Protocol Coding with Reordering of User Resources: Capacity Results for the Z-Channel

Zoran Utkovski* and Petar Popovski†
∗ Faculty of Computer Science, University Goce Delcev, Stip, Republic of Macedonia
† Department of Electronic Systems, Aalborg University, Denmark
Email: zoran.utkovski@ugd.edu.mk, petarp@es.aau.dk

Abstract— **We consider protocol coding that gives a rise to secondary communication channels, defined by combinatorial ordering of the user resources (packets, channels) in a primary (legacy) communication system. In general, the capacity analysis of the secondary communication channel depends on the way the errors are introduced in the communication. Here we extend the previous results for the capacity of secondary communication channels obtained for the binary erasure channel model, to the case of the Z-channel model. This error model is of practical relevance in secondary communication channels because they are often asymmetric, i.e the probability that a packet will not be detected is much higher than the probability that noise can produce detection of a valid packet, since packet existence is detected through very robust preamble/synchronization sequences. The capacity results are obtained by modelling the secondary channel by a cascade of channels, which proves to be an effective framework for capacity analysis.**

I. Introduction

While there are continuous efforts to introduce new communication systems and standards, it is of a significant practical interest to look for the opportunities to send additional bits by minimally changing the systems that are already operating. The place to look for such an opportunity is the communication protocol and we use the term *protocol coding* to refer to strategies for sending information by using the degrees of freedom available when one needs to decide the actions taken by a particular communication protocol [1], [2]. The concept of protocol coding gives a rise to secondary communication channels, defined by combinatorial ordering of the user resources (packets, channels) in a primary (legacy) communication system. Secondary communication channels arise as result of the inherent redundancy of communication systems and protocols.

In other word, protocol coding is a concept which unleashes latent or, what we call "secondary capacity". As an example, let Alice and Bob communicate by using a primary (legacy) communication protocol in which Alice sends packets of size $n = 50$ bits to Bob. Each packet uses four of the bits for a label, which is a number $0 \cdots 15$. If Alice, instead of sending the packets as they arrive, collects 16 packets and sends them in any of the possible 16! orderings, she can encode additional $\log_2(16!) \approx 44.25$ bits in the ordering of the packets. These bits are sent through a secondary communication channel, which leverages on the degrees of freedom left unused in the legacy protocol.

The concepts of secondary communication and protocol coding were introduced in [1] and [2], where communication models were described that enable us to compute the capacity of such secondary channels under suitable restrictions imposed by the primary systems. There, capacity analysis was performed in the cases without transmission errors as well as the case with packet erasures. Forms of protocol coding can be also found in other works that mention the possibility to send data by modulating the random access protocol, see for example [3], or the seminal work [4] where information is modulated in the arrival times of data packets. More recent works on possible encoding of information in relaying scenarios through *protocol–level* choice of whether to transmit or receive is presented in [5] [6] and [2]. At a conceptual level, protocol coding bridges information theory and networking [7]. Ideas for communication based on packet reordering have been presented in the context of covert channels [8] [9]. There is also a relation to [10], where a set of packets is randomly permuted and is useful in determining the rate/delay tradeoffs when transmitting temporally ordered content over multipath routed networks. The practical coding strategies are related to the frequency permutation arrays for power line communications [11], [12].

Besides the obvious practical importance of this concept, it is also of information-theoretic value. In [1] it is shown that there is a relation between the capacity of secondary channels to the capacity of channels with causal channel state information at the transmitter (CSIT), originally considered by Shannon. By using the specific communication setup, in [13] an alternative framework for achieving the capacity was developed. There, the secondary communication channel was represented through a cascade of channels and coding strategies that need to be used over the secondary channels were discussed.

The representation through a cascade of channels, brings modularity to the problem of finding the capacity under different scenarios. In general, the capacity analysis of the secondary communication channel depends on the way the errors are introduced in the communication, i.e. on the underlying model for the transmission errors. In the previous works [1], [2], [13], we focused on the case with packet erasures, based on the block BEC (binary erasure channel). While erasures are suitable for modelling the error process in secondary communication channels, other error models are possible, where an incorrectly received packet address can be confused with another address. Let us take the example when address 0 is an “empty” user, while address 1 means that there is a packet transmission (irrespective to which user it is addressed). In that case, there is a probability that a packet will not be detected, i. e. a probability that 1 is interpreted as 0. On the other hand, the probability that noise can produce detection of a valid packet, i. e. 0 interpreted as 1 is practically zero, since packet existence is detected through very robust preamble/synchronization sequences. This other error model corresponds to the Z-channel [15], which is a special form of an binary asymmetric channel, and is the model of interest in this paper.

In practice, a secondary channel can be defined over virtually any existing wireless system and it is of interest to find the coding strategies that are suited to a certain primary system. In this sense, this analysis is the first step towards an unified solution for the case of more general error models. Additionally, it can give an insight in the coding strategies that are approaching the capacity.

II. System Model

We introduce the following communication model. We consider a primary system in which a Base Station (BS) communicates with K primary users using Time Division Multiple Access (TDMA), i. e. only one user receives data form the BS at a time. The BS serves the users in scheduling frames. Each frame consists of F packet transmissions, addressed to $m \leq F$ primary users, where $m \leq F$. Each packet carries the address of a user to whom the packet is destined. We also allow for empty packet slots being addressed to an “empty” user address, such that an empty frame slot actually can be treated as a valid packet. We point out that, in general, we do not have to constrain ourselves on the above mentioned system. The concepts of secondary communication and protocol coding are generic in nature and can be applied to any system where resources such as packets and users and more general, such as time and frequency, can be combinatorially ordered.

Unless stated otherwise, in the sequel we will assume that the number of different packet types in a frame is $m \leq 2$, i. e. each packet in a frame is addressed either to user 0 or user 1. This setup is sufficient to illustrate the main concepts and ideas when capacity of these systems is analyzed.

The key assumption in the model is that the packets that are scheduled in a frame are decided by the primary communication system. This means that the primary system decides that s packets in a frame will be sent to user 0 and $(F-s)$ packets will be addressed to user 1, where $0 \leq s \leq F$. This assumption captures the fact that the secondary communication is restricted in selecting its communication symbol, i.e. the secondary communication does not affect the primary communication.

