Wiretap channel of type II

Seminar report for Research Seminar in Cryptography

Pille Pullonen

Supervisor: Vitaly Skachek

May 24, 2014

1 Introduction

A wiretapper on the network is an eavesdropper that listens to some part of the communication and intends to learn something about the content of the communication. It is trivial that if the communication channel is not secured and the eavesdropper can see all of the communication then it has the same information as the recipient of the communication. However, it is interesting to analyse how much can the eavesdropper learn if it is not possible to observe all of the communication.

This paper explores the strong model where the adversary can eavesdrop some fixed ratio of the transmitted bits, but can choose these bits on its own. For example, if we transmit two bits and an adversary can see one of them, then we could use a code based on the parity of the message to encode one bit and the adversary would learn nothing about the content of the message by just observing one bit. More specifically message 1 would be transmitted using 01 or 10 and correspondingly 0 can be encoded as 00 or 11. Therefore, if adversary just observes one of the two transmitted bits then both messages are still equally likely.

In fact, as long as we use N bits in the encoded string and have a K-bit message, we can achieve such perfect secrecy given that the adversary does not observe more than $N - K$ bits of the message. In addition to this result, a bound for the amount of information that the adversary can have if it sees more than $N - K$ bits is also derived. The following does not propose any specific codes for achieving perfect secrecy, but proves that such codes exist.

This work is based on paper [OW85]. The purpose of this report is to give a more thorough introduction to the information theory used in the original paper and to describe their main results in more detail. Especially, this document gives a thorough explanation of the two main results that prove the basic properties achievable for wiretap channels. We do not include the stronger results from the original paper, but a short discussion about them is given in Section 6.

In general, Section 2 gives an overview of the communication model and basics of information theory. After that Section 3 introduces the wiretap channel and states the

main definition and property of this channel. In the following, Section 4 shows that these properties are in fact necessary and Section 5 shows that if the desired properties are satisfied then there also exist codes that satisfy the achievability definition for the wiretap channel.

2 Preliminaries

2.1 Notation

We denote random variables by capital letters, for example X. If we want to specify that X has k bits we also write X^k. In addition, for a k-bit variable we can use X_i for $1 \leq i \leq k$ to denote the i-th bit of X. Analogous notation holds for fixed elements denoted by lower case letters. Moreover, for a set or bitstring S we use the common notation $|S|$ to refer to the cardinality of this set.

In addition, we use Greek letters for ratios of the numerical random variables and other numerical values. Furthermore, if we commonly use a random variable, for example R then in some contexts we still keep the notation R if the value of the variable is actually fixed. We commonly use lower case letter for denoting fixed integers.

In addition, we usually say *wiretap channel* and mean *wiretap channel of type II* for convenience throughout this document.

2.2 Model of communication

The paper follows the Shannon model of communication system on binary channels. This section gives a basic overview of this model, but the specifics of the wiretap channel are discussed later. In general, the model contains a source that generates random bits, an encoder that transforms the input according to some rules and sends it over a channel. Finally, the decoder analyses the channel output and tries to restore the initial data generated by the source.

A *source* outputs a sequence $\{S_k\}_1^\infty$ uniform binary random variables. In the following we denote this by S and will only consider fixed length blocks of these outputs. The source models the input of the system, for example a user.

An *encoder* is parametrised by (K, N) meaning that it uses a K-bit input alphabet $\{0,1\}^K$ and an N-bit output alphabet $\{0,1\}^N$, this defines a function $\mathcal{E} : \{0,1\}^K \to \{0,1\}^N$. Hence, in terms of random variables, the encoder $\mathcal{E}$ has an input S^K and an output X^N where for each specific input s there is a fixed probability for each possible output x.

In different cases the decoder either has to restore the output of the encoder or the source. For the purpose of this paper we consider the latter, as the channel properties ensure that the input of the decoder is the same as the output of the encoder. A *decoder* is a mapping $\mathcal{D} : \{0,1\}^N \to \{0,1\}^K$. The decoder with input x makes an error, if its output is not the same as the encoder input that produced x. Let $\hat{S} = (\hat{S}_1, \ldots, \hat{S}_K) = \mathcal{D}(X^N)$

be the output of the decoder, then the error rate of the decoder is defined as

$$\Pr_e = \frac{1}{K} \sum_{k=1}^{K} \Pr(S_k \neq \hat{S}_k) \ .$$

In addition, one commonly considers the channel that transmits the output of the encoder to the decoder. For example, such a channel may modify the message or delete parts of the message. For the purpose of the wiretap channel we assume that the whole message is transmitted without errors or erasures. As a more specific property of the *wiretap* channel we consider the case where someone might be eavesdropping on this channel. However, this is specified in the following, in Section 3 as this discussion requires some introduction to information theory.

2.3 Information theory

For a random variable X, the entropy measures the uncertainty about the value of this variable. This is also known as Shannon entropy and measures the amount of information in this variable. Entropy is commonly measured in bits and we only consider the binary case in this report.

For example, if X is a uniform n-bit string, then it has entropy n, but if we know that the first bit of X is always 0 and the rest are uniform, then it has entropy $n - 1$ which corresponds to the number of unknown bits.

