An IR-HARQ scheme for covert communications

Meritxell Lamarca
Universitat Politècnica de Catalunya Barcelona, Spain , Email: meritxell.lamarca@upc.edu

Abstract—**This paper explores the advantages of incremental redundancy hybrid-ARQ schemes for covert communications in which a feedback channel is available. The use of efficient incremental redundancy with sub-codewords of different sizes provides significant gains in terms of covertness while preserving the error rate performance. The convexity of covertness metrics is exploited to design the size and degree profile of each block when a concatenated code based on MDNL codes is employed in the binary symmetric channel.**

I. Introduction

In covert communications a transmitter (Alice) wants to send a message to a receiver (Bob) while keeping the communication undetectable to a second receiver (the Warden) under some covertness metrics. Several authors have focused on the study of the theoretical limits covert communications ([1], [2], [3], [4]). It has been shown that in many relevant scenarios when the Warden channel is worse than the Alice-Bob channel the maximum number of bits that can be reliably transmitted while keeping the transmission covert from the Warden grows asymptotically as $O(\sqrt{n})$, being n the codeword length, even if Alice and Bob do not share any secret key. Furthermore, a scheme based on pulse position modulation (PPM) that is asymptotically optimal has been proposed for binary-input discrete memoryless channels and the Gaussian channel ([5],[6],[7]). In all these works the focus has been on the amount of information bits that could be sent when the codeword length increased unbounded. Unfortunately, the behavior for asymptotically large n is not relevant for those applications where latency must be kept low and/or the transceivers have limited memory, computational capacity and power consumption. This is particularly important for covert communications, as a linear increase in the number of information bits leads to a quadratic increase in the codeword length.

This work focuses on the finite-length regime and highlights that the use of Incremental Redundancy Hybrid Automatic Repeat Request (IR-HARQ) schemes can provide great benefits in this scenario. The analysis considers the Binary Symmetric Channel (BSC). Roughly speaking, for a fixed codeword length (i.e. fixed latency) and a fixed amount of information bits (i.e. fixed rate) there is a trade-off between error performance and covertness in terms of the codebook Hamming weight distribution. The higher the codeword Hamming weights, the smaller the error rate and the easier it is for the Warden to detect the communication. As shown in this paper, the use IR-HARQ schemes allows to improve this trade-off when a feedback channel is available.

IR-HARQ is commonly used as a technique to increase throughput in time-variant channels (e.g. fading channels) ([8]). However, the application of HARQ schemes to covert systems remains unexplored. The benefits of IR-HARQ in terms of covertness are twofold. First, the Warden's uncertainty of the number of re-transmissions required for successful decoding increases covertness. Second, as subsequent re-transmissions occur scarcely, the average codeword length and Hamming weight can be lowered and covertness can be improved with no error performance penalty.

In this paper the design and performance of IR-HARQ for the BSC based on the non-linear covert codes proposed in [9] is addressed. As opposed to the aforementioned PPM scheme, these codes offer flexibility in codeword length and simplicity in the code design, as well as good performance for short and moderate codeword lengths.

In this paper covertness is measured in terms of variational distance and relative entropy. Their relationship with the parameters of the IR-HARQ scheme is very intricate, so this work exploits their convexity to simplify the design. The result is the optimization of a cost function that is a pessimistic bound on the covertness or it corresponds to the true covertness depending on the Warden's access to the feedback channel. The design aims at optimizing the covertness while keeping the error rate and the maximum number of channel uses fixed. The proposed approach allows to predict the system performance without the need of any IR-HARQ simulation. The design relies on the measurement by means of MonteCarlo simulation of the word error rate (WER) and Hamming weight distributions of a family of codes with different lengths and degree profiles.

Throughout the paper, vectors and matrices are indicated in boldface, $(\mathbf{v})_j$ indicates the j-th element in vector $\mathbf{v}$ and $\mathbf{v}_i^j$ indicates the stacking of a set of vectors : $\mathbf{v}_i^j = [\mathbf{v}_i \mathbf{v}_{i+1} \dots \mathbf{v}_j]$.

II. Scenario definition

We consider an IR-HARQ covert scheme as depicted in figure 1, in which the Alice-Bob and Alice-Warden channels are Binary Symmetric Channels (BSC) with crossover probabilities ϵ_b and ϵ_w respectively. There is also a Bob-Alice feedback channel through which positive/negative acknowledge (ACK/NACK) messages are sent. We consider this feedback channel to be noiseless. Two different scenarios arise depending on whether the Warden can or cannot have access to the data in this feedback channel, as described below.

This work has been supported by the Spanish Ministry of Science and Innovation through projects WINTER (TEC2016-76409-C2-1-R / AEI / FEDER, UE) and RODIN (PID2019-105717RB-C22 / AEI / 10.13039/501100011033) and by the Catalan Government (AGAUR) under grant 2017 SGR 578.

With probability P_t Alice sends a message to Bob. If a transmission occurs, Alice employs a mother code to encode the k information bits of the message $\mathbf{M}$ into a codeword of $n_{all} = n_1 + \ldots + n_T$ bits. Initially, the sub-codeword $\mathbf{C}_1$ containing n_1 coded bits is transmitted. If Bob's attempt to decode the codeword is successful Bob sends Alice an ACK message and the transmission is finished; otherwise Bob sends a NACK message and Alice sends n_2 additional coded bits in codeword $\mathbf{C}_2$, so Bob attempts to decode again using all the $n_1 + n_2$ received bits. The procedure is repeated after each re-transmission request until the message is successfully decoded or the maximum number of transmission attempts T is reached. The number of bits of each sub-codeword n_i can be different.

