An information theoretical view on reliability, efficiency, security, and stealthiness

Øyvind Ytrehus

Finse, May 4 & 5, 2015

Cast: Main characters

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

What do Alice and Bob want?

Alice wants to send messages to Bob...

- reliably and efficiently
  - Information theory
- securely: confidentially,authenticated
  - Cryptography
  - Other security techniques
  - Information Theory
- stealthily
  - Anonymizing networks?
  - Information Theory

Information theory and noisy channels

- What is Information?
- Consider a discrete stochastic variable X with set of possible outcomes $\mathcal{X}$ and probability mass function $p(x)$. The *entropy* of X, measured in bits, is

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

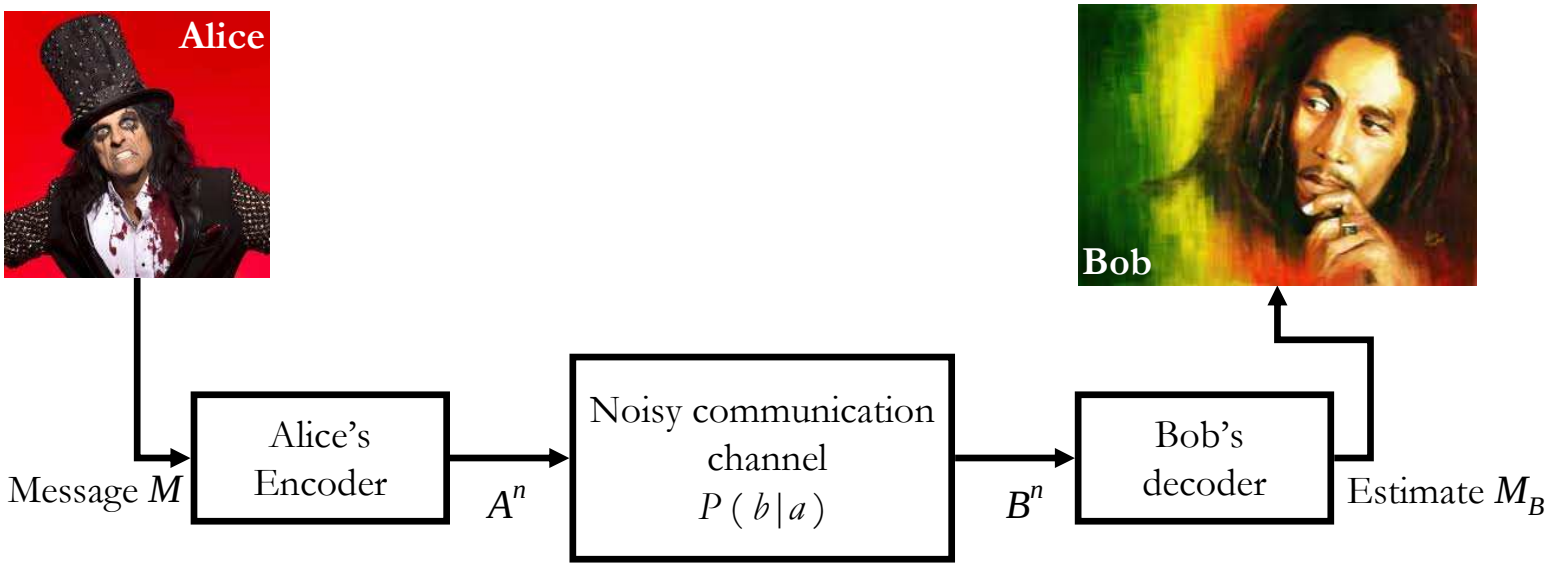
  In particular, in the (worst) case where $p(x)$ is the uniform distribution on $\mathcal{X}$, it holds that $H(X) = \log_2(|\mathcal{X}|)$.
- Consider two discrete stochastic variables X and Y with set of possible outcomes $\mathcal{X} = \mathcal{Y}$ and joint probability mass function $p(x, y)$. The *mutual information* between X and Y, measured in bits, 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”
- 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

Shannon's noisy channel



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

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

Shannon's noisy channel: Sketch of proof, $R < C$

- Message $\underline{m} = \{m_1, \ldots, m_k\}$, $k = 2^{nR}$
- Choose 2^{nR} *random* length-n codewords according to $p^*(a)$
- Choose message $\underline{m}_0$, send codeword $\underline{a}(\underline{m}_0)$, receive $\underline{b}$
- Decode received $\underline{b}$ to $\underline{m}$ iff $(\underline{a}(\underline{m}), \underline{b})$ typical
- Error if
  1. $(\underline{a}(\underline{m}_0), \underline{b})$ atypical, or
  2. $\exists \underline{m} \neq \underline{m}_0$ s. t. $(\underline{a}(\underline{m}), \underline{b})$ typical

  Probabilities of these events:
  1. $P((\underline{a}(\underline{m}_0), \underline{b})$ atypical $) \to_{n\to\infty} 0$
  2. $P() \leq 2^{nR} P($ random $\underline{a}, \underline{b})$ typical$) \lesssim 2^{n(R-C)}$

Shannon's noisy channel: Fano's lemma

Fano's lemma: A, B rand.var. $\in \{x_1, \ldots, x_L\}$.
Let $P_e = P(Z = 1) = P(A \neq B)$. Then

$$H(A|B) \leq H(P_e) + P_e \log_2(L-1)$$

Proof:

$$\begin{aligned} H(A|B) &= H(A|B) + H(Z|A,B) \\ &= H(A,Z|B) \\ &= H(Z|B) + H(A|B,Z) \\ &\leq H(Z) + H(A|B,Z) \\ &\leq H(Z) + P(Z=1)\log_2(L-1) \end{aligned}$$