Entropy of X is denoted by $H(X)$. If X has possible values $\{x_1, \ldots, x_n\}$, then we have

$$H(X) = -\sum_{i=1}^{n} \Pr(x_i) \cdot \log_2 \Pr(x_i) \ .$$

For example, if $X = 1$ always, then we have $H(X) = 0$ as there is no uncertainty. However, if X is a uniform bit, then $H(X) = -\frac{1}{2} \cdot \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2} = 1$. For any n-bit random variable X we have $0 \leq H(X) \leq n$.

For two random variables X and Y we can also measure their *joint entropy*. This can be thought of as the uncertainty about this pair of variables. Note that this can be extended to any set of random variables in a straightforward manner. Analogously to previous definition we have

$$H(X, Y) = -\sum_{i}^{n} \sum_{j}^{k} \Pr(x_i, y_j) \cdot \log_2 \Pr(x_i, y_j) \ .$$

We can also consider the uncertainty about X given some background information Y. This is measured using *conditional entropy* $H(X|Y)$. By definition, we have

$$H(X|Y) = -\sum_{i}^{n} \sum_{j}^{k} \Pr(x_i, y_j) \cdot \log_2 \Pr(x_i|y_j) \ .$$

In the following, we make use of the *chain rule*

$$H(X|Y) = H(X,Y) - H(Y)$$

that can be derived from these definitions.

Finally, we can also consider the relation between two random variables X and Y based on their dependency. For this purpose we define *mutual information* that measures dependency in bits. We denote it as $I(X;Y)$ and by definition

$$I(X;Y) = \sum_{i}^{n} \sum_{j}^{k} \Pr(x_i, y_j) \cdot \log_2 \frac{\Pr(x_i, y_j)}{\Pr(x_i) \cdot \Pr(y_j)} \ .$$

Based on this definition we can also write $I(X;Y) = H(X) - H(X|Y)$ and use the chain rule to obtain other relations. In addition, we also need *conditional mutual information* that is defined as

$$I(Z;Y|Z) = H(X|Z) - H(X|Y,Z).$$

Besides these notions we also need the *binary entropy function* $h(p)$. This is defined as the entropy $H(X)$ of a Bernoulli random variable X where $X = 1$ with probability p and $X = 0$ with probability $1 - p$. From the previous definitions we obtain

$$h(p) = H(X) = -p \log_2 p - (1-p) \log_2 (1-p) \ .$$

The binary entropy function and the conditional entropy are related through *Fano's inequality*. Namely, we have

$$H(X|Y) \leq h(\Pr(e)) + \Pr(e) \log_2 (|\mathcal{X}| - 1)$$

where $\mathcal{X}$ is the support of X meaning the set of values that X can take. It is commonly described in a context where X is sent on the network and Y is received and we need to estimate how much is uncertain about the original message, assume that Y decodes to some X. In addition, e means $X \neq Y$. If, on the other hand, X *does not agree with* Y, then X has to be among the $|\mathcal{X}| - 1$ other possible values that do not correspond to Y. In total $h(\Pr(e))$ is the average entropy we have for either having an error or not, and $\Pr(e) \log_2 (|\mathcal{X}| - 1)$ bounds the entropy of having some other value for X.

3 Wiretap channel of type II

3.1 Channel description

Recall the description of the communication model from the preliminaries. For a wiretap channel of type II the special property is that there is an intruder on the channel. For a parameter $\mu \leq N$ where N is the number of bits in the encoded word, the intruder can observe μ bits of the encoder output. Namely, the intruder picks a set $S \subseteq \{1, 2, \ldots, N\}$

such that $|S| = \mu$ and observes the bits of X^N corresponding to S. Let $Z^N = (Z_1, \ldots, Z_n)$ be what the intruder sees, then

$$Z_i = \begin{cases} X_i & i \in S \\ ? & otherwise \end{cases} . \tag{1}$$

Here ? denotes something unknown and the idea is that the intruder has some partial view of X^N.

The goal for the designer of a encoding and decoding algorithm for this channel is to maximise the uncertainty about the initial encoded word S^K for such an intruder. For a fixed Z^N this uncertainty is measured by $H(S^K|Z^N)$, however, we have to account for the worst possible Z^N. Define

$$\Delta = \min_{S:|S|=\mu} H(S^K|Z^N) .$$

The final goal is therefore to maximise Δ that is a bound on the uncertainty that any adversary has after observing μ bits.

Note that by previous definitions $0 \leq \Delta \leq K$. In here, $\Delta = K$ is the perfect state where no choice of the set S reveals any information about the encoded word. In the following we show that perfect security is attainable if $N \geq K + \mu$ and show how to achieve a bound for Δ in other cases.

Note that this setup is a common problem in cryptography. However, the current exposition has two significant differences from cryptographic approaches. Firstly, no cryptographic assumptions are made in this analysis and all of the following results are purely information theoretical. Specifically, we can consider adversaries with unbounded resources. Secondly, there is no secret information between the encoder and decoder meaning that all information required for decoding a string is public. In such setting, using encryption for securing the channel is not feasible, however, we see that very good results can be obtained using coding theory.