Let's denote as $\mathbf{C}_i$, $\mathbf{X}_i$, $\mathbf{Y}_i$ and $\mathbf{Z}_i$ $1 \leq i \leq T$ the sub-codewords, the encoder output and the observations at Bob's and Warden's end for each of the T transmission blocks. Let's also indicate Bob's decoding result (either success or failure) as $D_i \in \{s, f\}$, and the failure probability after the transmission of $\mathbf{C}_1^i$ as WER_i. When no transmission occurs or when Bob has already successfully decoded the message, in the i-th block the innocent all-zero symbol of length n_i is transmitted, indicated as $\mathbf{O}_i$. The Warden's task is to detect whether a transmission has taken place or his observation corresponds to a noisy replica of $\mathbf{O}_1^T$.

We assume that Alice, Bob and the Warden know the values ϵ_b and ϵ_w and the Warden has full knowledge of the coding scheme described in the next section except for the scrambling sequence. Regarding the feedback channel, in this paper two scenarios are considered in the covertness analysis. In the first one, the feedback channel is not accessible to the Warden. As the Warden does not know which of Bob's decoding attempts was successful, he does not how many re-transmissions were required. In the second scenario, the Warden has full access to the feedback channel, so in order to fool him Bob generates random ACK/NACK messages even when no information has been sent, breaking the policy to remain silent when no transmission occurs. In this case, the Warden does not know if Alice transmitted to Bob, but he knows how many sub-codewords would have been employed if she did.

III. Code description

In this paper the channel codes introduced in in [9] are extended to the IR-HARQ set-up.

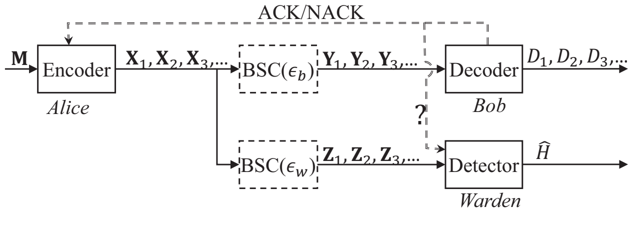
ACK/NACK
M
Encoder
Alice
$\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \ldots$
BSC(ϵ_b)
$\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3, \ldots$
Decoder
Bob
$D_1, D_2, D_3, \ldots$
?
BSC(ϵ_w)
$\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3, \ldots$
Detector
Warden
$\widehat{H}$

Figure 1. Block diagram of the IR-HARQ covert scheme.

A. Covert channel codes based on MDNL codes

The architecture proposed in [9] is depicted in Figure 2. It consists of the serial concatenation of a linear and an non-linear code that are jointly decoded using an iterative procedure. The non-linear stage is a Moderate Density Non-Linear (MDNL) code.

An information word of k bits is fed to a linear code to generate a codeword of l bits and its output is fed to the non-linear code of length n. The value of l is chosen to obtain the desired balance between the rate of the two codes; generally most of the rate is allocated to the nonlinear code.

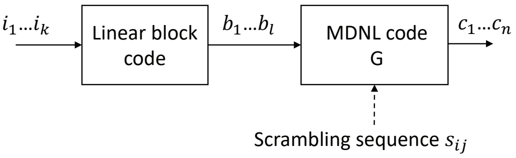

Figure 2. Block diagram of the proposed concatenated code.

In the simulations a low-density parity-check (LDPC) code ([10]) was employed for the linear stage, but it could have been replaced by any linear code that had good convergence properties when concatenated with the non-linear code.

The inner code is a MDNL code. They are graph-based codes in which the AND operator is employed to generate low-weight codewords. Specifically, each coded bit, c_j is obtained as the AND of a few possibly negated input bits, b_i:

$$c_j = \prod_{i \in \mathcal{I}_j \cup \mathcal{N}_j} (b_i \oplus s_{ij}) \;\; j = 1 \ldots n \;\; s_{ij} = \begin{cases} 0 & i \in \mathcal{I}_j \\ 1 & i \in \mathcal{N}_j \end{cases} \quad (1)$$

where '$\oplus$' stands for the XOR operation and s_{ij} is a scrambling sequence that implements the negation operation. This equation leads to the representation of the MDNL code with a bipartite graph in which the belief propagation decoding algorithm can be applied (see details in [9]).

The MDNL code is characterized by its moderately sparse generator matrix $\mathbf{G}$ of size $l \times n$, which has a +1 (-1) in the (i, j)-th element if the i-th input bit participates without (with) negation in the computation of the j-th coded bit. Analogously to LDPC codes, the generator matrix is chosen randomly from the ensemble of matrices having a certain amount of non-zero entries per row and per column. As in this paper sub-codewords of different lengths are employed in the HARQ scheme, we introduce a small change in the notation employed in [9] and in LDPC codes. We denote as AND degree profile of the MDNL code and indicate it by $(\mathbf{n}, \mathbf{d})$ the pair of vectors that store in the j-th component $(\mathbf{n})_j$ the number of columns in $\mathbf{G}$ that have $(\mathbf{d})_j$ ones, i.e. the number of AND nodes in the graph that have degree $(\mathbf{d})_j$. In all the simulations the rows had constant or almost constant number of ones.

In the non-ARQ scenario the proposed concatenated code does not attain the asymptotic theoretical limit because it requires the use of a pseudo-random sequence that remains unknown to the Warden, whereas the aforementioned PPM scheme is asymptotically optimal. Nevertheless, the simulations presented in [9] showed that for moderate codeword lengths the proposed code had better WER performance than the scheme based on PPM for the same covertness constraint and analogous design of the linear outer coding stage. Furthermore, in the PPM-based scheme the codeword size is constrained by the covertness specification, whereas the scheme based on MDNL codes has absolute flexibility. MDNL codes can be regarded as rate-compatible codes, since extra coded bits can be generated by adding columns to matrix $\mathbf{G}$. The lack of length constraints and the good performance for short blocklengths makes MDNL codes a good candidate for covert IR-HARQ schemes; this is the reason that motivated their use in this paper.