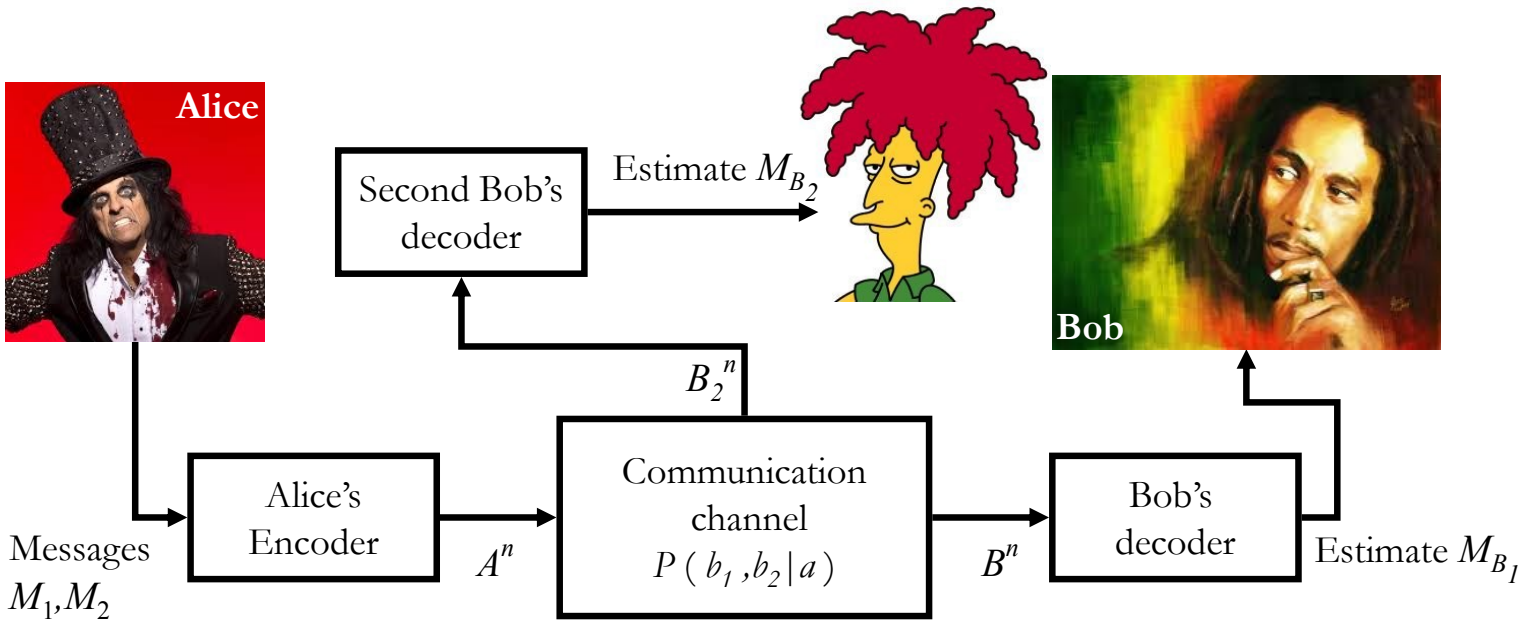
Shannon's noisy channel: Sketch of proof, $R > C$

M binary i. i. d., Message $\underline{m} = (m_1, \ldots, m_k)$,
decoded message $\underline{\hat{m}} = (\hat{m}_1, \ldots, \hat{m}_k)$
$Z = 1$ if $\underline{m} \neq \underline{\hat{m}}$, otherwise $Z = 0$.

$$\begin{aligned} H(P_e) + P_e(\log_2(2^k - 1) &\geq H(\underline{m}|\underline{\hat{m}}) \\ &= H(\underline{m}) - I(\underline{m}; \underline{\hat{m}}) \\ &\geq H(\underline{m}) - I(\underline{a}; \underline{b}) \\ &\geq k - nC = k(1 - C/R) \end{aligned}$$

