Eve
  - Bob wants a simple and efficient decoding to get the best decoding solution: Bit error rate reasonable metric
  - Eve willing to spend more efforts, maybe try out different option: Mutual information
- How to compute the mutual information (or, equivalently the equivocation)?
- Bob's channel is not noiseless
- Alice's message is not infinitely long
- How to deal with the tradeoff between secrecy, and power efficiency for Alice?
  That is, with a fixed redundancy r, devote r_B to help Bob and $r_E = r - r_B$ to confuse Eve. What is the optimum r_B?

The wiretap channel (Type II)



The wiretap channel (Type II)

How to encode?

- Choose a code with large generalized Hamming weights in the dual code
- Represent message by a syndrome vector
- Alice sends random codeword + corresponding coset leader

Outline

References

- T. M. Cover and J.A. Thomas, *Elements of Information Theory*, 2nd.ed., 2006.
- A. El Gamal and Y.-H. Kim, *Network Information Theory*, 2011.
- C. E. Shannon, “A mathematical theory of communication,” Bell syst. techn. j., vol. 27, no. 3, pp. 379-423, vol. 27, no. 4, pp. 623-656, 1948.
- C. E. Shannon, “Communication theory of secrecy systems,” Bell syst. techn. j., vol. 28, no. 4, pp. 656-715, 1949.
- A. D. Wyner, “The wire-tap channel,” Bell syst. techn. j., vol. 54, no. 8, pp. 1355-1387, 1975.
- D. Klinc, J. Ha, S. W.. McLaughlin, J. Barros, and B.-J, Kwak, “LDPC Codes for the Gaussian Wiretap Channel,” IEEE Trans.Inf. For. Sec., vol. 6, no. 3, Sept. 2011.
- Ahmed Abotabl and Aria Nosratinia, “Achieving the Secrecy Capacity of the AWGN Wiretap Channel via Multilevel Coding,” 55th Annual Allerton Conference, October 3-6, 2017.
- M. Bloch et al., ”An Overview of Information-Theoretic Security and Privacy: Metrics, Limits and Applications,” in IEEE Journal on Selected Areas in Information Theory, vol. 2, no. 1, pp. 5-22, March 2021, doi: 10.1109/JSAIT.2021.3062755.

Outline

"... as we know, there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns - the ones we don't know we don't know. ..."- *D.Rumsfeld*

Outline

Covert, deniable, subliminal, invisible, undetectable communication

Covert, deniable, subliminal, invisible, undetectable communication

- What if Alice and Bob do not want a listener to know that there is communication

Covert, deniable, subliminal, invisible, undetectable communication

- What if Alice and Bob do not want a listener to know that there is communication
- In general, communication can be *reliably detected* unless Alice and Bob has an advantage:

Covert, deniable, subliminal, invisible, undetectable communication

- What if Alice and Bob do not want a listener to know that there is communication
- In general, communication can be *reliably detected* unless Alice and Bob has an advantage:
  - shared randomness

Covert, deniable, subliminal, invisible, undetectable communication

- What if Alice and Bob do not want a listener to know that there is communication
- In general, communication can be *reliably detected* unless Alice and Bob has an advantage:
  - shared randomness
  - better channel

Covert, deniable, subliminal, invisible, undetectable communication

- What if Alice and Bob do not want a listener to know that there is communication
- In general, communication can be *reliably detected* unless Alice and Bob has an advantage:
    - shared randomness
    - better channel
    - more channels

Steganography

- Methods for encoding hidden messages in an apparently legitimate and apparently innocent host message

Steganography

- Methods for encoding hidden messages in an apparently legitimate and apparently innocent host message
- Alice may tattoo a hidden message on a messenger's shaved head

Steganography

- Methods for encoding hidden messages in an apparently legitimate and apparently innocent host message
- Alice may tattoo a hidden message on a messenger's shaved head
- Alice may write a message in invisible ink between the lines of an innocent-looking pretext letter.