B. *WER indicators*

Figure 3 depicts the codeword Hamming weight distribution for the configuration $k = 300, l = 400, n = 60000$, an LDPC code with bit node degree 2 and a regular MDNL code with AND degree profile (60000, 10) in a BSC with $\epsilon_b = 7e-3$. In this case the concatenated code had $WER = 0.15$. The distribution is shown for the codewords that were successfully/unsuccessfully decoded as well as for the all of them. As evidenced, codewords with low Hamming weight are more prone to errors than those with high weight. This strong dependency between the Hamming weight and the error probability of the concatenated code must be taken into account in the covertness analysis of the HARQ scheme, since sub-codewords with low Hamming weight are more likely to require transmission of additional bits.

Figure 4 illustrates another feature relevant for IR-HARQ code design. It depicts the WER performance for the same BSC and the same linear stage in figure 3 for MDNL codes with lengths in the range $40000 \leq n \leq 64000$ and AND nodes of degrees 9 and 10, having a fraction of nodes of degree 9 ranging from 0% to 20%. The same WER values are depicted as a function of the codeword length and as a function of the total mutual information provided by n independent bits generated by AND nodes with the same degree profile as the MDNL code, I. If we denote the degree of the j-th node as d_j then the probability of a zero channel output is given by $\mathrm{P}(Y_j = 0) = \left(1 - 2^{-d_j}\right)(1 - \epsilon_b) + 2^{-d_j}\epsilon_b$ and the total mutual information is given by

$$I = \sum_{1 \leq j \leq n} \left(H_b\left(\mathrm{P}(Y_j = 0)\right) - H_b\left(\epsilon_b\right)\right) \tag{2}$$

where H_b stands for the binary entropy. Hence, the plot evidences that the mutual information has a linear dependency on the code degree profile and it is an excellent indicator of the WER performance irrespective of the $(\mathbf{n}, \mathbf{d})$ values. This fact is exploited in the code design in section V.

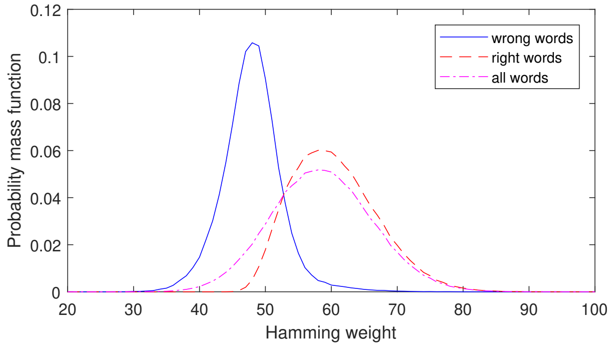

Figure 3. Codeword Hamming weight distribution.

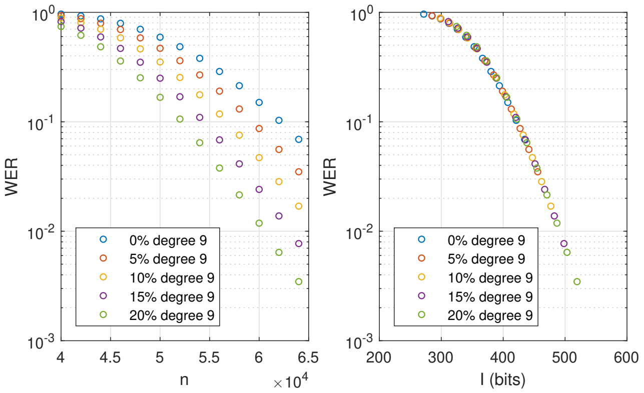

Figure 4. WER of the concatenated code for different MDNL degree profiles.

C. *Extension to IR-HARQ*

To extend the covert scheme proposed in [9] to IR-HARQ the linear code in figure 2 is concatenated with a mother MDNL code whose generator matrix of size $l \times n_{all}$ is created block wise as $\mathbf{G}_{all} = [\mathbf{G}_1 \ldots \mathbf{G}_T]$. Each block $\mathbf{G}_i$ has size $l \times n_i$, a degree profile $(\mathbf{n}_i, \mathbf{d}_i)$ and is randomly generated from the ensemble of matrices of the desired size and degree profile. Compared to the scheme without HARQ the main difference is that now each block has a regular row degree, whereas in the former case the row degree is only regular for $\mathbf{G}_{all}$ but not for its parts. The MonteCarlo simulations have shown that this difference does not have an impact on the code performance.

In the sequel we denote as $\left(\mathbf{n}_1^i, \mathbf{d}_1^i\right)$ the vector that stores the details of the degree profile in blocks $1 \ldots i$, and we denote as $(\mathbf{n}_{1:i}, \mathbf{d}_{1:i})$ the degree profile of the matrix $\mathbf{G}_1^i = [\mathbf{G}_1 \ldots \mathbf{G}_i]$ considered as a single code. Thus, vectors $\mathbf{n}_{1:i}$ and $\mathbf{d}_{1:i}$ have as many components as different column degrees are included in $\mathbf{G}_1^i$. Note that while the overall WER of the IR-HARQ scheme only depends on $\mathbf{G}_{all}$ and its degree profile $(\mathbf{n}_{1:T}, \mathbf{d}_{1:T})$, the covertness performance and the number of re-transmissions depends on $\left(\mathbf{n}_1^T, \mathbf{d}_1^T\right)$.