3.2 The main result

The main contributions of [OW85] are the relations between the various properties of the communication system. These properties include the length K of the plain word, length N of the encoded word, the minimal entropy Δ, the decoding error $\Pr_e$ and finally, the number μ of bits that the intruder can observe.

Definition 1 (Achievable)**.** We say that a triple (R, α, δ) is achievable if for all $\varepsilon > 0$ and all integers $N_0 > 0$ there exists an encoder and decoder pair with parameters $N \geq N_0$, $K \geq (R - \varepsilon)N$, $\mu \geq (\alpha - \varepsilon)N$, $\Delta \geq (\delta - \varepsilon)K$ and $\Pr_e \leq \varepsilon$.

The values R, α and δ can be seen as the normalised values corresponding to K, μ and Δ respectively. In the following these are easier to analyse as there are clear bounds for their values and the obtained results are more general than they would be for using the basic parameters.

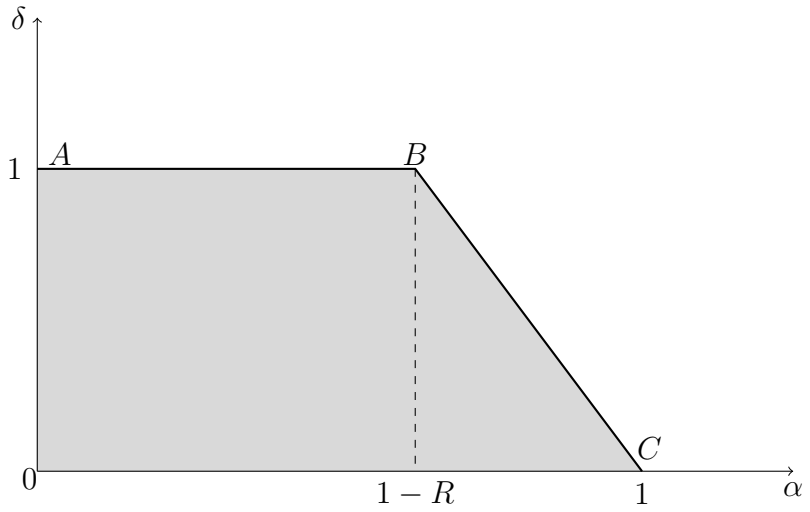

Figure 1: Achievable region for fixed R

Theorem 1 (Achievability conditions)**.** *A triple* (R, α, δ) *is achievable for* $0 \leq R, \alpha, \delta \leq 1$ *and*

$$\delta \leq \begin{cases} 1 & 0 \leq \alpha \leq 1 - R \\ \frac{1-\alpha}{R} & 1 - R \leq \alpha \leq 1 \end{cases}.$$

This theorem can be considered as a corollary of the two results proven in Sections 4 and 5. At first we show that these conditions are necessary to follow the definition of achievable triples and then we show that if these conditions hold then there also exists an encoder and decoder pair required by the achievability definition. However, before that we further analyse the meaning of this definition and the above conditions.

3.3 Analysis of achievability

The conditions in Theorem 1 can be pictured as shown on Figure 1 by the gray region if we fix a value for R. However, note that if the borders defined by points A, B and C are achievable then the region is clearly achievable. Assume that some point on this border is achievable, then, by decreasing δ we can use the same code and the same Δ as for the border point. Therefore, if there exists a code such that the border point is achievable by definition, then all points with smaller value of δ are also achievable. Thus, all points below this border point correspond to achievable triples.

In fact, the most crucial point is B. We know that the line (A, B) is defined by the perfectly secure part as in there $\alpha \leq 1 - R$ and the upper bound $\delta = 1$ is trivial. The following Lemma 1 specifies that if B is achievable then all the points on the line to C are achievable. Moreover, the point C is trivial as if the adversary can look at all the bits $\alpha \cdot N$, then trivially it will also know the encoded secret.

Intuitively, if an intruder can learn more bits, then the entropy about the original encoded string can only decrease. The following lemma however gives a bound on how

much the decrease actually is. The intuitive explanation for this is that seeing one new bit can give at most one additional bit of information.

Lemma 1. *Consider an encoder $\mathcal{E}$ and decoder $\mathcal{D}$ with parameters N, K and Pr_e. There are two intruders, first with parameters μ_1 and Δ_1 and second with μ_2 and Δ_2. Then, if $\mu_2 \geq \mu_1$, then*

$$\Delta_2 \geq \Delta_1 - (\mu_2 - \mu_1) \ .$$

Proof. Let the first intruder use indices S_1 and the second S_2 such that $S_1 \subseteq S_2 \subseteq \{1, 2, \ldots, N\}$ and $|S_i| = \mu_i$. Correspondingly, let Z_i be the string defined as in (1) for S_i. Let S^K be the random variable for the encoded word, then

$$\begin{aligned} H(S^K|Z_2) - H(S^K|Z_1) &= H(S^K|Z_2, Z_1) - H(S^K|Z_1) = -I(S^K; Z_2|Z_1) \\ &= -I(Z_2; S^K|Z_1) \geq -H(Z_2|Z_1) \geq -(\mu_2 - \mu_1) \end{aligned}$$

The first equality results from the fact that Z_2 already contains all the information of Z_1 and the second results from the definition of conditional mutual information. In the following we use the definition of conditional mutual information again, but swap the order of elements. The inequality follows from the fact that $H(Z_2|S^K, Z_1) \geq 0$ and we leave this part out of the definition. The final inequality follows from the fact that the entropy left about the extra bits in Z_2 after seeing Z_1 can not exceed $\mu_2 - \mu_1$ that is the number of additional fixed bits in Z_2.