The number of packets s addressed to user 1 in a given frame is called *state* of the frame. We assume that the process by which the primary system selects packets for transmission is a memoryless random process: in each frame, a packet is addressed 1 with probability q and 0 with probability $1-q$, independently of the other packets and the previous frames. Hence, the probability that a frame is in state s is:

$$P_S(s) = \binom{F}{s} q^s (1-q)^{F-s}. \qquad (1)$$

As already mentioned, the packets that are scheduled in a frame are decided by the primary communication system, leading to variable and unpredictable amount of information that can be sent over the secondary system. This is the key property of the secondary communication. The freedom left to the secondary transmitter is to re-arrange the packets in a frame.

Erasures are suitable to model the error process in the secondary communication system. We have addressed the problem in the case with erasures in [1] and [13]. Besides the erasure model, other error models are possible. We are going to address the case where the errors are generated according to the Z- channel model. As already discussed, this approach is justified since there is a certain probability that a packet will not be detected, i. e. a probability that 1 is interpreted as 0. On the other hand, the probability that noise can produce detection of a valid packet, i. e. 0 interpreted as 1 is practically zero, since packet existence is detected through very robust preamble/synchronization sequences. We use $\mathbf{x}$ and $\mathbf{y}$ to denote the transmitted and the received symbol, respectively, by using protocol coding for secondary communication. If not stated otherwise, we will always refer to the symbols, bits, etc. sent over the secondary communication channel. Both $\mathbf{x}$ and $\mathbf{y}$ are $F-$dimensional vectors, since each of them consists of F packets within the primary communication system. In the case of the Z-channel, $\mathbf{x}, \mathbf{y} \in \mathcal{X} = \{0, 1, \ldots K-1\}^F$.We use $\mathcal{S} = \{0, 1, \ldots F\}$ to denote the set of possible states.

III. Capacity Analysis: Preliminaries

A. Relation to the Shannon’s Model with Causal State Information at the Transmitter (CSIT)

The secondary communication channel can be represented by the framework that Shannon used to derive the capacity of channels with causal state information at the transmitter (CSIT) [14], as done in [1]. Shannon showed that instead of considering the original channel with CSIT, one can consider an ordinary, discrete memoryless channel with equivalent capacity that has a larger input alphabet. The input variable of the equivalent channel is denoted by T and each possible input letter t, termed *strategy* [14], represents a mapping from the state alphabet $\mathcal{S}$ to the input alphabet $\mathcal{X}$ of the original channel. Thus, a particular strategy $t \in \mathcal{T}$ is defined by the vector of size $|\mathcal{S}|$:

$$(t(1), \ldots t(|\mathcal{S}|)) \qquad \forall s, t(s) \in \mathcal{X}. \qquad (2)$$

We note that in this model, the cardinality $|\mathcal{T}|$ can be much larger than the cardinality $|\mathcal{X}|$ of the original system, for example in the order of $|\mathcal{X}|^{|\mathcal{S}|}$. The capacity of the equivalent channel can be found as:

$$C = \max_{P_T(\cdot)} I(T, \mathbf{Y}), \tag{3}$$

where $P_T(\cdot)$ is a probability distribution defined over the set $\mathcal{T}$ which is independent of the state S. The maximization is performed across all the joint distributions that satisfy [14]:

$$P_{S,T,\mathbf{X},\mathbf{Y}}(s,t,\mathbf{x},\mathbf{y}) = P_S(s)P_T(t)\delta(\mathbf{x},t(s))P_{\mathbf{Y}|\mathbf{X},S}(\mathbf{y}|\mathbf{x},s), \tag{4}$$

where $\delta(\mathbf{x},t(s)) = 1$ if $\mathbf{x} = t(s)$ and $\delta(\mathbf{x},t(s)) = 0$ otherwise. Following the properties of mutual information ([15], Section 8.3), in order to achieve the capacity in (3), the required cardinality of $\mathcal{T}$ is at most $|\mathcal{Y}|$.

When Shannon's results are applied to the model of a secondary communication channel, additional remarks are in order. The first thing to be noted is that, for a given state $S = s$ only a subset $\mathcal{X}_s \in \mathcal{X}$ of symbols $\mathbf{x}$ may be produced. For example, when $F = 4$ and the state is $s = 2$ it is not possible to send the symbol $\mathbf{x} = 1011$. In general, the set of transmittable secondary symbols $\mathcal{X}$ can be partitioned in $|\mathcal{S}| = F + 1$ different subsets, defined as:

$$\mathcal{X}_s = \left\{ \mathbf{x} | \sum_{i=1}^{F} x_i = s \right\}. \tag{5}$$

However, it should be noted that in the model with causal CSIT the distribution $P_{\mathbf{Y}|\mathbf{X},S}(\mathbf{y}|\mathbf{x},s)$ needs to be defined for *all pairs* $(\mathbf{x}, s)$, irrespective of the fact that in the original model some $\mathbf{x}$ are incompatible with s, i. e. when the state is $S = s$, the symbols $\mathbf{x} \notin \mathcal{X}_s$ cannot be sent. In order to deal with this situation, we need to extend the model.

We assume that the channel $\mathbf{X} - \mathbf{Y}$ is defined and thus the conditional distribution $P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$ is specified. Given $P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$, we define $P_{\mathbf{Y}|\mathbf{X},S}(\mathbf{y}|\mathbf{x},s)$ in the following way For each $\mathbf{x}_u \notin \mathcal{X}_s$ we take one $\mathbf{x}_v \in \mathcal{X}_s$ and define:

$$P_{\mathbf{Y}|\mathbf{X},S}(\mathbf{y}|\mathbf{x}_u,s) \equiv P_{\mathbf{Y}|\mathbf{X},S}(\mathbf{y}|\mathbf{x}_v,s) \qquad \forall \mathbf{y} \in \mathcal{Y}. \tag{6}$$

The idea to do this is based on the following: For example, when $F = 4$ and $s = 0$ only the symbol $\mathbf{x} = 0000$ can be sent; but we can look at it in another way — when $s = 0$ only the symbol $\mathbf{y} = 0000$ can be received when there are no errors (and the corresponding versions of 0000 when erasures occur). Thus, when $s = 0$, we can think that we can send any $\mathbf{x}$, but at the output we can receive only 0000 and the erroneous versions of it. In that case picking a strategy t'' in which $t''(s) = \mathbf{x}_u$ is equivalent to picking the strategy t' in which $t'(s) = \mathbf{x}_v$. In short, we define $P_{\mathbf{Y}|\mathbf{X},S}$ in order, for given s, to discourage selection of symbols $\mathbf{x}$ for which $\mathbf{x} \neq \mathbf{y}$ in absence of channel errors.