IV. Covertness Analysis

A. *Covertness analysis in IR-HARQ*

Let's consider the case of IR-HARQ with at most T transmissions. There are $T + 1$ possible disjoint events that can be described using the notation in section II as:

- E_0: No transmission is made. During all transmission blocks the transmitter sends the all-zero symbol, so $\mathbf{X}_{all} = \left[\mathbf{O}_1^T\right]$. $\mathrm{P}(E_o) = 1 - P_t$.
- E_1: sub-codeword $\mathbf{C}_1$ is transmitted and it is decoded successfully. Hence, $\mathbf{X}_{all} = [\mathbf{C}_1, \mathbf{O}_2^T]$ and $D_1 = s$. $\mathrm{P}(E_1) = P_t \cdot (1 - \mathit{WER}_1)$.
- E_i, $i = 2 \dots T-1$: The first $i-1$ decoding attempts fail, the i-th one is successful. Thus, $\mathbf{X}_{all} = [\mathbf{C}_1^i, \mathbf{O}_{i+1}^T]$, $D_{i-1} = f$ and $D_i = s$. $\mathrm{P}(E_i) = P_t \cdot (\mathit{WER}_{i-1} - \mathit{WER}_i)$.
- E_T: The maximum number of re-transmissions is needed. Thus, $\mathbf{X}_{all} = \left[\mathbf{C}_1^T\right]$, $D_{T-1} = f$. $\mathrm{P}(E_T) = P_t \cdot \mathit{WER}_{T-1}$.

We consider the Warden knows Alice's transmission scheme and the channel's quality ϵ_b and ϵ_w, so he knows $\{\mathit{WER}_i \;\, i = 1 \dots T\}$. The covertness analysis depends on the Warden's access to the feedback channel. We consider two cases:

1) The Warden has no access to the feedback channel: In this case the Warden does not know whether Bob's decoding attempts were successful or not. From the Warden's point of view, it is a binary hypothesis testing problem:

- H_0: event E_0 happened.
- H_1: Event E_1,... or E_T happened.

The Warden observes the signal received during the T possible transmission blocks and decides ($\hat{H}$) whether a message was sent or not. The Warden can make his decision based on the distributions $\mathrm{P}(\mathbf{Z}_1^T \mid H_0)$ and

$$\mathrm{P}(\mathbf{Z}_1^T \mid H_1) = \sum_{1 \le i \le T} \mathrm{P}(\mathbf{Z}_1^T \mid E_i)\, \mathrm{P}(E_i \mid H_1) \tag{3}$$

If $P_t = 0.5$ then the Warden's detection capability depends on the variational distance (VD) between these two distributions, as it is known that for any detector

$$\mathrm{P}(\hat{H} \neq H) \geq= \frac{1}{2}\left(1 - \mathrm{VD}[\mathrm{P}(\mathbf{Z}_1^T \mid H_0), \mathrm{P}(\mathbf{Z}_1^T \mid H_1)]\right) \tag{4}$$

Alternatively, the relative entropy between both distributions can also be employed to get an upper bound on the VD. Unfortunately, it is difficult to design the channel coding stage to optimize any of these two parameters, since there is no closed-form expression for their value for linear combinations of distributions such as the one in (3). Nevertheless, the convexity of the VD and the relative entropy can be exploited to upper-bound the attained covertness while keeping design complexity affordable. In terms of VD:

$$\begin{aligned} &\mathrm{VD}[\mathrm{P}(Z_1^T \mid H_1), \mathrm{P}(Z_1^T \mid H_0)] \leq \\ &\qquad \sum_{1 \le i \le T} \mathrm{P}(E_i \mid H_1)\, \mathrm{VD}[\mathrm{P}(Z_1^T \mid E_i), \mathrm{P}(Z_1^T \mid E_0)] \end{aligned} \tag{5}$$

and an analogous upper bound can be obtained for the relative entropy.

2) The Warden has full access to feedback channel: If the Warden has a noiseless observation of the feedback channel then covertness is not feasible with ARQ unless Bob sends ACK/NACK random messages also in the case that no information is transmitted. If so, when the Warden observes an ACK message after the time for the i-th transmission block has elapsed, he must decide between the two hypothesis:

- H_0: event E_0 happened.
- H_1: Event E_i happened.

It can be shown that in this case the overall probability of a Warden detection error is given by equation (4) replacing the VD in that equation by the right hand side of equation (5), so in this case the expression in (5) is exact, it is not a bound.

B. Covertness for the proposed codes

This section evaluates the right hand side of of equation (5) when the channel coding stage is designed using the concatenated code presented in section III. The analysis when covertness is measured in terms of relative entropy is analogous.

Let's denote by $\mathbf{h}_i$ (and $\mathbf{w}_i$) the vector of the same size as $\mathbf{d}_i$ whose coefficient $(\mathbf{d}_i)_j$ stores the Hamming weight of the bits in sub-codeword $\mathbf{C}_i$ (and of the corresponding Warden observation $\mathbf{Z}_i$) that are associated to AND nodes with the degree $(\mathbf{d}_i)_j$. Thus, the sum of the elements in $\mathbf{h}_i$ (and of those in $\mathbf{w}_i$) equals the Hamming weight of the sub-codeword $\mathbf{C}_i$ (and of $\mathbf{Z}_i$). Let's define $\mathbf{h}_1^T$ and $\mathbf{h}_{1:T}$ (and $\mathbf{w}_1^T$ and $\mathbf{w}_{1:T}$) accordingly.