Now, based on this we generalise it for all possible cases of adversaries. We have

$$H(S^K|Z_2) \geq H(S^K|Z_1) - (\mu_2 - \mu_1) \geq \Delta_1 - (\mu_2 - \mu_1)$$

based on the definition of Δ_1. This holds for any choice S_2, hence also for the case Δ_2. We have obtained

$$\Delta_2 \geq \Delta_1 - (\mu_2 - \mu_1)$$

as required. □

In the context of Figure 1 we have to analyse Lemma 1 for the normalised quantities δ, R and α. The statement can be rewritten as

$$\frac{\Delta_2}{K} \geq \frac{\Delta_1}{K} - \frac{(\mu_2 - \mu_1)}{K} = \frac{\Delta_1}{K} - \frac{(\mu_2/N - \mu_1/N)}{K/N} \ .$$

Hence, we can derive that a triple (R, α_1, δ_1) is achievable, then a triple (R, α_2, δ_2) is achievable for $\alpha_2 \geq \alpha_1$ and

$$\delta_2 = \delta_1 - \frac{\alpha_2 - \alpha_1}{R} \ . \tag{2}$$

If (R, α_1, δ_1) is achievable then there exists some code with parameters $K \geq RN$, $\mu_1 \geq \alpha_1 N$, and $\Delta_1 \geq \delta_1 K$. Assume, that for the second adversary and the same code we have $\mu_2 = \alpha_2 \cdot N$ and we need $\Delta_2 \geq \delta_2 K$. By Lemma 1 we have

$$\frac{\Delta_2}{K} \geq \frac{\Delta_1}{K} - \frac{(\mu_2/N - \mu_1/N)}{K/N} \geq \delta_1 - \frac{\alpha_2 - \mu_1/N}{K/N} \geq \delta_1 - \frac{\alpha_2 - \mu_1/N}{R} \geq \delta_1 - \frac{\alpha_2 - \alpha_1}{R} = \delta_2 \ .$$

With this we showed that $\Delta_2 \geq K \cdot \delta_2$ and therefore (R, α_2, δ_2) is achievable if (2) holds.

In the specific context of point B we have $\alpha_1 = 1 - R$ and $\delta_1 = 1$ and we obtain

$$\delta_2 = 1 - \frac{\alpha_2 - 1 + R}{R} = \frac{R - \alpha_2 + 1 - R}{R} = \frac{1 - \alpha_2}{R}$$

which exactly corresponds to the line (B, C) on Figure 1. Hence, in general we can prove only the achievability of point B to derive results for the whole region of achievable triples.

4 The necessity of achievability conditions

In this section we prove a theorem about the conditions of the encoder and decoder parameters. After the proof we also see how this corresponds to the achievability of the triples. In fact, this result shows that the conditions in Theorem 1 are necessary for a triple to satisfy the achievability definition. The first part of the condition is trivial, as we always have $\Delta \leq K$ as discussed before. However, the second part gives a more interesting upper bound for cases where we can not obtain perfect secrecy any more because the adversary sees too much information.

Theorem 2. *If $(K, N, \Delta, Pr_e, \mu)$ are the parameters of some code and adversary, then*

$$\Delta \leq \begin{cases} K & 0 \leq \mu \leq N - K \\ N - \mu + K \cdot h(Pr_e) & N - K \leq \mu \leq N \end{cases} .$$

Here, K is the length of the information, N is the length of the codeword, Δ is the advantage of the adversary, Pr_e is the decoding error rate and μ is the number of bits that the adversary can wiretap.

Proof. The proof mostly uses the chain rule for entropy given as

$$H(A|B) = H(A, B) - H(B)$$

in several combinations. We use these A, B and C to write out how exactly the chain rule is applied. Recall that $\hat{S}$ is the decoder output and S is a K-bit string and Z has length N. We can write

$$\begin{aligned} \Delta = H(S^K|Z^N) &= H(S, Z) - H(Z) \overset{(1)}{=} \\ &= H(S, X, Z) - H(X|S, Z) - H(Z) \overset{(2)}{=} \\ &= H(S|X, Z) + H(X, Z) - H(X|S, Z) - H(Z) \overset{(3)}{=} \\ &= H(S|X, Z) + H(X|Z) - H(X|S, Z) \overset{(4)}{=} \\ &= H(S|X, Z, \hat{S}) + H(X|Z) - H(X|S, Z) \ . \end{aligned}$$

In (1) the chain rule is applied to $H(S, Z)$ by introducing a new variable X, given the previous form we have $A = X$ and $B = (S, X)$. Secondly, for (2) the rule applies to $H(S, X, Z)$ as $B = (X, Z)$ and $A = S$. In (3) the rule is applied to $H(X, Z) - H(Z)$ as $B = Z$ and $A = X$. For the last equality (4) we just use the fact that knowing X also means knowing $\hat{S}$ as $\hat{S}$ is obtained by decoding X, therefore adding this as an extra condition does not decrease the entropy.