According to [13], the capacity of the secondary channel with memoryless state change across frames is given by (3) where the cardinality of the set of reduced strategies $\mathcal{T}$ satisfies:

$$|\mathcal{T}| = \prod_{s=0}^{F} \binom{F}{s}. \tag{7}$$

As pointed out in [14], expressing the capacity in terms of strategies might pose some conceptual and practical problems for code construction and implementation when F is large. Motivated by this observation, as well as by the specific way in which the set of states partitions the possible set of transmitted symbols $\mathcal{X}$, in [13] a different framework for computing the capacity for protocol coding based on reordering of user resources, was created. In the following we present the main aspects of this framework.

B. Capacity Analysis through a Cascade of Channels

The specific structure of the transition probabilities enable us to use models that can more easily lead to capacity characterization. Recall that T is the auxiliary random variable defined over the reduced set of possible strategies $\mathcal{T}$, where reduction is done according to Proposition 7. For given $T = t$ and each $s \in \mathcal{S}$ there is a single $t(s) \in \mathcal{X}_s$. Due to the randomized state change, each fixed $t \in \mathcal{T}$ induces a distribution on $\mathcal{X}$. In general, we can define the following transition probabilities:

$$P_{\mathbf{X}|T}(\mathbf{x}|t) = \sum_{s=0}^{F} \delta(\mathbf{x}, t(s)) P_S(s), \tag{8}$$

where $\delta(\mathbf{x}, t(s)) = 1$ if $\mathbf{x} = t(s)$ and is 0 otherwise. It is easily seen that:

$$P_{\mathbf{X}|T}(\mathbf{x}|t) = \begin{cases} P_S(s^*) & \text{if } \exists s^*, t(s^*) = \mathbf{x} \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

In this way, we do not need to explicitly consider state in the capacity analysis, but instead we model the secondary communication channel by using a cascade of two channels:

$$T - \mathbf{X} - \mathbf{Y}, \tag{10}$$

where $P_{\mathbf{X}|T}$ and $P_{\mathbf{Y}|\mathbf{X}}$ are well–defined. In order to express the mutual information $I(T; \mathbf{Y})$, we use:

$$\begin{aligned} I(T, \mathbf{X}; \mathbf{Y}) &= I(T; \mathbf{Y}) + I(\mathbf{X}; \mathbf{Y}|T) \\ &= I(\mathbf{X}; \mathbf{Y}) + I(T; \mathbf{Y}|\mathbf{X}) \end{aligned} \tag{11}$$

Due to the Markovian properties, T and $\mathbf{Y}$ are conditionally independent given $\mathbf{X}$, such that $I(T; \mathbf{Y}|\mathbf{X}) = 0$, which implies:

$$I(T; \mathbf{Y}) = I(\mathbf{X}; \mathbf{Y}) - I(\mathbf{X}; \mathbf{Y}|T). \tag{12}$$

Our objective is to maximize this mutual information. For this reason, we will perform individual analysis of the terms $I(\mathbf{X}; \mathbf{Y}$ and $I(\mathbf{X}; \mathbf{Y}|T)$.

To facilitate the discussion, in the rest of the paper we will use the terms "strategies" and "input symbols" interchangeably and we can equivalently treat the set $\mathcal{T}$ as consisting of the input symbols $\{1, 2, \ldots |\mathcal{T}|\}$.

We will use the following terminology: If the probability $P_{\mathbf{X}|T}(\mathbf{x}|t) > 0$, then $\mathbf{x}$ is a *representative* of t. According to the capacity results for channels with causal CSIT, each $T = t$ has a single representative in each $\mathcal{X}_s$, which will be denoted by $\mathbf{x}_s(t)$. In order to avoid further confusion and noticing that the ordering of input symbols $1, 2, \ldots \in \mathcal{T}$ is arbitrary, the following can be noted: instead of speaking about which strategies out of $\mathcal{T}$ that are chosen with non–zero probability, we can equivalently speak of which representatives to choose for given $T = t$. The set of representatives $\mathcal{M}_t = \{\mathbf{x}_s(t)\}$ for given t will be called a *multisymbol* of t. Additionally, the multisymbol which has representatives defined as follows:

$$x_s(t) = \begin{cases} 0 & \text{if } i \leq F - s \\ 1 & \text{otherwise} \end{cases} \tag{13}$$

will be called *basic* and will be denoted as $\mathcal{M}_b$.

IV. Capacity analysis with errors modelled by the Z-channel

In this section we present the main results in this paper. The aim is to find the capacity of the secondary channel represented as a cascade of the channels $T - \mathbf{X} - \mathbf{Y}$, in the case where the error process is modeled by the Z-channel model. We recall that the capacity of the binary Z-channel with crossover probability ϵ is given by

$$C_Z(\epsilon) = \log_2 \left(1 + (1-\epsilon)\epsilon^{\epsilon/(1-\epsilon)}\right) \tag{14}$$

and is achieved by non-uniform distribution over $\mathcal{X}$.

Our objective is to find the pair of distributions $\big(P_T(\cdot), P_{\mathbf{X}|T}(\cdot)\big)$ that maximizes $I(T; \mathbf{Y})$. Thus, the capacity of the secondary channel can be written as:

$$C = \max_{P_T(\cdot), P_{\mathbf{X}|T}(\cdot)} I(T; \mathbf{Y}). \tag{15}$$

For brevity, we will always assume that $P_{\mathbf{X}|T}(\cdot) \in \mathcal{P}_{\mathbf{X}|T}$, without noting it explicitly. The expression (15) can be upper–bounded as follows:

$$C \leq \max_{P_T(\cdot), P_{\mathbf{X}|T}(\cdot)} I(\mathbf{X}; \mathbf{Y}) - \min_{P_T(\cdot), P_{\mathbf{X}|T}(\cdot)} I(\mathbf{X}; \mathbf{Y}|T) \tag{16}$$

where the equality is achieved if and only if there is a pair of distributions $\big(P_T(\cdot), P_{\mathbf{X}|T}(\cdot)\big)$ that simultaneously attains the maximum and the minimum in the first and the second term, respectively. In the sequel we will decompose the problem (15) into two sub–problems, maximization of $I(\mathbf{X}; \mathbf{Y})$ and minimization of $I(\mathbf{X}; \mathbf{Y}|T)$.