The covertness analysis of the proposed codes relies on the following assumptions:

1) The Warden has full knowledge of the transmission scheme except for the scrambling sequence employed in the MDNL code. Thus, from his the point of view the bits in a sub-codeword are generated as independent Bernouilli random variables with probability of 1 equal to 2^{-d_p}, being d_p the degree of the associated AND node.
2) Given a mother code, all sub-codewords $\left[\mathbf{C}_1^i\right]$ with the same value of $\mathbf{h}_{1:i}$ have the same error probability. If the MDNL code were employed as a stand-alone code, this assumption would not be true, because the decoding success could depend on the location of the coded bits that are 1 and on the code graph. As the proposed scheme concatenates the MDNL code with an outer code and there are no constraints on the structure of that outer code, this assumption can be considered to be true for large codes with randomly picked generator matrices. This assumption is required to prevent the Warden from exploiting the knowledge of the generator matrix $\mathbf{G}_1^i$ to establish a dependency between the location of the coded bits that are 1 in $\mathbf{C}_1^i$ and the likelihood of a decoding failure, i.e. the likelihood of transmitting $\mathbf{C}_{i+1}$.
3) All codes with the same degree profile have the same WER performance. Since the MDNL code generator matrix is chosen randomly from the ensemble of matrices with the desired degree profile $\left(\mathbf{n}_1^T, \mathbf{d}_1^T\right)$, this assumption is valid with very high probability as long as the code is long enough. Thanks to this assumption the values of $\{\mathit{WER}_i\}_{i=1\dots T}$ depend only on $(\mathbf{n}_{1:i}, \mathbf{d}_{1:i})$ instead of depending on $\left(\mathbf{n}_1^i, \mathbf{d}_1^i\right)$.

Because of the first two assumptions, the only data the Warden can exploit to detect the transmission from the observation of $\mathbf{Z}_1^T$ is the knowledge of the probability of each one of the coded bits being 1, of $\left(\mathbf{n}_1^T, \mathbf{d}_1^T\right)$ and of $\mathrm{P}(E_i \mid H_1)$.

In the proposed scheme, the metrics for optimum detection at the Warden simplify to the evaluation of the Hamming weight of the observations $\mathbf{Z}_1^T$, computed separately for each sub-codeword and each set of coded bits being associated to AND nodes of same degree, i.e. the parameter $\mathbf{w}_1^T$ defined earlier. Thus, the covertness of the proposed scheme depends on the variational distance

$$\Phi_{VD} = \mathrm{VD}[\mathrm{P}(\mathbf{w}_1^T \mid H_1), \mathrm{P}(\mathbf{w}_1^T \mid H_0)] \tag{6}$$

As in the general case, the dependency of this VD on the code parameters is very intricatE. However, exploiting that

$$\mathrm{P}(\mathbf{w}_1^T \mid H_1) = \sum_{1 \leq i \leq T} \mathrm{P}(\mathbf{w}_1^T \mid E_i)\, \mathrm{P}(E_i \mid H_1) \tag{7}$$

and resorting to convexity an upper bound is obtained that can be employed for code optimization:

$$\begin{aligned} &\Phi_{VD} \leq \Phi_{VD}^{Bound} \\ &= \sum_{1, \leq i \leq T-1} \left(\mathit{WER}_{i-1} - \mathit{WER}_i\right) \Phi_{VD}^i + \mathit{WER}_{T-1} \Phi_{VD}^T \end{aligned} \tag{8}$$

where $\mathit{WER}_0 = 1$ and we have defined

$$\Phi_{VD}^i = \mathrm{VD}[\mathrm{P}(\mathbf{w}_1^T \mid E_i), \mathrm{P}(\mathbf{w}_1^T \mid E_0)] \qquad 1 \leq i \leq T. \tag{9}$$

As shown next, all the terms in equations (8)-(9) can be obtained from MonteCarlo simulations in a non-ARQ scenario of a code with appropriate degree profile. Let's consider that MonteCarlo simulations are run for the concatenated code in figure 2 with a fixed linear stage and different configurations of the MDNL code with degree profiles $\{(\mathbf{n}_{1:i}, \mathbf{d}_{1:i})\}_{i=1\ldots T}$, evaluating its WER and the Hamming weight distribution of its codewords.

Let's focus in one of the MonteCarlo simulations, for a generic degree profile $(\mathbf{n}, \mathbf{d})$. Let's denote as $\mathit{Bin}\,(w, p, n) = \binom{n}{w} p^w (1-p)^{n-w}$ the binomial distribution. When there is no transmission and the innocent symbol $\mathbf{O}$ is placed in the channel the multidimensional distribution of the Hamming weight at the Warden is given by $B_0\,(\mathbf{w}, \mathbf{n}) = \prod_j \mathit{Bin}\left((\mathbf{w})_j, \epsilon_w, (\mathbf{n})_j\right)$. When there is transmission, thanks to Assumption 1, the distribution of the Hamming weight $\mathbf{h}$ follows a Binomial distribution for each of the dimensions: $\mathrm{P}(\mathbf{h}) = B_1\,(\mathbf{h}, \mathbf{n}, \mathbf{d}) = \prod_j \mathit{Bin}\left((\mathbf{h})_j, 2^{-(\mathbf{d})_j}, (\mathbf{n})_j\right)$. However, the MonteCarlo simulation is required to evaluate the WER, denoted as $\mathit{WER}\,(\mathbf{n}, \mathbf{d})$, and to evaluate the Hamming weight distribution of the codewords that were successfully decoded and those that were decoded in error. Let's denote both distributions as $\mathit{Su}_C\,(\mathbf{h}, \mathbf{n}, \mathbf{d})$ and $\mathit{Un}_C\,(\mathbf{h}, \mathbf{n}, \mathbf{d})$. As indicated in section III-B, these two distributions are different. Note that $\mathrm{P}(\mathbf{h}) = \mathit{Su}_C\,(\mathbf{h}, \mathbf{n}, \mathbf{d})\,(1 - \mathit{WER}\,(\mathbf{n}, \mathbf{d})) + \mathit{Un}_C\,(\mathbf{h}, \mathbf{n}, \mathbf{d})\, \mathit{WER}\,(\mathbf{n}, \mathbf{d})$. These codeword Hamming weight distributions can be employed to find the corresponding ones for the Warden observations. For a generic set of n coded bits with Hamming weight distribution $\mathrm{P}(h)$ the distribution of the Hamming weight of the n corresponding Warden observations would be

$$\mathrm{P}(w) = \sum_h \mathrm{P}(h)\, \mathrm{P}(w \mid h) \tag{10}$$

$$\mathrm{P}(w \mid h) = \sum_k \mathit{Bin}\,(k, \epsilon_w, n-h)\, \mathit{Bin}\,(h - w + k, \epsilon_w, h) \tag{11}$$