At the final equation, we have $H(X|Z)$ as the entropy of the $N - \mu$ coordinates that the adversary did not observe, hence $H(X|Z) \leq N - \mu$ and also, trivially $H(X|S, Z) \geq 0$. For $H(S|X, Z, \hat{S})$ we can always drop some conditions and obtain $H(S|X, Z, \hat{S}) \leq H(S|\hat{S})$. Finally, this value is estimated using Fano's inequality to obtain $H(S|\hat{S}) \leq K \cdot h(\mathrm{Pr}_e)$. The details of deriving this estimation from the Fano's inequality are given in [**?**]. Namely, the Fano's inequality can also be written as $H(S|\hat{S}) \leq K \cdot (h(\mathrm{Pr}_e) + \mathrm{Pr}_e \cdot \log(|\mathcal{S}| - 1))$ where $\mathcal{S}$ is the set where each index of S is chosen from. For our case $\mathcal{S} = \{0, 1\}$, and therefore $H(S|\hat{S}) \leq K \cdot (h(\mathrm{Pr}_e) + \mathrm{Pr}_e \cdot \log 1) = K \cdot h(\mathrm{Pr}_e)$.

Putting it all together we obtain

$$\Delta \leq N - \mu + K \cdot h(\mathrm{Pr}_e) \ .$$

Note that this is also satisfied for $0 \leq \mu \leq N - K$ as in this case $N - \mu + K \cdot h(\mathrm{Pr}_e) \geq N - N + K \cdot h(\mathrm{Pr}_e) \leq K$. We have the case distinction in the theorem as K is a trivial upper bound for Δ and this distinction makes the statement stronger. □

This result is formed in terms of the code parameters, but for achievability we have to look at it based on the values of the respective parameters α, δ and R. By definition, for $\varepsilon > 0$ there has to exist a code with parameters N, $K \geq (R - \varepsilon)N$, $\mu \geq (\alpha - \varepsilon)N$, $\Delta \geq (\delta - \varepsilon)K$ and $\mathrm{Pr}_e \leq \varepsilon$.

The first half of the bound can be analysed trivially. For $0 \leq \mu \leq N - K$ we have $\Delta \leq K$. It also gives $(\delta - \varepsilon)K \leq K$ and trivially $\delta - \varepsilon \leq 1$. The respective condition $0 \leq \mu \leq N - K$ can be written as $0 \leq \alpha + O(\varepsilon) \leq 1 - R$ by replacing $\mu \geq (\alpha - \varepsilon)N$ and $K \geq (R - \varepsilon)N$ into the condition to obtain $\alpha - \varepsilon \leq 1 - R + \varepsilon$.

For the second half, we have to be more careful. We obtain

$$\delta - \varepsilon \leq \frac{N}{K} - \frac{\mu}{K} + h(\mathrm{Pr}_e) \leq \frac{1 - \alpha + \varepsilon}{R - \varepsilon} + h(\mathrm{Pr}_e)$$

for $1 - R \leq \alpha + O(\varepsilon) \leq 1$. However, this has to hold for all $\varepsilon > 0$ and we can get the closest bound for ε approaching 0, which gives us

$$\delta \leq \frac{1 - \alpha}{R} + h(\mathrm{Pr}_e) \leq \frac{1 - \alpha}{R} + h(\varepsilon) = \frac{1 - \alpha}{R}$$

and therefore coincides with the Theorem 1. Hence, Theorem 2 implies that the conditions stated in Theorem 1 are necessary for an achievable triple.

5 Achievability conditions are practical

In this section we show that for different cases we have an error free code that satisfies the definition of achievability. This shows that the conditions in Theorem 1 are sufficient meaning that if they are met, then there also exists a code such that the achievability definition is fulfilled.

Theorem 3. *Let $1 - R < \alpha < 1$ Then, for all $\varepsilon > 0$, $N_0 \geq 1$ there exists an $N \geq N_0$ and an encoder-decoder pair with parameters $K = R \cdot N$, $\mu = \alpha \cdot N$, $\Delta/K \geq \frac{1-\alpha}{R} - \varepsilon$ and $Pr_e = 0$.*

Proof. Let N and K be fixed, then define sets A_m for $a \leq m \leq 2^K$ that partition $\{0,1\}^N$ into subsets. Hence, $A_m \subseteq \{0,1\}^N$ such that $|A_m| = 2^{N-K}$ and they are disjoint $A_m \cap A_n = \emptyset$ if $m \neq n$. For each possible codeword we choose a unique set A_m and to encode this word we choose a uniform element from A_m. Due to disjointness of A_m we have $\Pr_e = 0$ as each codeword X^N is in only one set A_m and this set corresponds to a unique codeword S^K. Hence, $H(S|X,Z) = 0$ because X uniquely determines S.

