Communication through Reordering of Resources: Capacity Results and Trellis Code Design

Petar Popovski and Kasper F. Trillingsgaard
Department of Electronic Systems Aalborg University
Denmark
Email: {petarp,ktrill08@student.aau.dk}@es.aau.dk

Zoran Utkovski
Faculty of Computer Science
University Goce Delcev Stip
Republic of Macedonia
Email: zoran.utkovski@ugd.edu.mk

***Abstract*—[1] We define protocol coding as a way to encode information in the actions taken by a communication protocol. In this work we investigate strategies for protocol coding via combinatorial ordering of the labelled user resources (packets, channels) in an existing, *primary system*. This introduces a new, *secondary communication channel* in the existing system, which has been considered in the prior work exclusively in a steganographic context. Instead, we focus on the use of secondary channel for reliable communication with newly introduced secondary devices, that are low-complexity versions of the primary devices, capable only to decode the robustly encoded header information in the primary signals. We introduce a suitable communication model, capable to capture the constraints that the primary system puts on protocol coding. We derive the capacity of the secondary channel under arbitrary error models. The insights from the information–theoretic analysis are used to design practical error–correcting schemes for secondary channels based on trellis codes.**

I. Introduction

With the vast deployed infrastructure and variety of existing wireless systems, it is of significant practical value to introduce new features without changing the physical layer/hardware of the infrastructure, but only upgrade it in software. This can be achieved by a suitable, backward–compatible upgrade of the communication protocols. We use the term *protocol coding* to refer to techniques that convey information by modulating the actions of a communication protocol.

Consider the example on Fig. 1, where a cellular base station (BS) a group of *primary terminals* in its range. It is assumed that the cellular system is frame–based (WiMax [1], LTE [2], etc.). The metadata contained in the frame header informs the terminals how to receive/interpret the actual data that follows. The frame header is commonly encoded more robustly compared to the data, such that it can be reliably received in an area that is larger than the nominal coverage area, as depicted on Fig. 1. In such a context, while still using the same infrastructure, we can introduce new *secondary devices*, which are able to operate in the extended coverage area. These can be e. g. machine-type devices [3], such as sensors or actuators, that are controlled by the cellular BS. The secondary devices are simple and have a limited functionality, capable to decode only the frame header, but not the complex high–rate codebooks used for data. The main idea is that BS can send information to the secondary devices in the frame header. As another example, assume that there are two OFDMA channels, 1 and 2, defined in a diversity mode [1], such that if a user Alice is scheduled in a given frame, it is irrelevant whether it is assigned to channel 1 or 2. Hence, if BS schedules Alice and Bob in a given frame, then it can encode $1-$bit secondary information as follows: allocating Alice to channel 1 and Bob to channel 2 is a bit value 0, otherwise it is a bit value 1. Taking this simple example further, let there be three OFDMA channels, but still only two users, Alice and Bob. In a given frame, each of them can get from 0 up to 3 channels assigned, which is decided by the primary scheduling criterion; the secondary transmitter can encode information by assigning these channels to Alice/Bob in a particular way. If there are $2(1)$ packets for Alice (Bob), they can be assigned in 3 possible ways and in that particular frame, $\log_2(3)$ secondary bits can be sent. However, if all 3 packets are addressed to Alice, no secondary information can be sent in that frame. This variable amount of information due to the primary operation is the crux of the communication model considered in this work.

This paper investigates the fundamental properties of com-

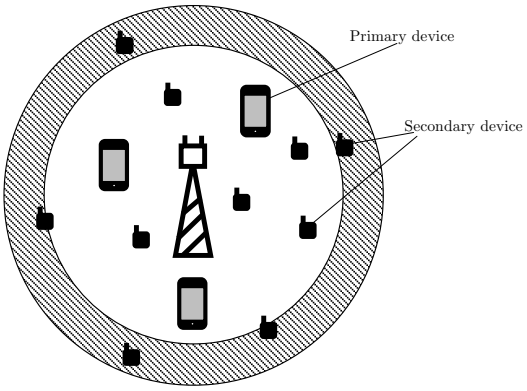


Fig. 1. Illustration of a secondary communication through protocol coding in cellular systems. A primary device can decode any information sent by the base station, while the secondary device has a limited functionality can only decode the information sent by protocol coding. The range of the primary communication system (white circle) is smaller than the range of the secondary information (shaded circle).

[1]Longer version of this work containing the proofs available as a 2-part paper in www.arxiv.org: P. Popovski and Z. Utkovski, "Protocol Coding through Reordering of User Resources, Part I: Capacity Results" and P. Popovski, Z. Utkovski, and K. F. Trillingsgaard, "Protocol Coding through Reordering of User Resources, Part II: Practical Coding Strategies".

munication systems that use protocol coding to send information, under restrictions imposed by a primary system. The secondary information is encoded in the ordering of labelled resources (packets, channels) of the primary (legacy) users. In this paper we introduce a suitable communication model that can capture the restrictions imposed by the primary system. The model captures the key feature of a secondary communication: in a given scheduling epoch, the primary system decides which packets/users to send data to, while secondary information can be sent by only rearranging these packets. Each primary packet is subject to an error (e. g. erasure), which induces a corresponding error model for secondary communication. For this model, we carry out information-theoretic analysis and devise suitable communication strategies.

Protocol coding can appear in many flavors. An early work that mentions the possibility to send data by modulating the random access protocol is [4], but in a rather "negative" context, since the model used *explicitly prohibits* to decide the protocol actions based on user data. The seminal work [5] uses a form of protocol coding: the 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 [6] [7] and [8]. At a conceptual level, protocol coding bridges information theory and networking [9]. The idea of communication based on packet reordering is not new per se and has been presented in the context of covert channels [10] [11] [12]. However, the big difference with our work is that our objective is not steganographic, but rather what kind of communication strategies can be used when the degrees of freedom for secondary communication are limited by a certain (random) process in the primary system. The practical coding strategies are related to the frequency permutation arrays for power line communications [13], [14].

Preliminary results of this work have appeared in [15] and [16]. In [15] we have introduced the notion of a secondary channel and sketched of the communication strategies when the primary packets are subject to an erasure channel, while in [16] we treated the case when the error model for the primary packets is represented by a Z–channel. In this paper we devise capacity–achieving strategies for arbitrary error model incurred on the primary packets. We first show the relation to the model of Shannon for channels with causal side information at the transmitter (CSIT) [17]. We then develop a new framework for computing the secondary capacity, which leads us to explicit specification of the communication strategies using trellis codes.

II. System Model

A. Communication Scenario

The communication model is depicted on Fig. 2. A Base Station (BS) transmits downlink data to a set of two users, addressed 0 and 1, respectively. The BS serves the users in scheduling frames with Time Division Multiple Access (TDMA). Each frame has a fixed number of F packets. Each packet carries the address of a user to whom the packet

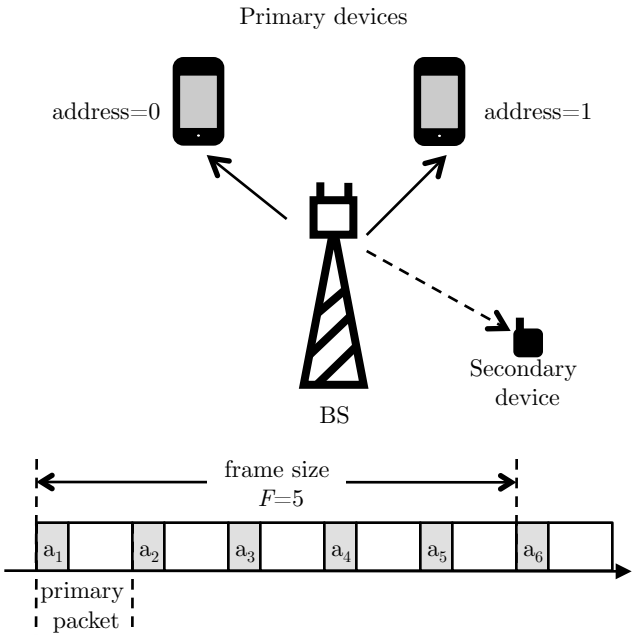


Fig. 2. The primary system consists of a Base Station (BS) and two primary devices. Each primary packet has a header that contains address $a_i \in \{0, 1\}$. The BS selects the orders of the packets in a frame in order to send information to the secondary device.

is destined, as well as data for that user. This is called *primary* data, destined to either user 0 or user 1. There is a third *secondary device*, that listens the TDMA frames sent by the BS. This device only records the address of each packet and ignores the packet data. Since this work is focused on the secondary communication, the notions "transmitter" and "receiver" will be used to refer to secondary transmitter and receiver, respectively. By addressing the packets in a given frame in a particular order, the BS sends *secondary information*. Thus, an input symbol for the secondary channel is an $F-$dimensional binary vector $\mathbf{x} \in \mathcal{X} = \{0, 1\}^F$.