Applying these equations to every dimension in the distributions $\mathrm{P}(\mathbf{h})$, $\mathit{Su}_C\,(\mathbf{h}, \mathbf{n}, \mathbf{d})$ and $\mathit{Un}_C\,(\mathbf{h}, \mathbf{n}, \mathbf{d})$ the corresponding distributions for the Warden observations $\mathrm{P}(\mathbf{w})$, $\mathit{Su}_Z\,(\mathbf{w}, \mathbf{n}, \mathbf{d})$ and $\mathit{Un}_Z\,(\mathbf{w}, \mathbf{n}, \mathbf{d})$ can be found.

All WER values and the VD involved in equations (8)-(9) can be obtained by the knowledge of these distribution families and the measured WER values in the non-ARQ scenario. For the WER:

$$\mathit{WER}_i = \mathit{WER}\,(\mathbf{n}_{1:i}, \mathbf{d}_{1:i}) \tag{12}$$

and for the VD the multidimensional distributions can be replaced by the following ones indicated by '$\Rightarrow$' because the VD for both is the same when assumption 2 holds.

$$\mathrm{P}(\mathbf{w}_1^T \mid E_0) = B_0\left(\mathbf{w}_1^T, \mathbf{n}_1^T\right) \Rightarrow B_0\,(\mathbf{w}_{1:T}, \mathbf{n}_{1:T}) \tag{13}$$

$$\begin{aligned} \mathrm{P}(\mathbf{w}_1^T \mid E_1) &= \mathit{Su}_Z\,(\mathbf{w}_1, \mathbf{n}_1, \mathbf{d}_1)\, B_0\left(\mathbf{w}_2^T, \mathbf{n}_2^T\right) \\ &\Rightarrow \mathit{Su}_Z\,(\mathbf{w}_1, \mathbf{n}_1, \mathbf{d}_1)\, B_0\,(\mathbf{w}_{2:T}, \mathbf{n}_{2:T}) \end{aligned} \tag{14}$$

$$\begin{aligned} &\mathrm{P}(\mathbf{w}_1^T \mid E_i) = \\ &\quad \mathrm{P}(\mathbf{w}_1^{i-1}, \mathbf{w}_i \mid D_{i-1,i} = (f, s))\, B_0\left(\mathbf{w}_{i+1}^T, \mathbf{n}_{i+1}^T\right) \\ &\Rightarrow \mathrm{P}(\mathbf{w}_{1:i-1}, \mathbf{w}_i \mid D_{i-1,i} = (f, s))\, B_0\,(\mathbf{w}_{i+1:T}, \mathbf{n}_{i+1:T}) \\ &\qquad 1 < i < T \end{aligned} \tag{15}$$

$$\begin{aligned} &\mathrm{P}(\mathbf{w}_1^T \mid E_T) \Rightarrow \mathrm{P}(\mathbf{w}_{1:T-1} \mid D_{i-1} = f) B_1\,(\mathbf{w}_T, \mathbf{d}_T, \mathbf{n}_T) \\ &= \mathit{Un}_Z\,(\mathbf{w}_{1:T-1}, \mathbf{n}_{1:T-1}, \mathbf{d}_{1:T-1})\, B_1\,(\mathbf{w}_T, \mathbf{d}_T, \mathbf{n}_T) \end{aligned} \tag{16}$$

After MonteCarlo evaluation of the Hamming weight distributions of the coding scheme without ARQ all the terms in these equations have known values except for $\mathrm{P}(\mathbf{w}_{1:i-1}, \mathbf{w}_i \mid D_{i-1,i} = (f, s))$. However, this term can also be computed from the known code statistics, as shown next. Let's consider the Hamming weight joint distribution at the Warden for blocks $i+1$ and the previous ones when $\mathbf{C}_1^i$ is transmitted. Let's consider the bits with a specific AND node degree d: $n_{1:i}$ bits in $\mathbf{C}_1^i$ and n_{i+1} bits in $\mathbf{C}_{i+1}$ (generalization to the multidimensional case is straightforward). It can be written as

$$\begin{aligned} &\mathrm{P}(w_{1:i}, w_{i+1}) = \\ &\quad \mathit{Su}\,(w_{1:i}, n_{1:i}, d)\, B_0\,(w_{i+1}, n_{i+1})\,(1 - \mathit{WER}_i) \\ &+ \mathrm{P}(w_{1:i}, w_{i+1} \mid D_{i,i+1} = (f, s))\,(\mathit{WER}_i - \mathit{WER}_{i+1}) \\ &+ \mathrm{P}(w_{1:i}, w_{i+1} \mid D_{i+1} = f)\, \mathit{WER}_{i+1} \end{aligned} \tag{17}$$

where the last term has been simplified taking into account that if the codeword cannot be successfully decoded with the bits in $\mathbf{C}_1^{i+1}$ bits then it cannot be decoded either with only $\mathbf{C}_1^i$. If $\mathrm{P}(w_{1:i}, w_{i+1} \mid D_{i+1} = f)$ is found then $\mathrm{P}(w_{1:i}, w_{i+1} \mid D_{i,i+1} = (f,s))$ can be isolated from this equation and can be found too because $\mathrm{P}(w_{1:i}, w_{i+1})$ is a binomial distribution and the other terms are also known. According to Assumption 2,

$$\begin{aligned}\mathrm{P}(w_{1:i}, w_{i+1} \mid D_{i+1} = f) = \\ \mathrm{P}(w_{1:i}, w_{i+1} \mid w_{1:i+1})\, Un_Z\left(w_{1:i+1}, n_{1:i+1}, d\right) = \\ \frac{\binom{n_{1:i}}{w_{1:i}}\binom{n_{i+1}}{w_{i+1}}}{\binom{n_{1:i+1}}{w_{1:i+1}}}\, Un_Z\left(w_{1:i+1}, n_{1:i+1}, d\right) \quad (18)\end{aligned}$$