Steganography

- Methods for encoding hidden messages in an apparently legitimate and apparently innocent host message
- Alice may tattoo a hidden message on a messenger's shaved head
- Alice may write a message in invisible ink between the lines of an innocent-looking pretext letter.
- alice may write a message So That a rEceiver GAN fOcus on larGe letteRs And PHorget anY small ones.

Steganography

- Methods for encoding hidden messages in an apparently legitimate and apparently innocent host message
- Alice may tattoo a hidden message on a messenger's shaved head
- Alice may write a message in invisible ink between the lines of an innocent-looking pretext letter.
- alice may write a message So That a rEceiver GAN fOcus on larGe letteRs And PHorget anY small ones.

Steganography

- Suppose Alice purports to send a message to Bob from the set {*Alice, Bob, Marilyn*}, representing the message as a picture. Let

Alice = {  ,  },

Bob = {  ,  }, and

Marilyn = {  ,  }

Steganography

- Suppose Alice purports to send a message to Bob from the set $\{$*Alice, Bob, Marilyn*$\}$, representing the message as a picture. Let

Alice = {  ,  },

Bob = { ,  }, and

Marilyn = { ,  }

- It follows that Alice may send one bit to Bob by selecting a pre-agreed image for each of the three possible cover messages.

Simmons' prisoner's problem

- Alice and Bob are prisoners who want to exchange information so that Willie is unable to detect the information transfer

Simmons' prisoner's problem

- Alice and Bob are prisoners who want to exchange information so that Willie is unable to detect the information transfer
- Using protocol redundancy

Simmons' prisoner's problem

- Alice and Bob are prisoners who want to exchange information so that Willie is unable to detect the information transfer
- Using protocol redundancy
- Concrete example: Using cryptographic signature schemes
    - Signature protocol uses random nonce

Simmons' prisoner's problem

- Alice and Bob are prisoners who want to exchange information so that Willie is unable to detect the information transfer
- Using protocol redundancy
- Concrete example: Using cryptographic signature schemes
  - Signature protocol uses random nonce
  - Alice and Bob sneakily agree to encode information into the choice of nonce

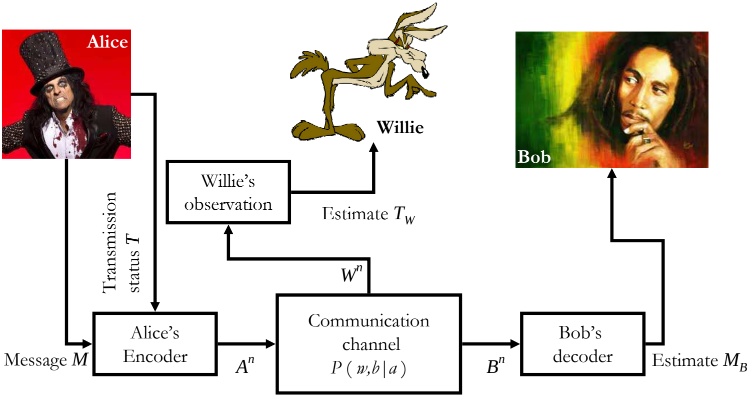
Simmons' prisoner's problem

- Alice and Bob are prisoners who want to exchange information so that Willie is unable to detect the information transfer
- Using protocol redundancy
- Concrete example: Using cryptographic signature schemes
  - Signature protocol uses random nonce
  - Alice and Bob sneakily agree to encode information into the choice of nonce
  - "Steganography", but hard for Willie to detect and prove