A. Analysis of $I(\mathbf{X}; \mathbf{Y})$

Let us define:

$$C_{XY} = \max_{P_{\mathbf{X}} \in \mathcal{P}_{\mathcal{X},\mathcal{S}}(\cdot)} I(\mathbf{X}; \mathbf{Y}). \tag{17}$$

Note that here the maximization is not done by considering all possible distributions $P_{\mathbf{X}}(\cdot) \in \mathcal{P}_{\mathcal{X}}$, but rather only distribution from the subset $\mathcal{P}_{\mathcal{X},\mathcal{S}} \subset \mathcal{P}_{\mathcal{X}}$ that satisfies the constraints posed on the input distribution by the primary packet scheduler. The set $\mathcal{P}_{\mathcal{X},\mathcal{S}}$ is defined as:

$$\mathcal{P}_{\mathcal{X},\mathcal{S}} = \left\{ P_{\mathbf{X}}(\cdot) | \sum_{\mathbf{x} \in \mathcal{X}_s} P_{\mathbf{X}}(\mathbf{x}) = P_S(s), \forall s = 0, 1, \cdots F \right\} \tag{18}$$

Clearly, the capacity C_{XY} is upper bounded by the capacity of the Z-channel. This is because the capacity–achieving distribution for the Z-channel requires a specific, non-uniform input distribution. In order to attain C_{XY}, the set $\mathcal{T}$, the distribution $P_T(\cdot)$ and the representatives of each $T = t$ (i. e. the distribution $P_{\mathbf{X}|T}(\cdot)$) should be carefully chosen.

In this text we are interested in channels $\mathbf{X} - \mathbf{Y}$ with a particular structure, where each single channel use $\mathbf{x}$ consists of F uses of a more elementary, identical channels. Therefore, the following symmetry takes place: the set of transition probabilities $\{P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})\}$ is identical for all $\mathbf{x} \in \mathcal{X}_s$, as they are all permutations of a vector with s 1s and $F - s$ 0s. This is valid irrespective of the the type of elementary channel that affects a single transmission of a primary packet. Such a symmetry is instrumental for making statements about C_{XY}. It can be shown that the distribution $P_{\mathbf{X}}(\cdot) \in \mathcal{P}_{\mathcal{X},\mathcal{S}}$ that achieves C_{XY} is:

$$P_{\mathbf{X}}(\mathbf{x}) = \frac{P_S(s)}{\binom{F}{s}} \tag{19}$$

i. e. all the inputs $\mathbf{x}$ that belong to the same $\mathcal{X}_s$ are equiprobable. Due to the lack of space, we present the result without proof.

As already discussed, in order to attain C_{XY}, the set $\mathcal{T}$, the distribution $P_T(\cdot)$ and the representatives of each $T = t$ (i. e. the distribution $P_{\mathbf{X}|T}(\cdot)$) should be chosen such that (19) is satisfied. These aspects are out of the scope of this paper and are a topic of current work.

B. Analysis of $I(\mathbf{X}; \mathbf{Y}|T)$

We recall that according to the analysis of the expression for mutual information (12), its second member can be written as:

$$I(\mathbf{X}; \mathbf{Y}|T) = \sum_{t=1}^{|\mathcal{T}|} P_T(T = t) I(\mathbf{X}; \mathbf{Y}|T = t). \tag{20}$$

With a slight notational abuse, we use $I(\mathbf{X}; \mathbf{Y}|\mathcal{M}_t)$ to denote the mutual information conditioned on $T = t$, where the multisymbol for $T = t$ is chosen to be $\mathcal{M}_t$. The question to be addressed is, how to choose the representatives $\mathbf{x}_s(t)$ that constitute the multisymbol $\mathcal{M}_t$ of t in order to minimize $I(\mathbf{X}; \mathbf{Y}|\mathcal{M}_t)$.

For the channel with erasures, this problem has been addressed in [13], where it has been proven that a necessary and sufficient condition to minimize $I(\mathbf{X}; \mathbf{Y}|\mathcal{M}_t)$ is that the multisymbol $\mathcal{M}_t$ of t is a permutation of the basic set of representatives.

Without going into details, we only say that the proof provided in [13] follows from two lemmas. The first lemma states that the value of $I(\mathbf{X}; \mathbf{Y}|\mathcal{M}_t)$ is minimized if the multisymbol $\mathcal{M}_t = \{\mathbf{x}_t(s)\}$ is chosen such that for the Hamming distance it holds $d_H(\mathbf{x}_s(t), \mathbf{x}_{s+1}(t)) = 1$ for each $s = 0 \ldots F-1$. The second lemma states that any multisymbol that achieves the minimal value of of $I(\mathbf{X}; \mathbf{Y}|\mathcal{M}_t)$ must be one of the $F!$ possible permutations $\Pi(\mathcal{M}_b)$ of the basic multisymbol $\mathcal{M}_b$.

Since the error model described by the Z-channel is different from the erasure channel, the question is if a similar conclusion holds in this case as well.Interestingly, we are going to show that the same conclusion holds also for the case when the errors are modelled by the Z-channel. In order to prove this statement, we will start by writing the mutual information $I(\mathbf{X};\mathbf{Y}|\mathcal{M}_t)$ in the form,

$$I(\mathbf{X};\mathbf{Y}|\mathcal{M}_t) = H(\mathbf{Y}|\mathcal{M}_t) - H(\mathbf{Y}|\mathbf{X},\mathcal{M}_t). \qquad (21)$$

Careful examination of the term $H(\mathbf{Y}|\mathbf{X},\mathcal{M}_t)$ reveals that this term does not depend on the choice of the multisymbol. Therefore, the multisymbol $\mathcal{M}_t = \{\mathbf{x}_s(t)\}$ should be chosen such that $H(\mathbf{Y}|\mathcal{M}_t)$ is minimized. This conditional entropy can be written as

$$H(\mathbf{Y}|\mathcal{M}_t) = -\sum_{\mathbf{y}\in\mathcal{Y}} P_{\mathbf{Y}|T}(\mathbf{y}|t)\cdot\log_2 P_{\mathbf{Y}|T}(\mathbf{y}|t), \qquad (22)$$

where

$$P_{\mathbf{Y}|T}(\mathbf{y}|t) = \sum_{s=0}^{F} P_{\mathbf{X}}(\mathbf{x}_s(t))P_{\mathbf{Y}|\mathbf{X},T}(\mathbf{y}|\mathbf{x}_s(t)). \qquad (23)$$