The source outputs uniform messages, hence, all 2^K messages are equally likely and therefore, also all codewords X^N are equally likely. Hence, a codeword is a uniformly distributed in X^N and we can use the common fact that in this case the coordinates of this codeword are independent uniform binary random variables. Hence, $H(X^N|Z^N) = N - \mu$ as all strings with these fixed bits are equally likely. Using the proof of Theorem 2 and the previous analysis, we obtain

$$\Delta = H(S|X,Z) + H(X|Z) - H(X|S,Z) = N - \mu - H(X^N|S^K, Z^N) \ .$$

In the following we want to show that there exists such a good partitioning that $H(X^N|S^K, Z^N)$ is small for each specific case. For this we require that each set A_m only has some small number of codewords that agree with each specific Z^K. In the following we specify that intuition.

Let $z \in \{0,1,?\}^N$ be a fixed information string of the adversary and let $x \in \{0,1\}^N$ be a codeword. We say that z is *consistent with* x if they agree on non ? symbols. Let $L \geq 1$ be an integer. We say that a partition is good, if for all m and all information strings $z \in \{0,1,?\}^N$ with $N - \mu$ unknown ? symbols we have

$$|\{x \in A_m : z \text{ is consistent with } x\}| < L \ .$$

Hence, if we have a good partition, then

$$H(X^N|S^K, Z^N) < \log_2 L$$

because knowing S^K fixes the set A_m and knowing Z^N we only consider those x that are consistent with Z^N. By definition, there are at most L and therefore the corresponding entropy is $\log_2 L$. In total, we get

$$\Delta \geq N - \mu - \log_2 L \ .$$

We state and prove the Lemma 2 that specifies that there indeed exists a good partition for

$$L > \frac{2N + K + 2\log_2 e}{K + \mu - N} \ .$$

We can define

$$\beta = \frac{1 + R + 3}{\alpha - (1 - R)}$$

whereas

$$\beta = \frac{N + NR}{N\alpha - (N - NR)} + \frac{3}{\alpha - (1 - R)} \geq \frac{N + NR}{N\alpha - (N - NR)} + \frac{2\log_2 e}{N\alpha - (N - NR)} = \\ = \frac{N + K + 2\log_2 e}{K + \mu - N}$$

where the inequality results from $2\log_2 e \leq 3$ and we use $K = RN$ and $\mu = \alpha N$. There exists a good partition with $L \leq \beta + 1$. We will show this in the following as Lemma 2 as this argument is quite long on its own and for now we use use this as a fact. Hence,

$$\Delta \geq N - \mu - \log_2(\beta + 1)$$

and we get

$$\frac{\Delta}{K} \geq \frac{N(1 - \alpha)}{RN} - \frac{\log_2(\beta + 1)}{RN} = \frac{(1 - \alpha)}{R} - \frac{\log_2(\beta + 1)}{RN} \ .$$

For satisfying the theorem we need

$$\frac{\log_2(\beta + 1)}{RN} \leq \varepsilon$$

so that $\frac{\Delta}{K} \geq \frac{(1-\alpha)}{R} - \varepsilon$. Note that, in fact β is a constant defined by R and α and we can always pick a suitable length N that satisfies the requirements.

As a remark, also note that $\beta < \infty$ and such partitioning is meaningful. Namely, for any fixed α and R we have $\beta \leq \frac{5}{\alpha - (1-R)}$ is less than some constant. Due to $\alpha > (1 - R)$ we also always have $\alpha - (1 - R) > 0$ and β is defined. □

Lemma 2. *For parameters K, N, μ such that $N - \mu - K < 0$ there exists a good partition for*

$$L > \frac{2N + K + 2\log_2 e}{K + \mu - N} \ .$$

Proof. Let $\{A_m\}$ be a partition as described in proof of Theorem 3. Let $\Psi(A_1, \ldots, A_{2^K})$ be a function on a partitioning such that $\Psi(A_1, \ldots, A_{2^K}) = 0$ if $\{A_m\}$ is good, according to the definition from Theorem 3 with respect to some value L and $\Psi(A_1, \ldots, A_{2^K}) = 1$ if the partitioning is not good.

We can write

$$\Psi(A_1, \ldots, A_{2^K}) \leq \sum_{m=1}^{2^K} \sum_{z} \phi(A_m, z)$$

where $z \in \{0, 1, ?\}^N$ is a variable that ranges over the strings that an adversary with a parameter μ sees. We have $\phi(A_m, z) = 1$, if this partition is bad, meaning

$$|\{x \in A_m : z \text{ is consistent with } x\}| \geq L$$

and in other cases $\phi(A_m, z) = 0$. In total, the inequality follows from the fact that if for a partitioning $\{A_m\}$ for some $1 \leq i \leq 2^K$ and z we have $\phi(A_i, z) = 1$ then we also have $\Psi(A_1, \ldots, A_{2^K}) = 1$. However, for this partitioning there may be more values where $\phi(A_i, z) = 1$. On the other hand, $\Psi(A_1, \ldots, A_{2^K}) = 0$ only if for all z all sets A_i are good and therefore, always $\phi(A_i, z) = 0$.