The model with only two primary is limiting, but extension to K primary addresses entails complexity that is outside the scope of this initial paper on the topic. Yet, the results with binary secondary inputs provide novel insights for the communication strategies and set the basis for generalizations to $K > 2$. Furthermore, the binary input captures the following practical setup. Consider the case in which the arrival of packets in the primary system is random and in a certain frame the BS has only $F' < F$ packets to send, then $(F - F')$ of the slots will be empty. In this case we can still use the binary input model. We assign address 0 to a the empty packet slots, such that these empty slots can be actually treated as valid secondary input symbols. On the other hand, the presence of a packet in a given slot is treated as a secondary symbol 1. The secondary receiver only needs to detect packet presence/absence, without decoding its header.

The key assumption in the model is that the packets that are scheduled in a frame are decided by the primary communication system: the primary system decides that s packets in

a frame will be addressed to user 1 and $(F-s)$ packets will be addressed to user 0, where $0 \leq s \leq F$. This assumption captures the essence of protocol coding: secondary communication is realized by modulating the degrees of freedom left over from the operation of the original, primary communication system. In other words, the operational requirements of the primary system are contained in the set of packets that the BS decides to send in a given frame.

The number of packets s addressed to user 1 in a given frame is called *state* of the frame. We assume that the primary system selects packets in a memoryless fashion: in each frame, a packet is addressed $1(0)$ with probability $a(1-a)$, independently of the other packets and the previous frames. Hence, the probability that a frame is in state s is binomial $P_S(s) = \binom{F}{s} a^s (1-a)^{F-s}$. With the state s decided by the primary system, the secondary transmitter is only allowed to rearrange the packets in the frame. Since s is a random variable over which the secondary transmitter has no control, a frame carries a variable amount of secondary information. For example, if $F = 4$ and the primary system decides $s = 3$, then the possible secondary symbols for the frame are $1110, 1101, 1011, 0111$. But, if $s = F = 4$, than in that frame the secondary transmitter cannot send any information.

B. Error Models for the Secondary Channel

From the perspective of a secondary transmitter/receiver, each packet is sent over a memoryless channel with binary inputs. Several suitable error models can be inferred from the physical setup. In an *erasure channel*, the receiver either correctly decodes the packet address 0 or 1 or the header checksum in incorrect, leading to erasure ϵ. In a *binary symmetric channel*, the receiver uses error-correction decoding to decide whether it is more likely that address 0 or 1 is received. This results in only two possible outputs and symmetric error events. Finally, the *Z-channel* is suitable if $0/1$ corresponds to packet absence/presence, respectively. The probability that, in absence of a packet, the noise produces a valid packet detection sequence, is practically 0, while the probability that packet transmission is not detected is $p_e > 0$.

In the general case of a channel with binary inputs, there can be J possible outputs from the set $\mathcal{J}$. The special cases above have $\mathcal{J} = \{0, 1, \epsilon\}$ and $\mathcal{J} = \{0, 1\}$. When $i = 0, 1$ is sent, there are J transition probabilities, represented by a vector:

$$\mathbf{q}_i = (q_{i1}, q_{i2}, \dots q_{iJ}) \qquad i = 0, 1 \tag{1}$$

where $q_{ij} = P(y = j | x = i)$ and some q_{ij} can be equal to 0. A secondary output symbol is $\mathbf{y} \in \mathcal{Y} = \mathcal{J}^F$. The input/output variables of the secondary channel are denoted by $\mathbf{X}$ and $\mathbf{Y}$, respectively. By denoting $\mathbf{x} = (x_1, x_2, \cdots x_F)$ with $x_f \in \{0, 1\}$ and $\mathbf{y} = (y_1, y_2, \cdots y_F)$ with $y_f \in \mathcal{J}$, we can define the channel $\mathbf{X} - \mathbf{Y}$ through the transition probabilities:

$$P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \prod_{f=1}^{F} q_{x_f y_f} \tag{2}$$

When there is no risk for confusion, we simply write $P(\mathbf{y}|\mathbf{x})$. Thus, the channel $\mathbf{X}-\mathbf{Y}$ is specified by the memoryless binary channel through which each packet is passed.

The following notation will be used. $\mathcal{S} = \{0, 1, \dots F\}$ to denote the set of possible states. The set of input and output symbols of the secondary channel is denoted by $\mathcal{X}$ and $\mathcal{Y}$, respectively. The set of input symbols is partitioned into $F+1$ subsets $\mathcal{X}_s$ defined as follows:

$$\mathbf{x} \in \mathcal{X}_s \Leftrightarrow \sum_{i=1}^{F} x_i = s \tag{3}$$

When the frame state is $S = s$, then only $\mathbf{x} \in \mathcal{X}_s$ can be sent over the secondary channel.

III. Framework for Analyzing the Capacity of a Secondary Channel

The secondary channel can be represented by the framework of Shannon for channels with causal state information at the transmitter (CSIT) [17]. 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 T and each possible input letter t, termed *strategy* [18], represents a mapping from the state alphabet $\mathcal{S}$ to the input alphabet $\mathcal{X}$ of the original channel. A particular strategy $t \in \mathcal{T}$ is defined by the vector of size $|\mathcal{S}|$: $(t(1), \dots t(|\mathcal{S}|))$, where $t(s) \in \mathcal{X}$. Therefore, if each $s \in \mathcal{S}$ can be mapped map to any $\mathbf{x} \in \mathcal{X}$, then the total number of possible strategies is $|\mathcal{X}|^{|\mathcal{S}|}$ and therefore $|\mathcal{T}| \leq |\mathcal{X}|^{|\mathcal{S}|}$. The capacity of the equivalent channel can be found as:

$$C = \max_{P_T(\cdot)} I(T, \mathbf{Y}) \tag{4}$$

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 [18]:

$$P_{S,T,\mathbf{X},\mathbf{Y}}(s, t, \mathbf{x}, \mathbf{y}) = P_S(s) P_T(t) \delta(\mathbf{x}, t(s)) P_{\mathbf{Y}|\mathbf{X},S}(\mathbf{y}|\mathbf{x}, s) \tag{5}$$

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 ([19], Section 8.3), the required cardinality of $\mathcal{T}$ is not more than $|\mathcal{Y}|$.

However, Shannon's result is for the general case of channels with causal CSIT. The secondary channel considered here has a specific structure that permits more explicit characterization of the communication strategies. As noted in relation to (3), 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 $s = 2$, it is not possible to send the symbol $\mathbf{x} = 1011$. Nevertheless, 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 extend the model: Given $P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$, we define $P_{\mathbf{Y}|\mathbf{X},S}(\mathbf{y}|\mathbf{x}, s)$ such that 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 behind this approach is the following. For example, let us assume $F = 4$ and the erasure model. When $s = 0$ only $\mathbf{x} = 0000$ can be sent. But we can look at it in another way: when $s = 0$ only $\mathbf{y} = 0000$ ore the versions of 0000 with erasures can occur. Hence, we can equivalently say that when $s = 0$, any $\mathbf{x}$ can be sent, but, in absence of errors, the output is always 0000. Picking a strategy t'' in which $t''(s) = \mathbf{x}_u$ is equivalent to picking t' in which $t'(s) = \mathbf{x}_v$. In short, for given s, we define $P_{\mathbf{Y}|\mathbf{X},S}$ in order to discourage selection of symbols $\mathbf{x}$ for which $\mathbf{x} \neq \mathbf{y}$ in absence of channel errors.

Expressing the capacity in terms of strategies might pose some conceptual and practical problems when F is large [18]. On the other hand, our objective is to use the specific way in which the set of states partitions the possible set of transmitted symbols $\mathcal{X}$ in order to provide insights in the capacity–achieving communication strategies. Therefore, a different framework for capacity analysis from will be used. Recall that T is an auxiliary random variable defined over the set of possible strategies $\mathcal{T}$. For given $T = t$ and each $s \in \mathcal{S}$ there is a single *representative of* t *in* s $\mathbf{x} = t(s) \in \mathcal{X}_s$. In the text that follows we use "strategies" and "input symbols" interchangeably. Hence, $\mathcal{T}$ consists of the input symbols $\{1, 2, \ldots |\mathcal{T}|\}$. The set of $F + 1$ representatives $\{\mathbf{x}_s(t)\}$ for given t will be called a *multisymbol* of t.