Given the Z-channel model, the transition probability $P_{\mathbf{Y}|\mathbf{X},T}(\mathbf{y}|\mathbf{x}_s(t))$ is given by

$$P_{\mathbf{Y}|\mathbf{X},T}(\mathbf{y}|\mathbf{x}_s(t)) = \epsilon^d(1-\epsilon)^{s-d}\cdot g_s(\mathbf{y},\mathbf{x}_s), \qquad (24)$$

where d is the Hamming distance between $\mathbf{y}$ and $\mathbf{x}_s(t)$, $d = d_H(\mathbf{y},\mathbf{x}_s(t))$and $g_s(\mathbf{y},\mathbf{x}_s(t)) = 1$ when $P_{\mathbf{Y}|\mathbf{X},T}(\mathbf{y}|\mathbf{x}_s(t)) \neq 0$ and 0 otherwise. In order to demonstrate how the conditional entropy $H(\mathbf{Y}|\mathcal{M}_t)$ depends on the choice of the multisymbol $\mathcal{M}_t$, we are going to present an example with frame length $F = 3$.

C. Example $F = 3$

We start with the **basic** multisymbol $\mathcal{M}_b = \{000,\ 001,\ 011,\ 111\}$, where the Hamming distance between two consecutive elements is 1. We concentrate for the moment on the set $\mathcal{Y}_1 = \{001, 010, 100\}$, i.e. the subset of $\mathcal{Y}$ corresponding to the state $s = 1$. The representation by the cascade of channels $T - \mathbf{X} - \mathbf{Y}$, together with the channel transitions, is shown in Fig. 1, where the left-hand part corresponds to the the basic multisymbol, illustrated with $T = 1$. For simplicity and for the purpose of the worked-out example, for the channel $\mathbf{X}-\mathbf{Y}$ only the transitions associated with the subset $\mathcal{Y}_1$ are shown We define

$$p_i^{(1)} = \frac{\epsilon^{i-1}(1-\epsilon)\binom{3}{i}}{2^3}, \qquad (25)$$

where the superscript "(1)" stands for $s = 1$. We note that $p_1^{(1)} \geq p_2^{(1)} \geq p_3^{(1)}$, $\forall\epsilon \in [0,1]$. With this, the involved conditional probabilities are given by $P_{\mathbf{Y}|T}(001|t) = p_1^{(1)} + p_2^{(1)} + p_3^{(1)}$, $P_{\mathbf{Y}|T}(010|t) = p_2^{(1)} + p_3^{(1)}$ and $P_{\mathbf{Y}|T}(100|t) = p_3^{(1)}$. Additionally, we denote $P_{1,2,3}^{(1)} = p_1^{(1)} + p_2^{(1)} + p_3^{(1)}$, $P_{2,3}^{(1)} = p_2^{(1)} + p_3^{(1)}$ and $P_3^{(1)} = p_3^{(1)}$. For the basic multisymbol, the contribution to the entropy $H(\mathbf{Y}|\mathcal{M}_b)$ which is a result of $\mathcal{Y}_1$ can be written as

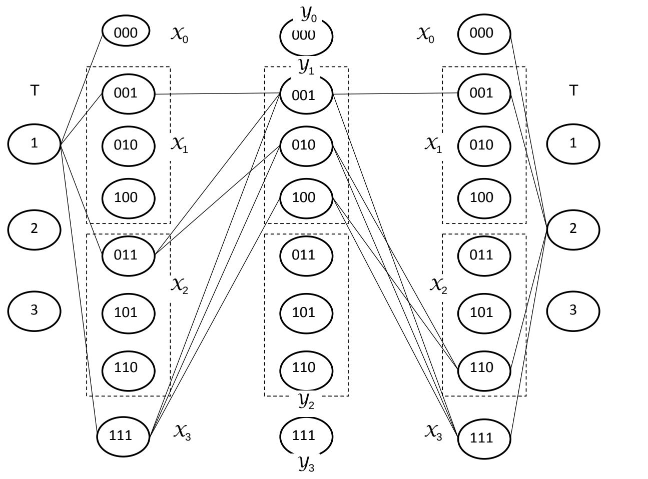

Fig. 1. Example: Choice of the multisymbols for $F = 3$, $T = 1$ and $T = 2$

.

$$\begin{aligned} H_{1,b}(\mathbf{Y}|\mathcal{M}_b) &= -\sum_{\mathbf{y}\in\mathcal{Y}_1} P_{\mathbf{Y}|T}(\mathbf{y}|t)\cdot\log_2 P_{\mathbf{Y}|T}(\mathbf{y}|t) \\ &= -\,(P_{1,2,3}^{(1)}\log_2 P_{1,2,3}^{(1)} + P_{2,3}^{(1)}\log_2 P_{2,3}^{(1)} \\ &\quad + P_3^{(1)}\log_2 P_3^{(1)}). \end{aligned} \qquad (26)$$

Now, let us take a multisymbol $\mathcal{M}_m$ which is a permutation of the basic multisymbol, i.e. for the Hamming distance between the consecutive symbols it holds

$$d_H(\mathbf{x}_s(t), \mathbf{x}_{s+1}(t)) = 1. \qquad (27)$$

The multisymbols that have the property (27) are of special interest and will be termed *minimal* multisymbols. For given F, there are $F!$ different minimal multisymbols. Indeed, any minimal multisymbol can be obtained from the basic multisymbol by a permutation π of size F, by defining the following operation:

$$\mathcal{M}_m = \gamma_\pi(\mathcal{M}_b), \qquad (28)$$

where the operation $\gamma_\pi(\cdot)$ is defined as follows: each element $\mathbf{x}' \in \mathcal{M}'$ is obtained from one element $\mathbf{x}$ of $\mathcal{M}$ by rearranging the components of $\mathbf{x}$ according to the same selected permutation. For example, if $F = 3$, the set of basic representatives is $\mathcal{M}_b = \{000, 001, 101, 111\}$ and the permutation used to define $\Pi(\mathcal{M}_b)$ is 132, then the obtained set of permuted representatives is $\mathcal{M}_m = \{000, 010, 110, 111\}$. It is easily verified that the operation $\gamma_\pi(\cdot)$ preserves the Hamming distances between any two representatives $\mathbf{x}_{s_1}, \mathbf{x}_{s_2}$:

$$d_H(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}) = d_H(\mathbf{x}'_{s_1}, \mathbf{x}'_{s_2}) = s_2 - s_1. \qquad (29)$$