In the following we consider the expected value of Ψ. As, by definition, Ψ is a binary function, then the expected value ranges from 0 to 1. In addition, if the expected value is less than one, then there has to exist at least one such partitioning that is good. We denote the expected value of Ψ by $\mathrm{E}\Psi$. By definition of Ψ and using the linearity of expectations, we obtain

$$\mathrm{E}\Psi \leq \sum_{m=1}^{2^K} \sum_{z} \mathrm{E}\phi(A_m, z) \ .$$

The rest of the proof computes this expectation and shows $\mathrm{E}\Psi < 1$.

Define some additional quantities for shorthands in the following proof. Let $Q(z) = \{x \subseteq \{0, 1\}^N : z \text{ is consistent with } x\}$ be a set of all possible bitstrings that agree with all the known values of z. Let $q = |Q(z)| = 2^{N-\mu}$ be the size of each such set. In addition, let $n = \{0, 1\}^N = 2^N$ be the size of the set of all possible codewords and $r = |A_m| = 2^{N-K}$ be the size of each set in the partitioning.

Each member of a set A_m is chosen uniformly and random from $\{0, 1\}^N$ without replacement, meaning that no element can be in more than one set. For some $\ell \in \mathbb{N}$, the probability that exactly ℓ members of A_m belong to $Q(z)$ can be expressed as

$$\pi_\ell = \frac{\binom{q}{\ell}\binom{n-q}{r-\ell}}{\binom{n}{r}} \ .$$

In total there are $\binom{n}{r}$ ways to choose the r elements of A_m from n possible bitstrings. The case where exactly ℓ are chosen from $Q(z)$ can be obtained in $\binom{q}{\ell}$ different ways. In addition, these can be combined with the $\binom{n-q}{r-\ell}$ ways to choose the elements not in $Q(z)$ into A_m. The probability of exactly ℓ elements from $Q(z)$ is therefore the number of these cases divided by the number of all possible cases for A_m. We can simplify this expression to give a bound for the probability π_ℓ that exactly ℓ members of A_m belong

to $Q(z)$ as

$$\pi_\ell = \frac{\binom{q}{\ell}\binom{n-q}{r-\ell}}{\binom{n}{r}} \le \frac{\binom{q}{\ell}\binom{n}{r-\ell}}{\binom{n}{r}} = \frac{q(q-1)\dots(q-\ell+1)}{\ell!} \cdot \frac{\binom{n}{r-\ell}}{\binom{n}{r}} \le \frac{q^\ell}{\ell!} \cdot \frac{\binom{n}{r-\ell}}{\binom{n}{r}} =$$
$$= \frac{q^\ell}{\ell!} \cdot \frac{n!}{(n-r+l)! \cdot (r-l)!} \cdot \frac{r!(n-r!)}{n!} = \frac{q^\ell}{\ell!} \cdot \frac{r!}{(r-l)!} \cdot \frac{(n-r!)}{(n-r+l)!} =$$
$$= \frac{q^\ell}{\ell!} \cdot (r(r-1)\dots(r-\ell+1)) \cdot \frac{1}{(n-r+\ell)(n-r+\ell-1)\dots(n-r+1)}$$
$$\le \frac{q^\ell}{\ell!} \cdot r^l \cdot \frac{1}{(n-r)^\ell} = \frac{q^\ell}{\ell!} \cdot \frac{(r/n)^\ell}{(1-r/n)^\ell} = \frac{(2^{N-\mu} \cdot 2^{N-K} \cdot 2^{-N})^\ell}{\ell!(1-2^{N-K} \cdot 2^{-N})^\ell} = \frac{(2^{N-\mu-K})^\ell}{\ell!(1-2^{-K})^\ell} \ .$$

This analysis mainly uses the basic definitions of combinations and basic estimations on these combinations that follow from the fact that for each $a, b \in \mathbb{N}$ such that $a > b$ we have $(a-b)^b \le a \cdot (a-1) \cdot \ldots \cdot (a-b+1) \le a^b$. At the final state we use the definition of these variables.

In the following we use the definition of expected value and the fact that $\phi(A_m, z) = 1$ if A_m has more than or equal to L elements from $Q(z)$ and it happens with probability π_L for each z. We obtain

$$\mathrm{E}\phi(A_m, z) = \sum_{\ell=L}^{2^{N-K}} \pi_\ell \le \sum_{\ell=L}^{2^{N-K}} \frac{(2^{N-\mu-K})^\ell}{\ell!(1-2^{-K})^\ell} \le \sum_{\ell=L}^{2^{N-K}} \frac{(2^{N-\mu-K})^\ell \cdot 2^\ell}{\ell!}$$
$$\overset{(1)}{\le} (2^{N-\mu-K})^L \sum_{\ell=L}^{2^{N-K}} \frac{2^\ell}{\ell!} \le (2^{N-\mu-K})^L \sum_{\ell=0}^{\infty} \frac{2^\ell}{\ell!} = (2^{N-\mu-K})^L \cdot e^2 \ .$$

In (1) we used the assumption that $N - \mu - K < 0$ and therefore $2^{N-\mu-K} < 1$. Hence, we have $2^{N-\mu-K}L \ge 2^{N-\mu-K}\ell$ for $\ell \ge L$. After that we also use the fact that for all $\ell \ge 0$ we have $\frac{2^\ell}{\ell!} \ge 0$ and adding more elements to the sum can only increase the total value. In the end we use the series representation for $e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!}$.