Thus, all distributions involved in (9) have closed-form expressions or can be computed by MonteCarlo evaluation of the distributions $Su_C\left(\mathbf{h}, \mathbf{n}, \mathbf{d}\right)$ and $Un_C\left(\mathbf{h}, \mathbf{n}, \mathbf{d}\right)$ for the codes with degree profiles $\{(\mathbf{n}_{1:i}, \mathbf{d}_{1:i})\}_{i=1\ldots T}$. Therefore, the bound on the VD of the IR-HARQ scheme can be obtained without carrying out any simulation of the IR-HARQ scheme.

V. Code design in IR-HARQ

This section addresses the design of the channel coding stages in the IR-HARQ transmission scheme. In the non-ARQ scenario covertness can be increased by reducing the codeword length and/or reducing the codeword Hamming weight. However, the impact of both actions on the IR-HARQ covertness is not clear because if they were applied to improve the covertness of the blocks $\mathbf{C}_1^i$ they would increase WER_i and the need for a re-transmission, which in its turn could degrade the covertness of the overall ARQ scheme. Hence, the design of the code for all transmission blocks must be done jointly.

We consider the design of the codes in section III so that the bound in (8)-(9) is optimized subject to a constraint on the maximum number of channel uses (n_{all}) and a constraint on the mutual information, which is essentially equivalent to a constraint on WER_T as seen in figure 4. Hence, as the linear stage is fixed and the MDNL code is randomly generated, the code design simplifies to the selection of the degree profile of each stage $\{(\mathbf{n}_i, \mathbf{d}_i)_{1 \leq i \leq T}\}$.

If we denote by $I\left(d\right)$ the mutual information of a coded bit associated to an AND node of degree d according to equation (2), then the optimization problem can be formulated as

$$\min_{(\mathbf{n}_i, \mathbf{d}_i)\ 1 \leq i \leq T} \quad \Phi_{VD}^{Bound} \qquad (19a)$$

$$\text{subject to} \quad \sum_{1 \leq i \leq T} \sum_{all\ j} \left(\mathbf{n}_i\right)_j \cdot I\left(\left(\mathbf{d}_i\right)_j\right) \geq I^{target}, \qquad (19b)$$

$$\sum_{1 \leq i \leq T} \sum_{all\ j} \left(\mathbf{n}_i\right)_j \leq n_{all}. \qquad (19c)$$

where Φ_{VD}^{Bound} in (19a) corresponds to equation (8) and I^{target} stands for the mutual information that corresponds to the desired specification on WER_T as dictated by Figure 4.

The constraints in (19b)-(19c) are both linear on the coefficients in $\mathbf{n}_i, i = 1 \ldots T$. Unfortunately, as the VD bound in (19a) is not a convex function of the design parameters, the optimization is a difficult task. Nevertheless, its complexity can be reduced taking into account that a regular or almost regular degree profile should be employed in each sub-code because otherwise covertness is penalized (proof is omitted due to lack of space). Hence, $\mathbf{d}_i$ contains at most two consecutive degrees, although they might not be the same ones for different sub-codes.

In the following section some results are presented where the search of the optimum degree profiles has been done by exhaustive search, a task that is feasible for IR-HARQ protocols including a small number of re-transmissions. The optimization results are presented for $T = 2$ and $T = 3$. In principle, $T = 2$ would require a four-dimensional search (at most two consecutive degrees per subcode), but it simplifies to a two-dimensional search when the linear constraints (19b)-(19c) are taken into account. In the case of $T = 3$ the complexity of the final four-dimensional search is further decreased by taking into account that the optimum values for $T = 2$ can be employed to reduce the search space for $T = 3$.

VI. Simulation results

The benefits of the IR-HARQ scheme have been evaluated taking as starting point one of the simulation scenarios in [9] and assessing the improvement in the trade-off WER performance vs VD obtained by the proposed scheme. In particular, the configuration $k = 300$, $l = 400$, $n = 61440$ with a regular (2,8) LDPC code and an MDNL regular code with AND node degree 10 was simulated for the $\epsilon_b = 0.007$ and $\epsilon_w = 0.33$ channels. In this scenario the non-HARQ original scheme obtained $WER = 0.114$ (corresponding to a mutual information $I = 417.324$) and a variational distance of 0.069728. The performance obtained by this scheme is indicated in figure 5 with '$\diamond$'. The performance of the non-HARQ schemes included in figure 4 is also depicted in dashed lines for the lengths $n = 6000$, 62000 and 64000. Note that the leftmost points (higher WER, better covertness) correspond to AND nodes of regular degree 10, whereas the rightmost points correspond to schemes having a 80% of AND nodes with degree 10 and the remaining ones with degree 9.

First, the covertness improvement thanks to IR-HARQ re-transmission was evaluated by constraining the MDNL code to keep an AND regular degree 10 and the same WER and length as the non-ARQ scheme ($I = 417.234$, $n_{all} = 61440$) and doing the optimization search in (19a)-(19c) over the length of each sub-codeword. The optimum values for $T = 2$ ($\mathbf{n}_1 =$ 52000, $\mathbf{n}_2 =$ 9440, $\Phi_{VD}^{Bound} = 0.067154$) and for $T = 3$ ($\mathbf{n}_1 =$ 48000, $\mathbf{n}_2 =$ 6215, $\mathbf{n}_3 =$ 7225, $\Phi_{VD}^{Bound} = 0.066336$) are shown in figure 5 with symbols $\times$ and $\circ$ respectively. The true weight multidimensional distribution of the $T = 3$ configuration was evaluated by MonteCarlo simulations and the true variational distance in equation (4) was evaluated for the case the warden has no access to the feedback channel obtaining $VD = 0.064184$; this result is depicted in the figure with the '$*$' symbol. Note that the optimized function in (19a) is effectively a loose upper bound on the true covertness, so the real gain provided by IR-HARQ is larger than the values indicated by the optimized bounds.