Any minimal multisymbol $\mathcal{M}_m$ can be obtained by specifying a permutation π_m of length F and applying $\mathcal{M}_m = \gamma_{\pi_m}(\mathcal{M}_b)$. Using contradiction, it can be proved that any minimal multisymbol must be obtained from the basic multisymbol by applying $\gamma_\pi(\cdot)$, such that there are in total $F!$ different minimal multisymbols.

As an example of a minimal multisymbol we take $\mathcal{M}_m = \{000, 010, 110, 111\}$. It is easy to check that the partial contribution to the entropy $H(\mathbf{Y}|\mathcal{M}_m)$ is given by

$$\begin{aligned} H_{1,m}(\mathbf{Y}|\mathcal{M}_m) =& - \sum_{\mathbf{y}\in\mathcal{Y}_1} P_{\mathbf{Y}|T}(\mathbf{y}|t) \cdot \log_2 P_{\mathbf{Y}|T}(\mathbf{y}|t) \\ =& - (P_{1,2,3}^{(1)} \log_2 P_{1,2,3}^{(1)} + P_{2,3}^{(1)} \log_2 P_{2,3}^{(1)} \\ &+ P_3^{(1)} \log_2 P_3^{(1)}), \end{aligned} \tag{30}$$

where $P_{1,2,3}^{(1)}, P_{2,3}^{(1)}$ and $P_3^{(1)}$ are the same as in the case of the basic multisymbol, leading to $H_{1,m}(\mathbf{Y}|\mathcal{M}_b) = H_{1,b}(\mathbf{Y}|\mathcal{M}_n)$, i.e $H(\mathbf{Y}|\mathcal{M}_m) = H(\mathbf{Y}|\mathcal{M}_b)$.

Now, let us take another multisymbol, $\mathcal{M}_n = \{000,\ 001,\ 110,\ 111\}$, which is not a permutation of the basic multisymbol (and thus not a minimal multisymbol), i.e. the Hamming distance between two consecutive symbols is not always 1. For this multisymbol, in the right-hand part of Fig. 1, we show the transitions for the cascade of channels $T - \mathbf{X} - \mathbf{Y}$. The choice of the multisymbol $\mathbf{M}_n$ corresponds to the strategy $T = 2$. In this case the contribution to the entropy $H(\mathbf{Y}|\mathcal{M}_n)$ is given by

$$\begin{aligned} H_{1,n}(\mathbf{Y}|\mathcal{M}_n) =& - (P_{1,3}^{(1)} \log_2 P_{1,3}^{(1)} + P_{2,3}^{(1)} \log_2 P_{2,3}^{(1)} \\ &+ P_{2,3}^{(1)} \log_2 P_{2,3}^{(1)}). \end{aligned} \tag{31}$$

In order to compare $H_{1,b}(\mathbf{Y}|\mathcal{M}_n)$ with $H_{1,n}(\mathbf{Y}|\mathcal{M}_n)$, we first present the following property of entropy

Property 1: Let $P = \{p_1, \ldots, p_i, \ldots, p_j, \ldots, p_n\}$ be a set such that $p_1 \geq p_2 \geq \cdots \geq p_n \geq 0$. Let us define $Q = \{q_1, \ldots, q_i, \ldots, q_j, \ldots, q_n\}$ such that $q_i = p_i - \Delta$, $q_j = p_j + \Delta$, where $\Delta \leq \mathrm{p_i - p_j}$ and $\mathrm{q_k = p_k, \forall\ k \neq i, j}$. Then $H(\mathcal{P}) \leq H(\mathcal{Q})$. The proof of this property is follows from the Jensen inequality.

With this it is straightforward to conclude that $H_{1,b}(\mathbf{Y}|\mathcal{M}_b) \leq H_{1,n}(\mathbf{Y}|\mathcal{M}_n)$. Indeed, we denote $\mathcal{P} = \{P_{1,2,3}^{(1)},\ P_{2,3}^{(1)},\ P_3^{(1)}\}$, where it holds that $P_{1,2,3}^{(1)} \geq P_{2,3}^{(1)} \geq P_3^{(1)}$. For the second multisymbol we denote $\mathcal{Q} = \{P_{1,3}^{(1)},\ P_{2,3}^{(1)},\ P_{2,3}^{(1)}\}$, where $P_{1,3}^{(1)} \geq P_{2,3}^{(1)}$. Since $P_{2,3}^{(1)} = P_{1,2,3}^{(1)} - P_2^{(1)} = P_3^{(1)} + P_2^{(1)}$, it follows from the above property that $H_{1,b}(\mathbf{Y}|\mathcal{M}_b) \leq H_{1,n}(\mathbf{Y}|\mathcal{M}_n)$. The same observation holds for all $s = 0, 1, 2, 3$.

Hence, the choice of the multisymbol $\mathcal{M}_t$ to be the basic multisymbol, $\mathcal{M}_t = \mathcal{M}_b$, or a permutation of it, minimizes the entropy $H(\mathbf{Y}|\mathcal{M}_t) = \sum_{s=0}^{3} H_s(\mathbf{Y}|\mathcal{M}_t)$. Consecutively, $I(\mathbf{X};\mathbf{Y}|\mathcal{M}_t)$ is minimized, which was required in order to maximize $I(T;\mathbf{Y})$.

D. The general case

Using the analogy with the presented example, we state the following result for the general case

Theorem 1: A necessary and sufficient condition to minimize $I(\mathbf{X};\mathbf{Y}|\mathcal{M}_t)$ is that the multisymbol of $\mathcal{M}_t$ is a permutation of the basic set of representatives.