$$P_e \geq 1 - C/R$$

The Broadcast Channel

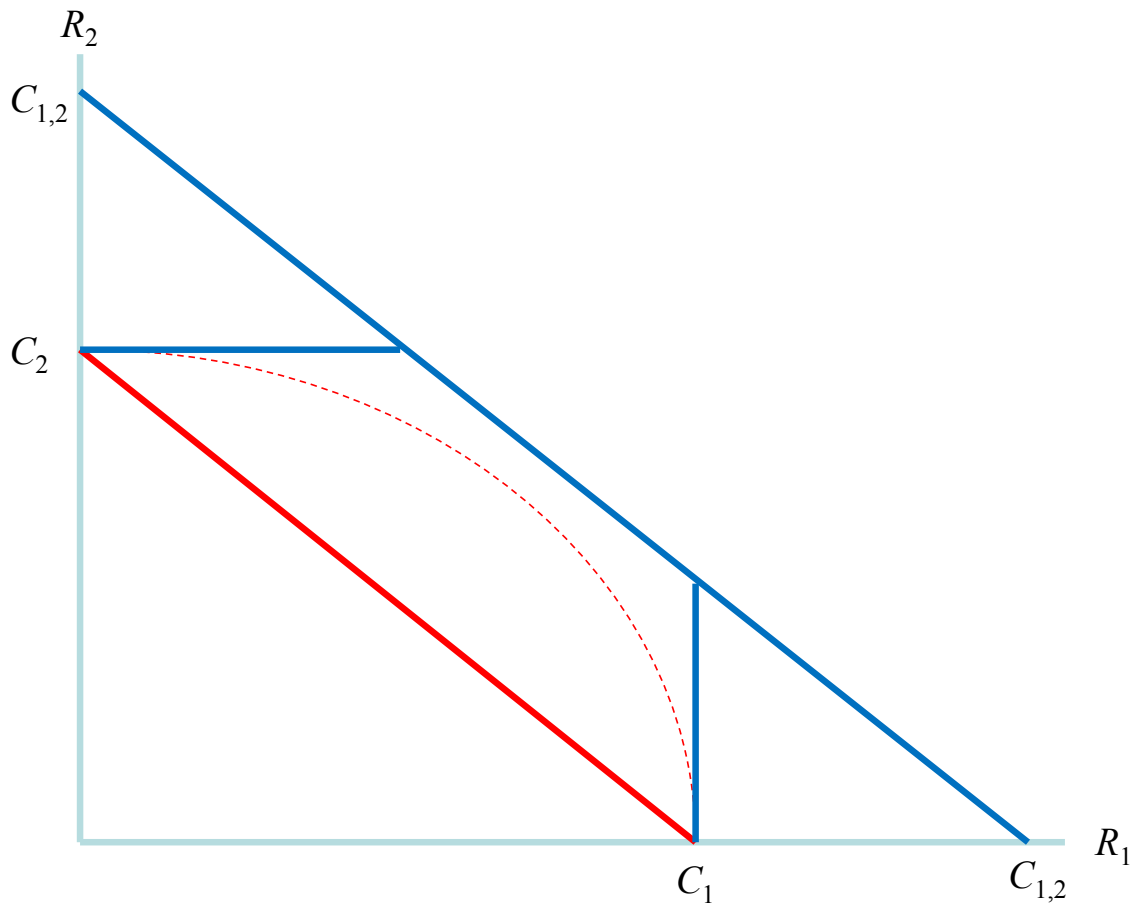


$$R_1 \leq C_1 = \max_{p_a} I(A; B_1), R_2 \leq C_2 = \max_{p_a} I(A; B_2),$$

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

The Broadcast Channel

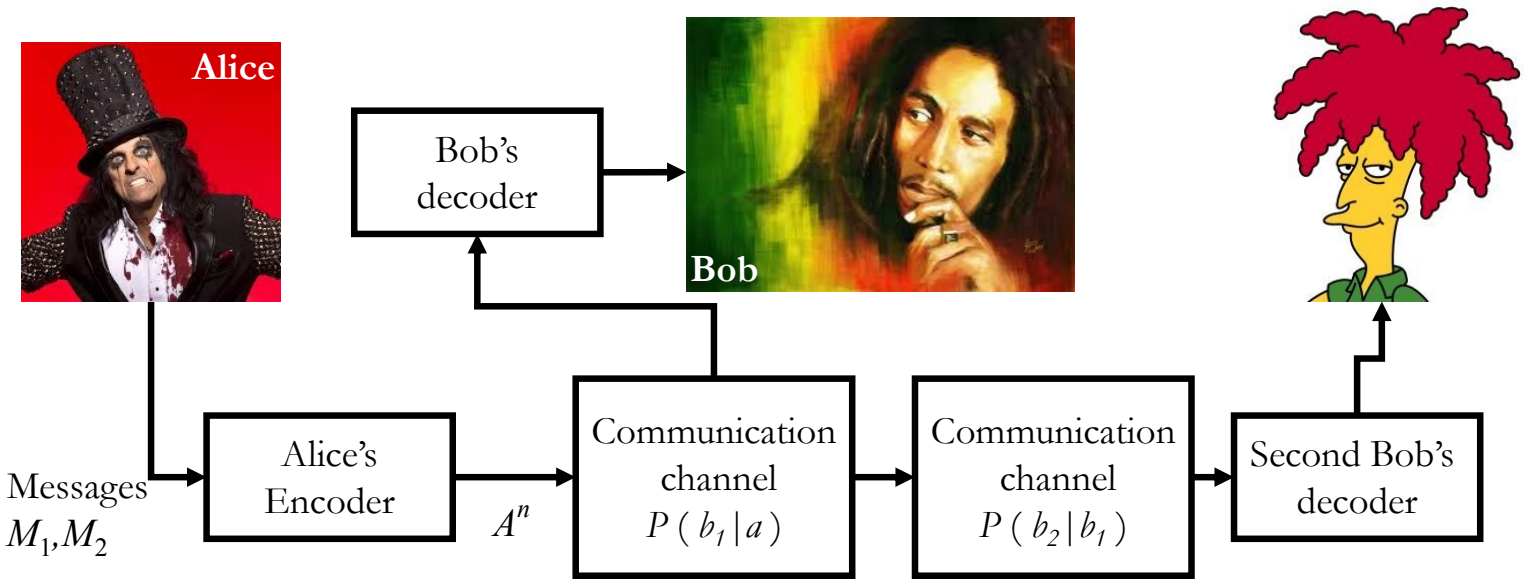
$$R_1 \leq C_1 = \max_{p_a} I(A; B_1), R_2 \leq C_2 = \max_{p_a} I(A; B_2),$$
$$R_1 + R_2 \leq C_{1,2} = \max_{p_a} I(A; B_1, B_2),$$