Due to the randomized state change, each $t \in \mathcal{T}$ induces a distribution on $\mathcal{X}$. For example, if $F = 2$ and the strategy is defined as $t(0) = 00, t(1) = 01, t(2) = 11$, then we can define $P_{\mathbf{X}|T}(\mathbf{x} = 00|t) = (1-a)^2 = P_S(0)$, $P_{\mathbf{X}|T}(\mathbf{x} = 11|t) = a^2 = P_S(2)$, $P_{\mathbf{X}|T}(\mathbf{x} = 01|t) = 2a(1-a) = P_S(1)$, and $P_{\mathbf{X}|T}(\mathbf{x} = 10|t) = 0$. In general, $P_{\mathbf{X}|T}(\cdot)$ should satisfy that for each $s \in \mathcal{S}$ there is a single $\mathbf{x} \in \mathcal{X}_s$ such that $P_{\mathbf{X}|T}(\mathbf{x}|t) = P_S(s)$. The set of such distributions is:

$$\mathcal{P}_{\mathbf{X}|T} = \left\{P_{\mathbf{X}|T}(\cdot)|\forall t \in \mathcal{T}, \forall s \in \mathcal{S}, \exists!\mathbf{x} \in \mathcal{X}_s, P_{\mathbf{X}|T}(\mathbf{x}|t) = P_S(s)\right\} \quad (7)$$

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 $T - \mathbf{X} - \mathbf{Y}$ and the primary constraints are reflected in the definition of $\mathcal{P}_{\mathbf{X}|T}$. In order to express the mutual information $I(T; \mathbf{Y})$, we write $I(T, \mathbf{X}; \mathbf{Y}) = I(T; \mathbf{Y}) + I(\mathbf{X}; \mathbf{Y}|T) = I(\mathbf{X}; \mathbf{Y}) + I(T; \mathbf{Y}|\mathbf{X})$ Using the Markov property for the cascade we get $I(T; \mathbf{Y}|\mathbf{X}) = 0$, which implies:

$$I(T; \mathbf{Y}) = I(\mathbf{X}; \mathbf{Y}) - I(\mathbf{X}; \mathbf{Y}|T) \quad (8)$$

Let $\mathcal{P}_T$ denote the set of all distributions $P_T(\cdot)$. Our objective is to find the pair of distributions $\big(P_T(\cdot), P_{\mathbf{X}|T}(\cdot)\big)$ that maximizes $I(T; \mathbf{Y})$. The capacity of the secondary channel is:

$$C = \max_{P_T(\cdot), P_{\mathbf{X}|T}(\cdot)} I(T; \mathbf{Y}) \quad (9)$$

We will always that $P_{\mathbf{X}|T}(\cdot) \in \mathcal{P}_{\mathbf{X}|T}$ always. The expression (9) can be upper–bounded:

$$C \leq \max_{P_T(\cdot), P_{\mathbf{X}|T}(\cdot)} I(\mathbf{X}; \mathbf{Y}) - \min_{P_T(\cdot), P_{\mathbf{X}|T}(\cdot)} I(\mathbf{X}; \mathbf{Y}|T) \quad (10)$$

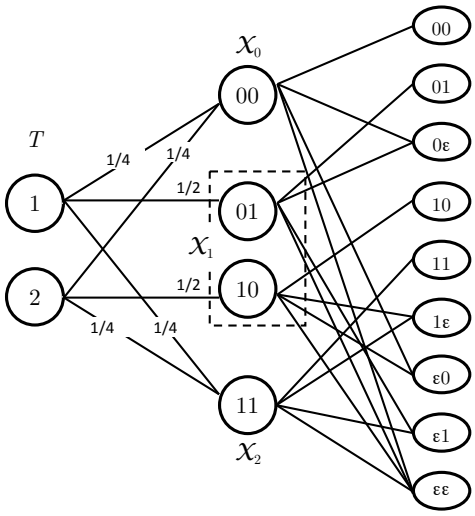


Fig. 3. Example choice of the probability distribution $P_{\mathbf{X}|T}$ with $F = 2$ and $\mathcal{T} = \{1, 2\}$. The transition probabilities on the channel $\mathbf{X} - \mathbf{Y}$ are not marked, but it is assumed that each packet 0 or 1 can become erased ϵ independently with probability p.

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 max/min in the first/second term, respectively. We will decompose the problem (9) into two sub–problems, maximization of $I(\mathbf{X}; \mathbf{Y})$ and minimization of $I(\mathbf{X}; \mathbf{Y}|T)$.

Fig. 3 illustrates the cascade of channels where $F = 2$ and erasure model for $\mathbf{X} - \mathbf{Y}$ with $\mathcal{J} = \{0, 1, \epsilon\}$ and $q_{00} = q_{11} = 1 - p$, while $q_{0\epsilon} = q_{1\epsilon} = p$. Let us assume that the primary constraint uses $a = \frac{1}{2}$. The two multisymbols, corresponding to $t = 1$ and $t = 2$ are $\{00, 01, 11\}$ and $\{00, 10, 11\}$, respectively. It is seen that uniform $P_T(\cdot)$ induces uniform $P_{\mathbf{X}}(\cdot)$. On the other hand, the capacity of the vector channel with erasures $\mathbf{X} - \mathbf{Y}$ is achieved when $P_{\mathbf{X}}(\cdot)$ is uniform. The reader can check that uniform $P_T(\cdot)$ and the choice of $P_{\mathbf{X}|T}(\cdot)$ according to Fig. 3 simultaneously maximizes $I(\mathbf{X}; \mathbf{Y})$ and minimizes $I(\mathbf{X}; \mathbf{Y}|T)$.

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

Each pair of distributions $\big(P_T(\cdot), P_{\mathbf{X}|T}(\cdot)\big)$ induces a distribution $P_{\mathbf{X}}$ on $\mathcal{X}$. Let $\mathcal{P}_{\mathbf{X}}$ denote the set of all possible distributions $P_{\mathbf{X}}(\cdot)$, while $\mathcal{P}_{\mathbf{X}}^T \subset \mathcal{P}_{\mathbf{X}}$ containing the distributions $P_{\mathbf{X}}(\cdot)$ that can be induced by all possible pairs $\big(P_T(\cdot), P_{\mathbf{X}|T}(\cdot)\big)$. Then the following holds (proof omitted):

Proposition 1: The set of distributions $\mathcal{P}_{\mathbf{X}}^T$ is a subset of $\mathcal{P}_{\mathbf{X},S}$, where $\mathcal{P}_{\mathbf{X},S} \subset \mathcal{P}_{\mathbf{X}}$ and:

$$\mathcal{P}_{\mathbf{X},S} = \left\{P_{\mathbf{X}}(\cdot)|\sum_{\mathbf{x}\in\mathcal{X}_s} P_{\mathbf{X}}(\mathbf{x}) = P_S(s), \forall s = 0, 1, \cdots F\right\} \quad (11)$$

The proposition implies $\max_{P_T(\cdot), P_{\mathbf{X}|T}(\cdot)} I(\mathbf{X}; \mathbf{Y}) \leq \max_{P_{\mathbf{X}}(\cdot)\in\mathcal{P}_{\mathbf{X},S}} I(\mathbf{X}; \mathbf{Y})$. We will first look for the distribution $P_{\mathbf{X}^*}(\cdot) \in \mathcal{P}_{\mathbf{X},S}$ that maximizes $I(\mathbf{X}; \mathbf{Y})$. Once $P_{\mathbf{X}^*}(\cdot)$ is known, we choose $\big(P_T(\cdot), P_{\mathbf{X}|T}(\cdot)\big)$ in order to induce the

desired $P_{\mathbf{X}^*}(\cdot)$. Let us define:

$$C_{XY} = \max_{P_{\mathbf{X}} \in \mathcal{P}_{\mathbf{X},S}(\cdot)} I(\mathbf{X};\mathbf{Y}) \tag{12}$$

which is never larger than the capacity of $\mathbf{X}-\mathbf{Y}$, achieved by selecting over all $P_{\mathbf{X}}(\cdot) \in \mathcal{P}_{\mathbf{X}}$. For example, if the probability $a \neq \frac{1}{2}$ and there are erasure–type errors, then $C_{XY} < F(1-p)$, where $F(1-p)$ is the capacity of F erasure channel uses. This is because uniform distribution $P_{U,\mathbf{X}}(\mathbf{x}) = 2^{-F}$ achieves the capacity of the erasure channel, which induces the necessary condition $\sum_{\mathbf{x}\in\mathcal{X}_s} P_{U,\mathbf{X}}(\mathbf{x}) = \binom{F}{s}2^{-F}$, but this is not equal to $P_S(s)$ if $a \neq \frac{1}{2}$.