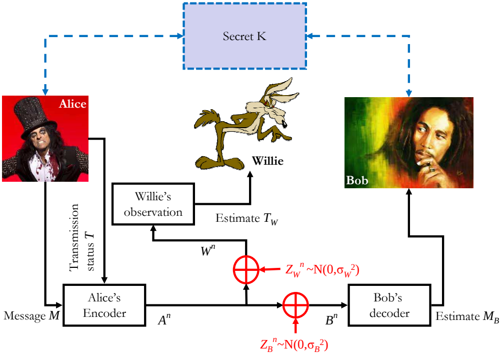
Simmons' prisoner's problem

- Alice and Bob are prisoners who want to exchange information so that Willie is unable to detect the information transfer
- Using protocol redundancy
- Concrete example: Using cryptographic signature schemes
  - Signature protocol uses random nonce
  - Alice and Bob sneakily agree to encode information into the choice of nonce
  - "Steganography", but hard for Willie to detect and prove
  - Can be blocked by zero-knowledge proofs etc, but still allows 1-bit subliminal channel (Desmedt)

Reliable deniable channels



Reliable deniable AWGN channels with randomness common to Alice and Bob



A^n, B^n, and W^n are real-valued n-dimensional vectors, and Z_B^n and Z_W^n are n-dimensional AWGN noise vectors. Alice and Bob need to share a secret key.

A reminder of complexity notation

- $f(n) = \mathcal{O}(g(n))$ if there exist constants $m, n_0 > 0$ such that $0 \leq f(n) \leq mg(n)$ for all $n \geq n_0$. This means that "$f(n)$ grows roughly at the same rate as $g(n)$".

A reminder of complexity notation

- $f(n) = \mathcal{O}(g(n))$ if there exist constants $m, n_0 > 0$ such that $0 \leq f(n) \leq mg(n)$ for all $n \geq n_0$. This means that "$f(n)$ grows roughly at the same rate as $g(n)$".
- $f(n) = o(g(n))$ if, for *any* constant $m > 0$ there exists a constant $n_0 > 0$ such that $0 \leq f(n) < mg(n)$ for all $n \geq n_0$. This means that "$f(n)$ grows slower than $g(n)$".

A reminder of complexity notation

- $f(n) = \mathcal{O}(g(n))$ if there exist constants $m, n_0 > 0$ such that $0 \leq f(n) \leq mg(n)$ for all $n \geq n_0$. This means that "$f(n)$ grows roughly at the same rate as $g(n)$".
- $f(n) = o(g(n))$ if, for *any* constant $m > 0$ there exists a constant $n_0 > 0$ such that $0 \leq f(n) < mg(n)$ for all $n \geq n_0$. This means that "$f(n)$ grows slower than $g(n)$".
- $f(n) = \omega(g(n))$ if, for *any* constant $m > 0$ there exists a constant $n_0 > 0$ such that $0 \leq mg(n) < f(n)$ for all $n \geq n_0$. This means that "$f(n)$ grows faster than $g(n)$".

Reliable deniable AWGN channels with randomness common to Alice and Bob: Results

1. For any $\varepsilon > 0$ and *unknown* σ_W^2, Alice can reliably transmit $o(\sqrt{n})$ information bits to Bob in n channel uses while lower-bounding Willie's sum of the probabilities of detection errors $\alpha + \beta \geq 1 - \varepsilon$.

Reliable deniable AWGN channels with randomness common to Alice and Bob: Results

1. For any $\varepsilon > 0$ and *unknown* σ_W^2, Alice can reliably transmit $o(\sqrt{n})$ information bits to Bob in n channel uses while lower-bounding Willie's sum of the probabilities of detection errors $\alpha + \beta \geq 1 - \varepsilon$.
2. If Alice knows a nontrivial lower bound $\hat{\sigma}_W^2 > 0$ on the noise power on Willie′s channel (*i.e.*, $\sigma_W^2 \geq \hat{\sigma}_W^2$), she can reliably transmit $\mathcal{O}(\sqrt{n})$ information bits to Bob in n channel uses while lower-bounding Willie′s sum of the probabilities of detection errors $\alpha + \beta \geq 1 - \varepsilon$.