The broadcast channel: Sketch of proof, and example

- Jointly typical sequences
- Superposition coding

The Broadcast Channel, Degraded

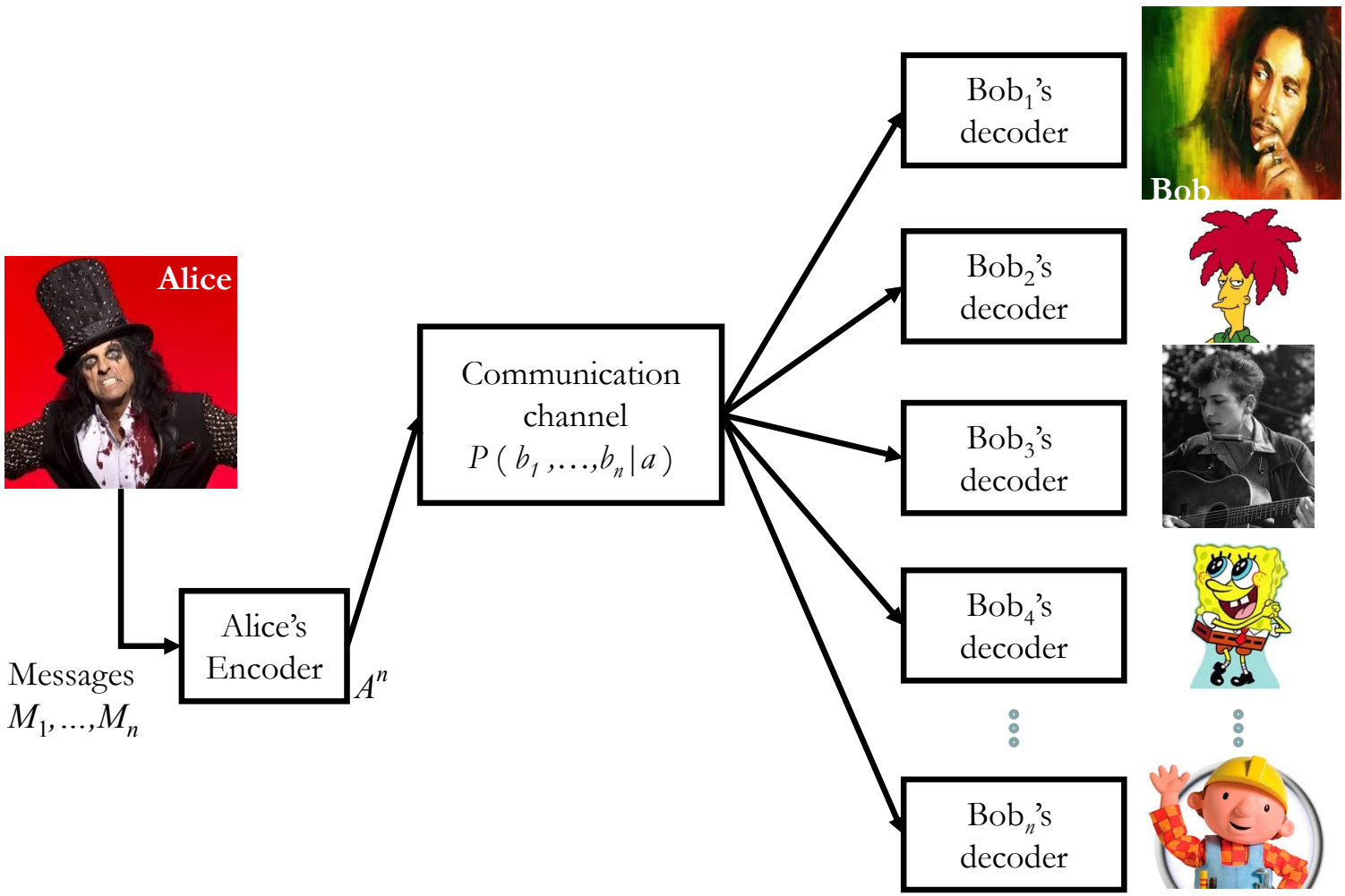


$$R_1 \leq I(A; B_1|U),$$

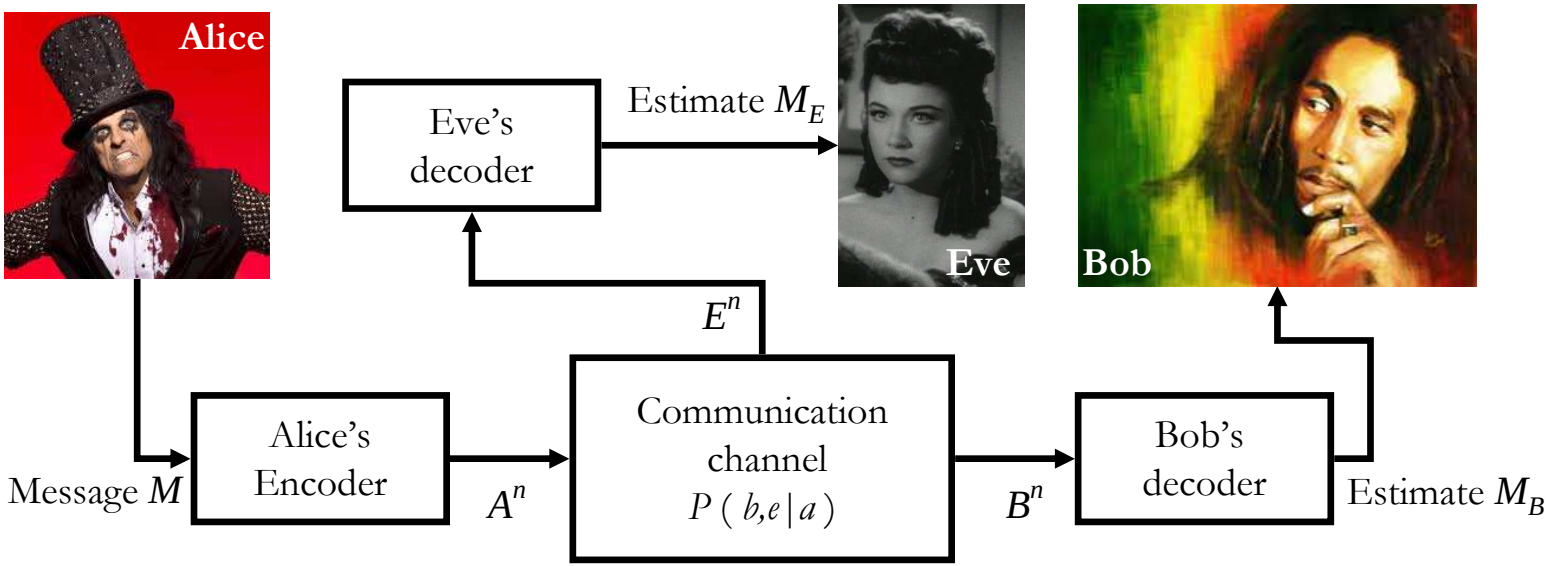
$$R_2 \leq I(U; B_2),$$

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