In this text we are interested in channels $\mathbf{X}-\mathbf{Y}$ where each single channel use $\mathbf{x}$ consists of F uses of a more elementary, identical channels, leading to the following symmetry: 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 used for a single primary packet. Such a symmetry is instrumental for making statements about C_{XY}.

Lemma 1: The distribution $P_{\mathbf{X}}(\cdot) \in \mathcal{P}_{\mathcal{X},\mathcal{S}}$ that achieves C_{XY} is, for all s and each $\mathbf{x} \in \mathcal{X}_s$:

$$P_{\mathbf{X}}(\mathbf{x}) = \frac{P_S(s)}{\binom{F}{s}} \tag{13}$$

Having found $P_{\mathbf{X}}(\cdot)$ that attains C_{XY}, it remains to find $\mathcal{T}$, $P_T(\cdot)$ and $P_{\mathbf{X}|T}(\cdot)$ (i. e. the representatives of each $T=t$) such that (13) is satisfied. For example, let $F=4$ and $|\mathcal{X}_s| = 1,4,6,4,1$ for $s=0,1,2,3,4$, respectively. Let at first take $|\mathcal{T}| = 4m$ and uniform $P_T(t) = \frac{1}{4m}$. Then each $\mathbf{x} \in \mathcal{X}_1$ can be a representative of exactly m different elements of $\mathcal{T}$, such that $P_{\mathbf{X}}(\mathbf{X}=\mathbf{x}) = P_S(1)\cdot m \cdot \frac{1}{4m} = P_S(1)/\binom{4}{1}$. In general, if $|\mathcal{T}| = \binom{F}{s}\cdot m$ and uniform $P_T(t)$, we can choose $\mathbf{x} \in \mathcal{X}_s$ to be a representative of exactly m elements from $\mathcal{T}$; i. e. $P_{\mathbf{X}|T}(\mathbf{x}|t) = P_S(s)$ for m different values t and zero otherwise. The resulting $P_{\mathbf{X}}(\cdot)$ satisfies (13). To satisfy this condition for all s simultaneously, $|\mathcal{T}|$ should be divisible with $\binom{F}{s}$ for all $s = 0\cdots F$, leading to the following lemma, stated without proof (lcm stands for "least common multiplier"):

Lemma 2: The distribution $P_{\mathbf{X}}(\cdot)$ that satisfies (13) can be achieved by choosing uniform $P_T(\cdot)$ over a set with a minimal cardinality of $|\mathcal{T}| = \text{lcm}\left(\binom{F}{0}, \binom{F}{1}, \ldots, \binom{F}{F}\right)$.

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

The multisymbol $\mathcal{M}_t = \{\mathbf{x}_0(t), \cdots \mathbf{x}_F(t)\}$ corresponding to t has one representative in each $\mathbf{x}_s(t) = \mathcal{X}_s$, such that $P_{\mathbf{X}|T}(\mathbf{x}_s(t)|t) = P_S(s)$ and is zero for the other $\mathbf{x}$. Since $I(\mathbf{X};\mathbf{Y}|T=t)$ depends on the choice of representatives in $\mathcal{M}_t$, we will denote it by $I(\mathbf{X};\mathbf{Y}|\mathcal{M}_t)$, such that:

$$I(\mathbf{X};\mathbf{Y}|T) = \sum_{t\in\mathcal{T}} I(\mathbf{X};\mathbf{Y}|\mathcal{M}_t) \tag{14}$$

For example, let $F=5$ with $\mathcal{M}_1 = \{00000, 00001, 00011, 00111, 01111, 11111\}$ and $\mathcal{M}_2 = \{00000, 00001, 00110, 11100, 10111, 11111\}$. Assuming a binary symmetric channel with $q_{00} = q_{11} = 0.8, q_{01} = q_{10} = 0.2$ it can be seen that $I(\mathbf{X};\mathbf{Y}|\mathcal{M}_1) < I(\mathbf{X};\mathbf{Y}|\mathcal{M}_2)$. For intuitive explanation, consider two representatives $\mathbf{x}_{s_i} \in \mathcal{X}_{s_i}$, $i=1,2$. From (3) the Hamming weight of $\mathbf{x}_{s_i}$ is s_i and, without loss of generality, assume $s_1 > s_2$. For the multisymbol $\mathcal{M}_1$, the Hamming distance between any two representatives is given by:

$$d_H(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}) = s_2 - s_1 \tag{15}$$

and is minimal possible. Informally, any two representatives from $\mathcal{M}_1$ are as similar to each other as possible since they represent the same input $T=1$, which is not the case for $\mathcal{M}_2$.

The multisymbols satisfying (15) will be termed *minimal* multisymbols. Among them, there is one termed *basic multisymbol* $\mathcal{M}^b$ with a particular structure: the representative in $\mathcal{X}_s$ is $00\cdots011\cdots1$ starts with $F-s$ consecutive zeros and s consecutive ones. It can be shown that any minimal multisymbol can be obtained from the basic one via permutation, such that there are $F!$ different minimal multisymbols. For example, let $\mathcal{M}^b = \{000, 001, 011, 111\}$ and we apply the permutation $\pi = 321$: the components of each $\mathbf{x} \in \mathcal{M}^b$ are permuted according to π to obtain $\mathcal{M}^m = \{000, 100, 110, 111\}$. In general, for a given permutation π we define $\gamma_\pi(\cdot)$:

$$\mathcal{M}' = \gamma_\pi(\mathcal{M}) \tag{16}$$

such that each $\mathbf{x}'_s \in \mathcal{M}'$ is obtained from the corresponding $\mathbf{x}_s \in \mathcal{M}$ by permuting the packets according to π and the Hamming distance between any two representatives is preserved $d_H(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}) = d_H(\mathbf{x}'_{s_1}, \mathbf{x}'_{s_2}) = s_2 - s_1$.

We write the mutual information $I(\mathbf{X};\mathbf{Y}|\mathcal{M}_t) = H(\mathbf{Y}|\mathcal{M}_t) - H(\mathbf{Y}|\mathbf{X},\mathcal{M}_t)$ and first consider:

$$H(\mathbf{Y}|\mathbf{X},\mathcal{M}_t) = \sum_{s=0}^{F} P_S(s) H(\mathbf{Y}|\mathbf{x}_s(t)) \tag{17}$$

Since each component of $\mathbf{x}_s$ uses identical memoryless channel, $H(\mathbf{Y}|\mathbf{x}_s(t))$ depends only on the Hamming weight s, and not on the arrangement of $0,1$s in $\mathbf{x}_s$. This is stated through:

Lemma 3: The conditional entropy for $\mathbf{x}_s \in \mathcal{X}_s$, having a Hamming weight of s, is given by:

$$H(\mathbf{Y}|\mathbf{X}=\mathbf{x}_s) = sH(\mathbf{q}_1) + (F-s)H(\mathbf{q}_0) = H_s \tag{18}$$

where $H(\mathbf{q}_i) = -\sum_{j=1}^{J} q_{ij}\log_2 q_{ij}$ for $i=0,1$ and $\mathbf{q}_i$ is given by (1).

Using the lemma, (17) can be rewritten as $H(\mathbf{Y}|\mathbf{X},\mathcal{M}_t) = \sum_{s=0}^{F} P_S(s) H_s$ and is not affected by the actual choice of $\mathcal{M}_t$, as long as there is a representative in each $\mathcal{X}_s$.