Proof: Let $\mathcal{M}_b$ be the basic set of representatives. We introduce the following notation. For the channel $\mathbf{X} - \mathbf{Y}$, we say that there is a link between $\mathbf{x}$ and $\mathbf{y}$ if $p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) \neq 0$. The elements of $\mathcal{X}$ and $\mathcal{Y}$ are ordered in ascending order (according to the binary notation, starting from $0 \cdots 0$). We denote by $l_{s,k}$ the number of links (defined by the Z-channel model) ending at the k-th element $\mathbf{y}_{s,k}$ of $\mathcal{Y}_s$.For the basic multisymbol, the number of links ending at $\mathbf{y}_{s,k}$ is given by

$$l_{s,k} = \left\{ \begin{array}{rl} F - s + 1 & \text{for } \ k = 1 \\ \vdots & \\ F - s - (j-1) & \text{for } \ \binom{s+j}{s} + 1 \leq k \leq \binom{s+j+1}{s} \\ \vdots & \\ 1 & \text{for } \ \binom{F-1}{s} + 1 \leq k \leq \binom{F}{s} \end{array} \right. \tag{32}$$

where $j = 0, 1, \ldots, F - s$. We note that each link $l_{s,k}$ arises from different symbol (representative) from the multisymbol $\mathbf{x}_s(t)$. The total number of links ending at the elements of $\mathcal{Y}_s$ is

$$L_s = \sum_{j=0}^{F-s} \binom{s+j}{s}. \tag{33}$$

We observe that the total number of links L_s is constant and independent on the choice of the multisymbol $\mathcal{M}_t$. However, the choice of the multisymbol affects the distribution of the number of links between the individual elements of $\mathcal{Y}_s$, $l_{s,k}$, and thus the entropy $H(\mathbf{Y}|\mathcal{M}_t)$. For example, in the case when $F = 3$ and $s = 1$, given the choice of the basic multisymbol $\mathcal{M}_b$, for the number of links at each element $\mathbf{y}_{1,k} \in \mathcal{Y}_1$, we have $\{l_{1,k}\} = \{3,\ 2,\ 1\}$, as represented in Fig. 1. The links at each element $\mathbf{y}_{1,k}$ arise from different symbols from the basic multisymbol $\mathcal{M}_b = \{000, 001, 011, 011, 111\}$. For example, the 3 links ending at 001 arise from 001, 011 and 111 respectively, the 2 links at 010 from 011 and 111 and the 1 link at 100 from 111.

In the case of the multisymbol $\mathcal{M}_n = \{001, 011, 110, 111\}$, for the number of links at each element $\mathbf{y}_{1,k} \in \mathcal{Y}_1$, we have $\{l_{1,k}\} = \{2,\ 2,\ 2\}$, also shown in Fig.1. We see that the total number of links in both cases is $L_s = 6$ and does not depend on the choice of the multisymbol. It is only the distribution of the links which differs in both cases.

Now, as in the example for $F = 3$, we define

$$p_i^{(s)} = \left\{ \begin{array}{rl} 0, & \text{for } 0 \leq i < s \\ \frac{\epsilon^{i-s}(1-\epsilon)^s \binom{F}{i}}{2^F}, & \text{for } s \leq i \leq F \end{array} \right. \tag{34}$$

It holds that $p_i^{(s)} \geq p_{i+1}^{(s)}$, since $\epsilon \in (0, 1)$. Additionally, we define

$$P_{\mathcal{I}}^{(s)} = \sum_{i\in\mathcal{I}} p_i^{(s)}. \tag{35}$$

For brevity of the notation, we denote $\mathcal{U} = \{u_1, \ldots, u_K\}$, where for the basic multisymbol $\mathcal{M}_b$

$$u_k = \left\{ \begin{array}{rl} P_{\{s,\ldots,F\}}^{(s)} & \text{for } \ k = 1, \\ P_{\{i+1,\ldots,F\}}^{(s)} & \text{for } \ \binom{i}{s} + 1 \leq k \leq \binom{i+1}{s} \end{array} \right. \tag{36}$$

and $i = s, s+1, \ldots, F$. For the elements of $\mathcal{U}$ obtained in this way it holds $u_k \geq u_{k+1}$, $k = 1, \ldots, K - 1$. For the basic multisymbol $\mathcal{M}_b$, the contribution at the entropy $H(\mathbf{Y}|\mathcal{M}_b)$ resulting from $\mathcal{Y}_s$ is given by

$$H_{s,b}(\mathbf{Y}|\mathcal{M}_b) = -\sum_{k=1}^{K} u_k \log_2 u_k, \tag{37}$$

which is a direct result of (32) and (35).

For example, in the case when $F = 4$ and $s = 2$ we have

$$H_{2,b}(\mathbf{Y}|\mathcal{M}_b) = -\sum_{k=1}^{6} u_k \log_2 u_k \tag{38}$$

where $\mathcal{U} = \{P^{(2)}_{\{2,3,4\}}, P^{(2)}_{\{3,4\}}, P^{(2)}_{\{3,4\}}, P^{(2)}_{\{4\}}, P^{(2)}_{\{4\}}, P^{(2)}_{\{4\}}\}$.

It can be easily shown that for any minimal multisymbol $\mathcal{M}_m$ which is a permutation of the basic multisymbol $\mathcal{M}_b$, i.e. the Hamming distance between the consecutive symbols is 1, the set $\mathcal{U}'$ is identical with the set $\mathcal{U}$ of the basic multisymbol. This is a direct consequence of the properties of the Z channel. For example, the choice of the multisymbol $\mathcal{M}_m = \{000, 001, 101, 111\}$ yields a link distribution equivalent to the one for the basic multisymbol $\mathcal{M}_b$, which yields the same set of distributions, $\mathcal{U}$.

Now, let us take a multisymbol $\mathcal{M}_n$ which is not a minimal multisymbol. For this multisymbol we define the set $\mathcal{V} = \{v_1, \ldots, v_K\}$. This set has the property that $v_k = u_k$, for all $k \in \{1, 2, \ldots, K\}/\mathcal{J}$ where $\mathcal{J}$ is a set containing an even number of indices, ordered in pairs, $\mathcal{J} = \{i_1, j_1, \imath_2, j_2, \ldots, i_m, j_m\}$. For the indices of the set $\mathcal{J}$ it holds $v_{i_m} = u_{i_m} - \Delta_m$ and $v_{j_m} = u_{j_m} + \Delta_m$, $\Delta_m \leq |u_{i_m} - u_{j_m}|$. By applying *Property 1* $|\mathcal{J}|/2$ times, it follows that the entropy associated with $\mathcal{V}$ is greater than the entropy associated with $\mathcal{U}$, $H(\mathcal{V}) \geq H(\mathcal{U})$. Hence, the choice of the set of representatives to be the basic multisymbol, or a permutation of it, minimizes the entropy $H(\mathbf{Y}|\mathcal{M}_t) = \sum_{s=0}^{F} H_{(s)}(\mathbf{Y}|\mathcal{M}_t)$. Consequently, $I(\mathbf{X}; \mathbf{Y}|\mathcal{M}_t)$ is minimized, which was required in order to maximize $I(T; \mathbf{Y})$, which concludes the proof.