The Broadcast Channel, Generalized



The wiretap channel (Type I)

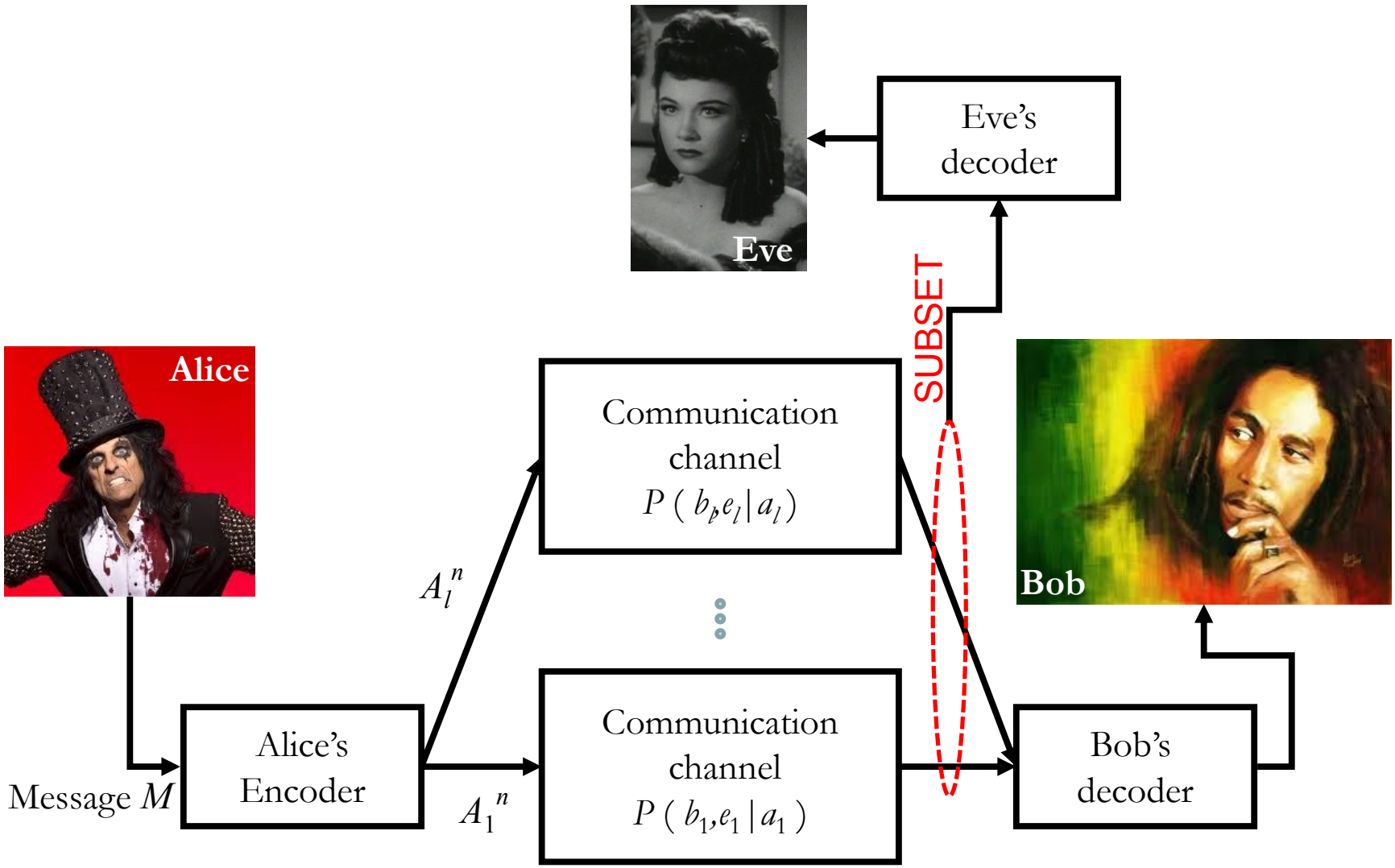


$$C_{DM-WTC} = \max\{0, \max_{p(u,a)}(I(U;B) - I(U;E))\},$$

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

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

The wiretap channel (Type II)