Second, the covertness improvement obtained allowing different degree profiles in each transmission block was evaluated by letting each sub-codeword to include a different amount of AND nodes with degree 9 and 10. Several target WER values were considered for $T = 2$ and $T = 3$. The optimum degree profile for each sub-codeword is indicated in table I (vector $\mathbf{d}_i$ is $[9\ 10]^T$ for all table entries, so it is is omitted for the sake of brevity) and the value of Φ_{VD}^{Bound} is depicted in figure 5. Note that IR-HARQ schemes operate in a region of the covertness-WER plot that cannot be attained by the non-ARQ scheme of the same number of channel uses n_{all}. Again, it is expected that the true covertness of the IR-HARQ will be significantly better than the bound in the plot.

Table I

Target WER	Target I	$T=2$ $\mathbf{n_1}$	 $\mathbf{n_2}$	$T=3$ $\mathbf{n_1}$	 $\mathbf{n_2}$	 $\mathbf{n_3}$
0.106	420	0 52143	418 8879	0 50000	0 5997	418 5025
0.058	440	0 54762	3439 3239	0 50000	439 7001	3000 1000
0.030	460	525 52915	5935 2065	0 52711	3310 1419	3150 850

VII. Conclusions and Future Work

This paper has explored the advantages that IR-HARQ presents in covert communications in those set-ups were a feedback channel is available. It has shown that the use of the feedback channel can significantly improve the trade-off between the Warden's detection capability and the error rate performance for finite wordlength.

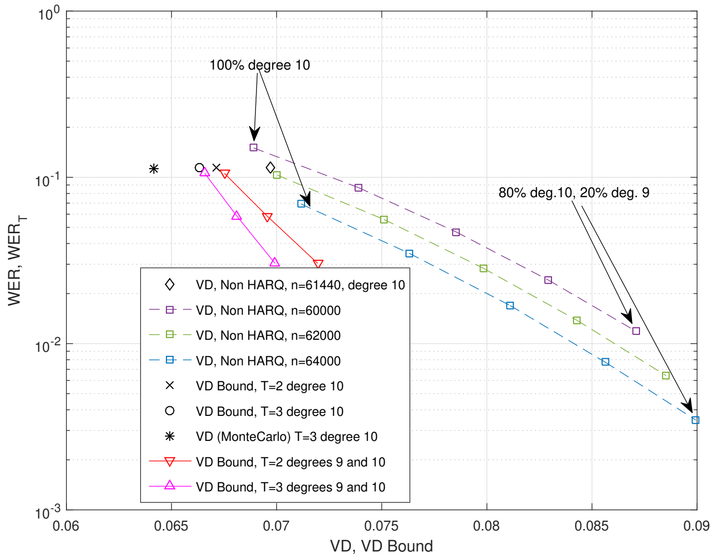

Figure 5. Variational distance vs WER for different schemes

The optimized variational distance corresponds to a bound of the true Warden's detection capability when he has no access to the feedback channel, and corresponds to the actual detection capability when he has noiseless observations of the feedback channel.

The IR-HARQ performance improvement has been evaluated for the concatenated codes proposed in [9], which are very flexible in terms of the design of the length and degree profile of each coding block. Because of the behavior of the non-linear stage, the decoding error probability in these codes depends on the codeword Hamming weight. Preliminary results not included in the paper due to the lack of space indicate that it is not clear whether this dependency is beneficial for covertness or not; this is a topic for future work.

References

[1] M. R. Bloch, "Covert Communication Over Noisy Channels: A Resolvability Perspective," *IEEE Transactions on Information Theory*, vol. 62 no. 5, pp. 2334-2354, May 2016.

[2] L. Wang, G. W. Wornell, and L. Zheng,"Fundamental limits of communication with low probability of detection," *IEEE Transactions on Information Theory* , vol. 62, no. 6, pp. 3493-3503, Jun. 2016.

[3] P.H. Che, M. Bakshi, and S. Jaggi. "Reliable deniable communication: hiding messages in noise," *arXiv preprint arXiv:1304.6693*, July 2016.

[4] M. Tahmasbi, M. R. Bloch, "First and Second Order Asymptotics in Covert Communication with Pulse Position Modulation," *.arXiv preprint: 1703.01362v2* December 2017.

[5] M. R. Bloch and S. Guha, "Optimal covert communications using pulse-position modulation," *2017 IEEE International Symposium on Information Theory (ISIT)*, 2017.

[6] I. A. Kadampot, M. Tahmasbi, and M. R. Bloch, "Multilevel-Coded Pulse-Position Modulation for Covert Communications," *2018 IEEE International Symposium on Information Theory (ISIT)*,2018.

[7] I. A. Kadampot, M. Tahmasbi, and M. R. Bloch, "Codes for Covert Communication over Additive White Gaussian Noise Channels," *2019 IEEE International Symposium on Information Theory (ISIT)*, 2019.

[8] S. Sesia, G. Caire, and G. Vivier, "Incremental redundancy hybrid ARQ schemes based on low-density parity-check codes," *IEEE Transactions on Communications*, vol. 52 no. 8, pp. 1311-1321, August 2004.

[9] M. Lamarca and D. Matas, "A non-linear channel code for covert communications," *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, 2019.

[10] T. Richardson and R. Urbanke, *Modern coding theory*, Cambridge Univ. Press, 2008