Reliable deniable AWGN channels with randomness common to Alice and Bob: Results

1. For any $\varepsilon > 0$ and *unknown* σ_W^2, Alice can reliably transmit $o(\sqrt{n})$ information bits to Bob in n channel uses while lower-bounding Willie's sum of the probabilities of detection errors $\alpha + \beta \geq 1 - \varepsilon$.
2. If Alice knows a nontrivial lower bound $\hat{\sigma}_W^2 > 0$ on the noise power on Willie′s channel (*i.e.*, $\sigma_W^2 \geq \hat{\sigma}_W^2$), she can reliably transmit $\mathcal{O}(\sqrt{n})$ information bits to Bob in n channel uses while lower-bounding Willie′s sum of the probabilities of detection errors $\alpha + \beta \geq 1 - \varepsilon$.
3. Conversely, if Alice attempts to transmit $\omega(\sqrt{n})$ bits in n channel uses, then, as $n \rightarrow \infty$, *either* $\alpha + \beta$ is arbitrarily close to zero *or* the communication to Bob is not reliable, regardless of the length of the shared secret.

Reliable deniable AWGN channels with randomness common to Alice and Bob: Interpretation

1. The capacity $\lim_{n\to\infty} \frac{\mathcal{O}(\sqrt{n})}{n} = 0$. But for finite codeword lengths n, a substantial amount $\mathcal{O}(\sqrt{n})$ of information may be reliably transmitted with low probability of detection.

Reliable deniable AWGN channels with randomness common to Alice and Bob: Interpretation

1. The capacity $\lim_{n\to\infty} \frac{\mathcal{O}(\sqrt{n})}{n} = 0$. But for finite codeword lengths n, a substantial amount $\mathcal{O}(\sqrt{n})$ of information may be reliably transmitted with low probability of detection.
2. Proof: A random coding argument, with actual code disguised by "key".

Reliable deniable AWGN channels with randomness common to Alice and Bob: Interpretation

1. The capacity $\lim_{n\to\infty} \frac{\mathcal{O}(\sqrt{n})}{n} = 0$. But for finite codeword lengths n, a substantial amount $\mathcal{O}(\sqrt{n})$ of information may be reliably transmitted with low probability of detection.
2. Proof: A random coding argument, with actual code disguised by "key".
3. Bob faces a noisy channel decoding problem.

Reliable deniable AWGN channels with randomness common to Alice and Bob: Interpretation

1. The capacity $\lim_{n\to\infty} \frac{\mathcal{O}(\sqrt{n})}{n} = 0$. But for finite codeword lengths n, a substantial amount $\mathcal{O}(\sqrt{n})$ of information may be reliably transmitted with low probability of detection.
2. Proof: A random coding argument, with actual code disguised by "key".
3. Bob faces a noisy channel decoding problem.
4. *The amount of randomness:* Simple scheme requires n coded bits, for an $\mathcal{O}(\sqrt{n})$-length message. A more refined scheme requiring $\mathcal{O}(\sqrt{n}) \log n$ is also presented.

Reliable deniable AWGN channels with randomness common to Alice and Bob: Interpretation

1. The capacity $\lim_{n\to\infty} \frac{\mathcal{O}(\sqrt{n})}{n} = 0$. But for finite codeword lengths n, a substantial amount $\mathcal{O}(\sqrt{n})$ of information may be reliably transmitted with low probability of detection.
2. Proof: A random coding argument, with actual code disguised by "key".
3. Bob faces a noisy channel decoding problem.
4. *The amount of randomness:* Simple scheme requires n coded bits, for an $\mathcal{O}(\sqrt{n})$-length message. A more refined scheme requiring $\mathcal{O}(\sqrt{n}) \log n$ is also presented.
5. The *constants* involved can become very small.