To gain intuition, we first consider a special type of $P_S(\cdot)$, in which only two states $s_1, s_2 \in \mathcal{S}$ occur with non-zero probability $P_S(s_1) = \lambda$ and $P_S(s_2) = 1-\lambda$, such that $\mathcal{M}_t = \{\mathbf{x}_{s_1}, \mathbf{x}_{s_2}\}$. Due to the symmetry implied by Lemma 3, without losing generality, we first pick an arbitrary $\mathbf{x}_{s_1} \in \mathcal{X}_{s_1}$. Then, how to select $\mathbf{x}_{s_2} \in \mathcal{X}_{s_2}$ in order to minimize the $H(\mathbf{Y}|\mathcal{M}_t)$? Slightly abusing the notation from (15), we use $d_H(\mathbf{x})$ to denote the Hamming weight of $\mathbf{x}$. Recall that $d_H(\mathbf{x}) = s$ for $\mathbf{x} \in \mathcal{X}_s$. Let $g_{uv}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2})$, where $u,v \in \{0,1\}$ denote the number of positions f at which $x_{s_1 f} = u$ and

$x_{s_2 f} = v$. For example, if $\mathbf{x}_{s_1} = 00110$, $\mathbf{x}_{s_2} = 11011$, then $g_{00} = 0$, $g_{01} = 3$, $g_{10} = 1$, and $g_{11} = 1$ (we write g_{uv} for brevity). Using similar arithmetics as in Lemma 3:

$$H(\mathbf{Y}|T=t) = g_{00}H(\mathbf{q}_0) + g_{11}H(\mathbf{q}_1) + g_{01}H(\lambda\mathbf{q}_0 \quad (19)$$
$$+(1\text{-}\lambda)\mathbf{q}_1) + g_{10}H((1\text{-}\lambda)\mathbf{q}_0 + \lambda\mathbf{q}_1) \quad (20)$$

The Hamming distance is $d_H(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}) = g_{01} + g_{10}$. The following lemma formalizes the intuition that $H(\mathbf{Y}|\mathcal{M}_t)$ is minimized when any two representatives are as similar to each other as possible.

Lemma 4: When $\mathcal{M}_t$ consists of only two representatives $\mathbf{x}_{s_1}, \mathbf{x}_{s_2}$, $H(\mathbf{Y}|\mathcal{M}_t)$ is minimized when the Hamming distance $d_H(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}) = |s_2 - s_1|$ is minimal possible.

We now consider a general $P_S(\cdot)$. As indicated above, $H(\mathbf{Y}|\mathcal{M}_t)$ can be written as:

$$H(\mathbf{Y}|\mathcal{M}_t) = \sum_{f=1}^{F} H(\mathbf{u}_f) \quad (21)$$

where $\mathbf{u}_f$ is the probability distribution that corresponds to the $f-$th position, defined as:

$$\mathbf{u}_f = \sum_{s=0}^{F} P_s \left[(1 - x_{s,f})\mathbf{q}_0 + x_{s,f}\mathbf{q}_1\right] \qquad \text{where } x_{s,f} \in \{0,1\} \quad (22)$$

Without losing generality, let us take the first value x_{s1} of each of the representatives $\mathbf{x}_s$ to create $(F+1)-$dimensional vector $\mathbf{z}_1$. In a similar way $\mathbf{z}_2$ is created, such that:

$$\mathbf{z}_1 = (x_{01}, x_{11}, \cdots x_{F1}) \qquad \mathbf{z}_2 = (x_{02}, x_{12}, \cdots x_{F2}) \quad (23)$$

The probability distribution vectors $\mathbf{u}_1, \mathbf{u}_2$ can be written as:

$$\mathbf{u}_1 = (Q_{00} + Q_{01})\mathbf{q}_0 + (Q_{10} + Q_{11})\mathbf{q}_1 \quad (24)$$
$$\mathbf{u}_2 = (Q_{00} + Q_{10})\mathbf{q}_0 + (Q_{01} + Q_{11})\mathbf{q}_1 \quad (25)$$

where $Q_{uv} = \sum_{s \in \mathcal{G}_{uv}(z_1,z_2)} P_s$ and the sets $\mathcal{G}_{uv}(z_1, z_2) = \{s | x_{s,1} = u, x_{s,2} = v\}$ for $u, v \in \{0, 1\}$.

Lemma 5: The contribution of the positions 1 and 2 to $H(\mathbf{Y}|\mathcal{M}_t)$ is minimal when one of the sets $\mathcal{G}_{01}, \mathcal{G}_{10}$ is empty.

This analysis leads us to the following theorem and corollary:

Theorem 1: When each individual packet in a frame is sent over an identical channel with binary inputs and general outputs, the minimal multisymbol minimizes $H(\mathbf{Y}|\mathcal{M}_t)$.

Corollary 1: The following mutual information is constant for all minimal multisymbols $\mathcal{M}^m$:

$$I(\mathbf{X}; \mathbf{Y}|\mathcal{M}^m) = H(\mathbf{Y}|\mathcal{M}^m) - H(\mathbf{Y}|\mathbf{X}, \mathcal{M}^m) = I_m \quad (26)$$

C. Achieving the Capacity of the Secondary Channel

Here we analyze (10) and find $\mathcal{T}$ and $\{\mathcal{M}_t\}$ (i. e. $P_T(\cdot)$) and $P_{\mathbf{X}|T}(\cdot)$, respectively) that simultaneously maximizes $I(\mathbf{X}; \mathbf{Y})$ according to Lemma 1 and minimizes $I(\mathbf{X}; \mathbf{Y}|T) = I_m$ according to (26). Recall that uniform T with $|\mathcal{T}| = \text{lcm}\left(\binom{F}{0}, \binom{F}{1}, \ldots, \binom{F}{F}\right) = L$ can achieve C_{XY}. Since there are $F! \geq L$ multisymbols, then in principle it should be possible to select L minimal multisymbols in order to have $I(\mathbf{X}; \mathbf{Y}|T) = I_m$ and maximize $I(\mathbf{X}; \mathbf{Y})$.

In order to show that it is always possible to select $\{\mathcal{M}_t\}$, with $|\{\mathcal{M}_t\}| = L$ and uniform T, we first take an example with $F = 4$. The set of $L = 12$ multisymbols can be selected as on Fig. 4(a). Multisymbols can be represented by a directed graph, see Fig. 4(b). Each node in the graph represents a particular $\mathbf{x} \in \mathcal{X}$. An edge exists between $\mathbf{x}_s \in \mathcal{X}_s$ and $\mathbf{x}_{s+1} \in \mathcal{X}_{s+1}$ if and only if the Hamming distance is $d_H(\mathbf{x}_s, \mathbf{x}_{s+1}) = 1$. The directed edge from $\mathbf{x}_s$ to $\mathbf{x}_{s+1}$ exists if they can both belong to a same minimal multisymbol $\mathcal{M}_t$. A multisymbol is represented by a path of length F that starts at $00\cdots0$ and ends at $11\cdots1$. To each edge we can assign a nonnegative integer, which denotes the number of multisymbols (paths) that contain that edge. On Fig. 4(b), each edge that starts from 0000 has a weight 3, each edge between an element of $\mathcal{X}_1$ and $\mathcal{X}_2$ has a weight 1, etc. The weight of each edge between $\mathbf{x}_s$ and $\mathbf{x}_{s+1}$ can be treated as an outgoing weight for $\mathbf{x}_s$ and incoming weight for $\mathbf{x}_{s+1}$. Using this framework, we need to prove that, for each $s = 0 \ldots F-1$, it is possible to match all outgoing weights from $\mathcal{X}_s$ to all incoming weights from $\mathcal{X}_{s+1}$.

Theorem 2: If $L = \text{lcm}\left(\binom{F}{0}, \binom{F}{1}, \ldots, \binom{F}{F}\right)$ and the distribution over $\mathcal{T}$ is uniform, then the multisymbols can be chosen such as to achieve the capacity of the secondary channel.

If $F = 4$ it turns out that $\frac{m_s}{F-s}$ is always an integer, such that all the outgoing/incoming weights to the same node are identical. This is not the case if, e. g., $F = 7$, then $L = 105$, $m_1 = 15$ and $\frac{m_1}{7-1} = \frac{15}{6}$, such that each node from $\mathcal{X}_1$ has 3 outgoing edges of weight 3 and 3 of weight 2.