Finally, we can use this in the estimation for the expected value of Ψ as

$$\mathrm{E}\Psi \le \sum_{m=1}^{2^K} \sum_z \mathrm{E}\phi(A_m, z) \le \sum_{m=1}^{2^K} \sum_z (2^{N-\mu-K})^L \cdot e^2$$
$$= \sum_{m=1}^{2^K} \sum_z 2^{(N-\mu-K)\cdot L + 2\log_2 e} = \sum_z \sum_{m=1}^{2^K} 2^{(N-\mu-K)\cdot L + 2\log_2 e}$$
$$= \sum_z 2^K \cdot 2^{(N-\mu-K)\cdot L + 2\log_2 e} \overset{(1)}{\le} 2^{2N} \cdot 2^K \cdot 2^{(N-\mu-K)\cdot L + 2\log_2 e}$$
$$= 2^{(N-\mu-K)\cdot L + 2\log_2 e + K + 2N} \ .$$

To obtain the inequality (1) we have to estimate the number of different z values for μ fixed coordinates. We know that we the adversary can choose $\binom{N}{N-\mu}$ locations where

to look at in the coded string. For each of these we have $2^\mu \leq 2^N$ different possible z values with fixed bits 0 or 1. Hence, in total we have $\binom{N}{N-\mu} \cdot 2^\mu \leq 2^N \cdot 2^N$ different values for z. Note that $\binom{N}{N-\mu} \leq 2^N$ due to the sum of the binomial coefficients that gives $2^N = \sum_{i=0}^{N} \binom{N}{i}$.

Finally, we have to analyse this value for $L > \frac{2N+K+2\log_2 e}{K+\mu-N}$. We know that $N-\mu-K < 0$ consider only the exponent

$$\begin{aligned}&(N-\mu-K)\cdot L + 2\log_2 e + K + 2N \\ &< (N-\mu-K)\cdot \frac{2N+K+2\log_2 e}{K+\mu-N} + 2\log_2 e + K + 2N\end{aligned}$$

We can analyse this in parts, using $N < \mu + K$ which also gives $N < \mu + K + 1$ and $N - 1 < \mu + K$. Using these as replacements, we get

$$\frac{N-\mu-K}{K+\mu-N} < \frac{1+K+\mu-K-\mu}{K+\mu-N} < \frac{1}{N-1-N} = -1$$

which can be replaced back into the exponent to obtain

$$\begin{aligned}&(N-\mu-K)\cdot L + 2\log_2 e + K + 2N \\ &< -1\cdot(2N+K+2\log_2 e) + 2N + K + 2\log_2 e = 0\ .\end{aligned}$$

In total we have obtained $\mathrm{E}\Psi < 2^0 = 1$ and from this we know that there has to exist at least some partitioning $\{\hat{A}_m\}$ such that $\Psi(\hat{A}_1, \ldots, \hat{A}_{2^K}) = 0$. This is the good partitioning required by the statement. □

6 Security is achievable using coset coding

The previous sections proved that secure codes exist, the original paper also proved a stronger result showing that these codes can be obtained using coset coding.

Theorem 4. *If a triple (R, α, δ) satisfies Theorem 1 then it is achievable using an encoder and decoder based on coset coding.*

For this result it is shown that the good partitioning defined in Theorem 3 is in fact achievable using coset coding. The idea is that we take some parity check matrix H that is a $K \times N$ matrix with full rank and the partitioning $\{A_m\}$ is defined by H and its cosets. Specially, for a word s all bitstrings of length N that have error syndrome equal to s are valid codewords for s. Namely, to encode s the encoder solves $H \cdot X^T = s^T$ where $X = (X_1, \ldots, X_N)$ is the codeword and finds a uniformly random solution X. In addition, the decoder gets x and just has to compute $H \cdot x^T$ to learn s^T, therefore we have $\Pr_e = 0$ as the decoder can never make an error.

The goal of the proof is to show that there exists a matrix H that defines a good partitioning as required by Theorem 2. In fact, only achievability of point B on Figure 1 is shown and the rest follows from Lemma 1 and previous reasoning.

7 Conclusion

This work showed that it is possible to design codes that preserve security against network eavesdroppers. More specifically, in the setting where the wiretapper has exactly the same information about the used code as the decoder, but can observe only limited ratio of bits on the network, it is possible to achieve perfect secrecy in many cases. Namely, as long as the adversary does not see K or more bits for a K-bit message, it is possible to achieve perfect secrecy. In other settings, it is possible to give an upper bound to the amount of information that the adversary does not know. In addition, it is feasible to design codes that satisfy these upper bounds.

References

[OW85] L.H. Ozarow and A.D. Wyner. Wire-tap channel ii. In Thomas Beth, Norbert Cot, and Ingemar Ingemarsson, editors, *Advances in Cryptology*, volume 209 of *Lecture Notes in Computer Science*, pages 33–50. Springer Berlin Heidelberg, 1985.