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 ofcommunication,” Bell system technical journal, vol. 27, no. 3, pp. 379-423, vol. 27, no. 4, pp. 623-656, 1948.
- C. E. Shannon, “Communication theory of secrecy systems,” Bell system technical journal, vol. 28, no. 4, pp. 656-715, 1949.
- A. D. Wyner, “The wire-tap channel,” Bell system technical journal, vol. 54, no. 8, pp. 1355-1387, 1975.

“... 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*

Covert, deniable, subliminal, invisible, undetectable communication

- What if Alice and Bob does 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
- 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

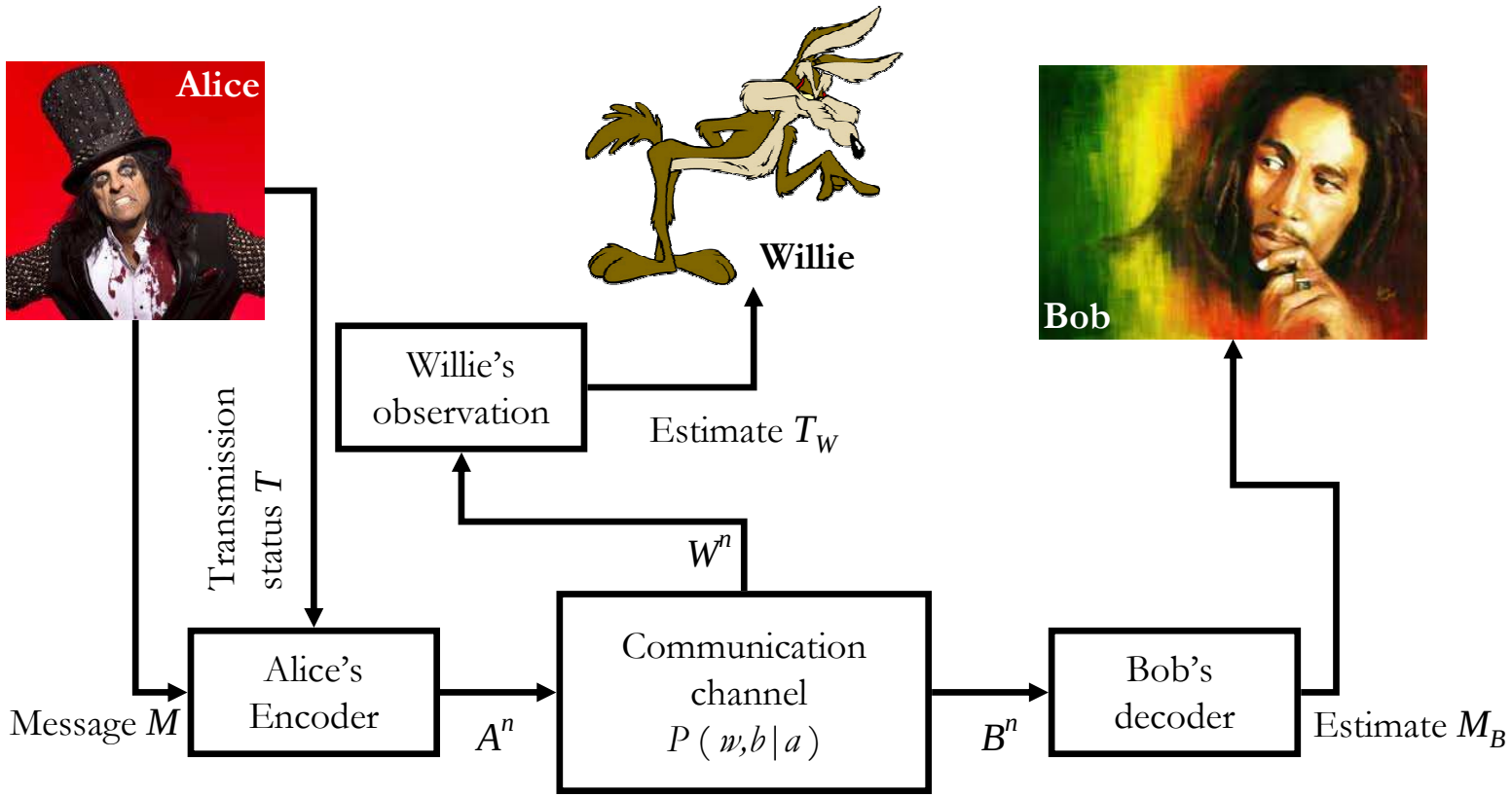
Marilyn = {  ,  }

- 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
- 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