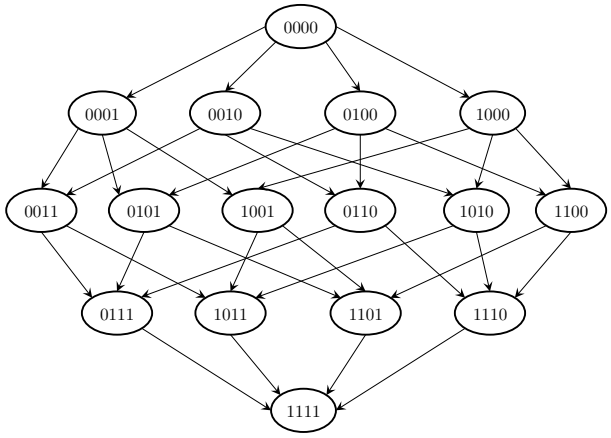
IV. Strategies for Trellis Coding

As a reference, we first look at a naïve coding strategy, which ignores the capacity–achieving analysis. We take any error-correction code of rate R and interleave the output of this code, e. g. by using a pseudo-random interleaver. The motivation for using an interleaver is to break the burst bit errors within one secondary symbol (frame). For example, for $F = 4$ we take 4 of the coded/interleaved bits and look at the current state of the channel. Then, we pick arbitrarily any of the possible frames, obtained by permuting the packets, that has the minimal possible Hamming distance. For example, let the coded bits are 0101 and let the state be $s = 3$. Then the Hamming distance of the "true information" 0101 from 0111, 1101 is 1 (minimal possible), while it is 3 from 1011 or 1110. Hence, when the system needs to send 0101 and $s = 3$, it chooses randomly between 0111 and 1101. One trouble with the described naïve strategy is that, even when the channel does not introduce errors, there will be decoding errors. Additionally, the results will show that the naïve strategy has overall poor error performance.

We propose a coding strategy, inspired by the capacity results for the secondary communication channel. From the viewpoint of capacity, the choice of the multisymbols is irrelevant, as long they are minimal and the distribution of X

t	$\{\mathbf{x}_s(t)\}$
1	$(0000, 0001, 0011, 0111, 1111)$
2	$(0000, 0001, 0101, 0111, 1111)$
3	$(0000, 0001, 1001, 1011, 1111)$
4	$(0000, 0010, 0011, 0111, 1111)$
5	$(0000, 0010, 0110, 0111, 1111)$
6	$(0000, 0010, 1010, 1011, 1111)$
7	$(0000, 0100, 0101, 0111, 1111)$
8	$(0000, 0100, 0110, 0111, 1111)$
9	$(0000, 0100, 1100, 1101, 1111)$
10	$(0000, 1000, 1001, 1101, 1111)$
11	$(0000, 1000, 1010, 1110, 1111)$
12	$(0000, 1000, 1100, 1110, 1111)$

(a)



(b)

Fig. 4. Selection of the representative sets for $F = 4$ that achieve the capacity. (a) Multisymbols for the 12 inputs (b) Graph representation of the process for selecting the multisymbols $\mathbf{x}_s(t)$.

fulfills the required condition. However, the choice of the multisymbols does affect the performance of the error-correcting code constructed based on the multisymbol framework. Our aim is to use the multisymbol framework in the construction of practical coding schemes which are better suited for the secondary communication channel than the naïve approach. The question to ask is which criterion, e. g. distance metric we are going to use in the selection of the multisymbols. We adopt a heuristic approach and take the *expected Hamming distance* as the metric of interest for the choice of the L multisymbols. This distance for two multisymbols t_1 and t_2 is defined:

$$E_{d_H}(t_1, t_2) = \sum_{s=0}^{F} P_S(s) d_H(x_s(t_1), x_s(t_2)), \qquad (27)$$

where d_H is the Hamming distance between the two vectors. This metric incorporates the state of the channel which can not be controlled by the secondary system. We can now define a trellis code. A state contains two outgoing branches, each of them corresponding to one possible input binary symbol. Also, each state has two incoming branches. In our case, the input symbol is binary and the output symbol is one of the L multisymbols. The trellis is chosen on purpose to have L branches, such that each multisymbol is used only once. The question is how we associate multisymbols with the transitions

t	$\{\mathbf{x}_s(t)\}$
1	$(0000, 0001, 0011, 0111, 1111)$
2	$(0000, 0001, 0101, 1101, 1111)$
3	$(0000, 0001, 1001, 1011, 1111)$
4	$(0000, 0010, 0011, 0111, 1111)$
5	$(0000, 0010, 1010, 1011, 1111)$
6	$(0000, 0010, 1010, 1110, 1111)$
7	$(0000, 0100, 0101, 1101, 1111)$
8	$(0000, 0100, 0110, 0111, 1111)$
9	$(0000, 0100, 0110, 1110, 1111)$
10	$(0000, 1000, 1001, 1011, 1111)$
11	$(0000, 1000, 1100, 1101, 1111)$
12	$(0000, 1000, 1100, 1110, 1111)$

Fig. 5. Selection of the representative sets for $F = 4$, Example 1.

in the trellis. We use the known rules from trellis coding: the output symbols on the branches exiting from the same state should be maximally separated in terms of the expected Hamming distance. The same is valid for the output symbols associated with the two branches that enter the same state. However, here these rules are only heuristic, as we have not rigorously related the error performance to the the expected Hamming distance. To illustrate the code construction, we set $F = 4$, where the minimal cardinality of the uniform auxiliary variable T is $L = \mathrm{lcm}\{\binom{4}{0}, \binom{4}{1}, \ldots, \binom{4}{4}\} = 12$.

There are multiple ways in which the multisymbols can be chosen. The choice is facilitated by the representation of the multisymbols as paths in the directed graph, see Fig. 5. In order to maximize the expected Hamming distance between multisymbols, the paths corresponding to the multisymbols should be as diverse as possible. To assure this, we have to choose the multisymbols such to avoid, as much as possible, having multisymbols with common edges, as in that case the expected Hamming distance is 0. The necessary condition to avoid a common edge between the nodes from $\mathcal{X}_s$ and $\mathcal{X}_{s+1}$, where $s \leq \lfloor F/2 \rfloor - 1$, is that $L/\binom{F}{s} \leq F - s$. In other words, the edge weight should be at most 1. Since in the general case it is difficult to control the code distance spectrum, we turn to the the heuristic of minimal expected Hamming distance as a simplified indicator of the code performance. However, as we are going to see in the next section, some of the simulation results indicate that the minimal expected Hamming distance is not the only factor which is decisive for the error performance. In the following we give three representative examples of the set of multisymbols $\mathcal{T}$, created for $F = 4$.

1) Choice of multisymbols, Example 1: In the first example, we choose the 12 multisymbols as given in Fig. 5 (a). This choice is capacity achieving, but we need to investigate its performance in terms of error rate when used to construct a channel code. Using the representation from Fig. 4(b), it is noted that some of the multisymbols have common edges which can be avoided. This, for example, is the case with the multisymbols $t_1 = \{0000, 0001, 0011, 0111, 1111\}$ and $t_4 = \{0000, 0010, 0011, 0111, 1111\}$. The expected Hamming distance profile for the above choice of the set of multisymbols reveals that the minimal distance is 0.5.

2) Choice of multisymbols, Example 2: In the second example, we choose the 12 multisymbols as given in Fig. 6. We observe that no two multisymbols are identical and

t	$\{\mathbf{x}_s(t)\}$
1	$(0000, 0001, 0011, 0111, 1111)$
2	$(0000, 0001, 0101, 1101, 1111)$
3	$(0000, 0001, 1001, 1011, 1111)$
4	$(0000, 0010, 0011, 1011, 1111)$
5	$(0000, 0010, 0110, 0111, 1111)$
6	$(0000, 0010, 1010, 1110, 1111)$
7	$(0000, 0100, 0101, 0111, 1111)$
8	$(0000, 0100, 0110, 1110, 1111)$
9	$(0000, 0100, 1100, 1101, 1111)$
10	$(0000, 1000, 1001, 1101, 1111)$
11	$(0000, 1000, 1010, 1011, 1111)$
12	$(0000, 1000, 1100, 1110, 1111)$

(a)

t	$\{\mathbf{x}_s(t)\}$
1	$(0000, 0001, 0011, 0111, 1111)$
2	$(0000, 0001, 0101, 0111, 1111)$
3	$(0000, 0001, 1001, 1011, 1111)$
4	$(0000, 0010, 0011, 1011, 1111)$
5	$(0000, 0010, 0110, 0111, 1111)$
6	$(0000, 0010, 1010, 1011, 1111)$
7	$(0000, 0100, 0101, 1101, 1111)$
8	$(0000, 0100, 0110, 1110, 1111)$
9	$(0000, 0100, 1100, 1101, 1111)$
10	$(0000, 1000, 1001, 1101, 1111)$
11	$(0000, 1000, 1010, 1110, 1111)$
12	$(0000, 1000, 1100, 1110, 1111)$

(b)

Fig. 6. Selection of the representative sets for $F = 4$ that achieve the capacity, Example 2. (a) Multisymbols for the 12 inputs with minimal expected Hamming distance 1. (b) Multisymbols for the 12 inputs with minimal expected Hamming distance 0.75.

the choice of the multisymbols is capacity achieving. The multisymbols are constructed by avoiding common edges as much as possible, except for the edges between $\mathcal{X}_0 = \{0000\}$ and $\mathcal{X}_1 = \{0001, 0010, 0100, 1000\}$. For example, we choose $t_2 = \{0000, 0001, 0101, 1101, 1111\}$ instead of $t_2 = \{0000, 0001, 0101, 0111, 1111\}$ in order to avoid a common edge with $t_1 = \{0000, 0001, 0011, 0111, 1111\}$ in the last section of the graph. The minimal expected Hamming distance for this choice of multisymbols is 1. We expect that this set will perform better compared to the set of multisymbols in Example 1, due to the better distance spectrum. This is confirmed by the simulation results.