$\blacksquare$

Hence, we can conclude that if the multisymbol for each t is a permutation of the basic multisymbol, then $I(\mathbf{X}; \mathbf{Y}|T = t)$ is minimal and constant (independent of t). We recall that according to (16) we had the following upper bound of the capacity

$$C \leq \max_{P_T(\cdot), P_{\mathbf{X}|T}(\cdot)} I(\mathbf{X}; \mathbf{Y}) - \min_{P_T(\cdot), P_{\mathbf{X}|T}(\cdot)} I(\mathbf{X}; \mathbf{Y}|T).$$

Having conducted the analysis of $I(\mathbf{X}; \mathbf{Y})$ and $I(\mathbf{X}; \mathbf{Y}|T)$, we can write (16) in the following form

$$C \leq C_{XY} - I_m. \tag{39}$$

C_{XY} is the capacity of the $\mathbf{X} - \mathbf{Y}$ channel, given by (17). We recall that this capacity is attained by the distribution (19) and is less or equal to the capacity of the Z-channel. I_m is the minimal value (constant) of $I(\mathbf{X}; \mathbf{Y}|T = t)$, achieved by the choice of the multisymbol as a permutation of the basic multisymbol. It can be explicitly calculated by (21), by substituting the discussed choice of the multisymbol. The equality is achieved if and only if there is a pair of distributions $\big(P_T(\cdot), P_{\mathbf{X}|T}(\cdot)\big)$ that simultaneously attains the maximum and the minimum in the first and the second term, respectively. Preliminary results show that it is always possible to find such distributions, which remains to be formally proven and is a topic of current work.

V. Conclusions and Future Work

We elaborated on the capacity of communication channels with protocol coding, where the information is modulated in the actions taken by the communication protocol of an existing, primary system. In general, the capacity analysis of the secondary communication channel depends on the way the errors are introduced in the communication, i.e. on the underlying model for the transmission errors. In the previous works [1], [2], [13], we focused on the case with packet erasures, based on the block BEC (binary erasure channel). Here, we extended the capacity results derived in [13], to the case when the errors are modelled by the Z-channel model, which is of practical relevance. We used the framework where the secondary communication channel was represented through a cascade of channels, which is an alternative to the Shannon's representation of channels with causal channel state information at the transmitter (CSIT). The alternative framework shows to be an effective tool for capacity computation, independent on the way the errors are introduced in the communication. Additionally, it can give an insight in the coding strategies that are approaching the capacity, which is a part of an ongoing work. A problem of current analysis is the computation of the capacity under more general error models. In practice, a secondary channel can be defined over virtually any existing wireless system and it is of interest to find the coding strategies that are suited to a certain primary system.

References

[1] P. Popovski and Z. Utkovski, On the Secondary Capacity of the Communication Protocols, in IEEE GLOBECOM, Honolulu, HI, USA, Dec. 2009.

[2] P. Popovski and O. Simeone, Protocol Coding for Two-Way Communications with Half-Duplex Constraints, in IEEE GLOBECOM, Miami, FL, USA, Dec. 2010.

[3] J. L. Massey, Channel Models for RandomAccess Systems, ser. Performance Limits in Communication Theory and Practice, NATO Advances Studies Institutes Series E142. Kluwer Academic, 1988, pp. 391402.

[4] V. Anantharam and S. Verdu, Bits through Queues, IEEE Trans. Inform. Theory, vol. 44, pp. 418, Jan. 1996.

[5] G. Kramer, Models and Theory for Relay Channels with Receive Constraints, in Proc. 42 Annual Allerton Conference on Communications, Control and Computing, Urbana-Champaign, IL, USA, Sep. 2004.

[6] T. Lutz, C. Hausl, and R. Kotter, Bits Through Relay Cascades with HalfDuplex Constraint, 2009, submitted, (arXiv:0906.1599).

[7] A. Ephremides and B. Hajek, Information Theory and Communication Networks: An Unconsummated Union, IEEE Trans. Inform. Theory, vol. 44, no. 6, pp. 24162434, Oct. 1998. June 29, 2011

[8] K. Ashan, Covert Channel Analysis and Data Hiding in TCP/IP. M. Sc. thesis, Dept. of Electrical and Computer Engineering, University of Toronto, August 2002.

[9] R. C. Chakinala, A. Kumarasubramanian, R. Manokaran, G. Noubir, C. P. Rangan, and R. Sundaram, Steganographic Communication in Ordered Channels, in Information Hiding, Lecture Notes in Computer Science, vol. 4437. Springer- Verlag, 2009.

[10] J. Walsh and S. Weber, Capacity Region of the Permutation Channel, in Proc. 46 Annual Allerton Conference on Communications, Control and Computing, Urbana-Champaign, IL, USA, Sep. 2008, pp. 646652.

[11] A. J. H. Vinck, Coded Modulation for Power Line Communications, AEU Journal, pp. 4549, Jan. 2000.

[12] W. Chu, C. J. Colbourn, and P. Dukes, Constructions for Permutation Codes in Powerline Communications, Designs, Codes and Cryptography, Kluwer Academic Publishers, vol. 32, pp. 5164, 2004.

[13] P. Popovski and Z. Utkovski, ”Protocol Coding through Reordering of User Resources: Applications and Capacity Results”, submitted to IEEE Trans. Commun.,http://arxiv.org/abs/1011.5739.

[14] C. E. Shannon, Channels with side information at the transmitter, IBM Journal of Research and Development, vol. 2, pp. 289293, Oct. 1958.

[15] T. Cover and J. Thomas, Elements of Information Theory. Wiley-Interscience, 2nd Edition, 2006.