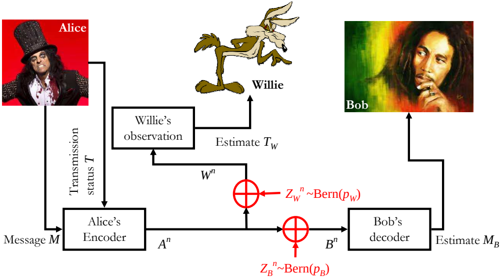
Reliable deniable AWGN channels with randomness common to Alice and Bob: Interpretation

1. The capacity $\lim_{n\to\infty} \frac{\mathcal{O}(\sqrt{n})}{n} = 0$. But for finite codeword lengths n, a substantial amount $\mathcal{O}(\sqrt{n})$ of information may be reliably transmitted with low probability of detection.
2. Proof: A random coding argument, with actual code disguised by “key”.
3. Bob faces a noisy channel decoding problem.
4. *The amount of randomness:* Simple scheme requires n coded bits, for an $\mathcal{O}(\sqrt{n})$-length message. A more refined scheme requiring $\mathcal{O}(\sqrt{n}) \log n$ is also presented.
5. The *constants* involved can become very small.
6. Prior probability distribution on T is assumed unknown, does it matter?

Reliable deniable AWGN channels with randomness common to Alice and Bob: Interpretation

1. The capacity $\lim_{n\to\infty} \frac{\mathcal{O}(\sqrt{n})}{n} = 0$. But for finite codeword lengths n, a substantial amount $\mathcal{O}(\sqrt{n})$ of information may be reliably transmitted with low probability of detection.
2. Proof: A random coding argument, with actual code disguised by "key".
3. Bob faces a noisy channel decoding problem.
4. *The amount of randomness:* Simple scheme requires n coded bits, for an $\mathcal{O}(\sqrt{n})$-length message. A more refined scheme requiring $\mathcal{O}(\sqrt{n}) \log n$ is also presented.
5. The *constants* involved can become very small.
6. Prior probability distribution on T is assumed unknown, does it matter?
7. Quantum channel version

Reliable deniable BSC channels without randomness common to Alice and Bob



The binary symmetric subliminal channel. Here A^n, B^n, and W^n are binary n-dimensional vectors, and Z_B^n and Z_W^n are binary n-dimensional noise vectors in which elements are generated independently according to their respective Bernoulli distributions.

Reliable deniable BSC channels without randomness common to Alice and Bob: Results

1. *Deniability.* When $T = 0$, Willie should observe a fraction of p_w 1's. So if Alice uses a code with codewords of weight larger than np_w, then Willie will suspect that $T = 1$.

Reliable deniable BSC channels without randomness common to Alice and Bob: Results

1. *Deniability.* When $T = 0$, Willie should observe a fraction of p_w 1's. So if Alice uses a code with codewords of weight larger than np_w, then Willie will suspect that $T = 1$.
2. *Reliability and deniability: upper bound on code rate.* If Bob's channel is noisy and reliable communication to Bob is required, any code selected by Alice can convey at most $\mathcal{O}(\sqrt{n})$ information bits per n channel uses.

Reliable deniable BSC channels without randomness common to Alice and Bob: Results

1. *Deniability.* When $T = 0$, Willie should observe a fraction of p_w 1's. So if Alice uses a code with codewords of weight larger than np_w, then Willie will suspect that $T = 1$.
2. *Reliability and deniability: upper bound on code rate.* If Bob's channel is noisy and reliable communication to Bob is required, any code selected by Alice can convey at most $\mathcal{O}(\sqrt{n})$ information bits per n channel uses.
3. *Reliability and deniability: lower bound on code rate*. If Bob's channel is *sufficiently much better* than Willie's, then there exist (random) codes that can convey to Bob $\mathcal{O}(\sqrt{n})$ information bits per n channel uses. If Bob's channel is noiseless, there exist (random) codes that can convey to Bob $\mathcal{O}(\sqrt{n}) \log n$ information bits per n channel uses.

Reliable deniable BSC channels without randomness common to Alice and Bob: Interpretations

1. This channel also has “zero capacity”, but still allows, in theory, a substantial reliable and undetectable information transfer.