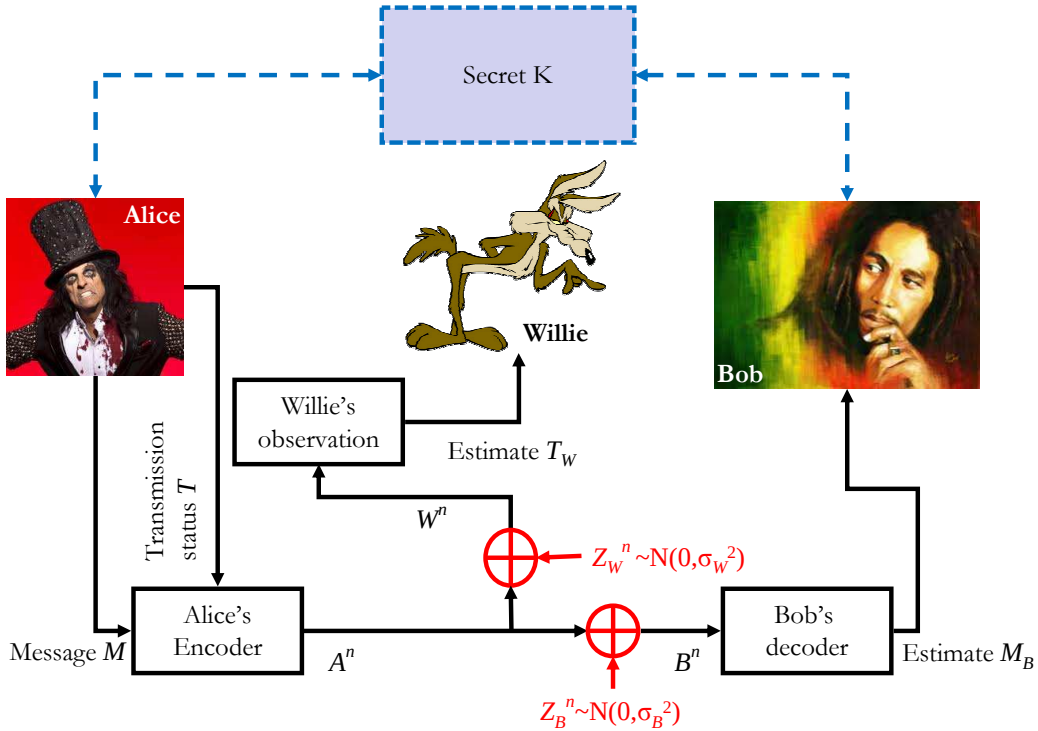


Figure: A^n, B^n, and W^n are real-valued n-dimensional vectors, and Z_B^n and Z_W^n are n-dimensional AWGN noise vectors. Alice and Bob need to share a secret key.

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 $g(n)$".
- $f(n) = o(g(n))$ 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 Willies 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 $n \to \infty$, *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 reliable 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

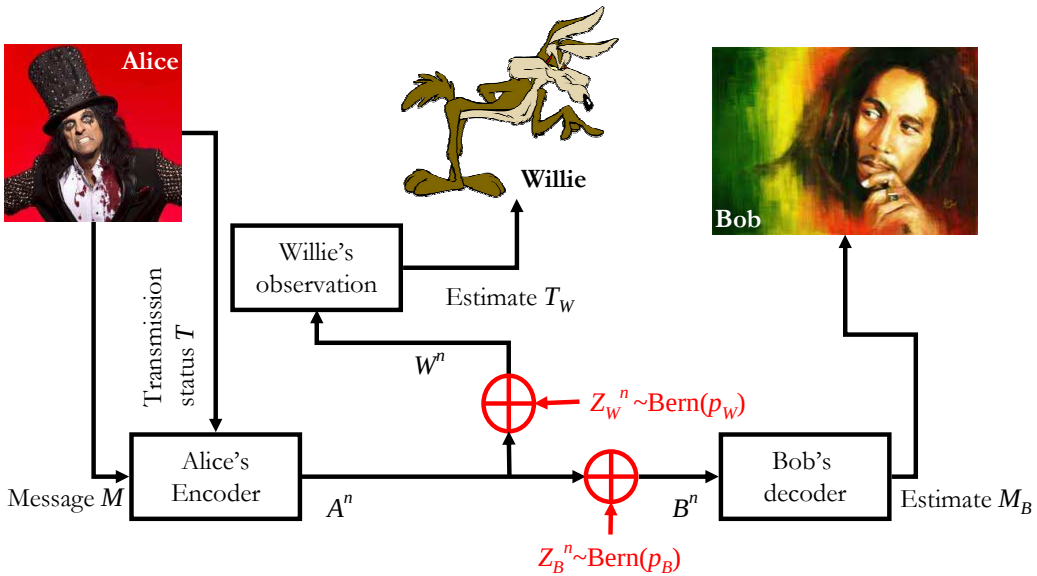


Figure: The binary symmetric subliminal channel. Here A^n, B^n, and W^n are binary n-dimensional vectors, and Z_B^n and Z_W^n are binary n-dimensional noise vectors in which elements are generated independently according to their respective Bernoulli distributions.

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 $T = 1$.
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.
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.

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?*
- *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 nonlinear codes.
- 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 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.
- 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?
- 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.
- *Rateless coding schemes?*
- Is there a reliable and deniable *network coding* scheme?