Surprisingly, we have been able to find a set of multisymbols with minimal expected Hamming distance 0.75 which performs better than the above set with minimal distance 1, as presented in the simulation results in the next section. The set of multisymbols is presented on Fig. 6 b). We suspect that the reason for this behavior is that the minimal distance itself is not decisive for the performance, even if the conjecture that the expected Hamming distance is the relevant metric for the error-control coding holds. Nevertheless, the performance in both cases is superior to the naïve scheme, which makes the case for the relevance of the multisymbol framework in the design of practical error-control schemes.

3) Choice of the multisymbols, Example 3: As a third example, we choose the 12 multisymbols as shown in Fig. 7

t	$\{\mathbf{x}_s(t)\}$
1	$(0000, 0001, 0011, 0111, 1111)$
2	$(0000, 0001, 0011, 0111, 1111)$
3	$(0000, 0001, 0101, 0111, 1111)$
4	$(0000, 0010, 0110, 1110, 1111)$
5	$(0000, 0010, 1010, 1011, 1111)$
6	$(0000, 0010, 1010, 1011, 1111)$
7	$(0000, 0100, 0101, 1101, 1111)$
8	$(0000, 0100, 0110, 1110, 1111)$
9	$(0000, 0100, 1100, 1101, 1111)$
10	$(0000, 1000, 1001, 1101, 1111)$
11	$(0000, 1000, 1001, 1101, 1111)$
12	$(0000, 1000, 1100, 1110, 1111)$

(a)

Fig. 7. Selection of the representative sets for $F = 4$, Example 3. The selection of the multisymbols yields non-uniform distribution of $\mathcal{T}$. (a) Multisymbols for the 12 inputs.

a). We notice that, according to this choice, the multisymbols t_1 and t_2 are identical. Actually, besides $t_1 = t_2$, we have also $t_5 = t_6$ and $t_{10} = t_{11}$. We note that this choice does not violate the conditions for minimal multisymbols and satisfies the target distribution over X, thus it is capacity achieving.

At the first sight, this result seems counterintuitive and reveals the following problem: why we do not lose capacity even if we are not using the highest possible diversification at the input (in this case we assign the same multisymbol to two input symbols t)? We note that the cardinality of the input symbols is not 12, but 9. However, they are non-uniformly distributed — for example, 6 have probability $\frac{1}{12}$ and 3 have probability $\frac{1}{6}$. Until now, we have constrained ourselves to uniform distribution over the input symbols. However, it can be shown that if non–uniform distribution is used over $\mathcal{T}$, then capacity can be achieved even with $|\mathcal{T}| < L = \text{lcm}\left(\binom{F}{0}, \binom{F}{1}, \ldots, \binom{F}{F}\right)$. For this particular instance with $F = 4$, it can be shown that in the above example the capacity can be achieved by a set $\mathcal{T}$ with cardinality $|\mathcal{T}| = 8$. The probability distribution of the input symbols $\mathcal{T}$ is $P_T(t) = \frac{1}{6}$ for $t = 1, 2, 3, 4$ and $P_T(t) = \frac{1}{12}$ for $t = 5, 6, 7, 8$. In the following we specify only the nonzero members $\mathbf{P}$, the transition matrix for the channel $T - \mathbf{X}$. Note that the notation is slightly abused, with e. g. $P(0001|T = 1, 5)$ meaning $P(0001|T = 1)$ or $P(0001|T = 5)$): $P(0000|T = t) = P(1111|T = t) = \frac{1}{16}$ for any $t = 1 \ldots 8$; $P(0001|T = 1, 5) = P(0010|T = 2, 6) = P(0100|T = 3, 7) = P(1000|T = 4, 8) = \frac{4}{16}$ $P(0011|T = 1) = P(0110|T = 2) = P(1100|T = 3) = P(1001|T = 4) = \frac{6}{16}$; $P(0101|T = 5, 7) = P(1010|T = 6, 8) = \frac{6}{16}$, and $P(0111|T = 1, 5) = P(1110|T = 2, 6) = P(1101|T = 3, 7) = P(1011|T = 4, 8) = \frac{4}{16}$. The general case of $T-\mathbf{X}$ with non–uniform distribution on $\mathcal{T}$ and minimal required size $|\mathcal{T}|$ to achieve the capacity is outside of the scope for this paper and is a topic of ongoing work.

In the following section we present worked-out examples of trellis codes based on the multisymbol framework.

V. Code Design and Simulation Results

The coding scheme is a concatenation of an outer error correcting code, an interleaver and an encoder, as given in Fig. 8 a). The outer error correcting code is a convolutional

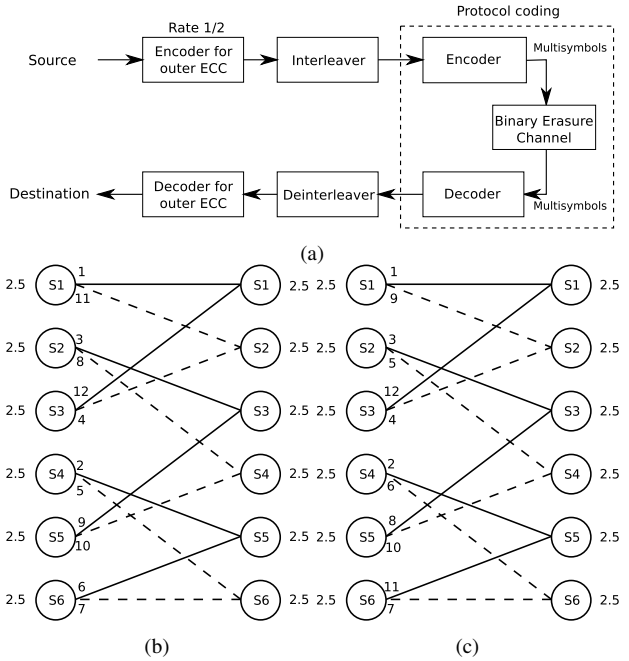


Fig. 8. Code Design. a) Block diagram of the code. b) Trellis construction for the set of multisymbols of Example 1. (c) Trellis construction for the set of multisymbols of Example 2.

code with rate $1/2$, thus $2n$ binary symbols are generated from n symbols. As already discussed, the inner code is trellis based, each branch in the trellis is associated with an input symbol (binary) and output symbol which is one of the L multisymbols. We associate multisymbols with the transitions in the trellis such that the output symbols on the branches exiting from the same state should be maximally separated in terms of expected Hamming distance. The same is valid for the output symbols associated with the two branches that enter the same state. For this evaluation, we assume a binary erasure channel, but it should be reiterated that the capacity results presented are valid for a wider class of channels.

The error bursts are broken by an interleaver, implemented as $\lambda \times \frac{2n}{\lambda}$ matrix, with $2n$ divisible by λ. The trellis based coding scheme for $F = 4$ defines a trellis with 12 branches. One option is to have 4 states and 3 branches from each state or a code, which implies that the source information is in ternary symbols. A more practical option is to have a trellis with 6 states and 2 branches from each state, such that one binary symbol is transmitted for each multisymbol. The trellis design the sets of multisymbols from Example 1 and Example 2 is given on Fig. 8 (b) and (c), respectively. In both cases, the multisymbols associated with the transitions in the trellis are chosen such that the output symbols on the branches exiting from the same state are maximally separated in terms of the expected Hamming distance. In both cases the distance between the incoming/outgoing multisymbols is maximal, 2.5 for all states.