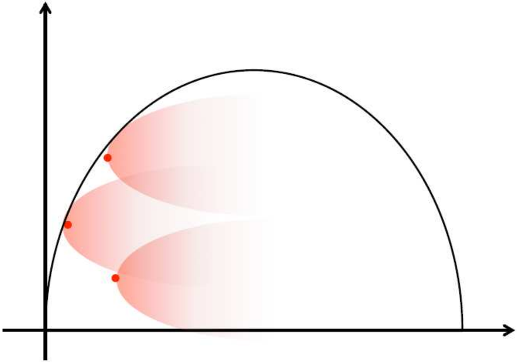
Reliable deniable BSC channels without randomness common to Alice and Bob: Interpretations

1. This channel also has "zero capacity", but still allows, in theory, a substantial reliable and undetectable information transfer.
2. When $T = 0$, Alice transmits nothing, and Willie observes only noise. For $T = 1$, Willie observes the (mod 2) sum of a codeword and random Bernoulli noise.

Reliable deniable BSC channels without randomness common to Alice and Bob: Interpretations

1. This channel also has “zero capacity”, but still allows, in theory, a substantial reliable and undetectable information transfer.
2. When $T = 0$, Alice transmits nothing, and Willie observes only noise. For $T = 1$, Willie observes the (mod 2) sum of a codeword and random Bernoulli noise.
3. Bob faces a (modified) BSC decoding problem. When $T = 0$, such decoding will be unsuccessful with overwhelming probability. Thus the channel will not produce “false information” to Bob. When $T = 1$, such decoding will be successful with overwhelming probability, provided that the code is appropriately selected.

Reliable deniable BSC channels without randomness common to Alice and Bob: Bob's decoder



Outline

Why Alice and Bob may have a harder time in practice than in theory

- Codeword synchronization

Why Alice and Bob may have a harder time in practice than in theory

- Codeword synchronization
- Key synchronization (for the AWGN case)

Why Alice and Bob may have a harder time in practice than in theory

- Codeword synchronization
- Key synchronization (for the AWGN case)
- For the AWGN channel: *How is Willie's observed signal to noise ratio obtained?*
  For the BSC channel: *How is p_w obtained?*

Why Alice and Bob may have a harder time in practice than in theory

- Codeword synchronization
- Key synchronization (for the AWGN case)
- For the AWGN channel: *How is Willie's observed signal to noise ratio obtained?*
  For the BSC channel: *How is* p_w *obtained?*
- Consider an example of a malware (software/hardware) agent that uses a "compromising emanations" secondary wireless channel for sending messages to Bob.
  In this case Willie typically will have a better SNR than Bob.

Why Alice and Bob may have a harder time in practice than in theory

- Codeword synchronization
- Key synchronization (for the AWGN case)
- For the AWGN channel: *How is Willie's observed signal to noise ratio obtained?*
  For the BSC channel: *How is p_w obtained?*
- Consider an example of a malware (software/hardware) agent that uses a "compromising emanations" secondary wireless channel for sending messages to Bob.
  In this case Willie typically will have a better SNR than Bob.
- *Implementation in practice?* Random coding is merely a theoretical tool and has no practical usage. What *practical coding schemes* can be used?
  AWGN: possible to use a normal LDPC code?
  Noisy BSC subliminal channel: Need constant (low) weight codes; nonlinear.

Why Willie may have a harder time in practice

- From Willie's perspective, the assumption of knowing the code agreed between Alice and Bob is a best-case scenario. Reasonable approach in cryptanalysis, maybe less so in the context of deniable channels?

Why Willie may have a harder time in practice

- From Willie's perspective, the assumption of knowing the code agreed between Alice and Bob is a best-case scenario. Reasonable approach in cryptanalysis, maybe less so in the context of deniable channels?
- For the previous issue, will a compressed sensing approach be sensible for Willie? That is, can we observe communication knowing that a code is used, but not which code is used?

Other issues

- In an AWGN channel where Bob has a better channel than Willie, do Alice and Bob need common randomness?

Other issues

- In an AWGN channel where Bob has a better channel than Willie, do Alice and Bob need common randomness?
- Schemes that require common randomness between Alice and Bob: can Alice and Bob use a hybrid scheme?