Conclusion, Single-path communication

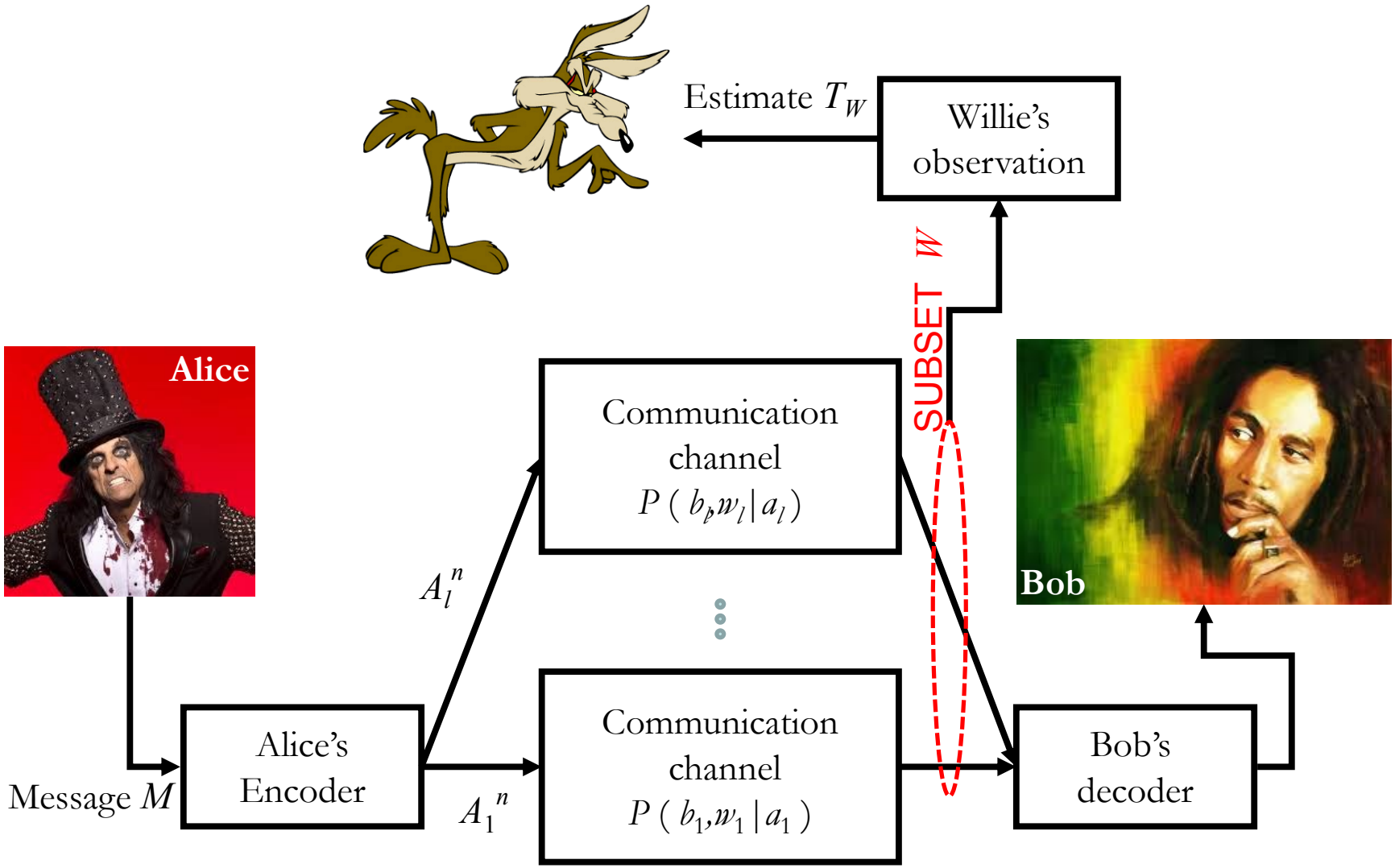
A covert entity Alice may use a communication channel to pass information to an accomplice Bob in a way that cannot be detected by a warden Willie.

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.
3. There are few results available in the open literature about the problem *from Willie's perspective,* and more research is required.

Reliable, deniable, hidable communication over Multipath networks

1. Multipath
2. Separation between Deniable and Hidable

Reliable, deniable, hidable communication over Multipath networks)



Reliable, deniable, hidable communication over Multipath networks

1. Reliable

$$P(\hat{T}_B = 1 | T = 0) + P(\hat{T}_B = 0 | T = 1) + P(\hat{M}_B \neq M | T = 1) \textit{ small}$$

2. Deniable

$$\mathbb{V}(p^{(i)}(), \hat{p}_W()) \textit{ small}, \text{ where}$$

$$\mathbb{V}(p_1(), p_2()) = 1/2 \sum_x |p_1(x) - p_2(x)|$$

3. Hidable (Secure)

$$\frac{P(\hat{M}_W = m | \mathcal{W}, T = 1)}{P(\hat{M}_W = m | T = 1)} \textit{ close to } 1, \forall m, \forall \mathcal{W}$$

Multipath example 1

Estimate T_W

Willie's observation

SUBSET W

0	**0**	**1**	**1**
0	1	0	1
0	½	¼	¼

A_2^n

Message M

Alice's Encoder

A_1^n

Bob's decoder

Multipath example 2

Estimate T_W

Willie's observation

SUBSET W

0	0	0	0	1	1	1	1
0	0	1	1	0	0	1	1
0	1	0	1	0	1	0	1
½	0	0	0	0	0	0	½

A_3^n

A_2^n

A_1^n

Message M

Alice's Encoder 1/3

Bob's decoder

Multipath example 1, revisited

Estimate T_W

Willie's observation

SUBSET W

0	**0**	**1**	**1**
0	1	0	1
0	½	½	0

A_2^n

Message M

Alice's Encoder

A_1^n

Bob's decoder

Conclusion, Multipath communication

- Nonzero capacity possible for reliable, deniable, hidable communication
- New area
- Many variations on the problem
- Research opportunities!!!

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.