The simulations have been performed with $\lambda = 8$ and the results are averaged over 10000 iterations. The simulation is performed for packet lengths $N = \{2, 6, 14, 30, 62\}$. These packet lengths are chosen such that $N + 2$ (two tail bits are added by the outer convolutional code) is divisible by $\lambda = 8$.

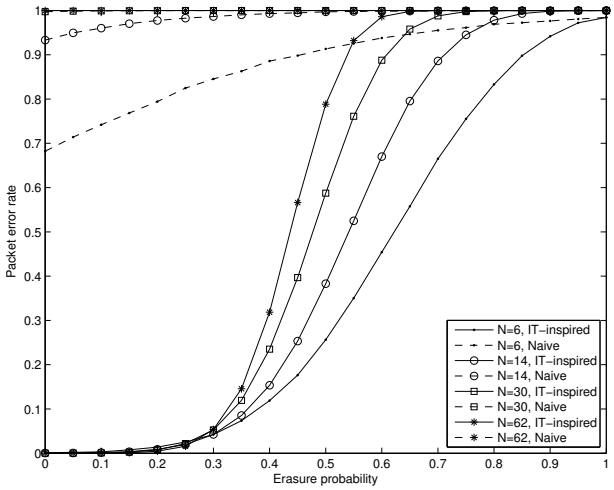
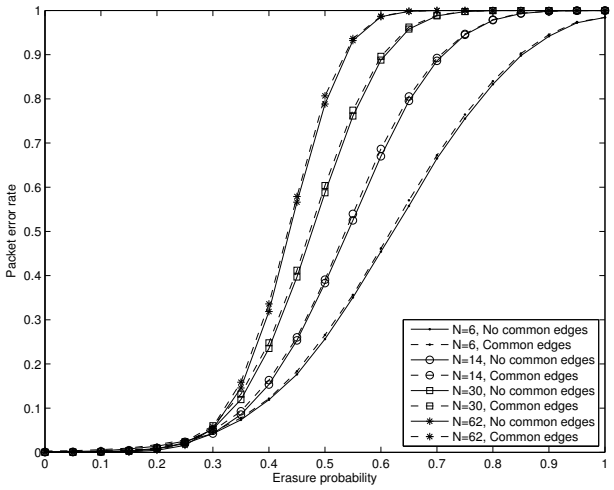


Fig. 9. Performance comparison between the naïve coding scheme and the scheme motivated from the multiuser framework

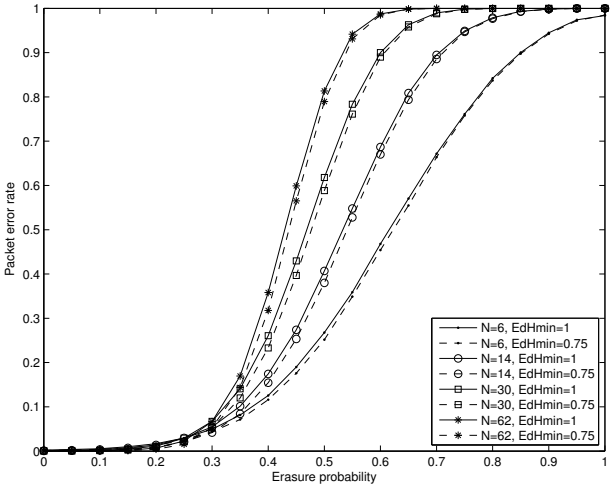
First, we compare the performance of the coding scheme inspired by the multisymbol framework and the naïve coding scheme, which does not account for the specifics of the secondary channel. The simulation results for the packet error rate (PER) for different erasure probability are shown in Fig. 9. This result present a clear evidence that the information-theoretic analysis carries a practical significance for the secondary communications channels. We also simulate the two different choices of the sets of multisymbols, with minimal expected Hamming distance 0.5 and 1 respectively, as presented in Section IV. As already commented, although the choice of the set with minimal distance 1 performs better than the set with minimal distance 0.5 (Fig. 10 a)), we were able to find another set with minimal distance 0.75 which outperforms both sets (Fig. 10 a)). This result indicates that besides the minimal expected Hamming distance, there are also other factors to be considered, notably the distance spectrum and the choice of the trellis transitions.

VI. Discussion

We used a simplified model, in which the set of packets sent in a given frame is independent from the other frames. In practice this is rarely satisfied, since buffering at the primary scheduler and/or packet retransmission due to errors creates dependencies between consecutive frames. In such a case, Shannon's result is not directly applicable and instead we need to use a more general model in which the sequence of frame states is not memoryless (see Section 6 in [18]). Another aspect is the freedom of in reordering user resources. For example, if in the case of WiMAX the scheduler puts each user on a channel where she can achieve a high data rate, then the freedom to permute users across channels becomes restricted. It is incorrect to say that protocol coding is not applicable once such restrictions are put by the primary system, but it should rather be observed that the secondary capacity is decreased.



(a)



(b)

Fig. 10. Performance of the error-correcting coding schemes. (a) Comparison between two sets with minimal distance 0.5 and 1 respectively. (b) Comparison between two sets with minimal distance 0.75 and 1 respectively.

This reiterates the observation that protocol coding can be used as a measure of how optimally given primary system operates.

VII. Conclusion and Future Work

We have introduced a class of communication channels with protocol coding, i. e. the information is modulated in the actions taken by the communication protocol of an existing, primary system. In particular, we have considered strategies in which protocol coding is done by combinatorial ordering of the labelled user resources (packets, channels) in the primary system. Differently from the previous works, our focus here is not on the steganographic usage of this type of protocol coding, but on its ability to introduce a new *secondary communication channel*. The communication model captures the constraints that the primary system operation puts on protocol coding i. e. the secondary information can only be sent by rearranging the set of packets made available by the primary system. The challenge is that the amount of information that can be sent in this way is not controllable by the secondary. We have derived the capacity of the secondary channel under arbitrary error models and provided practical trellis coding strategies.

As a future work, it is interesting to consider the case when the scheduling process in the primary system is generalized (e. g. buffering). Another direction is to compute the capacity under error models for channels with deletions/insertions. In practice, a secondary channel can be defined over virtually any existing wireless system and therefore it is of interest to find the coding strategies that are suited to the actual protocol specification in a certain primary system.

Acknowledgment

The authors would like to thank Prof. Osvaldo Simeone (NJIT) for useful discussions on the channels with causal channel state information at the transmitter.

References

[1] J. G. Andrews, A. Ghosh, and R. Muhamed, *Fundamentals of WiMAX*. Prentice-Hall, 2007.

[2] 3GPP, "LTE–Advanced," http://www.3gpp.org/article/lte-advanced.

[3] S.-Y. Lien, K.-C. Chen, and Y. Lin, "Toward Ubiquitous Massive Accesses in 3GPP Machine-to-Machine Communications," *IEEE Communications Magazine*, vol. 49, no. 4, pp. 66–74, Apr. 2011.

[4] J. L. Massey, *Channel Models for Random–Access Systems*, ser. Performance Limits in Communication Theory and Practice, NATO Advances Studies Institutes Series E142. Kluwer Academic, 1988, pp. 391–402.

[5] V. Anantharam and S. Verdu, "Bits through Queues," *IEEE Trans. Inform. Theory*, vol. 44, pp. 4–18, Jan. 1996.

[6] 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.

[7] T. Lutz, C. Hausl, and R. Kötter, "Bits Through Relay Cascades with Half–Duplex Constraint," 2009, submitted, (arXiv:0906.1599).

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

[9] A. Ephremides and B. Hajek, "Information Theory and Communication Networks: An Unconsummated Union," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2416–2434, Oct. 1998.

[10] 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.

[11] 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.

[12] A. El-Atawy and E. Al-Shaer, "Building covert channels over the packet reordering phenomenon," in *Proc. of IEEE INFOCOM*, Apr. 2009.

[13] A. J. H. Vinck, "Coded Modulation for Power Line Communications," *AEÜ Journal*, pp. 45–49, Jan. 2000.

[14] 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. 51–64, 2004.

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

[16] Z. Utkovski and P. Popovski, "Protocol Coding with Reordering of User Resources: Capacity Results for the Z-Channel," in *49th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, Illinois, USA, Sep. 2011.

[17] C. E. Shannon, "Channels with side information at the transmitter," *IBM Journal of Research and Development*, vol. 2, pp. 289–293, Oct. 1958.

[18] G. Keshet, Y. Steinberg, and N. Merhav, *Channel Coding in the Presence of Side Information*, ser. Foundations and Trends in Communications and Information Theory, 2007, vol. 4, no. 6.

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