Other issues

- In an AWGN channel where Bob has a better channel than Willie, do Alice and Bob need common randomness?
- Schemes that require common randomness between Alice and Bob: can Alice and Bob use a hybrid scheme?
- "$\mathcal{O}(\sqrt{n})$ information bits per n channel uses" $\Rightarrow$ asymptotic code rate zero. Normally, throughput improves as $n \to \infty$. Here, is there an optimum value of n?

Other issues

- In an AWGN channel where Bob has a better channel than Willie, do Alice and Bob need common randomness?
- Schemes that require common randomness between Alice and Bob: can Alice and Bob use a hybrid scheme?
- "$\mathcal{O}(\sqrt{n})$ information bits per n channel uses" $\Rightarrow$ asymptotic code rate zero. Normally, throughput improves as $n \to \infty$. Here, is there an optimum value of n?
- The concepts of *detectability* and *provability* are related, but they are not equivalent. Does this distinction matter?

Other issues

- In an AWGN channel where Bob has a better channel than Willie, do Alice and Bob need common randomness?
- Schemes that require common randomness between Alice and Bob: can Alice and Bob use a hybrid scheme?
- "$\mathcal{O}(\sqrt{n})$ information bits per n channel uses" $\Rightarrow$ asymptotic code rate zero. Normally, throughput improves as $n \to \infty$. Here, is there an optimum value of n?
- The concepts of *detectability* and *provability* are related, but they are not equivalent. Does this distinction matter?
- Some practical research problems: study typical emanating channels, or study theoretical channel models that may be forced into practice by a malware agent.

Other issues

- In an AWGN channel where Bob has a better channel than Willie, do Alice and Bob need common randomness?
- Schemes that require common randomness between Alice and Bob: can Alice and Bob use a hybrid scheme?
- "$\mathcal{O}(\sqrt{n})$ information bits per n channel uses" $\Rightarrow$ asymptotic code rate zero. Normally, throughput improves as $n \to \infty$. Here, is there an optimum value of n?
- The concepts of *detectability* and *provability* are related, but they are not equivalent. Does this distinction matter?
- Some practical research problems: study typical emanating channels, or study theoretical channel models that may be forced into practice by a malware agent.

Outline

Conclusion, Single-path communication

A covert entity Alice may use a communication channel to pass information to an accomplice Bob in a way that cannot be detected by a warden Willie.

1. undetectable low rate information transfer is feasible, but there remain serious challenges for Alice and Bob, having to do with implementation, with the set of parameters, and with the set of assumptions.

Conclusion, Single-path communication

A covert entity Alice may use a communication channel to pass information to an accomplice Bob in a way that cannot be detected by a warden Willie.

1. undetectable low rate information transfer is feasible, but there remain serious challenges for Alice and Bob, having to do with implementation, with the set of parameters, and with the set of assumptions.
2. For the warden Willie, there exist realistic scenarios that are worse than those assumed in the literature, and this creates extra problems.

Outline

More References

- P. H. Che, S. Kadhe, M. Bakshi, C. Chan, S. Jaggi, and A.Sprintson, “Reliable, Deniable and Hidable Communication: A Quick Survey,” in *Proc. ITW 2014*, Hobarth , Nov. 2014..
- G. Simmons, “The prisoners problem and the subliminal channel,” Proc. Crypto, 1983, pp. 51-67.
- B. A. Bash, D. Goeckel, and D. Towsley, “Limits of reliable communication with low probability of detection on AWGN channels,” JSAC, vol. 31, no. 9, pp. 1921-1930, 2013.
- P. Hou Che, M. Bakshi, and S. Jaggi, “Reliable deniable communication: Hiding messages in noise,” arXiv preprint arXiv:1304.6693, 2013.
- J. Hou and G. Kramer, “Effective secrecy: Reliability, confusion and stealth,” arXiv preprint arXiv:1311.1411, 2013.
- S. Kadhe, S. Jaggi, M. Bakshi, and A. Sprintson, “Reliable, deniable, and hidable communication over parallel link networks,” arXiv preprint arXiv:1401.4451, 2014.